

Sebastian Zaddach

**Zum Beitrag Bayesscher Schätzverfahren
in der Vergleichswertermittlung**

München 2016

**Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck**

ISSN 0065-5325

ISBN 978-3-7696-5184-3

**Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover
ISSN 0174-1454, Nr. 321, Hannover 2016**



Zum Beitrag Bayesscher Schätzverfahren in der Vergleichswertermittlung

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades
Doktor-Ingenieur (Dr.-Ing.)
genehmigte Dissertation

von

Dipl.-Ing. Sebastian Zaddach
geboren am 10. März 1981 in Osnabrück

München 2016

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

ISSN 0065-5325

ISBN 978-3-7696-5184-3

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover
ISSN 0174-1454, Nr. 321, Hannover 2016

Adresse der Deutschen Geodätischen Kommission:



Deutsche Geodätische Kommission

Alfons-Goppel-Straße 11 • D – 80 539 München

Telefon +49 – 89 – 23 031 1113 • Telefax +49 – 89 – 23 031 -1283 / - 1100

e-mail hornik@dgfi.badw.de • <http://www.dgk.badw.de>

Prüfungskommission

Vorsitzender: Prof. Dr.-Ing. Ingo Neumann

Referent: Prof. Dr.-Ing. Winrich Voß

1. Korreferent: Prof. Dr.-Ing. habil. Hansjörg Kutterer

2. Korreferent: Prof. Dr.-Ing. Hans Joachim Linke

Tag der mündlichen Prüfung: 12.02.2016

© 2016 Deutsche Geodätische Kommission, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

ISSN 0065-5325

ISBN 978-3-7696-5184-3

Selbständigkeitserklärung

Hiermit erkläre ich, die von mir eingereichte Dissertation selbständig verfasst, keine Textabschnitte von Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir verwendeten Hilfsmittel und Quellen in der Arbeit angegeben zu haben. Weiterhin habe ich Dritten weder unmittelbar noch mittelbar geldwerte Leistungen für Vermittlungstätigkeiten oder für die inhaltliche Ausarbeitung der Dissertation erbracht. Es wurde von mir keine gleiche oder in wesentlichen Teilen ähnliche Arbeit bei einer anderen Hochschule als Dissertation eingereicht.

Osnabrück, 13. April 2016

Kurzfassung

Die *Verkehrswertermittlung* als wesentlicher Bestandteil der Analyse des Grundstücksmarktes ist von jeher mit hohen Anforderungen verbunden, die sich in normativen Vorgaben zur Definition des Verkehrswertes sowie den Methoden zu seiner Bestimmung widerspiegeln. Hauptaufgabe einer Verkehrswertermittlung ist eine zutreffende, *modellhafte Abbildung realer Zusammenhänge*, die zu einer marktgerechten Einschätzung des Wertes einer Immobilie führt und weitestgehend frei von den auf einen Kaufpreis wirkenden subjektiven Einflüssen ist. Eine wichtige Rolle spielen die Angaben von *Unsicherheitsmaßen*, die die Plausibilität der Schätzergebnisse belegen sollen und einen Bezug zum realen Marktgeschehen herstellen. Das *Vergleichswertverfahren* wird oftmals als die Methode mit der größten Marktnähe beurteilt; mit der klassischen multiplen linearen Regressionsanalyse und der einfachen Kollokation als indirekte Vergleichswertverfahren wurden mathematisch-statistische Modelle entwickelt, mit deren Hilfe sich die wesentlichen Marktmechanismen nachvollziehen und abbilden lassen. Die Modellformulierungen beruhen seit ihrer Einführung in die Verkehrswertermittlung auf der Sichtweise klassisch frequentistischer Statistik. Während sich die Modelle für Aufgaben der Schätzung durchgesetzt haben, sind in der Praxis Defizite in der durchgreifenden Auseinandersetzung mit der Angabe von Unsicherheitsspannen in der Verkehrswertermittlung zu erkennen.

Kern der Arbeit ist es, durch die Einführung der *Bayes-Statistik* eine probabilistische Betrachtungsweise in die Vergleichswertermittlung zu integrieren. Modelle der Bayes-Statistik ermöglichen zum einen die Integration von Vorwissen über beobachtete Phänomene, sodass Informationen berücksichtigt werden können, die bereits vor einer Auswertung vorliegen. Zum anderen wird ein erweiterter Zugang zum Begriff der Unsicherheit ermöglicht. Im Mittelpunkt steht das Bayes-Theorem, welches über die Modellierung auf Basis von Wahrscheinlichkeitsverteilungen eine im Vergleich zu den frequentistischen Ansätzen alternative Modellbildung erlaubt. Auf Grundlage des Vergleichswertverfahrens werden die Modelle der *Bayesschen Regression* und der *Bayesschen Kollokation* formuliert. Als erstes Forschungsziel der Arbeit ist die rein *funktionale Modellierung* zu sehen, die durch die Integration und Fortpflanzung von Vorinformationen eine Verbesserung in der Schätzung und Prädiktion von Vergleichswerten ermöglichen soll. Das zweite Forschungsziel ist die erweiterte *Unsicherheitsmodellierung*, die für die Ergebnisse der Vergleichswertermittlung die Angabe von realistischen Unsicherheitsmaßen und zudem eine gegenüber den klassischen Ansätzen reduzierte Unsicherheit in den Schätzwerten erlaubt.

Da sich Aufgaben der Verkehrswertermittlung in einem Spannungsfeld aus normativen Anforderungen, stark differenzierten Teilmärkten und einer heterogenen bis lückenhaften Datengrundlage bewegen, wird für die Nutzung der Bayesschen Modelle eine *Auswertestrategie* entwickelt. Aufgabe ist es, die spezifischen Voraussetzungen des Grundstücksmarktes mit den Anforderungen der Modellbildung zu verknüpfen. Die Auswertestrategie wird als *rekursiver Algorithmus* aufgebaut, der die Fortpflanzung von Auswerteergebnissen als Vorinformation für folgende Auswerteepochen ermöglicht. Für eine Beurteilung der praktischen Anwendbarkeit wird die Auswertestrategie anhand von realen Datensätzen getestet. Die Ergebnisse zeigen, dass verbesserte funktionale Modellierungen durch die Bayesschen Ansätze nicht zu erreichen sind. Steigerungen der Approximationsgüte lassen sich ausschließlich durch die Erweiterung des Regressionsmodells um einen Kollokationsansatz erreichen. Die wesentliche Stärke der Bayesschen Modelle liegt in der Unsicherheitsmodellierung. Als Ergebnis der Untersuchungen zeigt sich sowohl für den Ansatz der Bayesschen Regression als auch der Bayesschen Kollokation eine deutliche Reduzierung in den Unsicherheitsspannen der Schätzwerte. Ein signifikanter Mehrwert der Bayesschen Modelle und der entwickelten Auswertestrategie für die Anwendung in der Vergleichswertermittlung wird auf diese Weise nachgewiesen. Mit der Arbeit wird ein grundlegender Baustein gegeben, der aufgabenspezifisch weiterentwickelt und damit als Ausgangspunkt für künftige Forschungsfragen der Verkehrswertermittlung verwendet werden kann.

Schlagnworte: Verkehrswertermittlung, Vergleichswertverfahren, Unsicherheit, Bayessche Regression, Bayessche Kollokation, Auswertestrategie

Abstract

Real estate valuation can be considered as an essential part in analyzing the real estate market and is therefore associated with high demands. These demands are reflected in particular by the normative requirements for the definition of the market value and the methods for its determination. The main task of real estate valuation is an *accurate modelling of real circumstances*, which leads to a fair market assessment concerning the value of a property and which is not dependent on subjective influences. In this context, a key role is associated to the measurements of *uncertainty* which prove the plausibility of the estimated results and help to connect the results to the observed market situation. The *sales comparison approach* is often considered as the method with the highest marketability. The classical multiple linear regression analysis as well as the collocation method as mathematical-statistical models have been developed in order to be used as so called indirect sales comparison approaches and to help understanding and reflecting the essential market mechanisms underlying the market value. The models introduced to applications of real estate valuation are based on the classical perception of frequentist probability. While the models have asserted themselves as standard methods for several purposes in real estate valuation, deficits may be recognized in expertise discussion regarding the specification of uncertainty measures.

The main objective of this thesis is to integrate an enhanced probabilistic approach to purposes of real estate valuation by introducing the *Bayesian approach*. On the one hand, the Bayesian approach enables the integration of prior knowledge about the observed phenomena, so that information can be included which is already given in preparation of an evaluation. On the other hand, the Bayesian point of view offers the possibility to allow for a comprehensive access to uncertainty measures. The fundamental Bayes theorem provides a concept, which is based on the use of probability density functions and therefore gives an alternative way of statistical modelling. In this context, the sales comparison approach is adapted by introducing a *Bayesian regression approach* as well as a *Bayesian collocation approach*. The first aim of this research is the *functional modelling*. An improvement concerning the quality of estimation is supposed to be reached by the integration and propagation of prior knowledge. The second aim of research is the extended *modelling of uncertainty*, which permits the specification of realistic uncertainty within the process of valuation. Furthermore, another aspect lies in the proof that the use of prior knowledge leads to the reduction of uncertainty in comparison to approaches, where no further information is considered.

As the determination of the market value is characterized by a complex field of normative demands, strongly differentiated submarkets and mostly heterogeneous and incomplete data basis, a *strategy for evaluation* is required. The main task of the strategy is the connection of the specific conditions of the property market with the demands of the mathematical-statistical modelling. The strategy is based on a *recursive algorithm*, which allows for the propagation of results as prior information for subsequent periods of evaluation. In order to be able to evaluate the practical applicability of the Bayesian approach, the strategy for estimating comparative values is tested based on a data set of purchases. The results show that an improved functional modelling can neither be reached by the use of the Bayesian regression approach nor the use of the Bayesian collocation approach. The quality of the functional approximation can only be reached by the enhancement of the regression model by the collocation method. However, the essential strength of the Bayesian approaches is the uncertainty modelling. As a result of the practical evaluation the Bayesian regression approach as well as the Bayesian collocation approach it can be stated, that the integration of prior information leads to a significant reduction regarding the uncertainty of the estimated values. In this way, the Bayesian approach in conjunction with the developed strategy for estimating comparative values is able to improve the application of the sales comparison approach and gives an additional benefit to purposes of real estate valuation. The methodology of this thesis can be modified due to specific tasks and therefore has to be considered as a starting point for future research questions concerning purposes of real estate valuation.

Keywords: Real estate valuation, sales comparison approach, uncertainty, Bayesian regression approach, Bayesian collocation approach, strategy for evaluation

Inhaltsverzeichnis

1. Einführung	15
1.1. Stand der Forschung	15
1.2. Forschungsziele	18
1.3. Methodik und Aufbau der Arbeit	19
2. Modellbildung und deren Unsicherheit im Kontext der Verkehrswertermittlung	23
2.1. Rahmenbedingungen der Verkehrswertermittlung	23
2.1.1. Charakterisierung von Immobilien und Immobilienmärkten	23
2.1.2. Wertbegriff und rechtliche Rahmenbedingungen	26
2.1.3. Normative Anforderungen an Genauigkeiten in der Verkehrswertermittlung	29
2.2. Modelle als Abbild der Realität und ihre Unsicherheit	31
2.2.1. Mathematische Modellbildung und deren Qualität	31
2.2.2. Definition des Unsicherheitsbegriffs	34
2.3. Modelle des Vergleichswertverfahrens	39
2.3.1. Allgemeine Modellbildung in der Verkehrswertermittlung	40
2.3.2. Multivariate Analyseverfahren in der Vergleichswertermittlung	43
3. Grundlagen und Modelle der Bayes-Statistik	49
3.1. Einführung in die Bayes-Statistik	49
3.1.1. Grundlagen und Bayes-Theorem	50
3.1.2. Verallgemeinertes Bayes-Theorem	52
3.1.3. Parameterschätzung und Konfidenzbereiche im Bayesschen Modell	55
3.2. Lineares Modell der Bayesschen Regression für konjugierte Verteilungsdichten	57
3.2.1. Modelldefinition	58
3.2.2. Analytische Lösung im Modell mit unbekanntem Varianzfaktor	60
3.2.3. Konfidenzregion der Parameter und Schätzwerte	64
3.3. Lineares Modell der Bayesschen Kollokation für konjugierte Verteilungsdichten	65
3.3.1. Modelldefinition	65
3.3.2. Analytische Lösung im Modell mit unbekanntem Varianzfaktor	67
3.3.3. Konfidenzregionen der Parameter und Schätzwerte	72
3.4. Numerische Lösung für nicht konjugierte Verteilungsdichten	73
3.4.1. Monte-Carlo Integration	73
3.4.2. Monte-Carlo-Markov-Chain-Methoden zur numerischen Approximation	74
4. Entwicklung einer Auswertestrategie für eine Bayessche Vergleichswertermittlung	77
4.1. Voraussetzungen und Formulierung der Auswertestrategie	77
4.1.1. Voraussetzungen aus Sicht der Verkehrswertermittlung	78
4.1.2. Allgemeine Verteilungs- und Modellannahmen	80
4.1.3. Schema der Auswertestrategie	85
4.2. Vorbereitung der Bayesschen Auswertung	86
4.2.1. Modelldefinition	87
4.2.2. Aufbereitung einer Stichprobe	88
4.3. Bayessche Auswertung einer Stichprobe	92
4.3.1. Ablauf der Bayesschen Modellschätzungen	93

4.3.2. Modellwahl mittels Hypothesentests	98
4.3.3. Rekursive Bayessche Schätzung	103
4.4. Validierung und Wertung der Modelle und Ergebnisse	105
4.4.1. Methodik der Kreuzvalidierung	105
4.4.2. Methodik des Paarvergleichs	106
5. Praktische Anwendung der Auswertestrategie in der Vergleichswertermittlung	109
5.1. Modelldefinition und Stichprobenaufbereitung der Initialepoche	110
5.1.1. Definition der Grundgesamtheit	110
5.1.2. Definition der Stichprobendaten	111
5.1.3. Stichprobenaufbereitung der initialen Epoche	115
5.2. Modellierung im Ansatz der Bayesschen Regression	119
5.2.1. Modellwahl mittels Bayes-Faktor	120
5.2.2. Schätzergebnisse der Bayesschen Regression	125
5.2.3. Unsicherheitsmodellierung mittels Bayesscher Regression	129
5.3. Modellierung im Ansatz der Bayesschen Kollokation	134
5.3.1. Ableitung von Autokorrelationsfunktionen	134
5.3.2. Schätzergebnisse der Bayesschen Kollokation	137
5.3.3. Unsicherheitsmodellierung mittels Bayesscher Kollokation	143
5.4. Validierung der Modelle	148
5.4.1. Prädiktionsergebnisse der Bayesschen Regression	148
5.4.2. Prädiktionsergebnisse der Bayesschen Kollokation	149
6. Mehrwert Bayesscher Ansätze im Vergleich	157
6.1. Diskussion der funktionalen Modellierung	157
6.1.1. Schätzung der Modellparameter	158
6.1.2. Schätzung und Prädiktion der Posteriori-Zielgröße	160
6.1.3. Fazit zum ersten Forschungsziel	168
6.2. Diskussion der Unsicherheitsmodellierung	169
6.2.1. Unsicherheit der Modellparameter	169
6.2.2. Unsicherheit in der Schätzung und Prädiktion der Posteriori-Zielgröße	172
6.2.3. Fazit zum zweiten Forschungsziel	177
6.3. Beitrag rekursiver Bayesscher Schätzverfahren zur Verkehrswertermittlung	178
6.3.1. Wertung der Ergebnisse im Kontext der Vergleichswertermittlung	179
6.3.2. Wertung der Ergebnisse im Kontext der Verkehrswertermittlung	181
7. Fazit und Ausblick	187
Literatur	191
A. Anhang – Stichprobenbeschreibungen	205
A.1. Deskriptive Statistiken der Stichproben 2006 bis 2014	205
A.2. Korrelationen der Einflussgrößen in den Stichproben 2004 bis 2014	206
B. Anhang – Ergebnisse der Bayesschen Regression	209
C. Anhang – Ergebnisse der Bayesschen Kollokation	219
Curriculum vitae	253

Abkürzungsverzeichnis

Abs.	Absatz
AK OGA	Arbeitskreis der Gutachterausschüsse und Oberen Gutachterausschüsse in der Bundesrepublik Deutschland
AKS	Automatisierte Kaufpreissammlung
ALT	(Modifiziertes) Alter
AST	Ausstattungsstandard
B-Plan	Bebauungsplan
BAJ	Baujahr
BauGB	Baugesetzbuch
BBauG	Bundesbaugesetz
BF	Bayes-Faktor
BGB	Bürgerliches Gesetzbuch
BGH	Bundesgerichtshof
BIC	Bayesian Information Criterion
BLUE	Beste lineare erwartungstreue Schätzung - Best Linear Unbiased Estimation
BRW	Bodenrichtwert
bspw.	beispielsweise
BVerwG	Bundesverwaltungsgericht
bzgl.	bezüglich
bzw.	beziehungsweise
ca.	circa
d. h.	das heißt
DAT	Datum des Kaufvertrages
DIC	Deviance Information Criterion
DIN	Deutsches Institut für Normung
eng.	englisch
EU	Europäische Union
f.	folgende
ff.	folgende
FNP	Flächennutzungsplan
G	Gamma-Verteilung
GAG	Gutachterausschuss für Grundstückswerte
GBO	Grundbuchordnung
GFZ	Geschossflächenzahl
ggf.	gegebenenfalls
GHM	Gauß-Helmert-Modell
GMM	Gauß-Markoff-Modell

GRF	Grundstücksfläche
GRZ	Grundflächenzahl
GUM	Guide to the Expression of Uncertainty in Measurement
HPDI	Highest Posterior Density Interval
i. d. R.	in der Regel
i. i. d.	identically and independently distributed
i. S.	im Sinne
IG	Invertierte Gamma-Verteilung
ImmoWertV	Verordnung über die Grundsätze für die Ermittlung der Verkehrswerte von Grundstücken - Immobilienwertermittlungsverordnung
IPK	Immobilien-Preis-Kalkulator
KEL	Unterkellerung
KPR	Kaufpreis
KPW	Kaufpreis pro Wohnfläche
LGLN	Landesamt für Geoinformation und Landesvermessung Niedersachsen
MCMC	Monte-Carlo-Markov-Chain
MkQ	Methode der kleinsten Quadrate
ML-Schätzung	Maximum-Likelihood-Schätzung
NBank	Investitions- und Förderbank Niedersachsen
NG	Normal-Gammaverteilung
Nr.	Nummer
NV	multivariate Normalverteilung
OGA	Obere Gutachterausschuss
RMSE	Root Mean Square Error
SW-RL	Richtlinie zur Ermittlung des Sachwerts - Sachwertrichtlinie
t	(multivariate) t-Verteilung
u. a.	unter anderem
u. U.	unter Umständen
vgl.	vergleiche
VKM	Varianz-Kovarianzmatrix
VKV	Niedersächsische Vermessungs- und Katasterverwaltung
VW-RL	Richtlinie zur Ermittlung des Vergleichswerts und des Bodenwerts - Vergleichswertrichtlinie
W	Wohnbaufläche
WA	allgemeines Wohngebiet

WertR	Richtlinien für die Ermittlung der Verkehrswerte (Marktwerte) von Grundstücken -Wertermittlungsrichtlinien 2006
WIS	Wertermittlungsinformationssystem
WOF	Wohnfläche
WR	reines Wohngebiet
z. B.	zum Beispiel

1. Einführung

Kaufpreise von Immobilien sind durch verschiedene heterogene Einflüsse geprägt. Mit der Ermittlung des Verkehrswertes soll die Kaufpreisbildung nachvollzogen werden. Zielsetzung ist die modellhafte Abbildung realer Zusammenhänge: Eine marktgerechte Einschätzung des Immobilienwertes und eine größtmögliche Unabhängigkeit von den auf einen Kaufpreis wirkenden subjektiven Einflüssen ist Ziel der Verkehrswertermittlung. Durch die Definition des Verkehrswertes nach § 194 Baugesetzbuch (BauGB) in Verbindung mit den subordinativen Rechtsgrundlagen wird der allgemeingültige Begriff des Wertes weitgehend objektiviert und der Möglichkeit einer transparenten, marktkonformen und nachvollziehbaren Ermittlung zugeführt. Auf diese Weise kann ein objektiver Verkehrswert aus subjektiv beeinflussten Kaufpreisen abgeleitet werden. Das BauGB und insbesondere die nachgeordnete Verordnung über die Grundsätze für die Ermittlung der Verkehrswerte von Grundstücken - Immobilienwertermittlungsverordnung (ImmoWertV) geben Modelle vor, die durch Abstraktion der Wirklichkeit die preisbestimmenden Mechanismen am Grundstücksmarkt nachvollziehen. Als Datengrundlage dienen vorrangig Marktbeobachtungen, deren wesentlichste Komponente die Kaufpreisinformationen sind. Bei der Abbildung der komplexen Zusammenhänge am Immobilienmarkt sind Unsicherheiten im Bewertungsprozess sowohl durch die Eingangsdaten als auch in den Modellen an sich nicht zu vermeiden. Den Modellen der Verkehrswertermittlung kommt damit eine hohe Bedeutung zu: Trotz der charakteristisch heterogenen Datengrundlage werden hohe Ansprüche an die Genauigkeit des Verkehrswertes gestellt, insbesondere durch die Rechtsprechung. Somit ist von besonderer Bedeutung, wie weit ein ermittelter Verkehrswert vom Kaufpreis abweicht oder abweichen darf.

In der Praxis der Verkehrswertermittlung haben sich für die reine Modellbildung mathematisch-statistische Analyseverfahren bewährt, die eine Abbildung kausaler Zusammenhänge zwischen Kaufpreisen und vermuteten wertbeeinflussenden Merkmalen ermöglichen. Die vorliegende Arbeit ordnet sich in den Kontext des Vergleichswertverfahrens als eines der normierten Verfahren nach ImmoWertV ein: Hier werden die funktionalen Zusammenhänge in der Bildung der Kaufpreise beziehungsweise (bzw.) der Verkehrswerte geeignet modelliert. Im Fokus steht neben der Approximation von Kaufpreisen durch geschätzte Funktionswerte insbesondere die Unsicherheitsmodellierung, die bislang in praktischen Anwendungen kaum Berücksichtigung findet. Die Ergebnisse nach dem Vergleichswertverfahren werden dadurch mit zusätzlichen Angaben über die Plausibilität der Modelle und der geschätzten Funktionswerte signifikant erweitert.

1.1. Stand der Forschung

Die Preisbildung von Immobilien vollzieht sich nach Kriterien, deren Einflüsse und Zusammenwirken oftmals schwer zu beurteilen sind. Zum einen spielen die immanenten Charakteristika einer Immobilie eine wesentliche Rolle, die zu einer Zersplitterung in eine Vielzahl von räumlichen und sachlichen Teilmärkten führen. Zum anderen ist der Einfluss von weichen Faktoren von hoher Bedeutung, die sich entgegen der individuellen Charakteristika nicht immer konkret erfassen lassen. Hierunter fallen beispielsweise (bspw.) der Zugang zu Informationen über den Immobilienmarkt für Anbieter und Nachfrager, die Zielsetzungen der Marktteilnehmer sowie Verhandlungsgeschick. Weiterhin sind auch übergeordnete Entwicklungen wie demografische Einflüsse und volkswirtschaftliche Rahmenbedingungen von Bedeutung. Durch das komplexe Zusammenspiel von beeinflussenden Faktoren ist eine exakte Zuordnung zu Kaufpreisen nahezu unmöglich.

Ein wesentliches Qualitätsmaß für die Verkehrswertermittlung stellt mit den gegebenen Voraussetzungen eine adäquate Abbildung realer Zusammenhänge am Grundstücksmarkt dar. Es lassen sich unter diesem Gesichtspunkt zwei Komponenten identifizieren: Zum einen die Approximation realer Zusammenhänge durch eine geeignete, abstrahierende Modellbildung, zum anderen die Entstehung von Unsicherheiten, verursacht durch den Prozess der Datengenese sowie der Modellbildung an sich. Als unverzichtbares Hilfsmittel haben sich die mathematisch-statistischen Methoden erwiesen, die den Ansprüchen der immobilienpezifischen Eigenarten genügen.

Die Modellbildung als Approximation der Realität soll eine mathematische Formalisierung kausaler Zusammenhänge durch reale Beobachtungen ermöglichen: In der Verkehrswertermittlung sind insbesondere ökonometrische Modelle von Bedeutung. Diese spiegeln den Grundstücksmarkt in Bezug auf konkrete sachliche und räumliche Teilmärkte wider und ermöglichen eine optimale Anpassung von Schätzergebnissen an reale Zusammenhänge und Beobachtungen. Wird die Betrachtungsweise auf das indirekte Vergleichswertverfahren als zentrale Methode der vorliegenden Arbeit übertragen, stellt sich die Frage, wie stark die geschätzten Vergleichswerte von den tatsächlichen Kaufpreisen abweichen (*Aufgabe der funktionalen Modellierung*). Diesbezüglich hat sich für die Ermittlung des Vergleichswertes die multiple lineare Regressionsanalyse bewährt. Die mathematische Formulierung ermöglicht die Darstellung einer Beobachtung – hier dem Kaufpreis – als Linearkombination von signifikanten wertrelevanten Einflussgrößen. Eine Erweiterung stellt die einfache Kollokation dar. Sie ist in der Lage, das lineare Trendmodell um die Ausschöpfung eines Informationsgehaltes aus stochastischen Restinformationen zu ergänzen.

Begründet durch die speziellen Voraussetzungen der Wertermittlung von Immobilien in Form von Datenheterogenität und subjektiv beeinflusster Kaufpreise sind diese Modelle nicht in der Lage, ein exaktes Abbild betrachteter Teilmärkte zu geben. Aus diesem Grund sind für die Vergleichswerte geeignete Bereiche anzugeben, in denen sich der reale Kaufpreis mit hoher Wahrscheinlichkeit bewegt (*Aufgabe der Unsicherheitsmodellierung*). In den Ergebnissen der Verkehrswertermittlung ist im Idealfall neben einer Punktschätzung von Zielgröße und Modellparametern auch deren jeweilige Unsicherheit anzugeben, um die unsichere Informationslage auszudrücken. Sowohl die funktionale Modellierung als auch die Unsicherheitsmodellierung sind als zentrale Aspekte der Arbeit aufzufassen und führen unter Berücksichtigung ihrer Wechselwirkungen zur Notwendigkeit der Auseinandersetzung mit Modellbildung und Unsicherheiten im Bewertungsprozess.

Wird die erste identifizierte Aufgabenstellung der *funktionalen Modellierung* betrachtet, baut diese auf den bis heute fundamentalen Grundlagen der Vergleichswertermittlung durch die Forschungen von ZIEGENBEIN (1976, S. 219 ff.) auf; ZIEGENBEIN (1977) entwickelt einen Auswertealgorithmus, der auf den spezifischen Daten der Immobilienwertermittlung aufsetzt und neben der allgemeinen Ableitung einer Regressionsfunktion die Themen der Optimalität, Gütekriterien, Ausreißersuche, Hypothesentests und Konfidenzbereiche für Vergleichswerte grundlegend behandelt. Bis heute gilt die Vorgehensweise als Standard für die Auswertung der Kaufpreissammlung nicht nur in Niedersachsen und steht darüber hinaus unter dem Namen „Automatisierte Kaufpreissammlung (AKS) Niedersachsen“ als kommerzielles Softwareprodukt bundesweit zur Verfügung (vergleiche (vgl.) unter anderem (u. a.) ZIEGENBEIN (1977), ZIEGENBEIN (1995), ZIEGENBEIN (2010)). Seit der Adaption der Regressionsanalyse hat sich das Modell im Zusammenhang mit der Untersuchung spezifischer Wertermittlungsprobleme in zahlreichen Publikationen als grundlegendes Instrument erwiesen. In den Folgejahrzehnten verfasste Publikationen aus Wissenschaft und Forschung nutzen die Regressionsanalyse als Instrument der Untersuchungen. Durch zahlreiche Veröffentlichungen zu praktischen Wertermittlungsfällen wird die hervorgehobene Bedeutung der Regressionsanalyse gefestigt (vgl. zum Beispiel (z. B.) KOCH (1987), KANNGIESER ET AL. (1990), KANNGIESER ET AL. (1994), MÜRLE (2006) sowie ZEISSLER (2012)). Auch im internationalen Kontext hat sich die Anwendung für unterschiedliche Aufgabenstellungen der Verkehrswertermittlung durchgesetzt. Als Beispiel dienen reine Grundlagenwerke wie WANG & WOLVERTON (2002) oder MUNIZZO & VIRRUSO MUSIAL (2009), die sich mit der Rolle der Regressionsanalyse im Vergleichswertverfahren beschäftigen. Einen umfassenden Überblick über Anwendungsgebiete in unterschiedlichen Ländern geben KAUKO & D'AMATO (2008). Aufgrund der Vielzahl von nationalen und internationalen Veröffentlichungen, die auf der Regressionsanalyse basieren, wird auf eine abschließende Nennung verzichtet.

Ergänzend zur klassischen Regressionsanalyse stellt PELZER (1976) das Vergleichswertverfahren als Prädiktionsproblem dar. Dem erweiterten Modellansatz liegt die Annahme zu Grunde, dass die verbleibenden Abweichungen zwischen den Schätzwerten einer Regressionsfunktion und den zu erklärenden Preisen nach wie vor Systematiken enthält, die über einen Filteransatz ermittelt werden können. Erste praktische Anwendungen werden von PELZER (1978) sowie ZIEGENBEIN & HAWERK (1978) aufgezeigt und von UHDE (1982) als Modell der Filterung und Prädiktion, der einfachen Kollokation, vertieft behandelt. Die Veröffentlichungen von ZADDACH & ALKHATIB (2012a), ZADDACH & ALKHATIB (2012b), ZADDACH & ALKHATIB (2013a) sowie ZADDACH & ALKHATIB (2014a) knüpfen an die Untersuchungen von UHDE (1982) durch die Übertragung der Kollokation auf unterschiedliche sachliche und räumliche Teilmärkte an. Im direkten Vergleich zur Regressionsanalyse ist die Relevanz der Kollokation eher

als rein wissenschaftlich einzustufen; sie konnte sich bislang in praktischen Anwendungen nicht durchsetzen. Als ein weiterer alternativer Modellansatz wurden durch HAWERK (1992) die Anwendungsmöglichkeiten eines dynamischen Kalman-Filters erforscht und auf die Vergleichswertermittlung übertragen. Trotz plausibler Ergebnisse und eines nachgewiesenen Mehrwerts hat sich auch hier die praktische Anwendung nicht durchsetzen können. Anhand der bisherigen Ansätze für die funktionale Modellierung in der Verkehrswertermittlung wird deutlich, dass weiterer Forschungsbedarf hinsichtlich der Erprobung neuer oder erweiterter Ansätze gegeben ist, die relevante Fortschritte auch für praktische Anwendungen mit sich bringt.

Für die zweite identifizierte *Aufgabenstellung der Unsicherheitsmodellierung* wird aus den bisherigen Erläuterungen deutlich, dass trotz normativer Definition des Verkehrswertes und gegebenen Wertermittlungsmodellen Unterschiede zwischen Kaufpreisen und den Verkehrswerten in Form natürlicher Variationsbreite bestehen. Unsicherheiten über den wahren (fiktiven) Wert liegen nach HILDEBRANDT (1999, S. 325) sowohl in den Unsicherheiten der Eingangsdaten als auch in unterschiedlichen Einschätzungen wertrelevanter Elemente durch Akteure wie Sachverständige begründet.

Die Tatsache, dass keine präzisen Verkehrswerte abgeleitet werden können, ist von jeher Untersuchungsgegenstand wissenschaftlicher und praxisbezogener Untersuchungen gewesen. Für eine vertiefte Literaturstudie wird in diesem Zusammenhang auszugsweise auf STREICH (2003), PETERSEN (2007), METZGER (2008) sowie ENGEL (2008) verwiesen. KLEIBER (2014, Teil IV Nr. 1.2.3 Rn. 15) fordert für die Vermeidung einer Scheingenauigkeit unabhängig von der Methodik die Rundung der Schätzergebnisse, die sich in praktischen Anwendungen bewährt hat – vgl. u. a. RÖSSLER ET AL. (2005).

In Bezug auf das im Fokus stehende Vergleichswertverfahren wird der Verkehrswert im direkten sowie indirekt evidenten Preisvergleich durch die Mittelbildung angepasster Vergleichsfälle bestimmt (KLEIBER, 2007), (MÜRLE, 2006). Eine Genauigkeitsabschätzung erfolgt durch die Standardabweichung des Mittelwertes auf Basis einer Varianzfortpflanzung. Die Annahme normalverteilter Daten ermöglicht die Ableitung von Konfidenzbereichen sowie Hypothesentests. Im Rahmen der multiplen linearen Regressionsanalyse und der einfachen Kollokation als Standardverfahren des indirekt statistischen Preisvergleichs werden zufällige Variabilitäten berücksichtigt. Eine Genauigkeitsabschätzung erfolgt im Fall der Regressionsanalyse durch das Bestimmtheitsmaß sowie den Variationskoeffizienten. Beide Maße leiten sich aus der Streuung der Stichprobe ab und sind u. a. abhängig von der Anzahl und der Wahl der Vergleichsfälle (PETERSEN, 2007). In beiden statistischen Modellen können sowohl für die Modellparameter als auch die Schätzergebnisse die Konfidenzbereiche angegeben werden.

Trotz des Zugeständnisses gewisser Variationsbreiten im Schätzergebnis durch die Auslegung diverser Rechtsprechungen hat sich die Angabe eines Genauigkeitsbereiches in der Wertermittlungspraxis bis heute nicht etabliert: JESTER & ROESCH (2006, S. 161) geben an, dass sich Sachverständige scheuen, ein Gütemaß für ein Wertermittlungsergebnis anzugeben. Die Ausweisung von Schwankungsbreiten im Ergebnis werden aus vornehmlich juristischen Gründen vermieden. Andererseits muss konstatiert werden, dass eine zielführende Methodik für eine zutreffende Ableitung von Unsicherheiten nicht pauschal gegeben ist. Tendenziell große Variationsbreiten in den Schätzungsergebnissen scheinen zudem ein Hemmnis für den einzelnen Sachverständigen sein, die Unsicherheit einem Gutachten hinzuzufügen. Im Gegenzug und als Ergebnis der intensiven Auseinandersetzung mit der Problematik der näherungsweisen Angabe von Verkehrswerten plädiert die Fachwelt in großen Teilen für die Angabe von Unsicherheiten. Ziel muss es sein, der Verkehrswertermittlung neben der bewährten bisherigen Methodik ein Qualitätsmaß hinzuzufügen, welches die Angabe von Unsicherheiten ermöglicht (vgl. JESTER & ROESCH (2006), PETERSEN (2007), ENGEL (2008), METZGER (2008)).

Neben der Formulierung in den klassischen Modellen ist die Berücksichtigung von Unsicherheiten zentraler Untersuchungsgegenstand in weiteren Publikationen. Während STREICH (2003) in dem vereinfachten Ansatz einer Sensitivitätsanalyse die Auswirkungen von Variationen einzelner Einflussgrößen auf den Verkehrswert untersucht, zeigt HAACK (2006) eine Möglichkeit der umfassenden Sensitivitätsanalyse mittels Varianzfortpflanzung für die normierten Wertermittlungsverfahren. Neuere Methoden wie die Monte-Carlo Simulation werden zunehmend eingesetzt, um eine Aussage über variable Einflussgrößen und deren Auswirkungen auf den Verkehrswert treffen zu können. Die benötigten Verteilungsfunktionen werden dabei nicht empirisch auf Basis einer gegebenen Datengrundlage abgeleitet, sondern anhand gutachterlichen Sachverständnisses bestimmt (SOMMER, 2000), (JANSSEN,

2002), (STROTKAMP & SAUERBORN, 2002), (SIMON, 2004). Eine erweiterte Unsicherheitsmodellierung unter Verwendung neuer statistischer Ansätze erfolgt in ALKHATIB & WEITKAMP (2012) sowie WEITKAMP & ALKHATIB (2012). Mit dem Ziel, statistische Analysen mit einem adäquaten Unsicherheitshaushalt in kaufpreisarmen Regionen zu ermöglichen, wird die Bayes-Theorie auf Anwendungen in der Verkehrswertermittlung übertragen. Aufbauende Auswertungen zeigen vielversprechende Ergebnisse für die Modellierung der zufällig bedingten Unsicherheiten ((ALKHATIB & WEITKAMP, 2013), (WEITKAMP & ALKHATIB, 2014)). Die Erweiterung der klassischen Regression um einen Bayesschen Ansatz im Standardfall und unabhängig von kaufpreisarmen Lagen erfolgt durch ZADDACH & ALKHATIB (2013b) und wird in ZADDACH & ALKHATIB (2014b) erstmalig auf eine rekursive Parameterschätzung in der Verkehrswertermittlung angewendet. Die gemeinsame Behandlung zufälliger und systematischer Komponenten in einem umfassenden Ansatz wird erstmals durch ZADDACH (2014c) entwickelt und in ALKHATIB ET AL. (2015) weitergeführt. Als für die Verkehrswertermittlung innovatives Verfahren werden die zufälligen Abweichungen in einem Bayesschen Ansatz modelliert, die Berücksichtigung systematischer Fehlerkomponenten erfolgt auf Grundlage der Fuzzy-Theorie. Durch die Kombination der beiden mathematischen Ansätze wird der Unsicherheitshaushalt in der Verkehrswertermittlung signifikant erweitert und ermöglicht eine realitätsnähere Beschreibung der auftretenden Abweichungen.

Zusammenfassend ist zu erkennen, dass sich im Bereich der praktischen Anwendung von Modellen der Verkehrswertermittlung und speziell des Vergleichswertverfahrens offene Fragestellungen ergeben, die bislang nicht umfassend untersucht wurden. Im Mittelpunkt steht die Behandlung von Unsicherheiten, die im Bewertungsprozess auftreten und eine Quantifizierung der Auswirkungen auf die Schätzergebnisse eines Modells. Im Sinne der Markttransparenz als übergeordnetem Ziel der Verkehrswertermittlung ist die Angabe von Unsicherheitsspannen sowohl von Modellen als auch Schätzergebnissen unerlässlich und erfordert die methodische Weiterentwicklungen mit der Prüfung auf alternative Zugänge zur funktionalen Modellierung und Unsicherheitsmodellierung. Die bisherigen Anwendungen der klassischen Regressionsanalyse und der einfachen Kollokation wurden seit ihrer Einführung als Vergleichswertverfahren nicht signifikant erweitert; eine explizite Unsicherheitsmodellierung ist zudem kein zentraler Aspekt, der sich in der Praxis durchgesetzt hat.

Bezüglich einer erweiterten Methodik hat sich gemäß der genannten Publikationen die Bayes-Theorie als geeignet erwiesen, in der die zufällige Variabilität sowohl in den resultierenden Schätzwerten als auch in den Modellen selbst abgebildet werden kann. Diese Möglichkeit bleibt den klassischen Ansätzen verwehrt. Ein genereller Mehrwert ist in der Tatsache zu erkennen, dass entgegen der klassischen Ansätze Vorinformationen in die Auswertungen integriert werden können, die vor der eigentlichen Analyse bereits vorliegen und daher einen Erklärungsbeitrag liefern. Bislang stützen die Forschungen sich dabei ausschließlich auf Einzelauswertungen, so dass die Fortpflanzung von Unsicherheiten über mehrere Auswertzeiträume noch nicht umfassend in ihren Auswirkungen und eventuellen Mehrwerten untersucht wurde: Die Möglichkeiten rekursiver Schätzverfahren wurden weder in den klassischen noch in den bisherigen Bayesschen Ansätzen beleuchtet. Zudem fehlt eine zielführende Auswertestrategie, die die Erfordernisse der immobilien-spezifischen Datengrundlage und der Bayes-Theorie miteinander verbindet.

Die vorliegende Arbeit knüpft an die offenen Forschungsfragen an und schließt so die Lücke zwischen der Weiterentwicklung der klassischen Regression sowie der einfachen Kollokation als Vergleichswertverfahren und der Bayes-Theorie als Methodik für eine erweiterte funktionale Modellierung und Unsicherheitsmodellierung.

1.2. Forschungsziele

Das Vergleichswertverfahren bedarf eines erweiterten statistischen Ansatzes, der es erlaubt, die zufällige Unsicherheit in den Schätzergebnissen zu quantifizieren, zu reduzieren und geeignet fortzupflanzen. Im Fokus stehen folglich zum einen die rein mathematische Modellierung des mittelbaren Vergleichswertverfahrens als Regressions- sowie Kollokationsmodell, zum anderen die Quantifizierung von Unsicherheiten und deren Einfluss auf Schätzergebnisse im jeweiligen gewählten Modell. Dabei beschränkt sich die Formulierung der Unsicherheiten nicht wie bislang nur auf die Schätzwerte, sondern umfasst auch die Unsicherheiten der Modellparameter und wird so der Tatsache gerecht werden, dass Unsicherheiten in Beobachtungen und Modellen vorliegen. Insbesondere werden die Aspekte der Objektivität und Reproduzierbarkeit von Wertermittlungsergebnissen berücksichtigt. Für die praktische Anwendung wird eine Auswertestrategie formuliert, die als Kern eine rekursive Schätzung für die Fortpflanzung

von Modellunsicherheiten auf nachfolgende Auswerteperioden enthält. Auf diese Weise wird ein in der Verkehrswertermittlung innovativer Ansatz gegeben, der entgegen der bisherigen Auswertepaxis Vorinformationen über die Genauigkeit der einzelnen Modellparameter und damit implizit Vorwissen über die Unsicherheit der Einflussgrößen berücksichtigen kann. In dieser Form liefern die Ergebnisse einen signifikanten Fortschritt für die Schätzung und Plausibilität von Ergebnissen der Verkehrswertermittlung und generieren einen Mehrwert für die Marktanalyse und -transparenz.

Daraus lassen sich zwei Forschungsziele ableiten. In ihrer Kombination sollen sie als zentrales Ergebnis zu Modellen mit abgeleiteten Unsicherheitshaushalten führen, welche den bislang gängigen Auswertansätzen in der Vergleichswertermittlung überlegen sind. Die Forschungsziele werden formuliert zu:

1. *Verbesserung der funktionalen Modellierung:* Die Modelle des Vergleichswertverfahrens sollen eine verbesserte Anpassung der geschätzten Vergleichswerte an die Realität ermöglichen.

Die Modelle des Vergleichswertverfahrens werden auf Basis der Modellansätze der multiplen linearen Regressionsanalyse und der einfachen Kollokation formuliert. Zur innovativen Erweiterung wird die Bayes-Theorie genutzt. Durch die erweiterte Berücksichtigung von Unsicherheiten sowie die Integration von Vorinformationen sollen Verbesserungen in der Schätzung der unbekannten Modellparameter und damit der Vergleichswerte als Zielgröße erreicht werden.

2. *Verbesserung der Unsicherheitsmodellierung:* Die Modelle des Vergleichswertverfahrens sollen die Reduzierung auftretender Unsicherheiten in den Schätzungen der Modellparameter und der geschätzten Vergleichswerte sowie durch rekursive Anwendung die Fortpflanzung auf folgende Auswerteperioden ermöglichen.

Durch die Formulierung der mathematischen Modellansätze sollen die im Bewertungsverfahren auftretenden Unsicherheiten und deren Fortpflanzung auf die geschätzte Zielgröße im Vergleich zur herkömmlichen Standardmethode signifikant verringert werden. Ebenso soll die durch die Einflussgrößen und das Modell bedingte Unsicherheit in den unbekannten Modellparametern signifikant verringert werden. Die Ansätze ermöglichen durch die Implementierung eines rekursiven Schätzverfahrens die Übertragung von Unsicherheiten vergangener Auswertepochen auf aktuelle Auswertungen.

1.3. Methodik und Aufbau der Arbeit

Die bisherigen in der Praxis genutzten Modellbildungen des indirekten Vergleichswertverfahrens basieren auf frequentistischen (traditionellen) Ansätzen. Der wesentliche Grundgedanke dieser Ansätze bezieht sich auf die Definition des Wahrscheinlichkeitsbegriffs: Aus einer frequentistischen Sichtweise wird die Wahrscheinlichkeit für das Eintreten eines Ereignisses als relative Häufigkeit interpretiert. Damit beruht die zu Grunde liegende Statistik auf der Definition einer Zufallsvariablen als Ergebnis eines Experiments, welches mathematisch modelliert wird. Der Zufallsvariablen wird eine Wahrscheinlichkeit für ihr Eintreten zugeordnet. Unsicherheiten über die Schätzwerte des Modells werden durch Konfidenzintervalle ausgedrückt, welche den wahren Wert mit einer vorgegeben Wahrscheinlichkeit überdecken.

Im Gegensatz zur frequentistischen Sichtweise bietet der Ansatz der Bayes-Theorie einen alternativen Zugang zu wahrscheinlichkeitstheoretisch basierten Modellen: Im Mittelpunkt steht der Grundgedanke, dass Wahrscheinlichkeit als Grad der Plausibilität von bestimmten Ereignissen interpretiert wird. Die Bayes-Theorie basiert auf der Formulierung von Wahrscheinlichkeitsverteilungen nicht nur für zufällige Ereignisse, sondern auch für unbekannte Parameter. Neben der Formulierung der Wahrscheinlichkeitsverteilungen für Daten und unbekannte Parameter ermöglicht die Bayes-Theorie die Integration von Vorinformation – der so genannten Priori-Information – über Daten und Parameter in das mathematische Modell. Ziel eines Bayesschen Ansatzes ist es daher, die vorhandenen Informationsquellen in einem einzigen Modell zu kombinieren. Die Priori-Information kann dabei auf unterschiedlichen Wegen gewonnen werden. KUTTERER (2002, S. 29) nennt drei Arten: Gesicherte Ergebnisse der entsprechenden Fachwissenschaft, empirische Erkenntnisse aus der Praxis sowie Ergebnisse aus vorherigen Untersuchungen. Während die Ansätze in ALKHATIB & WEITKAMP (2012) sowie WEITKAMP & ALKHATIB (2012) auf der Ableitung von Priori-Information basierend auf einer Befragung von Experten der Immobilienwertermittlung beruhen, wird

in der vorliegenden Arbeit ein empirisches Bayes-Verfahren genutzt. Dabei werden die Priori-Informationen direkt aus den Daten vorheriger Auswertungen abgeleitet. Erste Untersuchungen zu dieser Vorgehensweise lassen sich ZADDACH & ALKHATIB (2013b) sowie ZADDACH & ALKHATIB (2014b) entnehmen.

Die Forschungsziele sollen durch die Verknüpfung der theoretischen Grundlagen der Bayes-Theorie mit den Erfordernissen der Verkehrswertermittlung erreicht werden und lassen sich in vier aufeinander aufbauende Teilschritte gliedern, wobei diese jeweils durch die Nutzung unterschiedlicher Methoden gekennzeichnet sind:

1. Formulierung der theoretischen Modelle des Vergleichswertverfahrens

- Formulierung und Lösung der Regressions- und Kollokationsmodelle basierend auf der Bayes-Theorie

2. Formulierung der Auswertestrategie

- Aufbereitung von Stichprobendaten, Korrelationsanalysen
- Ausreißersuche und Hypothesentests basierend auf Bayes-Theorie
- Generierung von Vorinformation durch empirische Bayes-Verfahren
- Formulierung der Rekursion
- Ansatz zur Kreuzvalidierung

3. Fallbeispiel – Anwendung der Auswertestrategie,

- Schätzung der Ergebnisse basierend auf Bayesscher Regression und Bayesscher Kollokation
- Prädiktion basierend auf Kreuzvalidierung

4. Wertung der Ergebnisse

- Paarvergleiche der Modellvarianten anhand von Vergleichskriterien jeweils für funktionale Modellierung und Unsicherheitsmodellierung

Detaillierte Beschreibungen der einzelnen Schritte und Methoden erfolgen kontextbezogen im weiteren Verlauf. Abbildung 1.1 verdeutlicht den Aufbau der Arbeit. In Kapitel 2 werden zunächst die benötigten theoretischen Grundlagen aufbereitet. Im Fokus stehen die Erläuterung der Rahmenbedingungen für die Aufgaben der Verkehrswertermittlung und die für die Argumentation der Arbeit fundamentale Bedeutung des Unterschiedes zwischen Kaufpreis und Verkehrswert. Weiterhin werden die Hintergründe von allgemeinen Modellbildungen thematisiert, um ein Verständnis für die funktionale Modellierung und resultierende Unsicherheiten zu schaffen. Die Modelltheorie wird auf die ökonometrische Modellbildung übertragen, die als Grundlage für die Formulierung des indirekten Vergleichswertverfahrens dient. Auf diese Weise werden die Voraussetzungen der Modellbildung mit der Verkehrswertermittlung verknüpft und eine argumentative Grundlage für den Forschungsansatz gegeben. Kapitel 3 ist einer Einführung in die Bayes-Theorie gewidmet. Neben einer allgemeinen Erläuterung des Grundprinzips des zentralen Bayes-Theorems werden aufbauend auf den Erkenntnissen des ersten Kapitels sodann die Ansätze der Bayesschen Regression und der Bayesschen Kollokation vorgestellt. Gegeben werden Modelle sowohl für die Integration von empirischen Priori-Informationen als auch Modellansätze, für die keine Vorinformation vorliegt. Neben den umfassenden analytischen Lösungsmöglichkeiten werden mit einem kurzen Exkurs die Grundlagen der Monte-Carlo Simulationen als alternativer Lösungsansatz aufgezeigt. In Kapitel 4 werden die Anforderungen aus Sicht der Verkehrswertermittlung und die abstrakt formulierten Bayesschen Modellansätze zu einer auf die Vergleichswertermittlung spezifisch zugeschnittenen Auswertestrategie verknüpft. Ziel ist die Entwicklung eines innovativen rekursiven Auswertealgorithmus, dessen Kernpunkt die Fortpflanzungsmöglichkeiten empirischer Priori-Informationen ist. Ebenso werden erstmalig eine Ausreißersuche basierend auf der Bayes-Theorie und eine Variante für Hypothesentests zum Zweck der Modellwahl gegeben. Anhand einer Stichprobe von Kauffalldaten wird die Auswertestrategie in Kapitel 5 auf einen praktischen Anwendungsfall für die Vergleichswertermittlung übertragen und so eine Beurteilung über das Erreichen der Forschungsziele ermöglicht. Neben einer Beschreibung der konkreten Schritte anhand der Auswertestrategie stehen insbesondere die funktionale Modellierung sowie die Ergebnisse für die Unsicherheitsmodellierung im Blickpunkt. Abschließend wird ein Ansatz für eine Kreuzvalidierung formuliert, welche eine Beurteilung der Prädiktionsgüte in den Modellansätzen erlaubt. In Kapitel 6 werden die Ergebnisse anhand

der Methodik des Paarvergleichs diskutiert und gewertet, jeweils getrennt nach den Aspekten der funktionalen Modellierung und der Unsicherheitsmodellierung. Durch die gewonnenen Erkenntnisse wird abschließend das Erreichen der Forschungsziele bewertet und die Übertragungsmöglichkeit der entwickelten Bayesschen Auswertestrategie auf weitere Aufgaben der Verkehrswertermittlung diskutiert. Die Arbeit schließt in Kapitel 7 mit einem Fazit und Ausblick.

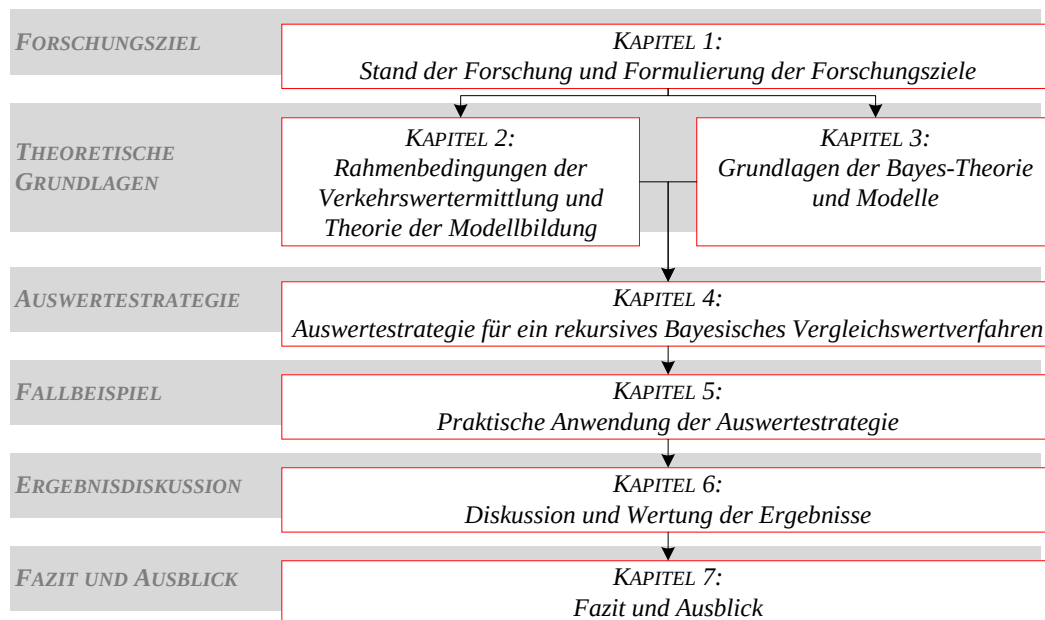


Abbildung 1.1: Aufbau der Arbeit

2. Modellbildung und deren Unsicherheit im Kontext der Verkehrswertermittlung

Der Gesetzgeber hat in Deutschland mit den §§ 192 bis 199 BauGB die Grundlagen für die Verkehrswertermittlung gelegt: Neben der Definition von Aufgaben und Organisation der Gutachterausschüsse und Oberen Gutachterausschüsse, der Führung einer Kaufpreissammlung sowie abzuleitenden Marktdaten ist insbesondere die fundamentale Definition des Verkehrswertes als Zielgröße von Wertermittlungsaufgaben zu nennen. Die Ermittlung des Verkehrswertes sowie dessen daten- und verfahrensbedingten Unsicherheiten sind ein wesentlicher Kernpunkt der vorliegenden Arbeit. In diesem Kontext werden im folgenden Kapitel die wesentlichen Grundlagen erläutern: In Abschnitt 2.1 werden zunächst die marktwirtschaftlichen Rahmenbedingungen von Immobilien aufgezeigt. Durch die Erläuterung der wesentlichen Charakteristika von Immobilien wird verdeutlicht, mit welchen Schwierigkeiten die Ermittlung von deren Wert verbunden ist. Ebenso werden die inhaltlichen Ansprüche an den Verkehrswert nach BauGB sowie die ergänzenden rechtlichen Grundlagen thematisiert. In Abschnitt 2.2 wird der Begriff des Modells zunächst abstrakt formuliert, um die Grundlagen für eine kontextbezogene Definition so genannter ökonomischer Modelle zu schaffen. Ausgehend von dem Modellbegriff wird im Anschluss der Term der Unsicherheit und seine Bedeutung für Modellbildungen erläutert. Abschnitt 2.3 gibt einen Überblick über die bislang gängigen Modelle der Verkehrswertermittlung mit einem besonderen Fokus auf dem direkten und indirekten Vergleichsverfahren. Abschließend werden die klassische multiple lineare Regressionsanalyse und die einfachen Kollokation als in praktischen Anwendungen bewährte Standardmethoden in ihren Grundzügen beleuchtet.

2.1. Rahmenbedingungen der Verkehrswertermittlung

Für ein erstes Verständnis der Komplexität von Aufgaben der Verkehrswertermittlung ist zunächst eine Definition des Begriffs der Immobilie selbst erforderlich. Die Herausforderungen für Wertermittlungsaufgaben können nur dann hinreichend beurteilt werden, wenn die besonderen Verhältnisse des Immobilienmarktes berücksichtigt werden: Durch die Charakterisierung der wesentlichen Eigenschaften von Immobilien sowie der Einordnung in den Kontext von Immobilienmärkten in Abschnitt 2.1.1 werden die Rahmenbedingungen für die Bewertung einzelner Immobilien aufgezeigt. Aufbauend auf der marktwirtschaftlich begründeten Sichtweise wird in Abschnitt 2.1.2 der insbesondere für mathematisch-statistisch basierte Schätzungen elementare Unterschied zwischen Kaufpreis und Wert eines Objekts erläutert. Durch die Definition des Verkehrswertes als maßgebende Größe des Bewertungsprozesses sowie den Erläuterungen der rechtlichen Vorschriften zu dessen Ermittlung wird der Objektivierung des stark subjektiv geprägten Immobilienmarktes Rechnung getragen.

2.1.1. Charakterisierung von Immobilien und Immobilienmärkten

Der Begriff der Immobilie ist per Gesetz nicht explizit definiert, der Gesetzgeber setzt den Begriff als bekannt voraus. Als inhaltlich gleichwertig kann der Term des Grundstücks gesehen werden, dessen Auslegung nach einem Urteil des Bundesverwaltungsgericht (BVerwG) aus der Grundbuchordnung (GBO) hervorgeht. Demnach bezeichnet ein Grundstück einen durch amtliche Vermessung räumlich begrenzten Bereich, der im Grundbuch auf einem Grundbuchblatt eingetragen oder unter einer eigenen Nummer im Bestandsverzeichnis auf einem gemeinschaftlichen Grundbuchblatt verzeichnet ist (BVerwG, IV C 82/69, vom 16.04.1971, NJW 1972, S. 701). Eine ergänzende Definition findet sich im deutschen Sachenrecht (§ 94 Absatz (Abs.) 1 Bürgerliches Gesetzbuch (BGB)): „Zu den wesentlichen Bestandteilen eines Grundstücks gehören die mit dem Grund und Boden fest verbundenen Sachen, insbesondere Gebäude, sowie die Erzeugnisse des Grundstücks, solange sie mit dem Boden zusammenhängen. (...)“. Wird der rein rechtlichen Ebene die wirtschaftliche Betrachtungsweise gegenüber gestellt, so steht die Immobilie als ökonomische Einheit im Vordergrund. Die Komplexität der Wertfindung für eine beliebige Immobilie kann

daher durch die Sichtweise der Immobilie als ein Wirtschaftsgut erklärt werden, welches mehr oder weniger marktwirtschaftlichen Gesetzmäßigkeiten unterworfen ist. Der Wert der Immobilie ist infolgedessen entscheidend davon abhängig, wie stark am Immobilienmarkt agierende Teilnehmer die durch eine Immobilie angebotenen Nutzungsmöglichkeiten nachfragen. Eine weitere immobilienökonomische Sichtweise geben SCHULTE ET AL. (2008, S. 13): „Immobilien sind Wirtschaftsgüter, die aus unbebauten Grundstücken oder bebauten Grundstücken mit dazugehörigen Außenanlagen bestehen. Sie werden von Menschen im Rahmen physisch-technischer, rechtlicher, wirtschaftlicher und zeitlicher Grenzen für Produktions-, Handels-, Dienstleistungs- und Konsumzwecke genutzt.“ Anhand dieser Definition ist zu erkennen, dass sich Immobilien von herkömmlichen Wirtschaftsgütern in ihren Charakteristika stark unterscheiden und dadurch auch entscheidend die Struktur des Immobilienmarktes beeinflussen.

Wirtschaftswissenschaftliche Theorien differenzieren generell zwischen vollkommenen und unvollkommenen Märkten, die auf den Kriterien der vollständigen und der unvollständigen Konkurrenz beruhen. In der Theorie drücken die Kriterien der vollständigen Konkurrenz die Idealvorstellung eines vollkommenen Marktes aus. Je weiter sich die Kriterien denen der unvollständigen Konkurrenz annähern, desto unvollkommener wird der Markt eingestuft. Als Kriterien für vollständige Konkurrenz gelten nach BRAUER (2013, S. 13 ff.) die Homogenität angebotener Wirtschaftsgüter, absolute Angebotselastizität, vollständige Markttransparenz sowie eine unendlich große Anzahl von Anbietern und Nachfragern. Die starken Abweichungen des Immobilienmarktes zu den Kriterien des Idealmarktes resultieren aus den Besonderheiten, die das Wirtschaftsgut Immobilie charakterisieren. Abbildung 2.1 zeigt einen Auszug der wesentlichsten Eigenschaften von Immobilien und deren Auswirkungen auf den Immobilienmarkt.

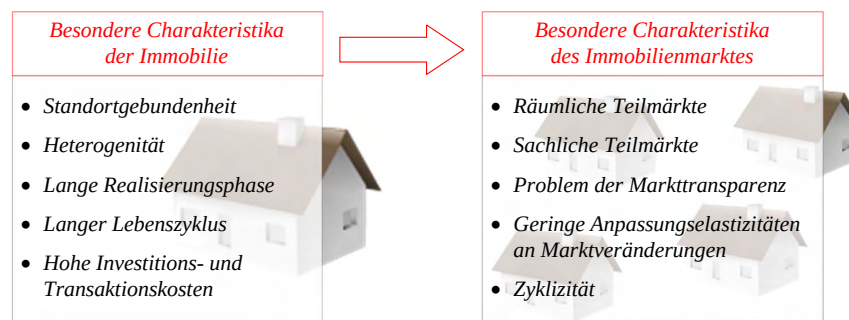


Abbildung 2.1: Wesentliche Eigenschaften von Immobilien und Immobilienmärkten

Anhand der dargestellten Merkmale lässt sich festhalten, dass Immobilienmärkte erheblich von vollkommenen Märkten im marktwirtschaftlichen Sinne abweichen (ENGEL, 2008, S. 2 f.):

- Die *Standortgebundenheit* gilt traditionell als die markanteste Eigenschaft einer Immobilie, da jede Lage einzigartig und nicht duplizierbar ist. Die Lage einer Immobilie wird in der Planungsphase festgelegt. Diese beeinflusst die späteren Nutzungsmöglichkeiten und den ökonomischen sowie in der Folge den subjektiven Wert der Immobilie. Die Lageeigenschaft einer Immobilie kann grundsätzlich nicht geändert werden, wodurch auch den unmittelbaren Umgebungseigenschaften eine erhebliche Bedeutung zukommt (GONDRING, 2009, S. 19) (BRAUER, 2013, S. 10).
- Sowohl unterschiedliche Nutzungsarten als auch die letztendliche architektonische Ausprägung eines Objekts führen dazu, dass es keine in allen Merkmalen exakt übereinstimmenden Immobilien geben kann, jedes Objekt kann als Unikat angesehen werden. Die *Heterogenität* einzelner Immobilien zeigt sich insbesondere in der Verschiedenartigkeit von Flächen (u. a. Grundfläche, Wohn- bzw. Nutzflächen, Mietflächen, ...), technischer Ausstattung, architektonischen Qualitäten, Alter, Instandhaltungszustand oder Modernisierungsgrad. Immobilien werden daher als einzigartiges, autonomes Wirtschaftsgut gesehen (GONDRING, 2009, S. 19) (SCHULTE, 2008, S. 15).
- Immobilien zeichnen durch eine *lange Realisierungsphase* aus. Bedingt durch den Projektumfang kann die Realisierung, gerechnet von der Planungsphase bis zur Fertigstellung, zwischen wenigen Monaten und mehreren Jahren liegen. Infolgedessen sind Reaktionsfähigkeit und Anpassungsflexibilität des Immobilienangebots auf Änderung in der Nachfragesituation als gering und zeitlich stark verzögert einzustufen (SCHULTE, 2008,

S. 15). Ergänzend zur langen Realisierungsdauer von Immobilien ist zudem deren *langer Lebenszyklus* charakteristisch. Während die Nutzungszeit von Grund und Boden keinerlei Einschränkungen unterliegt, wird bei Immobilien zwischen technischer und ökonomischer Nutzungsdauer unterschieden. Dabei übertrifft die erstgenannte die wirtschaftliche Nutzungsdauer, unter Umständen (u. U.) ergeben sich wirtschaftliche Nachnutzungen, wenn die technische Nutzbarkeit der Immobilie noch nicht überschritten ist. Beide Arten der Nutzungsdauer umfassen in der Regel (i. d. R.) mehrere Jahrzehnte.

- Die Neuerrichtung oder der Kauf von bestehenden Immobilien erfordert einen meist langfristigen Kapitaleinsatz. Anlagen am Immobilienmarkt sind daher ebenso durch *hohe Investitions- wie hohe Transaktionskosten* geprägt. Erstgenannter Aspekt bedingt die hohe Bedeutung von Fremdfinanzierung und in der Folge eine direkte Abhängigkeit der Neubautätigkeiten und der Liquidität des Immobilienmarktes vom Kreditvergabeverhalten von Banken und Investoren (WERNECKE & ROTTKE, 2001, S. 10). Unter die Transaktionskosten fallen Erwerbsnebenkosten, Grunderwerbsteuer, Gerichts-, Grundbuch- und Notargebühren. Als Resultat der erforderlichen Aufwendungen werden Immobilien gewöhnlich langfristig im Bestand gehalten, was wiederum zu schlechter Fungibilität und geringer Markttransparenz führt (GONDRING, 2009, S. 20).
- Für den Wert einer Immobilie spielen zusätzlich zu den genannten Aspekten *Anbieter und Nachfrager* eine bedeutende Rolle: Es bestehen zahlreiche persönliche Präferenzen wie bspw. Nähe einer Immobilie zum aktuellen Wohnort der Entscheidungsträger, Vorlieben für bestimmte Architektur oder die Lage der Immobilie zu Standorten wie Arbeitsplatz und öffentlichen Einrichtungen. Zudem besteht insbesondere im privaten Sektor oftmals eine eingeschränkte Markttransparenz: Anbieter und Nachfrager handeln demnach nicht zwangsläufig nach theoretisch fundierten marktwirtschaftlichen Gesichtspunkten.

In Bezug auf die genannten Eigenschaften von gehandelten homogenen Gütern auf einem vollkommenen Markt führen die Eigenschaften der Immobilien zu einem Mangel an unmittelbarer Vergleichbarkeit. In der Folge gibt es keine eindeutigen Marktpreise, vielmehr ist eine Streuung in gewissen Bandbreiten zu erwarten. SCHULTE (2008, S. 15) führt dazu an, dass Immobilienpreise daher nur über eine reduzierte Informationsfunktion verfügen. Als direkte Folge ergeben sich für den Immobilienmarkt die in Abbildung 2.1 aufgeführten Besonderheiten, die eine Abweichung von vollkommenen Märkten bewirken. Wie bereits angedeutet, stellt das bedeutendste Merkmal für die Bewertung von Immobilien regelmäßig deren Standortgebundenheit dar. Als unmittelbare Auswirkung auf den Immobilienmarkt ergibt sich, dass sich das Angebot lediglich auf einen räumlich begrenzten Ort bezieht. Die regional eingeschränkte Marktsituation resultiert in einer Zersplitterung des Immobilienmarktes in eine Vielzahl so genannter *räumlicher Teilmärkte*, deren wesentliches Charakteristikum nicht nur die geographische Lage ist, sondern auch die sich daraus ergebenden Abweichungen hinsichtlich Bevölkerungsstrukturen, Bebauungsdichte, Finanzkraft und bevorzugten Immobilientypen (GONDRING, 2009, S. 24). Die Nachfrage bezüglich (bzgl.) vergleichbarer Angebote in unterschiedlichen räumlichen Teilmärkten kann daher erheblich differieren: Immobilien an einem beliebigen Standort A können auf hohe Nachfrage treffen, während sie an einem Standort B nicht oder nur schwer gehandelt werden können. Die Nachfrage beruht auf den persönlichen Präferenzen der potenziellen Käufer und ist nicht oder lediglich bedingt substituierbar. Auch die erläuterte Heterogenität der Immobilien beeinflusst zusätzlich spezifische Angebote und Nachfragen. Je nach Betrachtungsweise und Präferenzen lassen sich lokale, regionale und überregionale Teilmärkte unterscheiden, die in ihrer Summe den Immobilienmarkt ergeben (SANDNER & WEBER, 2007, S. 329).

Eine weitere wesentliche Charakterisierung neben der räumlichen Trennung ist die Unterscheidung nach *sachlichen Teilmärkten*, die sich aufgrund verschiedener Nutzung und damit einhergehend einer Vielzahl von verschiedenen Immobilientypen ergeben. So lassen sich spezifische Märkte beispielsweise für Wohn-, Gewerbe und Sonderimmobilien unterscheiden, die wiederum jeweils individuellen Marktgegebenheiten und -gesetzmäßigkeiten unterworfen sind. Im Rahmen der vorliegenden Arbeit ist lediglich eine grobe Segmentierung in räumliche und sachliche Teilmärkte relevant, für einen tieferen Einblick in Teilmarktstrukturen wird auf weiterführende Literatur verwiesen (vgl. u. a. SCHULTE (2008), GONDRING (2009), BRAUER (2013), KLEIBER (2014)).

Nachdem die Definition von Immobilien gegeben ist, wird im folgenden Abschnitt der Begriff des Wertes sowie die Abgrenzung zum Preis mit den Gegebenheiten am Immobilienmarkt verknüpft.

2.1.2. Wertbegriff und rechtliche Rahmenbedingungen

Anhand der Charakteristika von Immobilien wird deutlich, dass für eine Wertfindung von Immobilien eine Vielzahl von Einflüssen zu berücksichtigen ist und sich diese daher als komplexer Prozess darstellt. Als Grundlage für die weiteren Ausführungen werden in den folgenden Unterabschnitten der Wertbegriff definiert sowie eine Abgrenzung zum Preis vorgenommen. Abschließend wird der Verkehrswert als für diese Arbeit maßgebende Größe erläutert.

2.1.2.1. Definition des Wertbegriffs und Abgrenzung zum Preis

Selbst bei streng objektiver und sorgfältigster Ermittlung des objektiven Wertes einer Immobilie ergeben sich aufgrund der in Abschnitt 2.1.1 erläuterten Aspekte mitunter erhebliche Differenzen zum tatsächlich ausgehandelten Kaufpreis. In einem objektiven Bewertungsprozess muss der (Immobilien)wert daher inhaltlich deutlich vom (Kauf)preis getrennt werden, da es sich zwar um zwei eng verknüpfte, aber in ihrer Aussage grundverschiedene Begriffe handelt. Dabei ist der Begriff der Bewertung - und damit auch allgemein der des Wertes - zunächst sehr weit gefasst und bedarf für eine zweckgebundene Anwendung einer tiefer gehenden Definition unter Berücksichtigung gängiger Werttheorien. FISCHER (2007, S. 721) führt in diesem Zusammenhang an, dass verschiedene Wertbegriffe existieren: Allgemeingültig könne der Vorgang der Bewertung in Bezug auf Immobilien definiert werden als die Zuordnung eines Wertes zu einem Objekt oder Rechten an Objekten. Das zu bewertende Objekt sei in diesem Fall der so genannte Wertträger, welches bestimmt werde durch die Summe seiner Funktionen, Merkmale oder Eigenschaften. Neben diesen Eigenschaften sei der Wert eines Objektes aber auch zusätzlich und entscheidend abhängig vom betrachtenden Subjekt bzw. dessen Bedürfnissen, woraus folge, dass ein Wert selbst keine objektive Eigenschaft sein könne. Für den so definierten Begriff des Wertes haben sich drei verschiedene Theorien herausgebildet: Die objektive, die subjektive sowie die funktionale Betrachtungsweise.

Die *objektive Werttheorie* bezieht sich auf einen Wert, dem bestimmte Objekteigenschaften anhaften. Im Kontext der Wertermittlung bedeutet dies einen Immobilienwert, der von jedermann realisiert werden kann. Die Ausrichtung oder Bezugnahme auf konkrete Käufer oder Verkäufer sowie deren individuelle Vorstellung und Verhandlungsgeschicke werden für die Ermittlung des Immobilienwertes ausgeblendet, lediglich die Erfolgspotenziale im Sinne (i. S.) des technischen und ökonomischen Nutzens einer Immobilie werden bewertet. Die Objekt-Subjekt-Beziehung nach Definition von FISCHER (2007) spielt in diesem Fall allenfalls eine untergeordnete Rolle, der resultierende Immobilienwert ist unabhängig von der Sichtweise aus Käufer- und Verkäufersicht (GERARDY ET AL., 2014, S. 1.1.2/3). Im Gegensatz zur objektiven Werttheorie stellt die *subjektive Werttheorie* eine streng auf die eigenen Interessen ausgerichtete Sichtweise eines handelnden Individuums in den Mittelpunkt. Im Ergebnis bedeutet dies für den resultierenden Wert je nach Käufer- oder Verkäufersicht, oder genauer nach deren subjektiven Interessen, eine Wertober- bzw. Wertuntergrenze. Im Vordergrund steht der einer Immobilie individuell beigemessene Gebrauchswert durch eine der Vertragsparteien. Die *funktionale Werttheorie* stellt sich als Symbiose der beiden erstgenannten Theorien dar. Während die Ermittlung eines Immobilienwertes auf den objektiven Kriterien gemäß der objektiven Werttheorie beruht, werden explizit die subjektiven Sichtweisen von Käufer und Verkäufer unter Berücksichtigung der jeweiligen subjektiven Absichten und Möglichkeiten mit berücksichtigt. KLEIBER (2007, S. 116) erläutert in diesem Kontext, bei der funktionalen Werttheorie soll es „(...) *entscheidend auf die Funktion des Werts als Entscheidungsgrundlage für eine konkrete subjektive Situation ankommen, wobei aber auch hier wirtschaftlich vernünftiges Handeln als Maßstab zu fordern ist* (...)“.

Wird für die Bewertung von Immobilien der strenge Maßstab der objektiven Werttheorie angehalten, zeigt sich, dass im Gegensatz hierzu ein Kaufpreis maßgeblich durch subjektive Kriterien beeinflusst wird. Gemäß § 433 Abs. 2 BGB definiert sich der Kaufpreis als Gegenleistung des Käufers für den Kauf einer Sache. Dies setzt die Einigung von Käufer- und Verkäuferseite über einen gültigen Kaufvertrag im Vorfeld voraus. Objektiver Wert und ausgehandelter Kaufpreis sind daher i. d. R. nicht gleichzusetzen. Abbildung 2.2 veranschaulicht den wesentlichen Unterschied: Der Kaufpreis bewegt sich innerhalb einer Spanne zwischen den subjektiv geprägten Grenzpreisen von Käufer und Verkäufer. In einem Urteil des Bundesgerichtshof (BGH) wurde der Unterschied weiterhin ausgeführt: „Der Preis einer Sache muss nicht ihrem Wert entsprechen. Er richtet sich gerade bei Grundstücken (...) nach Angebot und Nachfrage und wird jeweils zwischen Käufer und Verkäufer ausgehandelt. Marktpreis und objektiver Verkehrswert spielen keine entscheidende Rolle, vielmehr sind oft spekulative Momente (Kaufkraft, Geldwert usw.)

von erheblicher Bedeutung, häufig auch die persönlichen Vorstellungen und Wünsche des Kaufinteressenten. Der Verkäufer versucht den höchstmöglichen Preis zu erzielen, mag dieser auch unvernünftig sein. Der Käufer ist bestrebt, möglichst wenig zu zahlen, mag dabei das Grundstück auch „verschenkt“ sein.“ (BGH, Urt. Vom 25.10.1967 - VIII ZR 215/66 -, EzGuG 19.11; LG Darmstadt, Beschl. vom 16.10.1958 - 5 T 18/58 -, EzGuG 19.4).

Der Unterschied zwischen dem subjektiven Kaufpreis und einem objektivierten Wert ist für die Betrachtung von Unsicherheiten im Bewertungsprozess von entscheidender Bedeutung. Abweichungen zwischen realen Kaufpreisen und Verkehrswerten sind aufgrund der Komplexität der in den vorherigen Unterabschnitten genannten Eigenschaften einer Immobilie nicht zu vermeiden.

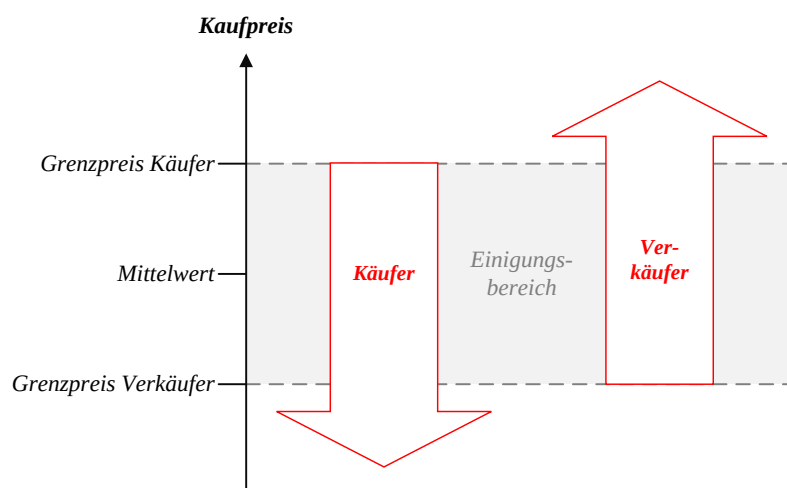


Abbildung 2.2: Prinzip der Preisbildung am Grundstücksmarkt nach ZEISSLER (2012, S. 5)

In Bezug auf das heutige Wirtschafts- und Rechtsleben existiert daher eine Vielzahl an Wertbegriffen, die aufgabenspezifisch zu betrachten sind. Eine umfassende Übersicht über die wichtigsten Begriffe geben KLEIBER (2007, S. 109 ff.) sowie BISCHOFF (2007, S. 446) und sollen an dieser Stelle nicht weiter vertieft werden. Der zentrale Wertbegriff für die vorliegende Arbeit beruht auf der Definition des Verkehrswertes nach § 194 BauGB:

„Der Verkehrswert (Marktwert) wird durch den Preis bestimmt, der in dem Zeitpunkt, auf den sich die Wertermittlung bezieht, im gewöhnlichen Geschäftsverkehr nach den rechtlichen Gegebenheiten und tatsächlichen Eigenschaften, der sonstigen Beschaffenheit und der Lage des Grundstücks oder des sonstigen Gegenstands der Wertermittlung ohne Rücksicht auf ungewöhnliche oder persönliche Verhältnisse zu erzielen wäre.“

In dieser materiell-rechtlichen Definition wird der Verkehrswert mit dem (Kauf)preis gleichgesetzt, um der Ermittlung von überindividuellen, marktkonformen und allgemein anerkannten Werten gerecht zu werden. Die Vermeidung rein subjektiv beeinflusster Werte erfolgt dabei über die normativen Einschränkungen für die Preisbildung, die sich allein an den objektiven Merkmalen eines Grundstücks orientieren. Der Verkehrswert kann daher auch als „verobjektivierter Tauschwert“ bezeichnet werden, wie er sich im freien Spiel aus Angebot und Nachfrage für jedermann ergibt (KLEIBER, 2014, Teil III, Rn 9).

Eine der wichtigsten normativen Bestimmungen ist die Angabe eines Wertermittlungsstichtages als *Zeitpunkt*, auf den sich die Ermittlung des Verkehrswertes bezieht. Von besonderem Interesse sind hier die allgemeinen Wertverhältnisse und in diesem Kontext die allgemeine Wirtschaftslage, die Verhältnisse am Kapitalmarkt sowie wirtschaftliche und demografische Entwicklungen. Anhand der genannten Kriterien wird deutlich, dass ein Verkehrswert i. d. R. selbst bei konstanten Zustandsmerkmalen des Wertermittlungsobjekts Schwankungen unterlegen ist (KLEIBER, 2014, Teil III, Rn 9). Der Zeitpunkt kann und soll jedoch nicht als absoluter Stichtagswert angesehen werden, sondern soll die allgemeinen Wertverhältnisse in einer eng begrenzten Zeitspanne (bis zu einigen Monaten) in Betracht ziehen (SANDNER & WEBER, 2007, S. 690).

Der Begriff des *gewöhnlichen Geschäftsverkehrs* als weitere normative Einschränkung geht von wirtschaftlich vernünftigem, marktgerechtem Handeln aus und schließt Umstände und Situationen aus, die für den Markt nicht üblich sind (ZIEGENBEIN, 2010, S. 426). Der Handel von Immobilien muss sich daher auf einem freien Markt vollziehen, auf dem ausschließlich Angebot und Nachfrage den Preis bestimmen. Gewöhnlicher Geschäftsverkehr

schließt gleichzeitig *ungewöhnliche oder persönliche Verhältnisse* aus; KLEIBER (2014, Teil III, Rn 18) weist auf die Tautologie der Begrifflichkeiten hin. Ausgeschlossen aus dem gewöhnlichen Geschäftsverkehr sind besondere Interessenslagen, ungeklärte Abhängigkeiten oder anormale Einflussgrößen. Als gängige Beispiele sind Erwerbsvorgänge in Zwangsversteigerungsverfahren oder bei Erbaueinandersetzungen zu nennen (SANDNER & WEBER, 2007, S. 690).

Rechtliche Gegebenheiten beziehen sich sowohl auf das öffentliche als auch das private Recht. Im Rahmen des öffentlichen Rechts sind die Regelungen des Bauplanungs- und Bauordnungsrechtes relevant, hier insbesondere die baurechtlichen Darstellung im Flächennutzungsplan (FNP) sowie Bebauungsplan (B-Plan). Im Privatrecht sind die auf einem Objekt lastenden Nutzungsrechte wie bspw. Nießbrauch, Verfügungs- und Erwerbsrechte sowie Sicherungs- und Verwertungsrechte von Belang (BISCHOFF, 2007, S. 466 f.).

Unter die *tatsächlichen Eigenschaften* eines Wertermittlungsobjekts fallen alle objektiv erfassbaren Merkmale. Eingeschlossen sind neben dem Gebäude auch die Gegebenheiten von Grund und Boden. Die regelmäßig bei einer Verkehrswertermittlung nach BauGB untersuchten wertbeeinflussenden Merkmale sind u. a. Grundstücksgröße und -zuschnitt, Erschließungszustand, Bodenbeschaffenheit und Baugrundeignung, Gebäudezustand und -alter sowie die Bauweise.

Der Begriff der *sonstigen Beschaffenheit* wird weder in der Rechtsprechung noch in der gängigen Fachliteratur abschließend definiert. Gemäß der Ausführungen von GERARDY ET AL. (2014, S. 1.2.3/20) ist unter dem unbestimmten Rechtsbegriff eine inhaltliche Zusammenfassung der *rechtlichen Gegebenheiten* und *tatsächlichen Eigenschaften* zu verstehen, mit der Maßgabe, dass alles, was nicht diesen Aspekten zuzurechnen ist, unter die *sonstige Beschaffenheit* fällt. Beispiele sind merkantile Minderwerte oder Beeinträchtigungen durch Immissionen (SANDNER & WEBER, 2007, S. 691).

Die *Lage* ist in jeder Verkehrswertermittlung als einer der wichtigsten Faktoren zu beurteilen. Wertermittlungsobjekte völlig gleicher Nutzungsart und Struktur können unterschiedlich wertvoll sein, wenn sie sich nicht am gleichen Ort befinden. Die räumliche Einordnung und damit die Eingrenzung von räumlichen Teilmärkten sind für eine Verkehrswertermittlung sorgfältig vorzunehmen.

2.1.2.2. Rechtliche Rahmenbedingungen

Aus der erläuterten Legaldefinition wird deutlich, dass sich die Ermittlung von Verkehrswerten nach objektiven Kriterien zu vollziehen hat. Neben der normativen Vorgabe des § 194 BauGB sind daher zusätzliche Vorschriften notwendig, die die objektive Bestimmung ermöglichen und den genannten Anforderungen gerecht werden. Die Grundlagen für die Verkehrswertermittlung sind in den §§ 192 bis 199 im Ersten Teil des Dritten Kapitels des BauGB geregelt. Inhalt sind neben der bereits erläuterten materiell-rechtlichen Definition des Verkehrswertes in § 194 die Einrichtung und Aufgaben der Gutachterausschüsse für Grundstückswerte sowie die Ermächtigung für den Erlass von ergänzenden Vorschriften durch Bundes- und Landesregierung. Auf Bundesebene wird die Regierung gemäß § 199 Abs. 1 BauGB ermächtigt, Vorschriften über die Anwendung gleicher Grundsätze bei der Ermittlung der Verkehrswerte und bei der Ableitung der für die Wertermittlung erforderlichen Daten einschließlich der Bodenrichtwerte zu erlassen. Erstmals wurden diese Grundsätze 1961 in Form der Wertermittlungsverordnung formuliert. Nach diversen Novellen ist am 1. Juli 2010 die bis heute gültige *ImmoWertV* als wichtigste Ergänzung zu den Wertermittlungsvorschriften im BauGB in Kraft getreten. Neben der Bundesregierung haben die Landesregierungen die Ermächtigung des § 199 Abs. 2 BauGB genutzt, um Rechtsverordnungen über die organisatorischen Voraussetzungen für die Tätigkeit der Gutachterausschüsse und ihrer Geschäftsstellen, für die Führung und Auswertung der Kaufpreissammlungen und für die Erledigung sonstiger Aufgaben festzulegen. Ergänzend zu den Rechtsverordnungen sind sowohl auf Bundes- als auch Landesebene zusätzliche Verwaltungsvorschriften erlassen worden. Inhaltlich soll durch die Richtlinien das Ziel einheitlicher Begriffsverwendungen und Verfahrensweisen bei der Ermittlung von Verkehrswerten erreicht werden (ZIEGENBEIN, 2010, S. 427). Für Anwendungsbereiche der Verkehrswertermittlung sind insbesondere die *Richtlinien für die Ermittlung der Verkehrswerte (Marktwerte) von Grundstücken - Wertermittlungsrichtlinien 2006 (WertR)* vom 1. März 2006 von Bedeutung. Durch die zeitlich nachgeordnete Novellierung der ImmoWertV wurde die Erforderlichkeit einer Überarbeitung und Anpassung erkannt, die schrittweise umgesetzt wird. Zu diesem Zweck werden zunächst einzelne Richtlinien erarbeitet und abschließend zu einer

Gesamtrichtlinie zusammengeführt. In diesem Zuge wurde am 20. März 2014 insbesondere die *Richtlinie zur Ermittlung des Vergleichswerts und des Bodenwerts - Vergleichswertrichtlinie (VW-RL)* veröffentlicht, die durch ihre Relevanz für die weitere Arbeit an dieser Stelle hervorgehoben werden soll.

In Ergänzung zu den rechtlichen Rahmenbedingungen werden im folgenden Unterabschnitt normative Anforderung an die Genauigkeit von Verkehrswerten durch die Rechtsprechung aufgeführt.

2.1.3. Normative Anforderungen an Genauigkeiten in der Verkehrswertermittlung

Die Beurteilung von Verkehrswerten in Bezug auf ihre Übereinstimmung mit dem tatsächlichen Marktgeschehen erfolgt über den Vergleich mit den Kaufpreisen. Eine besondere Rolle spielen normative Vorgaben: Da die Verkehrswertdefinition nach BauGB keine Grenzen vorgibt, innerhalb derer sich ein Verkehrswert bewegen darf, erfolgt i. d. R. eine Auslegung durch Gerichtsurteile sowie durch Auslegung und Konventionen in der Fachwelt.

In Unterabschnitt 2.1.2.1 wurde die materiell-rechtliche Definition des Verkehrswertes nach § 194 BauGB erläutert: Demnach ist der Verkehrswert begrifflich als der Wert aufzufassen, der im gewöhnlichen Geschäftsverkehr am wahrscheinlichsten zu erzielen wäre. Mit dieser Definition ist der Verkehrswert als ein *punktuellder Wert* aufzufassen, sodass das Ergebnis einer Verkehrswertermittlung stets als (hypothetischer) einzelner Schätzwert angegeben wird. Durch die unbestimmten Rechtsbegriffe im Rahmen der Verkehrswertdefinition wird jedoch deutlich, dass ein ermittelter Wert insbesondere von der Ermittlungsmethode und den Einschätzungen des jeweiligen Sachverständigen abhängt und sich letztlich innerhalb einer gewissen Variationsbreite bzw. Preisspanne bewegt. Ein wahrer Wert im Sinne rein statistischer Betrachtungsweise existiert nicht. Voneinander unabhängige Gutachter werden bei der Bewertung des gleichen Objekts regelmäßig zu einer abweichenden Einschätzung kommen, was den unterschiedlichen Erfahrungen, Ansichten und Auslegungen geschuldet ist. Aus diesem Grund kann keine Methodik nach den Vorgaben der ImmoWertV für sich in Anspruch nehmen, mit absoluter Gewissheit zu einem wahren Verkehrswert zu führen. Vielmehr geht es nach KLEIBER darum, den Verkehrswert so genau wie möglich unter Ausschöpfung aller zugänglicher Marktindikatoren – Kenngrößen, Einflussgrößen sowie sachverständigen Einschätzungen – zu ermitteln. Da der Verkehrswert gemäß Definition keine Spanne darstellt, sondern im Grunde einen „spitz“ berechneten Wert fordert, leitet KLEIBER (2014, Teil IV, Rn. 70) die Angabe von Verkehrswertspannen als unbrauchbar ab, auch wenn naturgemäß die Kaufpreise unterschiedlicher Vertragspartner streuen.

Basierend auf einem Urteil des BVerwG (4 C 56.76 vom 24.11.1978) im Zusammenhang mit Verkehrswertermittlung bei der Preisprüfung in städtebaulichen Fördermaßnahmen wirft SEELE (1982, S. 105) hingegen die grundsätzliche Frage auf, ob sich der Verkehrswert überhaupt als ein einziger, bestimmter (wenn auch fiktiver) Preis ermitteln lässt oder ob er ausschließlich als Preisspanne angegeben werden kann. Dabei kommt er zu dem Schluss, dass sich sowohl Verkehrswert als auch eine Preisspanne auf Grund der Unsicherheiten im Bewertungsprozess nicht genau angeben lassen und die rechtsstaatlichen Forderungen lediglich darauf abzielen können, die Ungewissheit eines ermittelten Wertes auf das unvermeidbare Maß zu beschränken. Dies ist nur unter den Aspekten einer zutreffenden qualitativen Einschätzung des Wertermittlungsobjekts, der Begründung der Wertermittlungsgrundlagen, der sorgfältigen Methodenauswahl und schlussendlich einer sachgerechten Einschätzung der Verfahrensergebnisse zu erreichen. Unter diesen Bedingungen sind die unvermeidbaren, verbleibenden Ungewissheiten als zumutbar anzusehen (SEELE, 1982, S. 110 f.).

Die Variationsbreite von Verkehrswerten ist auch durch die Rechtsprechung anerkannt und in diversen Urteilen behandelt worden. So hat das BVerwG bereits 1978 in einem Urteil ausgeführt (Urteil vom 24.11.78 (NJW 1979, S. 2578): „Werte sind in dem Sinne ungewiß – mehr oder weniger – daß sie sich nicht einfach „ausrechnen“ oder in ihrer Höhe einer Tabelle entnehmen lassen, sondern aus einem Ermittlungsverfahren hervorgehen, das zumindest praktisch vielfältig Gelegenheit bietet, so oder anders vorzugehen. Das ist insbesondere für den Verkehrswert offenkundig. Ganz abgesehen davon, daß einzelne Merkmale des § 141 Abs. 2 Bundesbaugesetz (BBauG) (gewöhnlicher Geschäftsverkehr, Eigenschaften, sonstige Beschaffenheit, Lage, ungewöhnliche Verhältnisse, persönliche Verhältnisse) Schwierigkeiten bereiten können, ist von Bedeutung vor allem, daß das Gesetz ausschlaggebend abstellt auf eine Hypothese zum Ablauf des gewöhnlichen Geschäftsverkehrs.“

Das vorstehende Urteil kann nach HILDEBRANDT (1999, S. 321) dahingehend interpretiert werden, dass der Verkehrswert zum einen durch den Bezug auf die Hypothese des gewöhnlichen Geschäftsverkehrs rechtstheoretisch

einen punktuellen Wert darstellt, sich zum anderen aber durch die Einflüsse der Akteure am Immobilienmarkt sowie durch die unterschiedliche Einschätzung von wertrelevanten Merkmalen durch Sachverständige nur innerhalb einer Variationsbreite bewegen kann. Weiterhin schließt HILDEBRANDT (1999, S. 324) aus dem Urteil, dass das BVerwG einerseits zwar die Unsicherheit bei der Verkehrswertermittlung beachtet wissen will, andererseits aber nicht die Legaldefinition des Verkehrswertes als punktuellen Wert durch einen Begriff wie Wert- oder Preisspanne ersetzen möchte (HILDEBRANDT (1999) nutzt dabei ursprünglich den Begriff der Ungewissheit, der jedoch für die Stringenz dieser Arbeit mit dem Begriff der Unsicherheit aufgefasst werden soll).

Die Auffassung des BVerwG wird durch weitere Urteile bestätigt. KLEIBER (2014, Teil III, Rn. 72) nennt eine ungefähre Genauigkeit von Verkehrswerten innerhalb einer Spanne von $\pm 20 - 30 \%$, die in der Rechtsprechung (unter der Voraussetzung des gewöhnlichen Geschäftsverkehrs und Sicherheit über künftige Nutzungsverhältnisse) gemeinhin akzeptiert wird. Die Toleranzgrenze darf jedoch keinesfalls als starr aufgefasst werden, sondern richtet sich nach dem Einzelfall und dem jeweiligen Streitgegenstand. Eine fester Prozentsatz für die Beurteilung der Genauigkeit existiert demnach nicht. Stellvertretend werden relevante Gerichtsurteile genannt, die die zum Teil stark voneinander differierenden Toleranzgrenzen verdeutlichen sollen:

- LG Hamburg, Urt. vom 31.10.1960 - 10 O 30/60 -; BBauBl. 1961, 376 = EzGuG 18.15
Erfahrene Sachverständige haben dem Gericht immer wieder bestätigt, dass Abweichungen in der Bewertungshöhe um 10 % nach oben und nach unten auch bei sorgfältigen Feststellungen und Bemühungen gar nicht zu vermeiden wären. Häufig ist der Ungenauigkeitsbereich noch größer.
- BGH, Urt. vom 28.6.1966 - VI ZR 287/64 -; WM 1966, 1150 = EzGuG 11.51
Eine Abweichung von 78 % ist nach Ansicht des Gerichts „grob falsch“. Das Gericht führt aus: „Der Revision kann nicht zugegeben werden, dass sich der Unterschied im Rahmen des Vertretbaren halte. Der Sachverständige hat ausgeführt, bei Schätzungen von Altbau-Grundstücken, deren Gebäude kriegsbeschädigt und wie hier teilweise mit alten Baustoffen wiederhergestellt sind, könne nur ein Genauigkeitsgrad von ± 20 Prozent bis 30 Prozent erwartet werden.“
- BGH, Urt. vom 1.4.1987 - IVa ZR 139/85 -, EzGuG 11.163k
Der Revision ist darin zuzustimmen, dass die Erheblichkeit oder Unerheblichkeit einer Schätzungsabweichung nicht schematisch nach dem Prozentsatz der Schadensabweichung beurteilt werden darf. (...) Bei gutachterlichen Wertermittlungen und Schätzungen liegt indessen – wie bereits das BG ausführt – die Möglichkeit eines gewissen Spielraums, eines Streubereichs in der Natur der Sache. Daher ist erst dann von der Unverbindlichkeit einer Sachverständigenfeststellung auszugehen, wenn die Feststellung erheblich außerhalb des an sich üblichen Toleranzbereichs entsprechender Schätzungen liegt. Das wird bei Abweichungen in einer Größenordnung von unter 15 % (Berechnungsmethode: geforderter höherer Betrag mit 100 % gleichgesetzt) regelmäßig zu verneinen sein.
- BGH, Urt. vom 26.4.1991 - V ZR 61/90 -, GuG 1992, 165 = EzGuG 7.114
Zwischen dem vom Schiedsgutachter ermittelten jährlichen Erbbauzins von 12.320 € und der im Berufungsurteil festgestellten Höhe von 10.250 € besteht eine Differenz von 2.070 €, also von 16,79 %. Dazu verweist die Revision auf eine im Schrifttum vertretene Auffassung, wonach Abweichungen bis zu 25 % im Allgemeinen hinzunehmen seien.
- OLG Brandenburg, Urt. vom 10.06.2008 - 11 U 170/05 -; NJOZ 2008, 3324 = EzGuG 11.644a
Ein Verkehrswertgutachten ist falsch, wenn es um ca. 50 % von dem wirklichen Verkehrswert abweicht.

Eine Auflistung weiterer Urteile kann insbesondere HAACK (2006) sowie KLEIBER (2014) entnommen werden. Deutlich wird an den genannten Auszügen, dass neben der punktuellen Bestimmung des Verkehrswertes eine ergänzende *Variationsbreite* unerlässlich ist. Zwar kann die Definition einer Preisspanne den punktuellen Verkehrswert auf Grund seiner hervorgehobenen Bedeutung im Rechts- und Wirtschaftsleben als solche nicht ersetzen, jedoch wichtige Anhaltspunkte für die Beurteilung der (Schätz)Genauigkeit liefern. Dabei ist zu differenzieren, dass die in den Gerichtsurteilen genannten Preisspannen nicht ausschließlich auf die Methodik der Wertermittlung zurückzuführen sind, sondern ebenso in dem zu ermittelnden Verkehrswert selbst begründet sind. Nach KLEIBER (2014)

liege die Unsicherheit nicht nur in der natürlichen Streuung von Vergleichspreisen begründet, sondern auch in den in die Verkehrswertermittlung eingehenden Parametern, die seiner Auffassung nach als das Resultat der allgemeinen im gewöhnlichen Geschäftsverkehr hinzunehmenden aleatorischen Preismechanismen seien. Für die letztliche Schätzung des Verkehrswertes bedeutet dies nach Auffassung von Kleiber, dass im Streitfall der Wert anzusetzen sei, der die größte Gewähr der Richtigkeit biete.

Um den Anforderungen an den Verkehrswert nach BauGB unter Berücksichtigung der Charakteristika von Immobilien und den Anforderungen nach der Rechtsprechung gerecht zu werden, sind insbesondere durch die ImmoWertV Modelle gegeben, die den Grundstücksmarkt geeignet abbilden sollen. Aufgrund der bisherigen Ausführungen ist ohne eine tiefergehende Definition des Modellbegriffs bereits zu schließen, dass eine Abbildung der Verhältnisse des Immobilienmarktes derart vorgenommen werden soll, dass eine möglichst hohe Übereinstimmung zwischen Kaufpreis und Verkehrswert erreicht wird. Im folgenden Unterabschnitt werden daher zunächst die theoretischen Grundlagen von Modellbildungen und den nicht vermeidbaren Unsicherheiten in der Approximation der Realität erläutert.

2.2. Modelle als Abbild der Realität und ihre Unsicherheit

Die Ermittlung von Verkehrswerten erfolgt i. d. R. durch die Formulierung theoretischer Modelle, die vorrangig durch die ImmoWertV und die ergänzenden Richtlinien vorgegeben werden. Grundprinzip der Modelle ist es, die wesentlichen wertbestimmenden Mechanismen am Immobilienmarkt geeignet abzubilden. Damit kommt dem Modell zum einen die Aufgabe zu, die Komplexität des Zusammenspiels wertbeeinflussender Umstände nach Abschnitt 2.1.1 auf einige wenige Determinanten zu reduzieren. Zum anderen soll trotz der notwendigen Reduktion eine bestmögliche Approximation zwischen der Theorie und der Realität erreicht werden, was sich durch eine möglichst hohe Übereinstimmung zwischen Kaufpreis und ermitteltem Verkehrswert interpretieren lässt. Von besonderer Bedeutung für den Forschungsansatz der Arbeit ist die Tatsache, dass trotz einer geeigneten und zutreffenden Modellformulierung stets eine Unsicherheit über den nicht zu vermeidenden Unterschied zwischen Realität und ihrer Abstraktion verbleibt. Die Unsicherheit resultiert zum einen aus der inhärenten Unsicherheit der Beobachtungen, zum anderen aus dem gewählten Modell als unvollkommenem Abbild der Realität. Diese ist insbesondere geprägt durch die starke Heterogenität des Immobilienmarktes, bedingt durch eine Subjektivität, die sich oftmals nicht logisch erklären lässt. Als zentrale Aspekte sind zunächst die Begriffe des Modells und der Unsicherheit zu definieren. Als Grundlage für die weiteren Ausführungen erfolgt in Abschnitt 2.2.1 zunächst eine Definition des Modellbegriffs und dessen Übertragung auf die Verkehrswertermittlung. In Abschnitt 2.2.2 wird der Begriff der Unsicherheit auf Grundlage von Normen und Standards behandelt.

2.2.1. Mathematische Modellbildung und deren Qualität

Wie bereits angedeutet, werden für die Ermittlung von Verkehrswerten Modelle benötigt, die den Immobilienmarkt geeignet abbilden. Um den Grad der Approximation sowie die im Bewertungsprozess auftretenden Unsicherheiten beurteilen zu können, werden in den folgenden Abschnitten die Theorie der Modellbildung allgemein und im Bezug auf die Verkehrswertermittlung aufgezeigt und die sich daraus ergebenden Rahmenbedingungen für die Qualität der mathematischen Modellierung erläutert.

2.2.1.1. Definition des Modellbegriffs

Als ein erster Zugang zum Modellbegriff kann eine allgemeingültige Definition herangezogen werden (BROCKHAUS, 2005): Demnach ist ein Modell „*ein Abbild bestimmter, der empirischen Erfahrung durch Beobachtung oder Experiment zugänglicher Aspekte der unbelebten oder belebten Natur (...), bei dem die für wesentlich erachteten Eigenschaften hervorgehoben und die als nebensächlich angesehenen Attribute außer Acht gelassen werden.*“ Ein Modell dient in diesem Sinne als Mittel zur Beschreibung und Interpretation eines realen Sachverhaltes und zur Begriffsbildung durch Idealisierung und Abstraktion der Wirklichkeit. Das Ziel, eine bestmögliche Approximation der Realität durch Modelle zu erreichen, findet sich neben Fragestellungen der Verkehrswertermittlung in zahlreichen weiteren Fachdisziplinen. Eher allgemein gehaltene Erläuterungen geben u. a. ROMMELFANGER (1994) sowie BIE-

WER (1997). Eine ausführliche Diskussion des allgemeinen Modellbegriffs lässt sich KUTTERER (2002) entnehmen, der neben grundsätzlichen Erläuterungen explizite Modelldefinitionen aus unterschiedlichen Fachdisziplinen aufgreift. Für den Rahmen der Arbeit soll in Verbindung mit den Ausführungen in Abschnitt 2.1.1 eine ökonometrische Sichtweise in den Mittelpunkt gerückt werden.

Die Fachdisziplin der Ökonometrie definiert sich als Teilgebiet der Wirtschaftswissenschaften und stellt sich als Synthese aus ökonomischer Theorie, mathematischen Methoden sowie statistischen Daten dar (STÖWE, 1977, S. 1). Nach HÜBLER (1989, S. 12) entspricht der Modellbegriff der Ökonometrie ökonomischen Modellbeziehungen, die in mathematischer Sprache definiert sind. So besteht die Hauptaufgabe darin, Regelmäßigkeiten in den Beziehungen zwischen ökonomischen Variablen – den Daten im STÖWE'schen Sinne – aufzudecken. Weiterhin verweist er darauf, dass die gefundenen Beziehungen zwischen den Variablen keine Gesetzmäßigkeiten im naturwissenschaftlichen Sinne sein können, da ökonomische Prozesse durch menschliche Verhaltensweisen geprägt werden. Das gewählte Modell soll daher nicht nur die Identifikation möglichst weniger Determinanten einer reduzierten Realität ermöglichen, sondern im Hinblick auf die Daten auch den Anforderungen der Objektivität (Bestimmung eines Merkmals frei von subjektiver Wahrnehmung), Reliabilität (Erfassung unter gleichen Randbedingungen führt zum gleichen Ergebnis) und Validität (Erhebung der Daten, die der zu Grunde liegenden Theorie entsprechen) genügen. Abbildung 2.3 zeigt die Grundstruktur ökonometrischer Modelle mit deren wesentlichen Komponenten basierend auf Definitionen nach HÜBLER (1989, S.12 ff.) und erweitert um kontextrelevante Zusammenhänge.

Durch einen gegebenen *Modellrahmen* wird festgelegt, über welche Daten bzw. welches Objekt Aussagen getroffen und mit dem Modell erklärt werden sollen. Die Bedeutung einzelner Eingangsgrößen kann dabei nicht a priori festgelegt werden, vielmehr ist eine iterative Prüfung auf deren Bedeutung im endgültigen Modell notwendig. Als wesentliche Eingangsgrößen sind demnach die *Beobachtungen* festzuhalten, die mehr oder weniger direkt in der Realität erfasst werden und die den Zustand des Objekts von Interesse beschreiben. Als Beobachtungen sollen dabei im Sinne der Erläuterungen von BANDEMER (2005, S. 19) sämtliche Größen verstanden werden, die eine Aussage über ein zu betrachtendes Objekt ermöglichen und als begriffliche Äquivalenz zu den Daten nach STÖWE (1977) sowie den Variablen nach HÜBLER (1989) aufzufassen sind.

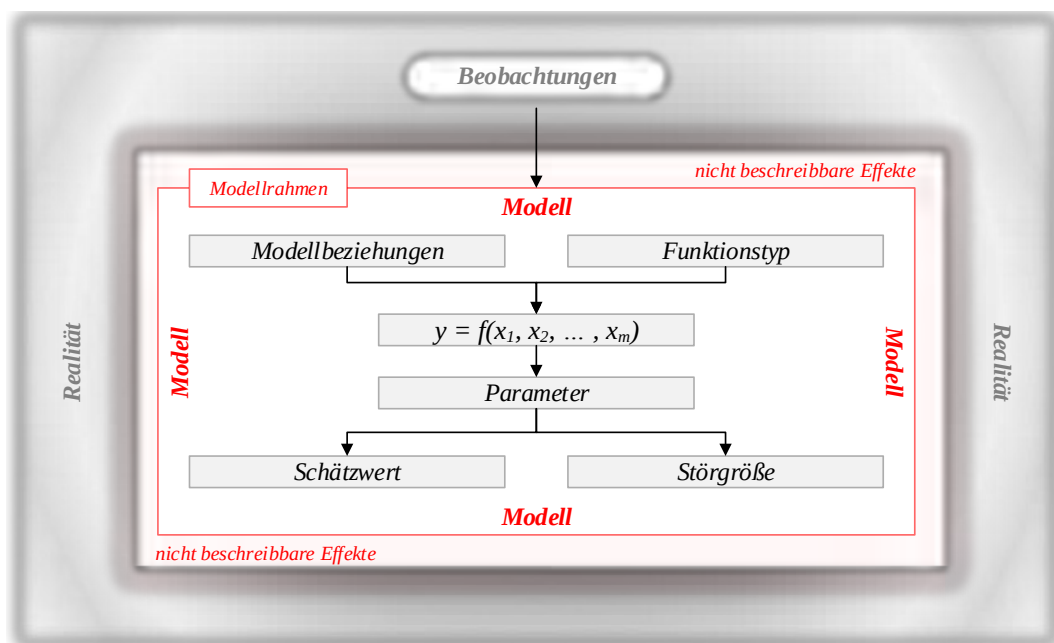


Abbildung 2.3: Grundstruktur ökonometrischer Modelle

Gemäß der Bezeichnungen in Abbildung 2.3 fallen unter diesen Begriff die Variablen y und $x_1, \dots, x_m$. Sind die Beobachtungen erfasst, erfolgt die Definition von *Modellbeziehungen*. Zielsetzung von deren Ableitung ist es, Zusammenhänge zwischen den Beobachtungen aufzuzeigen, die nicht nur für den konkreten Einzelfall gelten, sondern vielmehr einen gewissen Grad an Stabilität im Zeitverlauf aufweisen bzw. für verschiedene Realisierungen der Beob-

achtungen angesetzt werden können. Erforderlich sind hier Sachverstand und Erfahrung des Auswerterers bezüglich der zu Grunde liegenden Problemstellung.

Die *Parameter* eines Modells stellen die Verbindung zwischen den beobachteten Größen her und sind so als Konstanten zu betrachten. Als wesentlicher Unterschied zu den variablen Eingangsgrößen sind die Parameter i. d. R. nicht beobachtbar, sodass im Vorfeld der Datenanalyse Annahmen getroffen werden. Parameter sind aus den Beobachtungen zu schätzen und im Anschluss auf Verträglichkeit mit der zuvor formulierten Hypothese zu prüfen. Die Unsicherheit der Beobachtungen und die Unsicherheit über eine adäquate Modellformulierung drückt sich u. a. durch die Streuung von Modellparametern aus. Für die Schätzung der unbekannten Parameter ist daher ein geeigneter mathematischer Ansatz zu wählen. Zudem ist die Festlegung eines *Funktionstyps* erforderlich. Hier lassen sich im Wesentlichen lineare und nicht-lineare Typen unterscheiden, wobei erstgenannte als Regelfall zu betrachten sind (HÜBLER, 1989, S. 12 ff.).

Sind die unbekannten Parameter ermittelt, können auf Grundlage der Beobachtungen *Schätzwerte* berechnet werden. Liegt zudem eine geeignete mathematische Formulierung mit einem entsprechenden Modellhintergrund wie der Stochastik vor, können Maße für die Wahrscheinlichkeit der Schätzwerte angegeben werden. Die Schätzwerte sind eines der wesentlichen Ergebnisse der Datenanalyse und sollen eine möglichst hohe Approximation in Bezug auf die ursprünglichen Beobachtungen aufweisen.

Die Realisierungen von Beobachtungen orientieren sich nicht an Modellen, sondern sind Ergebnis realer Prozesse. Die Diskrepanz zwischen theoretischen Modellen und der Realität wird insbesondere in der *Störgröße* als Übergang von Theorie zu Modell aufgefangen. Wie in Abbildung 2.3 verdeutlicht wird, kann eine vollkommene Übereinstimmung mit der Realität nicht erreicht werden. Zum einen liegt dies in nicht beschreibbaren Effekten begründet, die einer rein funktionalen Modellierung in einem mathematischen Modell nicht zugänglich sind. Zum anderen pflanzen sich sowohl Beobachtungs- als auch Modellunsicherheiten im Rahmen der Datenanalyse auf das Schätzergebnis fort, sodass das Ergebnis stets mit Unsicherheiten behaftet sein wird. HÜBLER (1989, S. 7) verdeutlicht, dass die Begriffe Modell und Beobachtungen gedanklich nur schwer zu trennen sind: Die Modellspezifikation gibt an, welche Daten benötigt werden. Die Verfügbarkeit der Daten hingegen beeinflusst die Spezifikation des Modells. Aufgrund der Interdependenzen stellt sich die Datenanalyse als komplexer Prozess dar, in deren Verlauf u. U. eine iterative Betrachtung notwendig ist. Die Erforderlichkeit der erneuten Auswertung vorheriger Stufen bei Vorliegen neuer Erkenntnisse ist möglich und zumeist notwendig (KUTTERER, 2002, S. 15). Auf diese Weise soll den genannten Gründen für die notwendige Berücksichtigung einer Störgröße möglichst limitierend entgegengewirkt werden.

Nach der erläuterten Definition eines Modells aus ökonometrischer Sichtweise lässt sich im Hinblick auf Abschnitt 2.1.1 eine enge Verbindung zwischen ökonometrischen Modellen und Fragestellungen der Verkehrswertermittlung herstellen: Die Verknüpfung ergibt sich durch die bereits eingeführte Betrachtungsweise der Immobilie als Wirtschaftsgut, welches speziellen marktwirtschaftlichen Gesetzmäßigkeiten unterworfen ist. Der Modellrahmen legt dabei fest, welcher räumliche und sachliche Teilmarkt untersucht werden soll. Implizit sind damit auch die Einflussgrößen verbunden, die in den Modellansatz eingebracht werden. Für jede Immobilienart existiert aus bisherigen Erfahrungen Vorwissen über die vermuteten wertrelevanten Merkmale. Als Beobachtung wird regelmäßig der Kaufpreis genutzt, der als Zielgröße der Verkehrswertermittlung aufgefasst werden kann. Die Modellbeziehungen zwischen Kaufpreisen und wertrelevanten Merkmalen ergeben sich durch die unterstellten Kausalzusammenhänge – ebenfalls generiert aus sachverständigem Vorwissen. Unter diesen Voraussetzungen lassen sich gemäß der durch die rechtlichen Rahmenbedingungen vorgegeben Modelle der Verkehrswertermittlung die Funktionstypen und Parameter für die Modellschätzung festlegen. Die Festlegungen variieren in Abhängigkeit von der zu bewertenden Immobilienart sowie der Datenverfügbarkeit und können ausschließlich für den Einzelfall konkretisiert werden.

Bedingt durch diese Sichtweise und durch die Ausführungen in Abschnitt 2.1.1 wird deutlich, dass sich die Ermittlung eines Verkehrswertes mitunter als sehr komplexer Vorgang darstellen kann. Neben der Tatsache, dass es trotz der Legaldefinition des Verkehrswertes nach § 194 BauGB aus den genannten Gründen keinen allgemeingültigen, einzigen Wert für eine Immobilie geben kann, ist für die Modellbildung und dabei insbesondere die Auswahl der zutreffenden Daten ein hohes Maß an Sachverstand und subjektiver Einschätzung notwendig. Der folgende Abschnitt erläutert die im Kontext der Modellformulierung auftretenden Unsicherheiten im Hinblick auf Abweichungen zwischen Modell und Realität.

2.2.1.2. Qualität der mathematischen Modellierung

Die Qualität einer mathematischen Modellierung spiegelt sich im Grad der Annäherung zwischen dem ursprünglich formulierten Ziel und dem Modellergebnis sowie in der Güte der Modellierung wieder. Für unterschiedliche Aufgabenstellungen stellt die Mathematik dabei Werkzeuge zur Verfügung, die problembezogen und sorgfältig ausgewählt und dem formulierten Ziel und der anvisierten Modellgüte gerecht werden müssen. Die Modellierung gliedert sich in drei Bestandteile: Die Wahl eines mathematischen Modells, die eines mathematischen Lösungsverfahrens sowie die Beschreibung einer konkreten Situation durch mathematisch spezifizierte Daten. Für die Qualität der mathematischen Modellierung gilt in der Folge das Gesetz vom schwächsten Kettenglied: *„Das Ergebnis ist im konkreten Fall nur höchstens so gut in seiner Aussagekraft (relevant, zuverlässig, zutreffend, genau), wie die schwächste Qualität der drei beteiligten Kettenglieder“* (BANDEMER, 1997, S. 14 f.).

Die *Qualität des Modells* wird durch dessen Adäquatheit erfasst. Nach BANDEMER (1997, S. 14) bezeichnet die Adäquatheit die hinreichende Übereinstimmung des Modells mit der Wirklichkeit. Da es für ein reales Problem keine abschließende Möglichkeit an zutreffenden Modellen für dessen Approximation gibt, ist die Adäquatheit entscheidend von der Formulierung von Ziel und Zweck der Untersuchung abhängig. Einflussfaktoren sind ferner die unterschiedliche Kompliziertheit, mathematische und numerische Handhabbarkeit sowie praktische Durchschaubarkeit und Deutbarkeit der zur Auswahl stehenden Modelle. Als weitere Qualitätskriterien können darüber hinaus auch Korrektheit, Vollständigkeit, Effizienz oder Benutzerfreundlichkeit aufgefasst werden (KRUSE ET AL., 1995, S. 2). Die Beurteilung der Adäquatheit obliegt letztlich dem Anwender, eine Einschätzung durch Mittel der Mathematik ist prinzipiell unmöglich (BANDEMER, 1997, S. 15).

Die *Qualität des Lösungsverfahrens* bezeichnet die Eigenschaft, die Lösung eines Problems mit ausreichender Genauigkeit unter einem vertretbaren Aufwand zu liefern. Während die Genauigkeit als einer der zentralen Punkte der Arbeit im folgenden Abschnitt inhaltlich dargelegt wird, ist der vertretbare Aufwand wiederum durch den Anwender zu beurteilen und beeinflusst damit implizit auch die Qualität des Modells. Die Einschätzung der Qualität erfolgt über die Deutung des Ergebnisses. Insbesondere in der mathematischen Statistik werden die verwendeten Schätzmethoden nach ihrer Erwartungstreue und Varianz bewertet.

Die *Qualität der Daten* wird daran gemessen, wie verlässlich sie ein vorliegendes, konkretes Problem beschreiben (BANDEMER, 1997, S. 14 ff.). Neben der grundlegenden Voraussetzung, dass es sich um die richtigen, zutreffenden Daten zur Problembeschreibung handelt, ist insbesondere eine Beurteilung hinsichtlich grober, systematischer oder zufälliger Fehler erforderlich, denn gerade die Beschaffenheit der Daten legt den zu verwendeten Modellhintergrund – das mathematische Lösungsverfahren – für die Behandlung von Unsicherheiten im Bewertungsprozess nahe (KUTTERER, 2002, S. 15). Dabei beruhen grobe Fehler auf Irrtümern in der Datenerfassung, bspw. Zahlendrehern oder falschen Übernahmen von Eintragungen in eine Datenbank. Aus Sicht der Verkehrswertermittlung seien falsche Einträge in der AKS bei der Übernahme von Informationen aus Kaufverträgen oder der örtlichen Erfassung genannt. Grobe Fehler können durch genügende Aufmerksamkeit und sorgfältige Auswertungen i. d. R. vermieden werden; das Auftreten grober Fehler wird für den Kontext der Arbeit daher ausgeschlossen und nicht weiter betrachtet. Entscheidend für die weitere Datenanalyse ist nach BANDEMER (2005, S. 18 ff.) insbesondere die Qualität der Beobachtungen, die daran gemessen wird, wie verlässlich ein vorliegendes Problem konkretisiert wird. Von besonderer Bedeutung ist dabei, dass bei der Spezifizierung der Beobachtungen, wenn auch teilweise unbewusst, meist eine Genauigkeit der einzelnen Daten eingeführt wird. Die Tatsache der nicht exakt zu beobachtenden Realität drückt sich in Unsicherheit über das beobachtete Objekt aus. Die Modellwahl und Validierung ist daher nur auf der Grundlage einer realistischen Unsicherheitsbeschreibung der Beobachtungen erfolgversprechend, wohingegen eine reine Anpassung eines Modells an die Beobachtungen ohne Berücksichtigung von Unsicherheiten zu kurz greift (NEUMANN, 2009, S. 16). Der Begriff der Unsicherheit wird im folgenden Unterabschnitt aufbereitet.

2.2.2. Definition des Unsicherheitsbegriffs

Die Definition der Unsicherheit wird in Anlehnung an die Ausführungen von BANDEMER (1997), KUTTERER (2002), SCHÖN (2003), BANDEMER (2005) sowie NEUMANN (2009) gewählt. Insbesondere bei KUTTERER (2002) findet sich eine ausführliche Diskussion unterschiedlicher Begriffsauffassungen für Unsicherheit mit der Ungewissheit als abstrakterem Oberbegriff, die einer breit gefächerten Anwendung in unterschiedlichen Fachdisziplinen geschuldet

ist. Die allgemeinen Definitionen sollen an dieser Stelle nicht vorgestellt werden, für eine vertiefte Behandlung der Thematik wird auf die genannten Literaturstellen verwiesen. Im Vordergrund der folgenden Ausführungen steht die Definition der Unsicherheit, die sich insbesondere in ingenieurgeodätischen Anwendungen durchgesetzt hat.

Unsicherheiten treten regelmäßig dann auf, wenn erhobene Daten mit theoretischen Modellen zusammengeführt werden: Das Modell erklärt zwar einen gewissen Anteil der Daten, unerklärbare Anteile lassen sich jedoch zum einen auf Unsicherheiten in der Datengenese selbst zurückführen, zum anderen auf die Unsicherheit durch die Unvollkommenheit eines Modells. Eine strenge Trennung in Daten- und Modellunsicherheit ist jedoch nicht immer möglich (KUTTERER, 2002, S. 16). Die Begründung lässt sich nach NEUMANN (2009, S. 15 f.) darin finden, dass zum einen Modelle und deren Modellparameter aus Daten abgeleitet werden, zum anderen häufig feste Modellparameter in Modelle eingeführt werden, die zuvor selbst aus Daten abgeleitet wurden. Eine gemeinsame Behandlung der Unsicherheit der Daten und der Unsicherheiten des Modells sei daher naheliegend. In diesem Sinne kann die Beurteilung, ob die aus Daten gewonnen Informationen als gesichert gelten – bzw. für welche Aussagen sie sich eignen – lediglich dann erfolgen, wenn eine zutreffende Modellierung von auftretenden Unsicherheiten in den Daten, dem zugrunde liegenden Objekt sowie den angesetzten Modellen mit deren Modellparametern erfolgt. Im Rahmen der Datenanalyse ist daher die allgemeine Betrachtung von Unsicherheiten von entscheidender Bedeutung für die Güte der Modellierung.

Für die Beschreibung von Unsicherheiten sollen im Folgenden allgemein anerkannte Normen und Standards herangezogen werden (Deutsches Institut für Normung (DIN)). Von besonderer Bedeutung sind *DIN 1319* und *DIN 55350*, die sich mit den grundlegenden Definitionen von Unsicherheitsbegriffen befassen. In Ergänzung beziehen sich *DIN 18709* und *DIN 18710* auf eine Auslegung für die speziellen Anforderungen rein geodätischer Anwendungen. Als weitere wesentliche Quelle ist die *DIN V ENV 13005* zu nennen, die sich als international anerkannte Richtlinie *Guide to the Expression of Uncertainty in Measurement (GUM)* auch im deutschsprachigen Bereich unter dem Titel *Leitfaden zur Angabe der Unsicherheit beim Messen* durchgesetzt hat (NEUMANN, 2009, S. 17). Eine erste Definition nach GUM (2008, S. 38 f.) führt u. a. die Begriffe Messgröße, Einflussgröße sowie Messabweichung ein: Demnach handelt es sich bei der *Messgröße* um eine spezielle Größe, die Gegenstand einer Messung ist. Ziel ist die Ermittlung eines *Messergebnisses* als Realisierung der *Messgröße* durch Erfassung unter zusätzlicher Berücksichtigung der *Einflussgrößen*. Diese sind definiert als Größen, die keine Messgrößen sind, jedoch das Messergebnis beeinflussen. Wird der Grundgedanke auf die Verkehrswertermittlung übertragen, ist die Messgröße i. d. R. als Kaufpreis aufzufassen, zu dem ergänzend die Einflussgrößen in Form der vermuteten wertbeeinflussenden Merkmale erfasst werden. Die Unsicherheit eines Messergebnisses spiegelt sich durch eine unzureichende Kenntnis des Wertes der Messgröße wieder; übertragen auf die Verkehrswertermittlung bedeutet dies, dass ein einzelner Kaufpreis eine mögliche Realisierung aus einer unendlichen Menge darstellt. Weiterhin wird in GUM (2008, Nr. 3.3) ausgeführt, dass die Unsicherheit des Messergebnisses sich aus zufälligen Einflüssen und unvollkommener Berichtigung des Ergebnisses in Bezug auf systematische Einflüsse ergibt. Eine Unterscheidung ist durch die Angabe qualitativer und quantitativer Größen gegeben, die in den folgenden Unterabschnitten näher betrachtet werden.

2.2.2.1. Qualitative Maße

Gemäß DIN 55350 Teil 13 Nr. 1.3 bis 1.5 bezeichnet die *Genauigkeit* das Ausmaß der Annäherung von Ermittlungsergebnissen an einen Bezugswert, wobei dieser nach individueller Festlegung der wahre, der richtige oder der Erwartungswert sein kann. Der *wahre Wert* ist als tatsächlicher Merkmalswert zu verstehen, der oftmals einen rein ideellen Wert darstellt. Gemäß der zugrunde liegenden Norm ließe er sich nur dann feststellen, wenn sämtliche Abweichungen der Ermittlungsergebnisse vom wahren Wert vermieden werden könnten oder er sich aus rein theoretischen Überlegungen ergibt. Deshalb wird durch die Vorgaben des GUM (2008, S. 40) explizit darauf verwiesen, dass der wahre Wert nicht festgestellt werden kann. Insofern wird i. d. R. der *richtige Wert* als ein durch Vereinbarung anerkannter und einer speziellen Bezugsgröße zugeordneter Wert als Näherungswert für den wahren Wert verwendet. Die Abweichungen von wahren Wert und richtigem Wert werden als vernachlässigbar betrachtet. Der *Erwartungswert* mit dem Erwartungswertoperator $E(\cdot)$ spiegelt das mittlere Ermittlungsergebnis wider, welches aus der unablässig wiederholten Anwendung des unter vorgegebenen Bedingungen angewendeten Ermittlungsverfahrens gewonnen werden könnte und sich durch $E(y) = \bar{y}$ formulieren lässt. Hierbei bezeichnet y ein beliebiges

Ermittlungsergebnis und $\bar{y}$ den Erwartungswert der Beobachtung; im Kontext der Verkehrswertermittlung kann das Ermittlungsergebnis entsprechend als ein beliebiger Kaufpreis y_i angesehen werden. Der Erwartungswert bezeichnet dann den Wert, der sich theoretisch aus einer unablässig wiederholten Veräußerung des betrachteten Objektes ergeben würde.

Allgemein setzt sich die Genauigkeit nach DIN 55350 Teil 13 Nr. 2.1 aus zwei Unsicherheitskomponenten zusammen: Den systematischen Abweichungen und den zufälligen Abweichungen. Die *zufälligen Abweichungen* – im Folgenden ϵ_i – beziehen sich auf den Unterschied zwischen Erwartungswert $\bar{y}$ und Ermittlungsergebnis y_i . Gemäß der Definition nach DIN 1319-1 Nr. 3.5.1 ist die zufällige Abweichung nicht genau bekannt, ihr liegen nicht einseitig gerichtete Streuungen der Ermittlungsergebnisse um den Erwartungswert zugrunde. Sie können weder in ihrer Größenordnung noch in ihrem Vorzeichen vorhergesagt werden, sodass für den Erwartungswert $E(\epsilon_i) = E(y_i - \bar{y}) = 0$ gesetzt werden kann. In der Konsequenz sind alle theoretisch möglichen zufälligen Abweichungen über Zufallsvariablen und zugehörige Wahrscheinlichkeitsdichtefunktionen zu modellieren. *Systematische Abweichungen* Δ hingegen sind nach DIN 1319-1 Nr. 3.5.2 definiert als Unterschied zwischen Erwartungswert $\bar{y}$ und wahren Wert $\tilde{y}$. Wiederholte Ermittlungsergebnisse enthalten die systematische Abweichung als konstanten, gleichgerichteten Wert oder ändern sich in einer vorhersagbaren Art und Weise (NEUMANN, 2009, S. 18). In der Summe setzt sich die qualitativ definierte Genauigkeit eines beliebigen Wertes y_i als Ermittlungsergebnis aus den mittleren zufälligen und systematischen Abweichungen zu

$$y_i = \tilde{y} + \epsilon_i + \Delta \quad (2.1)$$

zusammen. Die systematische Abweichung Δ lässt sich wiederum zerlegen in einen bekannten Anteil Δ_b und einen unbekannten Anteil Δ_u , sodass $\Delta = \Delta_b + \Delta_u$ gilt. Während die zufälligen Abweichungen nicht reduzierbar oder zu vermeiden sind, müssen die Ermittlungsergebnisse um sämtliche bekannte systematische Abweichungen korrigiert werden (vgl. z. B. DIN 18710 Teil 1 A.2.3). Daraus ergibt sich als Forderung für die Genauigkeit in der Theorie

$$y_i^* = \tilde{y} + \epsilon_i + \Delta_u \quad (2.2)$$

mit y_i^* als berichtigtes Ermittlungsergebnis, welches keine bekannten systematischen Anteile mehr enthält (NEUMANN, 2009, S. 18 f.).

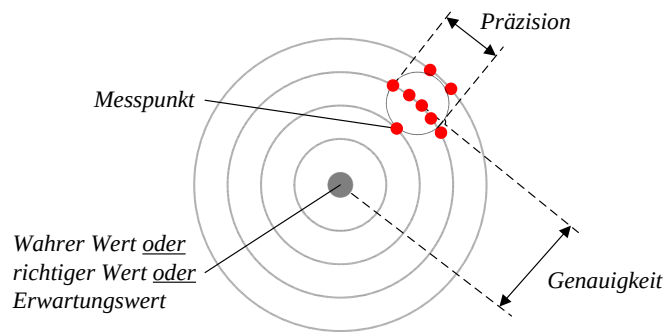


Abbildung 2.4: Prinzip von Genauigkeit und Präzision nach NEUMANN (2009, S. 18)

Neben der Genauigkeit beschreibt DIN 55350 Teil 13 Nr. 2.1.2 als weiteres qualitatives Maß die *Präzision*. Diese ist definiert als Ausmaß der gegenseitigen Annäherung voneinander unabhängiger Ermittlungsergebnisse unter mehrfacher Anwendung des Ermittlungsverfahrens unter den gegebenen Bedingungen, wobei diese sich weiter in Wiederhol- oder Vergleichsbedingungen untergliedern lassen. Zu beachten ist, dass die Definition der Präzision keinen Bezug zum wahren oder richtigen Wert enthält und daher Genauigkeit und Präzision streng voneinander zu trennen sind. Damit ist eine hohe Genauigkeit nicht gleichbedeutend mit einer hohen Präzision oder eine schlechte Genauigkeit mit einer geringen Präzision (NEUMANN, 2009, S. 18). Abbildung 2.4 zeigt schematisch den Unterschied zwischen Genauigkeit und Präzision.

Werden die Begrifflichkeiten auf die Verkehrswertermittlung übertragen, lässt sich die Genauigkeit als Differenz zwischen Kaufpreis und Verkehrswert auffassen, die den Einflüssen zufälliger Abweichungen ϵ_i und systematischen

Einflüssen unterliegt. Die zufälligen Abweichungen ergeben sich dabei durch die in Abschnitt 2.1.1 erläuterten Charakteristika des Immobilienmarktes und werden im weiteren Verlauf eingehender behandelt. Die systematischen Abweichungen werden, wie bereits angedeutet, im Weiteren ausgeschlossen. Die Präzision in der Verkehrswertermittlung lässt sich als die Streuung in der Schätzung von Verkehrswerten auffassen, die bspw. aus verschiedenen Einschätzungen eines konkreten Objektes durch unterschiedliche Sachverständige resultieren.

2.2.2.2. Quantitative Maße

Die im vorherigen Abschnitt dargestellten Begriffe sind rein qualitativ auszulegen und dürfen nicht mit den quantitativen Begriffen gleichgesetzt werden. Quantitative Begriffe sind geeignet, Genauigkeit und Präzision mit numerischen Angaben zu belegen. Für die Angabe eines quantitativen Maßes der Präzision werden nach DIN 55350 Teil 13 Nr. 3 *Standardabweichungen* herangezogen, die als positive Quadratwurzel der Varianz definiert sind (KOCH, 2004, S. 105):

$$\sigma_y = \sqrt{E((y_i - \bar{y}))^2} = \sqrt{E((\epsilon_i)^2)} . \quad (2.3)$$

Liegt eine ausreichende Anzahl von Beobachtungen y_i vor, lässt sich die empirische Standardabweichung angeben zu (DIN 1319-1 Nr. 3.9):

$$\hat{\sigma}_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} . \quad (2.4)$$

Als quantitativer Begriff für die Genauigkeit wird in DIN 55350 Teil 13 Nr. 4.1 die *Ergebnisunsicherheit* eingeführt. Demnach handelt sich um einen geschätzten Betrag zur Kennzeichnung eines Wertebereichs, innerhalb dessen der gewählte Bezugswert (wahrer Wert, richtiger Wert oder Erwartungswert) liegt. Die Ergebnisunsicherheit ist als Unterschied zwischen der oberen Grenze u_o bzw. unteren Grenze u_u eines Wertebereichs und den um systematische Abweichungen korrigierten Ermittlungsergebnis y_i^* aufzufassen. Gilt zudem $u_o = u_u = \sigma_u$ mit σ_u als Ergebnisunsicherheit, ergibt sich eine Weite der Ergebnisunsicherheit von $2\sigma_u$ und entsprechend für die Untergrenze des Wertebereichs $y_i^* - \sigma_u$ sowie für die Obergrenze $y_i^* + \sigma_u$. Die Ergebnisunsicherheit beruht, wie bereits durch den qualitativen Begriff der Genauigkeit deutlich wird, auf einer zufälligen und einer systematischen Komponente. In DIN 18710 Teil 1 A.4.3 wird eine quadratische Addition der jeweiligen Anteile vorgeschlagen, sodass

$$\sigma_u^2 = \sqrt{\sigma_y^2 + \sigma_{\Delta}^2} . \quad (2.5)$$

erhalten wird. Während die zufällige Komponente nach (2.3) ermittelt wird, kann die systematische Komponente nicht empirisch ermittelt werden, sondern muss als Varianz bzw. Standardabweichung einer Wahrscheinlichkeitsverteilung für die möglichen Werte der systematischen Abweichung abgeschätzt werden. Die Abschätzung basiert dabei entweder auf vorliegenden oder aus der Erfahrung ableitbaren Informationen über die Messgröße (KUTTERER, 2002, S. 18). Für den weiteren Verlauf der Arbeit soll die Standardabweichung als wesentliche Größe für die Angabe der Modellunsicherheit dienen und damit eine Aussage über die Präzision von geschätzten Verkehrswerten ermöglichen. Da systematische Abweichungen für die hier thematisierte Verkehrswertermittlung ausgeschlossen werden, ergibt sich die Ergebnisunsicherheit zudem nach (2.5) äquivalent zu (2.4).

Abschließend kann als Ergänzung der Konfidenzbereich für den Erwartungswert angegeben werden. Liegt die Schätzung eines Parameters oder einer Messgröße vor, gibt das Konfidenzintervall einen Bereich an, in dessen Intervallgrenzen der jeweilige Wert mit einer vorgegebenen Wahrscheinlichkeit liegt. Zu diesem Zweck ist die Irrtumswahrscheinlichkeit α geeignet zu wählen, sodass sich die Wahrscheinlichkeit P für die Überdeckung des Erwartungswert $\bar{y}$ durch die Intervallgrenzen zu $P = 1 - \alpha$ formulieren lässt (KOCH, 2004, S. 323). Unter Berücksichtigung der unteren und oberen Vertrauensgrenzen C_u bzw. C_o ergibt sich dann

$$P \{C_u \leq \bar{y} \leq C_o\} = 1 - \alpha . \quad (2.6)$$

NEUMANN (2009, S. 19) weist darauf hin, dass eine vertafelte Berechnung des Konfidenzintervalls nur dann möglich ist, wenn für die Wahrscheinlichkeitsdichtefunktion $p(y)$ eine Normalverteilung der Form $y \sim N(\bar{y}, \sigma_y^2)$ mit

$$p(y|\bar{y}, \sigma_y^2) = \frac{1}{\sigma_y \sqrt{2\pi}} e^{-(y-\bar{y})^2 / 2\sigma_y^2} \quad (2.7)$$

angenommen wird. Unter Berücksichtigung eines Schätzwertes $\hat{y}_i$ ergibt sich das Konfidenzintervall dann zu

$$P\{\hat{y}_i - k\sigma_y \leq \bar{y} \leq \hat{y}_i + k\sigma_y\} = 1 - \alpha, \quad (2.8)$$

wobei der Faktor k von der gewählten Irrtumswahrscheinlichkeit α abhängt und in Form von Standardwerten der DIN 18710 Teil 1 entnommen werden kann (NEUMANN, 2009, S. 20).

Die bisherigen Ausführungen erlauben die Beschreibung von Unsicherheiten in qualitativer wie quantitativer Hinsicht. Im folgenden Unterabschnitt werden die theoretischen Möglichkeiten betrachtet, Unsicherheiten durch mathematische Ansätze geeignet zu berücksichtigen.

2.2.2.3. Mathematische Ansätze für die Berücksichtigung von Unsicherheiten

Als Ergebnis der bisherigen Erläuterungen lässt sich zusammenfassen, dass die Qualität der mathematisch begründeten Modellierung realer Phänomene entscheidend durch die verwendeten Modelle und die Eingangsdaten als Beobachtungen definiert wird. Durch den lediglich approximativen Charakter sind die Schätzwerte der Modelle grundsätzlich mit Unsicherheiten behaftet, deren Nichtberücksichtigung im Auswerteprozess zu unzureichenden und gegebenenfalls (ggf.) verzerrten Ergebnissen führen kann. Dabei steigt das Problem einer adäquaten Berücksichtigung von Unsicherheiten mit der Zahl der unbekannten Parameter oder, allgemeiner formuliert, der Komplexität der Modellierung. Werden grobe Fehler in den Beobachtungen von vornherein ausgeschlossen, verbleiben Abweichungen in der Modellierung, die entweder den zufälligen oder den systematischen Effekten zugeschrieben werden müssen und als Resultat die Diskrepanzen zwischen Modell und Realität bewirken. Im Folgenden sollen ausschließlich die Unsicherheiten in den Daten sowie deren Fortpflanzung auf die unbekannten Parameter und die Schätzwerte des Modells betrachtet werden; Fragen der modellinduzierten Unsicherheiten bleiben künftigen Forschungen vorbehalten.

Nach BANDEMER (2005, S. 38 ff.) sowie NEUMANN (2009, S. 28 ff.) lässt sich die Unsicherheit von Beobachtungen durch Vagheit, zufällige Variabilität und Unschärfe beschreiben. Unter *Vagheit* wird die Unsicherheit verstanden, die durch verbal formulierte Ausdrücke einer Umgangs- oder Fachsprache entsteht. Nach der Definition ist nicht genau bekannt, was gemeint ist. Bezüglich metrischer Skalen handelt es sich um vage Beobachtungen, wenn ein grober Fehler nicht sicher ausgeschlossen werden kann. Für die Betrachtungen im Rahmen der Arbeit wird vorausgesetzt, dass sowohl sprachlich bedingte Vagheit als auch grobe Fehler durch eine sorgfältige Datenvorbereitung ausgeschlossen werden. Die *zufällige Variabilität* entsteht nach BANDEMER (2005, S. 34) durch „die Möglichkeit der zeitlichen Veränderung des Objekts und/oder seiner Eigenschaften oder durch die räumliche Veränderung unter gleichartigen Objekten und/oder deren Eigenschaften“ und ist daher dem Gebiet der Wahrscheinlichkeitstheorie zuzuordnen. Die Beobachtungen ergeben sich als Resultate, die ausschließlich dem Zufall unterliegen und daher als Ergebnisse eines Zufallsexperiments aufgefasst werden können. Dabei wird jeder Beobachtung eine Wahrscheinlichkeitsverteilung für deren Auftreten zugeordnet und als wesentliche mathematische Grundlage das Gebiet der Stochastik genutzt (KOCH, 2004, S. 81). Der Begriff der *Unschärfe* von Beobachtungen wird durch die Repräsentativität der Unschärfe für die Realität und deren Reichtum an Informationen definiert (NEUMANN, 2009, S. 29). Charakteristisch ist dabei die Unzulänglichkeit eines Zugangs zu wahrscheinlichkeitstheoretischen Interpretationen, bspw. Restfehlern zwischen Modell und Realität, Rundungsfehlern in den Eingangsgrößen oder einseitig gerichtete Fehleinschätzungen in den Ausprägungen einzelner Einflussgrößen. KUTTERER (2002, S. 25) erweitert die Unschärfe um den Begriff der *Impräzision* als Spezialfall, welche als Maß für die systematische Abweichung zwischen Beobachtungen und Modell zu verstehen sind: Demnach resultiert sie aus der Überlagerung einer Vielzahl von Einflussgrößen, die bei

einem Zufallsexperiment nicht vollständig ausgeschlossen werden können und die mangelnde Kenntnis über Art und Größe systematischer Einflüsse wiedergibt.

Die mathematische Behandlung der genannten Unsicherheitskomponenten setzt deren Identifikation in den Daten voraus. Neben der Trennung von Einflussgrößen, die in den meisten Fällen leicht entweder der zufälligen Variabilität oder der Unschärfe/Impräzision zugeordnet werden können, ist ebenfalls eine Überlagerung beider Unsicherheitskomponenten möglich. Besonders deutlich wird dies an den Kaufpreisen als originären Beobachtungen. Spielen sowohl zufällige Variabilität als auch Unschärfe in den Einflussgrößen eine Rolle, ist deren kombinierte Auswirkung auf den Kaufpreis offensichtlich.

Generell stellt die Mathematik die Modellhintergründe der numerischen Approximationstheorie, der Wahrscheinlichkeitstheorie, der Intervallmathematik sowie der Fuzzy-Theorie zur Verfügung (BANDEMER, 2006, S. 7). Die Wahl eines der Ansätze setzt voraus, dass die Qualität der Daten entsprechend eingeschätzt wird, denn diese reflektiert Art und Grad der Exaktheit der Daten in Form der Eigenschaften als fehlerfrei, zufällig oder impräzise (KUTTERER, 2002, S. 26 ff.): Während die numerische Approximationstheorie einen Ansatz darstellt, der Unsicherheiten weitgehend ausschließt, eignet sich die Wahrscheinlichkeitstheorie mit dem Instrument der Stochastik für die Beschreibung der zufälligen Variabilität. Stochastisch motivierte Ansätze finden eine weite Verbreitung in geodätischer Grundlagenliteratur, stellvertretend kann auf die Werke von GRAFAREND & SCHAFFRIN (1993), KOCH (2004), CASPARY & WICHMANN (2007) sowie NIEMEIER (2008) verwiesen werden. Die Wahrscheinlichkeitstheorie ist zudem als mathematisch-statistische Grundlage für sämtliche bisherige Modelle in der Verkehrswertermittlung zu sehen. Die Intervallmathematik als weiterer Ansatz basiert auf der Quantifizierung von Impräzision durch die Festlegung von minimalen und maximalen Grenzwerten der Beobachtungen, die durch Intervallradien spezifiziert werden; zufällige Variabilität wird komplett ausgeschlossen. Eine Einführung in die mathematischen Grundlagen bieten u. a. ALEFELD & HERZBERGER (1983), BAUCH ET AL. (1987), KUTTERER (1994) und KUTTERER (1999). Die Fuzzy-Theorie nutzt die Definition von unscharfen Mengen für die Berücksichtigung impräziser Beobachtungen und ist damit als Erweiterung der klassischen Mengenlehre zu sehen. Entwickelt wurde die Fuzzy-Theorie maßgeblich von ZADEH (1965), eine Einführung in unscharfe Mengen und Fuzzy-Theorie geben KRUSE ET AL. (1995), BIEWER (1997), WÜNSCHE & NÄTHER (2002) sowie VIERTL (2011).

Die Aufzählung der verschiedenen mathematischen Ansätze soll verdeutlichen, dass die Modellierung von Unsicherheiten als ganzheitlicher Prozess zu sehen ist, der neben der reinen Datenerfassung eine intensive Auseinandersetzung mit der Qualität der Daten erfordert. Nur unter dieser Voraussetzung kann eine zutreffende Berücksichtigung aller Unsicherheitskomponenten erfolgen. Hingewiesen werden soll in diesem Kontext auch explizit auf die Tatsache, dass die mathematischen Ansätze sich nicht gegenseitig ausschließen. Vielmehr folgt aus der bisherigen Argumentation, dass eine gemeinsame Berücksichtigung zufälliger Variabilität und Impräzision erfolgen kann. Gemeinsame Modellansätze können insbesondere den Arbeiten von KUTTERER (2002), SCHÖN (2003) sowie NEUMANN (2009) entnommen werden.

Nachdem bislang die Rahmenbedingungen der Verkehrswertermittlung sowie die Modellbildung und deren Unsicherheiten behandelt wurden, sollen im folgenden Abschnitt die für die Verkehrswertermittlung relevanten Modelle mit besonderem Fokus auf das indirekte Vergleichswertverfahren erläutert werden.

2.3. Modelle des Vergleichswertverfahrens

Die Modellbildung in Anwendungen der Verkehrswertermittlung hat sich in Form mathematisch-statistischer Verfahren insbesondere seit den Forschungen von ZIEGENBEIN (1977) etabliert. Wird die Preisbildung von Immobilien aus einer rein mathematisch-statistisch fundierten Sichtweise betrachtet, hängt sie von Faktoren ab, die in ihrem Zusammenwirken oftmals schwer zu beurteilen sind. In der Summe ergibt sich aus der Vielzahl von Faktoren, dass eine exakte Nachvollziehung von Kaufpreisen nahezu unmöglich ist. Gutachterlicher Sachverstand fließt dann in ein Modell ein, wenn es interpretiert und auf die Realität übertragen werden soll. Auf diese Weise lassen sich zumindest approximativ Kaufpreise und Preisbildung nachvollziehen. In den folgenden Abschnitten werden die Erläuterungen der Abschnitte 2.1 und 2.2 verknüpft, indem in Abschnitt 2.3.1 zunächst die allgemeine Modellbildung in der Verkehrswertermittlung mit besonderem Fokus auf das Vergleichswertverfahren aufgezeigt wird. In Abschnitt 2.3.2 werden die klassische multiple lineare Regressionsanalyse sowie die einfache Kollokation als Modelle des indirek-

ten Vergleichswertverfahrens in ihren Grundzügen behandelt. Über die Erläuterungen hinausgehende Ansätze und Auswertemethoden bleiben der weiterführenden Literatur vorbehalten, vgl. u. a. CHATTERJEE & PRICE (1995), URBAN & MAYERL (2006), FAHRMEIR ET AL. (2009).

2.3.1. Allgemeine Modellbildung in der Verkehrswertermittlung

Für die Ermittlung von Verkehrswerten unter Berücksichtigung der gegebenen rechtlichen Rahmenbedingungen nach Abschnitt 2.1.2 steht eine Reihe von Methoden zur Verfügung, deren Anwendung sich sowohl nach der Art des Wertermittlungsobjekts als auch der verfügbaren Datengrundlage richtet. Die Methoden lassen sich generell in die Gruppe der normierten und der nicht normierten Verfahren einteilen. Die normierten Verfahren werden in der ImmoWertV genannt und beschrieben: Gemäß § 8 Abs. 1 ImmoWertV sind für die Verkehrswertermittlung das Vergleichswertverfahren (§§ 15 und 16), das Ertragswertverfahren (§§ 17 bis 20), das Sachwertverfahren (§§ 18 bis 23) oder eine geeignete Kombination der Verfahren anzuwenden. Der Verkehrswert nach § 194 BauGB ist sachverständig aus dem Ergebnis des oder der angewendeten Verfahren unter besonderer Berücksichtigung der Lage auf dem Grundstücksmarkt zu bemessen. In Bezug auf die Gruppe der nicht normierten Verfahren sind u. a. Residualwertverfahren, Discounted-Cash-Flow-Methoden, Monte-Carlo Simulationen sowie Investmentverfahren zu nennen, deren Anwendung dann als gleichrangig zu den normierten Verfahren zu sehen ist, wenn letztere rechtlich nicht explizit zur Anwendung vorgeschrieben sind.

Trotz der innerhalb der genannten Methoden unterschiedlichen Herangehensweisen ist es das Ziel, einen normierten, einheitlichen Wert zu ermitteln, der sich auch bei Anwendung mehrerer Verfahren für ein Objekt zumindest approximativ entspricht. Sämtliche Methoden setzen voraus, dass das Grundstücksmarktgeschehen und die Preise von Grundstücken zu beobachten und zu analysieren sind. Erkenntnisse aus den Analysen können dann auf das aktuelle Wertermittlungsobjekt übertragen werden. Unter diesem Gesichtspunkt ist die Ermittlung von Verkehrswerten in besonderem Maße abhängig von der Verfügbarkeit von Preisen vergleichbarer Grundstücke auf dem betrachteten Teilmarkt, die mitunter sehr gering ausfallen kann. Nach ZIEGENBEIN (2010, S. 445) ergibt sich daraus die Konsequenz, dass je geringer die Informationsdichte ist, umso komplexer ist das Vorgehen für die Übertragung der Erkenntnisse aus dem Marktgeschehen auf das Wertermittlungsobjekt und je größer die Informationsdichte ist, umso einfacher ist das Vorgehen.

Aufgrund der Anbindung jeder Verkehrswertermittlung an das Marktgeschehen können die normierten Wertermittlungsverfahren als vergleichende Betrachtung oder Preisvergleiche aufgefasst werden: „*Ohne vergleichende Betrachtung des Wertermittlungsobjekts mit dem Marktgeschehen kann kein Preis, der im gewöhnlichen Geschäftsverkehr zu erzielen wäre, ermittelt werden.*“ (ZIEGENBEIN, 2010, S. 445). Nach REUTER (1989, S. 383) kann bei Preisvergleichen zwischen kaufpreisbezogenen ganzheitlichen Preisvergleichen und faktorpreisbezogenen Preisvergleichen differenziert werden. Bei erstgenanntem Vergleich werden die Preise der Grundstücke auf das zu bewertende Grundstück transformiert, bei letztgenanntem werden Preise und aus dem Markt abgeleitete Faktoren genutzt, um den Verkehrswert zu bestimmen. Abbildung 2.5 gibt einen Überblick über die Wertermittlungsverfahren nach ImmoWertV im Kontext des Preisvergleichs nach REUTER (1989).

In der ImmoWertV wird keine Rangfolge für die Anwendung der Verfahren genannt. Nach § 8 Abs. 1 ist das geeignete Verfahren unter Berücksichtigung der im gewöhnlichen Geschäftsverkehr bestehenden Gepflogenheiten und der besonderen Umstände des konkreten Wertermittlungsfalls sachgerecht zu wählen. Insbesondere die zur Verfügung stehenden Daten müssen hinreichend berücksichtigt werden. Damit ergibt sich in der praktischen Anwendung keine Präferenz für eines der drei Verfahren, sondern eine einzelfallbezogene Auswahl.

Dem Vergleichswertverfahren wird trotz der nominellen Gleichrangigkeit zum Ertrags- und Sachwertverfahren von jeher eine gewisse Vorrangigkeit eingeräumt, die nach KLEIBER (2010, S. 1232) auf dessen besonderer Überzeugungskraft und Plausibilität beruht. Beide Eigenschaften resultieren aus dem Grundgedanken des Vergleichswertverfahrens: Eine Sache ist regelmäßig so viel wert, wie im gewöhnlichen Geschäftsverkehr dafür als Preis erzielt werden kann. In Bezug auf die Verkehrswertermittlung wird für ein konkretes Objekt der Frage nachgegangen, was für ähnliche Objekte zum gleichen Zeitpunkt gezahlt worden ist. Durch den Vergleich mit sich ähnelnden Objekten ergibt sich im Gegensatz zum Ertrags- oder Sachwertverfahren direkt ein enger Bezug zum Immobilienmarkt. Die Grundzüge des Vergleichswertverfahrens werden durch § 15 ImmoWertV gegeben und können entsprechend Ab-

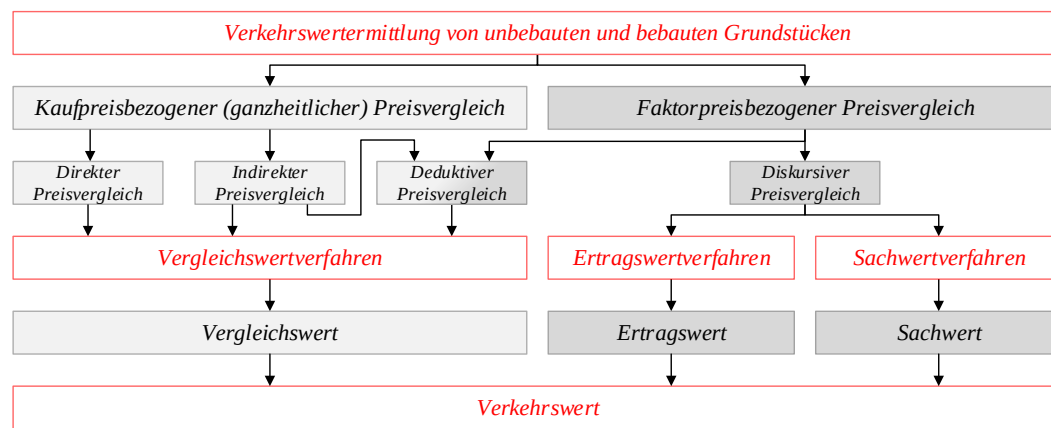


Abbildung 2.5: Übersicht der Wertermittlungsverfahren als Preisvergleiche (nach ZIEGENBEIN (2010, S. 445))

bildung 2.6 zunächst in zwei Arten des Preisvergleichs differenziert werden: Den direkten Preisvergleich und den indirekten Preisvergleich.

Die regelmäßig überzeugendste und anschaulichste Art der Wertermittlung ist der *direkte Preisvergleich*. Hier wird der Vergleichswert unmittelbar aus Vergleichspreisen abgeleitet, die zum Wertermittlungsstichtag vereinbart worden sind und eine hinreichende Übereinstimmung mit dem Wertermittlungsobjekt aufweisen. Der Begriff der *hinreichenden Übereinstimmung* ist nach ImmoWertV zunächst ein unbestimmter Rechtsbegriff, der weiterer Auslegung bedarf und mit Erlass der VW-RL weiter definiert wurde: „Eine hinreichende Übereinstimmung mit dem Wertermittlungsobjekt liegt vor, wenn die Vergleichsgrundstücke hinsichtlich ihrer wertbeeinflussenden Grundstücksmerkmale nur solche Abweichungen aufweisen, die unerheblich sind oder deren Auswirkungen auf die Kaufpreise in sachgerechter Weise berücksichtigt werden können.“ (BUNDESMINISTERIUM FÜR UMWELT, NATURSCHUTZ, BAU UND REAKTORSICHERHEIT, 2014, Nummer (Nr.) 3 Abs. 2)). Als wertbeeinflussende Grundstücksmerkmale, auch i. S. der ImmoWertV, werden weiterhin die in Tabelle 2.1 aufgelisteten Aspekte genannt, die jedoch keine abschließende Aufzählung darstellen können.

Tabelle 2.1.: Potenziell wertbeeinflussende Grundstücksmerkmale nach Nr. 4.2 bis 4.4 VW-RL

Beitrags- und abgabenrechtlicher Zustand	Maß der baulichen Nutzung
Grundstücksgröße	Lage (Klassifizierung, Stadtteil, Ecklage)
Grundstückstiefe	Anbauart (freistehend, Doppelhaushälfte, etc.)
Grundstücksbreite	Baujahr
Grundstückszuschnitt	Ausstattung (Klassifizierung oder Einzelmerkmale)
Wohnfläche, Anzahl der Wohnungen	Acker- und Gründlandzahl
Verpachtung, Vermietung	Kaufzeitpunkt

Neben der Übereinstimmung von Wertermittlungsobjekt und Vergleichsobjekten bezüglich der in Tabelle 2.1 genannten Grundstücksmerkmale ist nach ImmoWertV eine *ausreichende Zahl* von Vergleichspreisen notwendig. Hierbei handelt es sich um einen weiteren unbestimmten Rechtsbegriff, der jedoch im Gegensatz zu der hinreichenden Übereinstimmung nicht weiter definiert wird. Für eine praktische Anwendung hat ZIEGENBEIN (1977) Untersuchungen über verschiedene sachliche Teilmärkte in Niedersachsen durchgeführt und aus den Ergebnissen die erforderliche Anzahl von Vergleichsobjekten je Teilmarkt abgeleitet. Für eine Ergebnisdarstellung wird auf die entsprechende Literaturstelle verwiesen.

Trotz der Verfügbarkeit einer umfassenden Datenbank (AKS) werden in den seltensten Fällen Vergleichsobjekte vorhanden sein, die einem Wertermittlungsobjekt bezüglich der strengen Forderung nach mehr oder weniger völliger Übereinstimmung in den wertbeeinflussenden Merkmalen gerecht werden. KLEIBER (2010, S. 1237) beurteilt den unmittelbaren Preisvergleich daher als „idealtypische Wunschvorstellung“, die praktisch kaum je Bedeutung erlange. Begründet wird die Aussage mit den in Abschnitt 2.1.1 erläuterten Charakteristika einer Immobilie: Die Tatsache, dass es sich bei jeder Immobilie - selbst in unmittelbarer Nachbarschaft - um ein Unikat mit individuellen Eigenschaften handele, erlaube keine einfache Mittelwertbildung nach Abbildung 2.6. GERARDY ET AL. (2014,

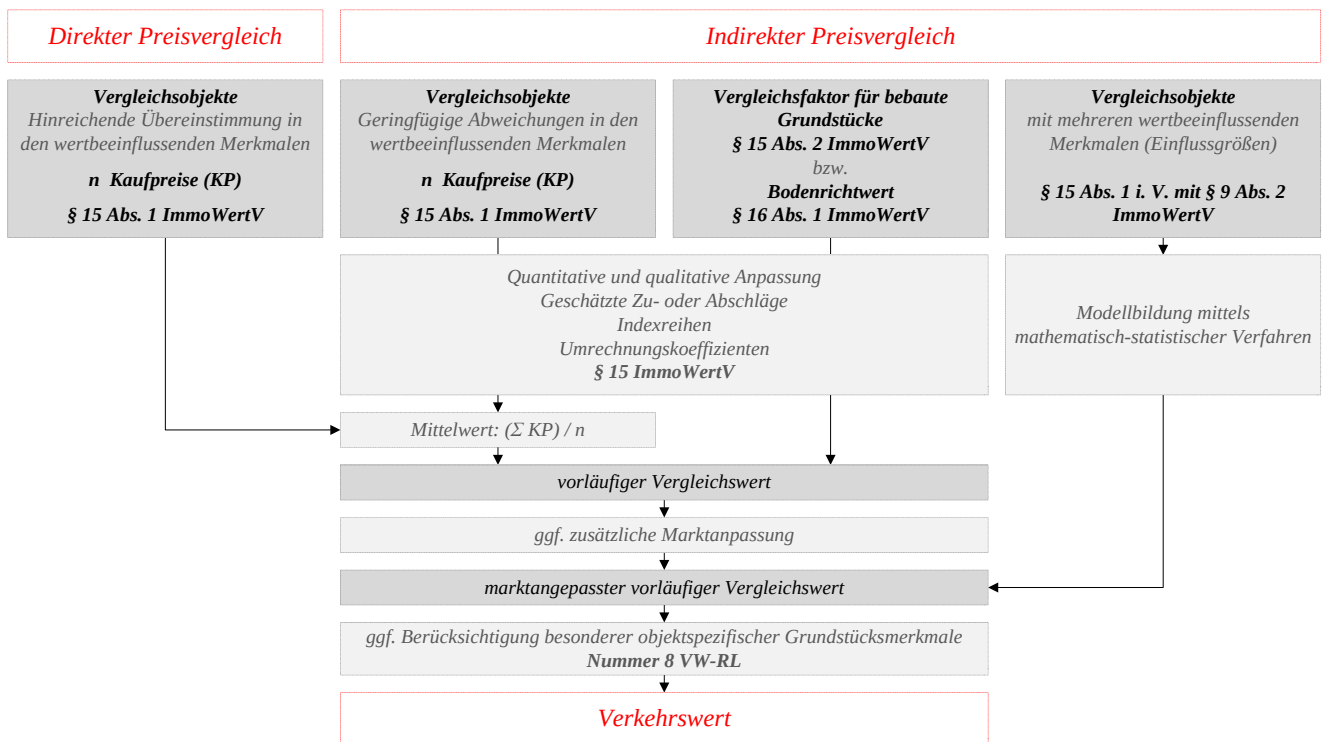


Abbildung 2.6: Arten des Vergleichswertverfahrens nach GERARDY ET AL. (2014, S. 1.3.7/8)

S. 1.3.7/6) stimmt der Aussage im Grundsatz zu, jedoch nur, sofern es sich um eine sehr enge Sichtweise handelt, in der die Vergleichsobjekte exakt in allen Merkmalen übereinstimmen müssen und die Kaufpreise in absoluter Höhe genutzt werden. Wird hingegen die Sichtweise in der Form definiert, dass die wertbeeinflussenden Merkmale von Wertermittlungs- und Vergleichsobjekten soweit übereinstimmen, dass die auf einen preisstabilen Zeitraum um den Wertermittlungstichtag und auf eine einheitliche Bezugsgröße transformierten Kaufpreise ohne weitere Korrekturen verwendet werden können, sei die Anwendung des direkten Preisvergleichs durchaus möglich.

In der praktischen Anwendung hat aufgrund der seltenen Verfügbarkeit direkt vergleichbarer Kaufpreise die Anwendung des *indirekten Preisvergleichs* nach Abbildung 2.6 eine ungleich höhere Bedeutung erlangt. Liegen keine direkt vergleichbaren Kaufpreise vor, können nach § 15 Abs. 1 Satz 4 ImmoWertV Kaufpreise herangezogen werden, die geringfügige Abweichungen untereinander und gegenüber dem Wertermittlungsobjekt aufweisen: „*Änderungen der allgemeinen Wertverhältnisse auf dem Grundstücksmarkt oder Abweichungen einzelner Grundstücksmerkmale sind in der Regel auf der Grundlage von Indexreihen oder Umrechnungskoeffizienten zu berücksichtigen.*“ In einem intertemporären und interqualitativen Preisvergleich werden die Zustandsmerkmale der Vergleichsobjekte auf das Wertermittlungsobjekt umgerechnet, es wird eine Anpassung auf die allgemeinen Wertverhältnisse am Wertermittlungstichtag vorgenommen. Voraussetzung für den so genannten *indirekten evidenten Preisvergleich* ist sowohl die Kenntnis über die wesentlichen wertbeeinflussenden Merkmale, die die Wertunterschiede hervorrufen, als auch die Verfügbarkeit von Indexreihen bzw. Umrechnungskoeffizienten für die konjunkturelle bzw. qualitative Anpassung. Nach der Anpassung beziehen sich die umgerechneten Kaufpreise der Vergleichsobjekte auf die Verhältnisse des Wertermittlungsobjekts und führen wie im unmittelbaren Preisvergleich durch das arithmetische Mittel zum Vergleichswert. Ein alternatives Vorgehen ist die Nutzung von Vergleichsfaktoren nach § 15 Abs. 2 ImmoWertV und Bodenrichtwerten als Vergleichsfaktoren für unbebaute Grundstücke nach § 16 Abs. 1 ImmoWertV, die umgerechnet auf den Wertermittlungsfall durch Multiplikation mit der Bezugsgröße zum Vergleichswert führen.

Neben den genannten Vorgehensweisen lässt § 9 Abs. 2 ImmoWertV die Wahl geeigneter, jedoch nicht näher definierter Methoden zur Ermittlung des Vergleichswertes offen: „*Kaufpreise solcher Grundstücke, die in ihren Grundstücksmerkmalen voneinander abweichen, sind im Sinne des Absatzes 1 Satz 1 nur geeignet, wenn die Abweichungen 1. in ihren Abweichungen auf die Preise sich ausgleichen, 2. durch Zu- oder Abschläge oder 3. durch andere geeignete Verfahren berücksichtigt werden können.*“ Mit der normativen Vorgabe wird der Verwendung mathematisch-statistisch basierter Verfahren Rechnung getragen, die in der Praxis regelmäßig zum Einsatz kom-

men. Wesentlicher Vorteil solcher Modelle ist die simultane Berücksichtigung von Abweichungen mehrerer wertbeeinflussender Merkmale in einem Rechengang. Ein kurzer Überblick über die theoretischen Grundlagen der in der indirekten Vergleichswertermittlung genutzten Modelle wird im folgenden Abschnitt gegeben. Für eine Vereinfachung in der Terminologie wird im weiteren Verlauf der allgemeine Begriff des Vergleichswertverfahrens genutzt; auf den expliziten Hinweis der indirekten Variante wird verzichtet.

2.3.2. Multivariate Analyseverfahren in der Vergleichswertermittlung

Das Modell, welches in der Praxis für Zwecke des Vergleichswertverfahrens am häufigsten Verwendung findet, ist die multiple lineare Regressionsanalyse. Durch die Ableitung einer optimalen Schätzfunktion können Vergleichswerte sowohl für individuelle Wertermittlungsobjekte als auch für Massenbewertungen großer Immobilienbestände ohne die Zwischenschritte der konjunkturellen und qualitativen Anpassung ermittelt werden. Als weiteres Modell ist die einfache Kollokation zu nennen, die im besten Fall im Vergleich zur Regressionsanalyse eine verbesserte funktionale Approximation ermöglicht. In den folgenden beiden Unterabschnitten werden die jeweiligen Grundzüge der beiden Modelle erläutert und die grundlegenden Formalismen beschrieben. Auf die verkürzte Darstellung und eine ausführliche Behandlung in den genannten Literaturquellen wird hingewiesen.

2.3.2.1. Grundzüge der multiplen linearen Regressionsanalyse

Grundgedanke der Regressionsanalyse ist die Beschreibung von Abhängigkeiten zwischen einer Zielgröße y (Beobachtung) und einer oder mehrerer zugehöriger Einflussgrößen $x_1, \dots, x_m$ (mit m = Anzahl der Einflussgrößen), denen ein Erklärungsbeitrag zu y unterstellt wird. Im Kontext des indirekten Vergleichswertverfahrens entsprechen die Einflussgrößen der Definition gemäß Abschnitt 2.2.1 (S. 31 folgende (ff.)). Die Zielgröße, zu der als abhängige Variable ein funktionaler Zusammenhang hergestellt wird, ist regelmäßig der Kaufpreis bzw. ein auf eine geeignete Bezugsgröße normierter Faktor. Charakteristisch für jede Art von Regressionsmodellen ist die Tatsache, dass die Abhängigkeit der Zielgröße y zu den Einflussgrößen nicht als exakte Funktion $y = f(x_1, \dots, x_m)$ ausgedrückt werden kann. Hauptziel der Regressionsanalyse ist unter diesem Gesichtspunkt die Untersuchung, welchen Einfluss die Merkmale x_k mit $k \in \{1, \dots, m\}$ auf den Mittelwert der Zielgröße haben.

Im Vergleichswertverfahren wird überwiegend die Klasse der linearen Regressionsmodelle genutzt, in der die funktionale Beziehung zwischen Ziel- und Einflussgrößen durch lineare Regressionkoeffizienten β_k mit $k \in \{0, \dots, m\}$ ausgedrückt wird und sich die zugehörige Beobachtung y als bedingter Erwartungswert durch

$$y = E(y | x_1, \dots, x_m) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \epsilon \quad (2.9)$$

modellieren lässt. Der Koeffizient β_0 repräsentiert einen konstanten Term und wird als Offset bzw. Intercept bezeichnet, während $\beta_1, \dots, \beta_m$ die Anteile der Einflussgrößen an der Zielgröße wiedergeben. Bei den Regressionskoeffizienten handelt es sich um die unbekannten, theoretischen Modellparameter nach Abbildung 2.3 (S. 32), welche aus den Daten geschätzt werden müssen (ZIEGENBEIN, 1977, S. 27).

Wie ersichtlich ist, wirkt jede der in das Modell eingeführten Kovariablen linear auf die Zielgröße; die Effekte wirken additiv auf die Beobachtung y . Die funktionale Modellierung durch die Regressionskoeffizienten wird auch *systematische Komponente* oder *Trend* genannt; die Koeffizienten werden daher im Folgenden als Trendkoeffizienten bezeichnet. In der Variablen ϵ als *stochastische Komponente* werden hingegen die durch den funktionalen Zusammenhang nicht erklärbaren, unsystematischen Abweichungen von der Zielgröße zusammengefasst (FAHRMEIR ET AL., 2009, S. 19 f.). Die stochastische Komponente wird auch Residuum genannt. Da die Zielgröße y eine Zufallsvariable darstellt, ist auch das Residuum ϵ als eine Zufallsvariable zu modellieren.

In Matrix-Vektor Notation lassen sich mehrere Beobachtungen y_i mit $i \in \{1, \dots, n\}$ und n als Anzahl der Beobachtungen in dem Vektor $\mathbf{y}$ zusammenfassen. Der Vektor der Trendkoeffizienten kann in diesem Kontext durch $\boldsymbol{\beta}$ dargestellt werden, der Vektor der Residuen durch $\boldsymbol{\epsilon}$ sowie die zugehörigen Einflussgrößen je Beobachtung durch die Designmatrix $\mathbf{X}$:

$$\underset{[n \times 1]}{\mathbf{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{[(m+1) \times 1]}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \underset{[n \times 1]}{\boldsymbol{\epsilon}} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad \underset{[n \times (m+1)]}{\mathbf{X}} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad (2.10)$$

sodass sich (2.9) ausdrücken lässt als

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \Leftrightarrow \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (2.11)$$

Für die Formulierung eines funktionalen Zusammenhangs bedeutet (2.11) eine lediglich approximative Annäherung an den realen Sachverhalt; zu komplex sind die wertrelevanten Einflussgrößen und deren Zusammenwirken in der marktorientierten und akteursabhängigen Verkehrswertermittlung für eine mathematisch zutreffende Formulierung (ZADDACH & ALKHATIB, 2013a, S. 145). Bezüglich der Unsicherheiten basierend auf zufälliger Variabilität wird das funktionale Modell nach (2.11) komplettiert durch das stochastische Modell, welches sich allgemein formulieren lässt zu

$$\underset{[n \times n]}{\boldsymbol{\Sigma}_{\mathbf{yy}}} = \underset{[n \times n]}{\boldsymbol{\Sigma}_{\mathbf{ee}}} = \sigma^2 \underset{[n \times n]}{\mathbf{P}}^{-1}, \quad (2.12)$$

mit $\boldsymbol{\Sigma}_{\mathbf{yy}}$ als $[n \times n]$ Varianz-Kovarianzmatrix (VKM) der Beobachtungen $\mathbf{y}$ und $\boldsymbol{\Sigma}_{\mathbf{ee}}$ als VKM der Residuen. σ^2 symbolisiert die unbekannte, theoretische Varianz der Gewichtseinheit unter der Voraussetzung der Varianzgleichheit. Abschließend bezeichnet $\mathbf{P}$ die bekannte, positiv definite Gewichtsmatrix der Beobachtungen. (2.11) in Verbindung mit dem Dispersionsmodell nach (2.12) wird als allgemeines Gauß-Markoff-Modell (GMM) bezeichnet. Der Beobachtungsvektor $\mathbf{y}$ in (2.11) trägt Informationen über die unbekannten Parameter $\boldsymbol{\beta}$ des Modells, welche mit Hilfe einer Schätzfunktion ermittelt werden müssen. Als Methode der Parameterschätzung sind unterschiedliche Ansätze denkbar: Neben der klassischen *Methode der kleinsten Quadrate (MkQ)* lassen sich auch bspw. die *Beste lineare erwartungstreue Schätzung - Best Linear Unbiased Estimation (BLUE)*, die *Maximum-Likelihood-Schätzung (ML-Schätzung)* oder Ansätze zur robusten Parameterschätzung nennen. Anwendungen in der Verkehrswertermittlung beruhen bislang ausschließlich auf der MkQ, deren Ziel die Minimierung der Quadratsumme der Abweichungen zwischen den Beobachtungen $\mathbf{y}$ und den Schätzwerten $s[E(\mathbf{y})]$ ihrer Erwartungswerte $E(\mathbf{y})$ als Funktion der unbekannten Parameter ist. Mathematisch formuliert ergibt sich die Minimierungsaufgabe für den Term $(\mathbf{y} - s[E(\mathbf{y})])' \boldsymbol{\Sigma}_{\mathbf{yy}}^{-1} (\mathbf{y} - s[E(\mathbf{y})]) \rightarrow \min$. Die Minimierung stellt ein Optimierungsproblem dar, welches sich als Extremwertaufgabe formulieren lässt. Die Lösung der Extremwertaufgabe steht nicht im Fokus der Arbeit; die Herleitung der optimalen Schätzung für die Modellparameter (Trendkoeffizienten $\boldsymbol{\beta}$ und Varianz σ^2) lassen sich der weiterführenden Literatur entnehmen (vgl. u. a. KOCH (2004, S. 171 f.)). Mit der Parameterschätzung und hier nicht thematisierten Hypothesentests für die Prüfung der Modellsignifikanz ist die klassische Normalregression vollständig spezifiziert und kann für die Prädiktion von Vergleichswerten genutzt werden. Weiterführend lassen sich u. a. in SCHÖNFELD (1971), CHATTERJEE & PRICE (1995), URBAN & MAYERL (2006), KOCH (2004) und FAHRMEIR ET AL. (2009) vertiefte Ausführungen zum Thema der Modelldiagnose finden, deren Methoden genutzt werden können, um ein optimales Regressionsmodell zu entwickeln.

Die Nutzung der Regressionsanalyse in der Vergleichswertermittlung erfordert eine Auswertestrategie, die den Eigenarten der immobilien-spezifischen Datengrundlage gerecht wird: In der Verkehrswertermittlung hat sich daher ein praktischer Verfahrensablauf zur Auswertung etabliert, bei dem in einem iterativen Prozess eine optimale Regressionsfunktion geschätzt wird. Nach der grundsätzlichen Adaption für Wertermittlungszwecke durch ZIEGENBEIN (1977) wurde im Laufe der Folgejahre eine programmgesteuerte Analysestrategie in der AKS implementiert, deren Grundlagen ausführlich in ZIEGENBEIN (1986) sowie ZIEGENBEIN (1995) beschrieben sind. Die Automatisierung der Kaufpreisauswertungen zum Gesamtsystem *Wertermittlungsinformationssystem (WIS) Niedersachsen* mit umfassenden Analyse-möglichkeiten wird in ZIEGENBEIN (1999) und MÜRLE (2006) thematisiert. Der bisherige in der AKS implementierte Ansatz zur regressionsgestützten Vergleichswertermittlung beruht auf einer iterativen

Auswertestrategie, die so lange durchlaufen wird, bis eine Regressionsfunktion erhalten wird, die allen geforderten Ansprüchen und Güterkriterien genügt. Für eine detaillierte Beschreibung wird auf die weiterführende Literatur verwiesen (vgl. ZIEGENBEIN (1977), ZIEGENBEIN (1995)). Im folgenden Abschnitt wird das Modell der einfachen Kollokation als Erweiterung zur Regressionsanalyse vorgestellt.

2.3.2.2. Grundzüge der einfachen Kollokation

Die Regressionsanalyse gibt die Möglichkeit, Variationen der Zielgröße durch die Variabilität der Einflussgrößen zu erklären, um so einen funktionalen Zusammenhang aufzuzeigen. Gemäß der bisherigen Ausführungen sind die Zusammenhänge zwischen Ziel- und Einflussgrößen jedoch nur vage bekannt, Kenntnisse über eine exakte Funktion sind nicht gegeben und können durch den mathematisch beschriebenen Ablauf der Regressionsanalyse i. d. R. nicht erreicht werden. Für eine verbesserte Approximation bietet sich die *einfache Kollokation* als stochastische Verfeinerung des funktionalen Trendmodells an: Grundgedanke ist, dass die Residuen nach Abspaltung des Trends gewisse Systematiken aufweisen, die funktional nicht sinnvoll erfasst werden können. Das Modell der Kollokation bietet über die Verknüpfung der stochastischen Eigenschaften der Beobachtungen einen Zugang für die Ausschöpfung eines zusätzlichen Informationsgehaltes. Die Methode der Kollokation (Modell der Prädiktion und Filterung) wurde bereits in den 1970er Jahren entwickelt und ursprünglich für Fragestellungen der Erdschwerefeldmodellierung in der physikalischen Geodäsie eingesetzt (vgl. KRARUP (1969), MEISSL (1976), MORITZ (1973a), MORITZ (1973b), MORITZ (1980), RUMMEL (1976), TSCHERING (1978)). Gegen Ende der 1970er Jahre wurde das Verfahren erstmals auf Fragestellungen der Verkehrswertermittlung übertragen (vgl. PELZER (1976), PELZER (1978), ZIEGENBEIN & HAWERK (1978)). UHDE (1982) zeigt aufbauend eine detaillierte Analyse der Anwendungsmöglichkeiten in der Verkehrswertermittlung mit der Empfehlung einer Auswertestrategie sowie den Ergebnissen praktischer Untersuchungen. Einen Überblick über praktische Untersuchungen bzgl. der Übertragbarkeit auf unterschiedliche räumliche und sachliche Teilmärkte geben in Ergänzung ZADDACH & ALKHATIB (2012a), ZADDACH & ALKHATIB (2012b), ZADDACH & ALKHATIB (2013a) sowie ZADDACH & ALKHATIB (2014a).

Die Grundidee der Kollokation beruht auf der Überlegung, dass sich jede Beobachtung zerlegen lässt in einen regelmäßig systematischen Anteil (Trend), einen unregelmäßig systematischen Anteil (Signal) sowie einen unregelmäßig zufälligen Anteil (Rauschen). Der *Trend* ermöglicht die Erklärung eines Kaufpreises durch meist einige wenige dominierende Wertmerkmale und ist vergleichbar mit dem Trendmodell $\mathbf{X}\beta$ der klassischen Regressionsanalyse nach (2.11). Das *Signal* enthält in Form von unregelmäßigen, aber systematischen Anteilen einen Informationsgehalt, den es für die Verbesserung der Trendfunktion und damit der Modellierung der Zielgröße auszuschöpfen gilt. Im *Rauschen* verbleibt letztlich die zufällige, durch das Modell nicht erklärbare Reststreuung. Abbildung 2.7 zeigt das Grundprinzip der Kollokation im Vergleich zur einfachen Trendbestimmung in der klassischen Regressionsanalyse nach (2.11).

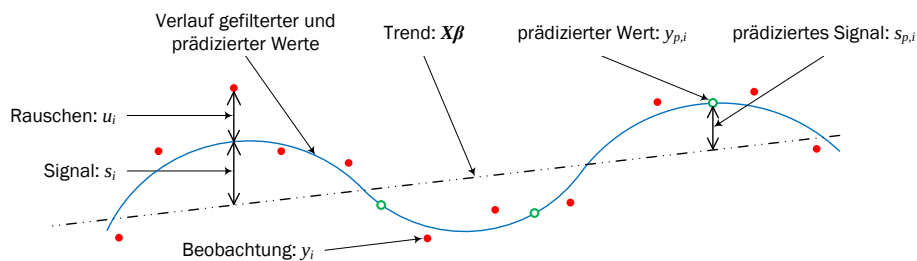


Abbildung 2.7: Prinzipskizze der Kollokation im zweidimensionalen Fall.

Wie die Abbildung verdeutlicht, ermöglicht die Berücksichtigung einer Signalkomponente im Idealfall eine verbesserte (Funktions)Approximation des mathematischen Modells an die gegebenen Daten. Die Interpretation des Signals in Anwendungen der Verkehrswertermittlung basiert auf der Modelldefinition der Regressionsanalyse: In (2.11) wird unterstellt, dass zwischen den Kaufpreisen als Beobachtungen und den wertbeeinflussenden Merkmalen als Einflussgrößen ein deterministischer Zusammenhang besteht, der durch feste, nichtstochastische Parameter ausgedrückt wird. Wie aber in den Abschnitten 2.1.1 und 2.1.2 verdeutlicht, können zum einen bestimmte wertbe-

einflussende Umstände nicht numerisch ausgedrückt und so in eine funktionale Beziehung zum Kaufpreis gebracht werden. Zum anderen ergibt sich die Schwierigkeit, dass manche Umstände oder Merkmale, die den Kaufpreis beeinflusst haben, nicht bekannt sind und daher nicht in einem Modell berücksichtigt werden können. UHDE (1982, S. 15) nennt als Beispiele die zeitliche Komponente und den Lageeinfluss, die einer funktionalen Modellierung nur schwer zugänglich sind. In ihrer Gesamtheit wirken sich die Effekte durch nicht berücksichtigte Einflussgrößen in der Signalkomponente aus, welche sich aus einer Linearkombination von Einzelgrößen zusammensetzt und die einem rein deterministischen Modell nicht zugänglich ist. Eine Interpretation des stochastischen Signals folgt daraus, dass bei ähnlichen Kauffällen die nach Abspaltung des deterministischen Anteils verbleibenden Residuen ϵ ein gewisses Maß an Autokorrelation aufzeigen (UHDE, 1982, S. 16). Das Signal kann daher als eine nach Trendabzug verbleibende Restsystematik ausgedrückt werden, die nicht in weitere einzelne Anteile bestimmter wertbeeinflussender Merkmale zerlegt werden kann. Die Interpretation erfolgt lediglich als Gesamteffekt, der sich aus dem rein stochastischen Verhalten der Daten ergibt, aber einen Beitrag zur Erklärung der Zielgröße in sich trägt.

Für die funktionale Modellierung wird das Ausgleichungsmodell nach (2.11) als gemischtes Modell definiert, in dem die Residuen ϵ durch eine Linearkombination unbekannter Parameter γ als Zufallsvariable ersetzt werden:

$$E(\mathbf{y}) = \mathbf{X}\beta + \mathbf{Z}\gamma. \quad (2.13)$$

Hier bezeichnet $\mathbf{X}\beta$ wie bisher den Trendanteil. Die Matrix $\mathbf{Z}$ mit $rg(\mathbf{Z}) = n$ und der Dimension $[n \times r]$ stellt eine bekannte Koeffizientenmatrix dar, der Vektor γ der Dimension $[r \times 1]$ beinhaltet die unbekannten Zufallsparameter mit $E(\gamma) = \mathbf{0}$ und der VKM $\Sigma_{\gamma\gamma} = \sigma^2 \mathbf{Q}_{\gamma\gamma}$. Der Term $\mathbf{Z}\gamma$ symbolisiert in diesem Zusammenhang den Signal(anteil). Das nach (2.13) definierte Modell wurde von HELMERT (1872) als Allgemeinfall der Ausgleichungsrechnung eingeführt und wird als Gauß-Helmert-Modell (GHM) bezeichnet (WOLF, 1977, S. 237), (KOCH, 2004, S. 231). Die Spezialisierung des Modells nach (2.13) ergibt

$$\mathbf{y} = \mathbf{X}\beta + \bar{\mathbf{Z}}\bar{\gamma} \quad \text{mit} \quad E(\bar{\gamma}) = \mathbf{0} \quad (2.14)$$

sowie den substituierten Größen

$$\bar{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}, & -\mathbf{I}_n \end{bmatrix}, \quad \bar{\gamma} = \begin{bmatrix} \gamma', & \mathbf{v}' \end{bmatrix}', \quad \Sigma_{\bar{\gamma}\bar{\gamma}} = \sigma^2 \begin{bmatrix} \Sigma_{\gamma\gamma} & \mathbf{0} \\ \mathbf{0} & \Sigma_{vv} \end{bmatrix}. \quad (2.15)$$

Im Vergleich zum allgemeinen Fall bedeutet $\mathbf{Z}$ wiederum eine bekannte Koeffizientenmatrix, die nun der Dimension $[n \times (r - n)]$ folgt, γ einen $[(r - n) \times 1]$ Vektor und $\mathbf{u}$ einen $[n \times 1]$ Vektor, der die zufälligen, durch das Modell nicht erklärbaren Rauschanteile enthält (KOCH, 2004, S. 240 ff.). Aus (2.14) ergibt sich dann das spezielle Modell der Prädiktion und Filterung zu

$$E(\mathbf{y}) = \mathbf{y} + \mathbf{v} = \mathbf{X}\beta + \mathbf{Z}\gamma \quad \text{mit} \quad E(\gamma) = \mathbf{0} \quad \text{und} \quad E(\mathbf{v}) = \mathbf{0}. \quad (2.16)$$

Die exakten Zusammenhänge zwischen den Koeffizienten in $\mathbf{Z}$ und den Zufallsparametern in γ , sind in Anwendungen der Verkehrswertermittlung nicht bekannt, weshalb die Signalkomponente alternativ durch $\mathbf{s} = \mathbf{Z}\gamma$ mit $E(\mathbf{s}) = \mathbf{0}$ ausgedrückt werden kann. Das Modell der einfachen Kollokation ergibt sich daher aus (2.16) zu

$$E(\mathbf{y}) = \mathbf{y} + \mathbf{v} = \mathbf{X}\beta + E(\mathbf{s}) = \mathbf{X}\beta + \mathbf{s}. \quad (2.17)$$

Gemäß der einleitenden Ausführungen in Verbindung mit Abbildung 2.7 ist die wesentliche Aufgabe der Kollokation als mathematisches Modell die Trennung der einzelnen Anteile in (2.17) durch drei Schritte (MORITZ, 1973a, S. 140), (MÖSER ET AL., 2000, S. 205):

1. *Ausgleichung*: Die Trendparameter werden so bestimmt, dass sich die Trendfunktion $\mathbf{X}\beta$ den Beobachtungen bestmöglich anpasst.
2. *Filterung*: Für die Beobachtung y_i sollen die Trendwerte um den Signalanteil s_i verbessert und das Rauschen u_i herausgefiltert werden, um den unregelmäßigen, aber systematischen Informationsgehalt voll auszuschöpfen.

3. *Prädiktion*: Es sollen an Stellen, an denen keine Beobachtungen vorliegen, die Werte des prädizierten Trendanteils $\mathbf{X}_p \hat{\boldsymbol{\beta}}$ (mit $\mathbf{X}_p$ als Designmatrix der Einflussgrößen der zu prädizierenden Objekte) und das zugehörige prädizierte Signal $\hat{\mathbf{s}}_p$ genutzt werden, um die Ermittlung von Schätzwerten $\mathbf{y}_p$ zu ermöglichen. Daher kann (2.17) formal erweitert werden zu

$$E(\mathbf{y}) = \mathbf{y} + \mathbf{v} = \mathbf{X}\boldsymbol{\beta} + E(\mathbf{s}) + \mathbf{0}E(\mathbf{s}_p) = \mathbf{X}\boldsymbol{\beta} + \mathbf{s} + \mathbf{0}\mathbf{s}_p. \quad (2.18)$$

Mathematisch handelt es sich bei dem letztgenannten Schritt der Prädiktion um eine Interpolationsaufgabe, bei der ein Wert an einer beliebigen Stelle aus den vorhandenen umliegenden Größen bestimmt wird. Zu diesem Zweck werden in der Kollokation im Gegensatz zur Regression stochastische Informationen der Beobachtungen genutzt, die sich letztlich in einem ähnlichen Verhalten benachbarter Residuen widerspiegeln. Da die Residuen in einen informativen Anteil sowie ein nichtinformatives Rauschen getrennt werden, erfolgt die Prädiktion bzgl. des Signalanteils über die Korrelationen der Beobachtungen untereinander. Diese werden durch deren Kovarianzen beschrieben, sodass sich die Beziehungen zwischen den Beobachtungen und dem Signal nach (2.17) allein durch deren stochastischen Zusammenhang ergeben. Das stochastische Modell der Kollokation setzt sich im Gegensatz zum Regressionsansatz aus zwei Komponenten zusammen. Die Kovarianzmatrix der Beobachtungen wird analog zur Regressionsanalyse mit $\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{vv} = \sigma^2 \mathbf{P}^{-1}$ formuliert. Als zweite Komponente wird die VKM $\boldsymbol{\Sigma}_{ss}$ des Signalanteils aufgestellt, wobei die Verteilungsaussage $\mathbf{s} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{ss})$ angenommen wird. Darüber hinaus wird für die Kreuzkorrelationen des Signals und der VKM der Beobachtungen Unkorreliertheit angenommen ($\boldsymbol{\Sigma}_{sy} = \boldsymbol{\Sigma}_{ys} = \mathbf{0}$). Unter diesen Bedingungen und unter Einbeziehung des prädizierten Signals $\mathbf{s}_p$ ergibt sich das stochastische Modell der Kollokation zu

$$\bar{\boldsymbol{\Sigma}} = \begin{vmatrix} \boldsymbol{\Sigma}_{yy} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{ss} & \boldsymbol{\Sigma}_{ss_p} \\ \mathbf{0} & \boldsymbol{\Sigma}_{s_p s} & \boldsymbol{\Sigma}_{s_p s_p} \end{vmatrix} \quad \text{bzw.} \quad \bar{\mathbf{Q}} = \frac{1}{\bar{\sigma}_0^2} \bar{\boldsymbol{\Sigma}} = \begin{vmatrix} \mathbf{Q}_{yy} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{ss} & \mathbf{Q}_{ss_p} \\ \mathbf{0} & \mathbf{Q}_{s_p s} & \mathbf{Q}_{s_p s_p} \end{vmatrix}, \quad \bar{\mathbf{P}} = \bar{\mathbf{Q}}^{-1}, \quad (2.19)$$

wobei $\bar{\boldsymbol{\Sigma}}$ die VKM, $\bar{\sigma}_0^2$ den a priori Varianzfaktor und $\bar{\mathbf{Q}}$ die Kofaktormatrix bezeichnen. Eine der wichtigsten Aufgaben in der Kollokation ist die zutreffende Besetzung der positiv definiten VKM des Signals $\boldsymbol{\Sigma}_{ss}$, die a priori bekannt sein muss. Die Lösung erfolgt i. d. R. über so genannte *Autokovarianzfunktionen*, die auf der Analyse der Residuen des Trendmodells beruhen. Die Kofaktormatrix $\mathbf{Q}_{ss} \left(= \frac{1}{\bar{\sigma}_0^2} \boldsymbol{\Sigma}_{ss} \right)$ des Signals ergibt sich zu

$$\mathbf{Q}_{ss} = \begin{vmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{vmatrix}, \quad (2.20)$$

in der sich die Korrelationskoeffizienten ρ_{ik} auf den Zusammenhang der Signalkomponenten s_i und s_k beziehen. Von besonderer Bedeutung ist in diesem Zusammenhang der Ansatz einer Funktion $\rho_{ik} = f(s_i, s_k)$, die den stochastischen Zusammenhang zutreffend darstellt und durch die Autokovarianzfunktionen erfolgt. Die Tendenz, dass beliebige unmittelbar benachbarte Werte y_i ein ähnliches Verhalten aufweisen, wird dabei als Erhaltensneigung bezeichnet und ist gleichbedeutend mit der stochastischen Abhängigkeit aufeinanderfolgender Werte (NIEMEIER, 2008, S. 45). Der Zusammenhang wird allgemein durch die Autokovarianzfunktion

$$C(t) = \frac{1}{n-t} \sum_{i=1}^{n-t} (y_i - \bar{y})(y_{i+t} - \bar{y}) \quad \Leftrightarrow \quad C(t) = \frac{1}{n-t} \sum_{i=1}^{n-t} \epsilon_i \epsilon_{i+t} \quad (2.21)$$

ausgedrückt, in der n die Anzahl beliebiger Messwerte y_i , t (mit $t \in \{0, 1, 2, \dots, n-t\}$) den Abstand zwischen den Messwerten y_i und $\bar{y}$ den Mittelwert der Messreihe bezeichnen. Numerisch entspricht (2.21) der rekursiven Anwendung der Formel für die Bestimmung von Kovarianzen. Für $t = 0$ ergibt sich die empirische Varianz der Messreihe, die Ergebnisse für $t > 0$ sind numerisch abnehmend mit $C(0) > C(1) > \dots > C(n-t)$. Durch die Normierung von (2.21) mit der Varianz $C(0)$ ergibt sich die *Autokorrelationsfunktion*

$$K(t) = \frac{C(t)}{C(0)}, \quad (2.22)$$

die mit einem Wertebereich zwischen 0,0 bis 1,0 direkt die Korrelation zwischen zwei Messwerten ausdrückt und daher die Besetzung der Signalmatrizen nach (2.21) ermöglicht (HOWIND, 2005, S. 14 f.). Für eine vertiefte Behandlung der Thematik stochastischer Prozesse wird auf die weiterführende Literatur verwiesen (vgl. u. a. KOCH & SCHMIDT (1994), HOWIND (2005), TEUSCH (2006), NIEMEIER (2008), HEUNECKE ET AL. (2013)). Nach der a priori Ableitung der empirischen Autokovarianzfunktion lassen sich mittels der stochastischen Verknüpfung der Signalanteile die Schritte der Ausgleichung, Filterung und letztlich der Prädiktion durchführen. Eine detaillierte Herleitung der Parameterschätzung im einfachen Kollokationsmodell auf Basis des GHM ist in der weiterführenden Fachliteratur zu finden und soll an dieser Stelle nicht vertieft werden (vgl. u. a. MORITZ (1973b), MORITZ (1980), KOCH (2004), NIEMEIER (2008)).

Die Modellunsicherheiten sowohl der klassischen linearen Regression als auch des Kollokationsansatzes beruhen auf stochastischen Ansätzen; auf die zufällige Variabilität in der Modellierung der Beobachtungen und der Residuen wurde hingewiesen. Die Unsicherheiten der als frequentistisch bezeichneten Ansätze werden im Rahmen des Kapitels 3 aufgegriffen. Zusammenfassend kann zunächst festgehalten werden, dass durch die Modelle der multiplen linearen Regression und der einfachen Kollokation die Schätzung von Vergleichswerten im indirekten Vergleichswertverfahren ermöglicht wird und sich damit eine direkte Übertragung der Definition ökonometrischer Modelle auf Problemstellungen der Verkehrswertermittlung erkennen lässt.

3. Grundlagen und Modelle der Bayes-Statistik

Allgemein wird als mathematische Statistik das Teilgebiet bezeichnet, das die Methoden und Verfahren der Statistik mit mathematischen Mitteln analysiert bzw. mit ihrer Hilfe erst begründet. Die Wahrscheinlichkeitstheorie ist dabei ein Teilgebiet der Mathematik, das aus der Formalisierung der Modellierung und der Untersuchung von Zufallsgeschehen hervorgegangen ist. Gemeinsam mit der mathematischen Statistik, die anhand von Beobachtungen zufälliger Vorgänge Aussagen über das zugrunde liegende Modell trifft, bildet sie das mathematische Teilgebiet der Stochastik. Die Anwendung der Wahrscheinlichkeitstheorie dient der mathematischen Modellierung von zufälligen Ereignissen, deren Ergebnisse nicht mit Sicherheit vorhergesagt werden können. Die Wahrscheinlichkeitstheorie bietet nach der Sichtweise der traditionellen Statistik eine Möglichkeit, Ergebnisse von Zufallsexperimenten mathematisch zu modellieren und ihnen ein Wahrscheinlichkeitsmaß für ihr Eintreten zuzuordnen (TAPPE, 2013, S. 1) (KOCH, 2000, S. 3). Entgegen des Grundgedankens der traditionellen Statistik ermöglicht die Bayes-Statistik einen umfassenderen Zugang zum Umgang mit Wahrscheinlichkeiten, da sich die Modellierung nicht auf die Wahrscheinlichkeit des Eintreffens von Ergebnissen eines Zufallsexperimentes beschränkt, sondern auch die Wahrscheinlichkeit von Aussagen berücksichtigt. Generell können sich Aussagen auf zufällige Ereignisse beziehen, sind aber sehr viel allgemeiner gefasst als die einem Ergebnis zugeordneten Wahrscheinlichkeiten mittels traditioneller Statistik. Das Wahrscheinlichkeitsmaß kann in diesem Kontext mit dem Grad der Plausibilität für das Eintreten eines konkreten Ergebnisses gleichgesetzt werden, was im Umkehrschluss direkte Rückschlüsse auf dessen Unsicherheit zulässt (KOCH, 2000, S. 3). Im Mittelpunkt steht das Bayes-Theorem, welches über die Modellierung auf Basis von Wahrscheinlichkeitsverteilungen einen Zugang zur Verarbeitung allgemeiner Aussagen bietet.

In Bezug auf die Forschungsziele ist die Integration der Bayes-Theorie in die Verkehrswertermittlung als zentraler Punkt der Arbeit zu sehen. Abschnitt 3.1 ist der Erläuterung des Grundgedankens Bayesscher Ansätze gewidmet. Insbesondere werden das Bayes-Theorem und die für die Untersuchungen wesentlichen Aspekte behandelt. Aufbauend werden in den Abschnitten 3.2 und 3.3 die in Kapitel 2 vorgestellten klassischen Modelle der Regression und einfachen Kollokation aus Bayesscher Sichtweise formuliert und Lösungsmöglichkeiten aufgezeigt. Abschnitt 3.4 ist als Exkurs aufzufassen und zeigt alternative Lösungsmöglichkeiten der Bayesschen Modelle durch die Nutzung von Monte-Carlo Methoden auf.

3.1. Einführung in die Bayes-Statistik

Ebenso wie bisherige traditionelle statistische (frequentistische) Ansätze ermöglicht die Anwendung der Bayes-Theorie die Berechnung von Schätzwerten für unbekannte Parameter sowie die Angabe von Unsicherheiten. Zu diesem Zweck wird das verallgemeinerte Bayes-Theorem genutzt, welches auf Wahrscheinlichkeitsdichten basiert und je nach Anwendungsfall durch analytische oder numerische Verfahren zu lösen ist.

Für ein Verständnis der Bayes-Theorie ist es notwendig, den Grundgedanken darzulegen und die wesentlichen Unterschiede zu der traditionellen Sichtweise aufzuzeigen. Der Grundgedanke bezieht sich insbesondere auf eine generelle Darstellung von Wahrscheinlichkeiten und allgemeinen Aussagen, die in ihrer Kombination zur Formulierung bedingter Wahrscheinlichkeiten und letztlich zum Bayes-Theorem als zentralem Baustein der Bayes-Inferenz führen.

Die Ausführungen in den folgenden Abschnitten sind lediglich als eine erste Einführung zu verstehen. Herleitungen und vertiefte Darstellungen können und sollen nicht gegeben werden. Die Grundlagen werden insoweit behandelt, wie sie für das Verständnis der Arbeit relevant sind. Vertiefende Darstellungen lassen sich u. a. KOCH (2000), VIERTL (2003), GELMAN ET AL. (2004), KOCH (2007), CHRISTENSEN (2011) sowie TSCHIRK (2014) entnehmen.

3.1.1. Grundlagen und Bayes-Theorem

Die Frage der zutreffenden Anwendung der frequentistischen oder der Bayesschen Statistik lässt sich nicht pauschal beurteilen, sondern stellt vielmehr zwei unterschiedliche Betrachtungsweisen gegenüber. Grundlegende Unterschiede lassen sich in der Inferenz sowie der Interpretation von Wahrscheinlichkeiten und Parametern finden.

Das statistische Schließen aus Daten in einem Modell wird als *Inferenz* bezeichnet. In rein frequentistisch begründeten Ansätzen basieren sämtliche Schlüsse auf den vorliegenden Daten. Potenzielle Fehlerquellen finden sich gemäß den Ausführungen in Kapitel 2 in der Stichprobenziehung sowie in Fehlern in den Beobachtungen (BANDEMER, 1997, S. 113). Als weiteres Charakteristikum sind zudem einzelne Auswertungen i. d. R. unabhängig von weiteren Analysen. Die vertiefte Behandlung mathematischer Inferenz ist an dieser Stelle nicht zielführend und bleibt der weiterführenden Fachliteratur vorbehalten (vgl. u. a. PAPULA (2008), TAPPE (2013)). In der Bayesschen Statistik liegt bereits vor der Auswertung Wissen vor und damit implizit eine bestimmte Erwartung über die aktuellen Ergebnisse. Die Art des Vorwissens ist dabei zunächst nicht relevant; allgemein formuliert kann ein Lernprozess durch die wiederholte Integration von Vorwissen aus zurückliegenden Auswertungen festgehalten werden. Damit stellt sich das Lernen in Bayesschen Anwendungen als kumulativer Prozess dar, der stets auf vergangenen Erfahrungen beruht.

Die Unterscheidung in der Inferenz beruht implizit auf der Betrachtung von *Wahrscheinlichkeiten*, präziser sowohl für frequentistische als auch Bayessche Ansätze in der Angabe von Wahrscheinlichkeiten für Zufallsvariablen. Aus frequentistischer Sichtweise werden Wahrscheinlichkeiten ausschließlich für zufällige Ereignisse eingeführt, die aus Zufallsexperimenten resultieren (KOCH, 2000, S. 1). Damit beziehen sie sich auf eine Interpretation als relative Häufigkeit eines Ereignisses, resultierend aus beliebig oft wiederholbaren Zufallsexperimenten: Die Wahrscheinlichkeit ist dann definiert als die relative Häufigkeit des Ereignisses im Verhältnis zur Anzahl der Wiederholungen.

Im Gegensatz dazu ist die Wahrscheinlichkeit im Rahmen der Bayesschen Statistik als Messung von Unsicherheit aufzufassen. Demnach können auch einzelnen Ereignissen Eintrittswahrscheinlichkeiten zugeordnet werden und sind so losgelöst vom Zwang der wiederholten Zufallsexperimente zu sehen. Der Begriff der Wahrscheinlichkeitsverteilung spielt dabei eine zentrale Rolle und wird im folgenden Abschnitt eingehender betrachtet.

Eng verknüpft mit der Erweiterung des Wahrscheinlichkeitsbegriffs ist die Betrachtung der unbekannten *Parameter* eines Modells: Diese werden aus frequentistischer Sichtweise grundsätzlich als feste Werte angenommen, sodass eine Berücksichtigung lediglich solcher Informationen erfolgt, die mittels klassischer Erwartungswert- und Dispersionsmodelle beschrieben werden (KUTTERER, 2002, S. 29). In der Folge treten Unsicherheiten lediglich in den Schätzern der Parameter auf, nicht aber im eigentlichen Parameter. In der Bayes-Statistik hingegen werden Wahrscheinlichkeiten nicht nur für zufällige Ereignisse, sondern ganz allgemein für Aussagen eingeführt. Werden unbekannte Parameter in einem Modellansatz als allgemeine Aussage aufgefasst, ermöglicht diese Vorgehensweise die Zuordnung einer Eintrittswahrscheinlichkeit. Die unbekannten Parameter werden dadurch als Zufallsvariable definiert; eine Betrachtungsweise, die frequentistischen Anwendungen verschlossen bleibt. Wahrscheinlichkeiten beziehen sich dann auf die Werte der Realisierungen der Parameter und sind nicht lediglich als relative Häufigkeiten zu interpretieren, sondern als Grad der Plausibilität. KOCH (2000, S. 1) weist darauf hin, dass eine allgemeine Aussage zudem nicht ausschließt, dass die Parameter Konstanten repräsentieren dürfen: Allgemein drückt die Wahrscheinlichkeit lediglich den Zustand von Wissen über eine Aussage aus.

Der Begriff der *Aussage* wird analog zur Theorie der Aussagenalgebra definiert: Demnach handelt es sich um einen abstrakten Begriff, der in der Alltagssprache in Sätzen ausgedrückt wird. Die Aussagenalgebra dient der Überführung dieser umgangssprachlichen Beschreibungen in eine wahrscheinlichkeitstheoretisch handhabbare Form. Aussagen können durch *Aussagenvariablen* formalisiert werden, die jeweils eine Aussage abstrakt darstellen und sich im Folgenden durch die Buchstaben A , B und C symbolisieren lassen. Der konkrete Inhalt einer Aussagenvariable ist dabei nicht relevant. In Erweiterung der Grundprinzipien der Aussagenalgebra wird häufig angenommen, dass eine Aussage A wahr ist unter der Bedingung, dass eine weitere Aussage B wahr ist. Ausgedrückt wird die Abhängigkeit durch $A|B$, was den direkten Anknüpfungspunkt zur Wahrscheinlichkeitstheorie darstellt. Grundidee der Wahrscheinlichkeitstheorie ist die Möglichkeit, die so genannte *bedingte Wahrscheinlichkeiten* $A|B$ in ein mathematisches Modell zu integrieren. Der Begriff der bedingten Wahrscheinlichkeit beschreibt die Tatsache, inwiefern bereits vor der Durchführung eines Zufallsexperimentes vorhandene Informationen aktuelle Ergebnisse beeinflussen

können, und wird mit $P(A|B)$ bezeichnet. Dabei bezeichnet P die Wahrscheinlichkeit der Aussage A unter der Bedingung, dass die Aussage B bereits eingetreten ist. P wird auch als Plausibilität von A bezeichnet und ist ein direktes Maß für die Unsicherheit der Aussage. Eine ergänzende Interpretation gibt KOCH (2000, S. 5): Demnach sind bedingte Wahrscheinlichkeiten geeignet, um empirisches Wissen auszudrücken, da durch $P(A|B)$ die Wahrscheinlichkeit angegeben wird, mit der vorhandenes Wissen für weiteres Wissen relevant ist. COX (1946) sowie mit Ergänzungen JAYNES (2003) haben gezeigt, dass sich allein durch logisches und konsistentes Schließen auf Basis von bedingten Wahrscheinlichkeiten die Grundlagen für alle Gesetze der Wahrscheinlichkeitstheorie, die für die Bayes-Statistik benötigt werden, ableiten lassen. Während COX einen Grad für die Plausibilität der Aussage $A|B$ einführt, formuliert JAYNES ergänzend drei grundsätzlichen Anforderungen an die Plausibilität (KOCH, 2000, S. 6):

- a) Der Grad der Plausibilität wird durch reelle Zahlen repräsentiert.
- b) Die qualitative Übereinstimmung mit dem gesunden Menschenverstand wird gefordert.
- c) Das Schließen muss konsistent sein.

Basierend auf der Einführung der Wahrscheinlichkeit als Funktion der Plausibilität stellen COX und JAYNES eine Beziehung zwischen der Plausibilität des Produkts AB und den Plausibilitäten der einzelnen Aussagen A und B unter der Bedingung her, dass eine Aussage C wahr ist. Die Aussage C enthält Zusatzinformationen oder Hintergrundwissen über den Zusammenhang, in dem A und B getroffen werden. Als wesentliches Ergebnis der umfangreichen Ableitungen lässt sich das *Produktgesetz* der Wahrscheinlichkeit

$$P(AB|C) = P(A|C) P(B|AC) = P(B|C) P(A|BC) \quad (3.1)$$

formulieren. Wird mit $P(S|C) = 1$ die Wahrscheinlichkeit der sicheren Aussage (mit Sicherheit ist Aussage A wahr unter der Bedingung, dass C wahr ist) bezeichnet, folgt mit

$$P(A|C) + P(\bar{A}|C) = 1 \quad (3.2)$$

das *Summengesetz* der Wahrscheinlichkeit mit $\bar{A}$ als Negation der Aussage A . Bezeichnet zudem $P(\bar{S}|C) = 0$ die unmögliche Aussage, wird schnell deutlich, dass die Wahrscheinlichkeit $P(A|C)$ gemäß den JAYNES'schen Forderungen eine reelle Zahl zwischen $0 \leq P(A|C) \leq 1$ ist. Wird in einem weiteren Schritt das Produktgesetz (3.1) nach der bedingten Wahrscheinlichkeit $P(A|BC)$ aufgelöst, ergibt sich direkt das fundamentale *Bayes-Theorem* zu

$$P(A|BC) = \frac{P(A|C) P(B|AC)}{P(B|C)}. \quad (3.3)$$

In diesem Zusammenhang wird A als allgemeine Aussage über ein unbekanntes Phänomen interpretiert, B repräsentiert Aussagen, die Informationen über das unbekannte Phänomen enthalten, und die Aussage C stellt Information über zusätzliches (Hintergrund)Wissen dar.

Die Wahrscheinlichkeit $P(A|C)$ wird in der Bayes-Statistik als *Priori*-Wahrscheinlichkeit bezeichnet und gibt das Wissen über ein Phänomen wieder, bevor neue, zusätzliche Informationen erfasst werden. Die Wahrscheinlichkeit $P(B|AC)$ wird als *Likelihood* charakterisiert und beschreibt die Wahrscheinlichkeit der Informationen in B unter der Bedingung der Aussage A über das Phänomen. Mit $P(A|BC)$ wird die *Posteriori*-Wahrscheinlichkeit angegeben, die aus der Modifikation der Priori-Wahrscheinlichkeit durch die Likelihood resultiert und damit die Wahrscheinlichkeit der Aussage über das unbekannte Phänomen wiedergibt, unter der Bedingung, dass die Informationen vorliegen (KOCH, 2000, S. 14). Der Term $P(B|C)$ stellt lediglich eine Normierungskonstante dar, die keine Informationen zu der Posteriori-Wahrscheinlichkeit beiträgt. Aus diesem Grund wird die Normierungskonstante i. d. R. weggelassen, sodass sich das Bayes-Theorem (3.3) formulieren lässt zu

$$P(A|BC) \propto P(A|C) P(B|AC) \quad (3.4)$$

mit der wesentlichen Aussage, dass sich die Posteriori-Wahrscheinlichkeit proportional zu dem Produkt aus Priori-Wahrscheinlichkeit und Likelihood verhält. Das Bayes-Theorem nach (3.3) bzw. (3.4) ist bislang allgemein formuliert

und bezieht sich abstrakt auf Aussagen. Im folgenden Abschnitt soll das verallgemeinerte Bayes-Theorem eingeführt werden, welches durch die Berücksichtigung von Wahrscheinlichkeitsdichten anstatt abstrakter Aussagen die Berücksichtigung numerischer Variablen ermöglicht.

3.1.2. Verallgemeinertes Bayes-Theorem

Für eine praktische Anwendung im Zusammenhang mit der Verkehrswertermittlung sind die bislang abstrakten Aussagen auf numerische Werte von Variablen in Form von reellen Zahlen zu beziehen. Während es sich nach frequentistischer Betrachtungsweise dabei um Zufallsvariablen handelt, gilt dies für Variablen in der Bayes-Statistik nicht, da die allgemeinen Aussagen die Werte beliebiger Variablen beinhalten können. Gemäß den Erläuterungen nach KOCH (2000, S. 17 ff.) soll der Begriff der Zufallsvariable auch in den folgenden Ausführungen zur Bayes-Statistik beibehalten werden, um eine in sich konsistente Terminologie zu gewährleisten. Zufallsvariablen umfassen dann zum einen die unbekannten Parameter eines Modells, die zunächst allgemeingültig durch θ repräsentiert werden sollen. Der Begriff der Aussage nach Abschnitt 3.1.1 bezieht sich dann auf die numerischen Werte der Parameter. Zum anderen sind als Zufallsvariablen die Beobachtungen zu sehen, die nach Abschnitt 2.2.1 (S. 31 ff.) definiert sind.

3.1.2.1. Wahrscheinlichkeitsverteilungen und verallgemeinertes Bayes-Theorem

Bezüglich der Zufallsvariablen ist eine Unterteilung in diskrete und stetige Größen erforderlich. Als diskrete Größen werden Variablen definiert, die ausschließlich endlich viele oder abzählbar unendlich viele Werte annehmen können, sodass für eine beliebige diskrete Zufallsgröße X mit den Werten x_i die Beziehung $x_i \in \mathbb{R}$ für $i \in \{1, \dots, m\}$ gültig ist. Stetige Größen unterliegen in dieser Hinsicht keiner Einschränkung und werden nicht als diskrete Werte, sondern in einem Intervall definiert. Theoretisch kann eine stetige Größe jeden Wert in einem Intervall annehmen, wobei für die folgenden Ausführungen eine stetige Zufallsvariable X mit den Werten $x \in \mathbb{R}$ im Intervall $-\infty < x < \infty$ angenommen wird. Auf Grund der Tatsache, dass sowohl die unbekannten Parameter als auch die Beobachtungen in Anwendungen der Verkehrswertermittlung als stetige Größen anzusehen sind, beziehen sich die folgenden Ausführungen ausschließlich auf stetige Zufallsvariablen. Eine vertiefte Auseinandersetzung mit Verteilungen von diskreten Zufallsvariablen geben z. B. KOCH (2000), KOCH (2004) oder NIEMEIER (2008).

Soll die Wahrscheinlichkeit $P(X < x|C)$ der von x abhängigen Aussage $X < x$ unter der Bedingung der Aussage C ausgedrückt werden, ergibt sich die Beziehung $F(x) = P(X < x|C)$, in der $F(x)$ die *Verteilungsfunktion* der Zufallsvariablen bezeichnet. Die Verteilungsfunktion gibt die Wahrscheinlichkeit an, mit der ein Wert x einer Zufallsvariablen X kleiner ist als ein gegebener Grenzwert und ist als Wahrscheinlichkeit analog zu den Ausführungen des vorherigen Abschnitts definiert im Bereich $0 \leq F(x) \leq 1$. Zudem werden die Eigenschaften der Stetigkeit und Differenzierbarkeit für $F(x)$ unterstellt (NIEMEIER, 2008, S. 17 ff.). Im Vergleich zur Verteilungsfunktion stellt die *Wahrscheinlichkeitsdichte* eine Funktion der Zufallsvariablen X dar, die die Wahrscheinlichkeit angibt, mit der X in den Intervallgrenzen g_u und g_o liegt, sodass $g_u \leq X < g_o$ gilt. Diese Beziehung lässt sich alternativ in drei Aussagen aufteilen: $(X < g_u)$, $(X < g_o)$ sowie $(g_u \leq X < g_o)$. Nach einer Umformung lässt sich über das Summengesetz nach (3.2)

$$P(g_u \leq X < g_o|C) = P(X < g_o|C) - P(X < g_u|C) \quad (3.5)$$

erhalten. Unter Berücksichtigung der allgemeinen Definition der Verteilungsfunktion ergibt sich daraus

$$P(g_u \leq X < g_o|C) = F(g_o) - F(g_u) \quad (3.6)$$

und unter Verwendung des Hauptsatzes der Infinitesimalrechnung

$$F(g_o) - F(g_u) = \int_{-\infty}^{g_o} p(x|C) dx - \int_{-\infty}^{g_u} p(x|C) dx = \int_{g_u}^{g_o} p(x|C) dx, \quad (3.7)$$

woraus

$$\frac{\partial F(x)}{\partial x} = p(x|C) \quad (3.8)$$

folgt (KOCH, 2000, S. 20). Die Bezeichnung $p(\cdot|\cdot)$ wird dabei als Notation für die Wahrscheinlichkeitsdichte von Zufallsvariablen stetiger Verteilungen für die folgenden Erläuterungen eingeführt. Der Zusammenhang zwischen Verteilungsfunktion und Wahrscheinlichkeitsdichte ergibt sich daher aus

$$F(x_g) = \int_{-\infty}^{x_g} p(x|C) dx, \quad (3.9)$$

wobei x_g den Grenzwert bezeichnet (KOCH, 2000, S. 20). NIEMEIER (2008, S. 18) nennt ergänzend die wesentlichen zu erfüllenden Bedingungen:

1. Wahrscheinlichkeiten sind definiert für die Grundgesamtheit aller theoretisch möglichen Beobachtungen (Zufallsvariable).
2. Die Wahrscheinlichkeit für ein Ereignis liegt stets zwischen 0 und 1, sodass $0 \leq p(x|C) \leq 1$.
3. Da die Wahrscheinlichkeit nie größer als 1 sein kann, folgt $\int_{-\infty}^{\infty} p(x|C) dx = 1$.

Die bisherigen Ausführungen beziehen sich auf die Darstellung von Verteilungs- und Dichtefunktion lediglich einer einzelnen Zufallsvariablen. Für die Anwendung der Bayes-Statistik in der Verkehrswertermittlung ist aufgrund der gleichzeitigen Auswertung von n Beobachtungen und der Ermittlung von m unbekannten Parametern die Erweiterung auf mehrdimensionale – sogenannte multivariate – stetige Verteilungen notwendig. Allgemein besitzt eine n -dimensionale Zufallsvariable $X_1, \dots, X_n$ Werte $x_1, \dots, x_n$ in den Intervallen $-\infty < x_k < \infty$ mit $k \in \{1, \dots, n\}$. Die Verteilungsfunktion ist dann gegeben durch

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n|C). \quad (3.10)$$

Aufbauend auf (3.8) folgt für die Dichtefunktion im mehrdimensionalen Fall direkt

$$\frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} = p(x_1, \dots, x_n|C), \quad (3.11)$$

die entsprechend die Bedingungen

$$p(x_1, \dots, x_n|C) \geq 0, \quad (3.12)$$

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_1, \dots, x_n|C) dx_1, \dots, dx_n = 1 \quad (3.13)$$

erfüllen muss (NIEMEIER, 2008, S. 18 f). Eine kompaktere Schreibweise ergibt sich, wenn die Zufallsvariablen x_i mit $-\infty < x_i < \infty$ und $i \in \{1, \dots, n\}$ in einem Vektor $\mathbf{x} = |x_1 \dots x_n|'$ zusammengefasst werden: Unter dieser Voraussetzung vereinfacht sich $p(x_1, \dots, x_n|C)$ zu $p(\mathbf{x}|C)$. Sollen aus einer gemeinsamen multivariaten Dichtefunktion $p(\mathbf{x}|C)$ die Wahrscheinlichkeiten einzelner Zufallsvariable oder einer Gruppe von Zufallsvariablen bestimmt werden, muss die *Randverteilung* ermittelt werden. Randverteilungen ermöglichen die Reduzierung eines mehrdimensionalen Zufallsvektors $\mathbf{x} = |x_1 \dots x_n|' \in \mathbb{R}^n$ auf einen Zufallsvektor $\mathbf{x}_r = |x_{n-i+1} \dots x_n|' \in \mathbb{R}^i$ mit der verringerten Dimension $i < n$. Für die Rückführung auf eine geringere Dimension ist die Integration über die zu eliminierenden Zufallsvariablen des Ausgangsvektors $\mathbf{x}$ notwendig (NEUMANN, 2009, S. 37):

$$\begin{aligned} p(\mathbf{x}_r|C) &= p(x_{n-i+1}, \dots, x_n) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_1, \dots, x_{n-i}, x_{n-i+1}, \dots, x_n) dx_1 \dots dx_{n-i}. \end{aligned} \quad (3.14)$$

Von Bedeutung sind Randverteilungen insbesondere für die Herleitung von Verteilungen einzelner Parameter. Dieser Aspekt wird in den Abschnitten 3.2 und 3.3 sowie in Kapitel 4 nochmals aufgegriffen.

Werden im Folgenden zwei mehrdimensionale Zufallsvektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ an Stelle der allgemeinen Aussagen A und B eingeführt, ergibt sich aus dem Produktgesetz nach (3.1) für den Vektor $\mathbf{x}_1$ die bedingte Wahrscheinlichkeitsdichte

$$p(\mathbf{x}_1|\mathbf{x}_2, C) = \frac{p(\mathbf{x}_1, \mathbf{x}_2|C)}{p(\mathbf{x}_2|C)}. \quad (3.15)$$

Die Wahrscheinlichkeit $p(\mathbf{x}_1|\mathbf{x}_2, C)$ wird als bedingte Dichte für $\mathbf{x}_1$ bezeichnet unter der Bedingung, dass $\mathbf{x}_2$ sowie die Hintergrundinformation C gegeben sind. Wie zu ersehen ist, wird die bedingte Verteilung von $\mathbf{x}_1$ erhalten, indem die gemeinsame Verteilung $p(\mathbf{x}_1, \mathbf{x}_2|C)$ durch die Randverteilung für $\mathbf{x}_2$ dividiert wird (KOCH, 2000, S. 27).

Die Anwendungen der Bayes-Statistik in der Verkehrswertermittlung zielen darauf ab, die Wahrscheinlichkeitsdichten der unbekannten Parameter als Maß für die Unsicherheit zu bestimmen. Zugrunde liegen dabei die Beobachtungen, die in diesem Fall konkret durch (normierte) Kaufpreise gegeben sind. Werden unbekannte Modellparameter zunächst allgemein in dem Zufallsvektor $\boldsymbol{\theta}$ und die Beobachtungen in dem Zufallsvektor $\mathbf{y}$ zusammengefasst, kann (3.15) konkreter formuliert werden zu

$$p(\boldsymbol{\theta}|\mathbf{y}, C) = \frac{p(\boldsymbol{\theta}, \mathbf{y}|C)}{p(\mathbf{y}|C)} \quad \text{bzw.} \quad p(\mathbf{y}|\boldsymbol{\theta}, C) = \frac{p(\mathbf{y}, \boldsymbol{\theta}|C)}{p(\boldsymbol{\theta}|C)}. \quad (3.16)$$

Werden die jeweiligen Gleichungen nach $p(\boldsymbol{\theta}, \mathbf{y}|C)$ aufgelöst und gleichgesetzt, ergibt sich nach Umformung das für stetige Zufallsvariablen *verallgemeinerte Bayes-Theorem*

$$p(\boldsymbol{\theta}|\mathbf{y}, C) = \frac{p(\boldsymbol{\theta}|C) p(\mathbf{y}|\boldsymbol{\theta}, C)}{p(\mathbf{y}|C)}, \quad (3.17)$$

wobei der Term $p(\mathbf{y}|C)$ wiederum eine Normierungskonstante bedeutet, die unabhängig von den Parametern $\boldsymbol{\theta}$ ist und daher vernachlässigt werden kann. Wird aus Gründen der Lesbarkeit und Vereinfachung zudem auf die Darstellung der Aussage C verzichtet, ergibt sich das verallgemeinerte Bayes-Theorem zu

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}). \quad (3.18)$$

Der Term $p(\boldsymbol{\theta}|\mathbf{y})$ wird entsprechend der vorherigen Ausführungen als Posteriori-Dichte bezeichnet, aus der die unbekannten Parameter zu schätzen sind. Sie ergibt sich proportional aus dem Produkt der Priori-Dichte $p(\boldsymbol{\theta})$, – die Informationen über die Parameter enthält, bevor die Daten $\mathbf{y}$ erhoben werden – und der Likelihood-Funktion $p(\mathbf{y}|\boldsymbol{\theta})$, die die vorliegenden aktuellen Daten enthält (VIERTL, 2003, S. 171 f.). Da die Daten vorliegen, wird die Likelihood-Funktion nicht als Funktion der Daten $\mathbf{y}$ interpretiert, sondern als Funktion der Parameter $\boldsymbol{\theta}$ (KOCH, 2000, S. 33). Durch die Priori- und Posteriori-Dichten wird für die unbekannten Parameter damit die Wahrscheinlichkeit bestimmt, dass die numerischen Werte in gewissen Bereichen liegen. Dies spiegelt somit die Plausibilität der Parameter wider.

3.1.2.2. Eigenschaften von Priori-Dichten

Um mittels Bayesscher Inferenz verlässliche Ergebnisse zu erhalten, ist die Auswahl der Priori-Dichte von besonderer Bedeutung. Wie in den folgenden Abschnitten gezeigt wird, entsprechen sich die Ergebnisse aus frequentistischer und Bayesscher Berechnung, sofern keine Vorinformation in Form einer Priori-Dichte vorliegt. Der wesentliche Unterschied zwischen den Ansätzen lässt sich daher an der Integration von Priori-Wissen festmachen.

Bayessche Inferenz beruht im Wesentlichen auf dem Prozess, einen Übergang von der Priori-Dichte $p(\boldsymbol{\theta})$ zu der Posteriori-Dichte $p(\boldsymbol{\theta}|\mathbf{y})$ zu formulieren. Unterstellt wird dabei, dass die jeweiligen Dichten in einem generellen Zusammenhang stehen. Als Beispiel führen GELMAN ET AL. (2004, S. 36) an, dass unter Berücksichtigung von Vorinformation für die Posteriori-Dichte eine verringerte Varianz erwartet werden kann. Ohne Herleitung ergibt sich für Erwartungswert $E(\cdot)$ und Varianz $V(\cdot)$ unbekannter Parameter aus bedingten Dichten

$$E(\boldsymbol{\theta}) = E(E(\boldsymbol{\theta}|\mathbf{y})), \quad V(\boldsymbol{\theta}) = E(V(\boldsymbol{\theta}|\mathbf{y})) + V(E(\boldsymbol{\theta}|\mathbf{y})). \quad (3.19)$$

Insbesondere die Berechnung der Varianz zeigt, dass die Posteriori-Varianz im Durchschnitt geringer ist als die Priori-Varianz, und zwar um einen Betrag, der von der Variation der durchschnittlichen Posteriori-Ergebnisse abhängt (GELMAN ET AL., 2004, S. 37). Da die Priori-Dichte den Kenntnisstand über die unbekannten Parameter θ wiedergibt, bevor die Beobachtungen $\mathbf{y}$ erfasst werden, kann die Kombination von Priori-Dichte mit Likelihood-Funktion eine signifikante Verringerung von Unsicherheiten bewirken. Generell lassen sich für Priori-Dichten verschiedene Eigenschaften definieren, von denen im Kontext der Arbeit eine Unterscheidung in nichtinformative und informative sowie konjugierte und nichtkonjugierte Dichten erfolgen soll.

Nichtinformative Priori-Dichten finden dann Anwendung, wenn keinerlei Vorinformationen über ein betrachtetes Phänomen vorliegen. Mathematisch formuliert, können die unbekannten Parameter sich dann in dem Wertebereich $-\infty < \theta < \infty$ bewegen. Für die Priori-Dichte folgt daraus, dass sie mit $p(\theta) \propto c$ angenommen wird, wobei c eine Konstante repräsentiert. Im Hinblick auf die Forderung (3.13) wird diese Art der Dichte als uneigentliche Priori-Dichte bezeichnet, da sich

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(\theta_1, \dots, \theta_n) d\theta_1, \dots, d\theta_n \neq 1 \quad (3.20)$$

ergibt. Diese Eigenschaft deutet auf eine Nichtnormierbarkeit des Integrals hin. JEFFREYS (1961, S. 179) sowie KOCH (1990, S. 11) entwickeln zur Problemlösung allgemeine Formeln für die Ableitung nichtinformativer Priori-Dichten, welche an dieser Stelle nicht weiter vertieft werden sollen. Im Gegensatz zu nichtinformativen Priori-Dichten werden *informative Priori-Dichten* dann eingeführt, wenn Vorwissen über die unbekannten Parameter vorhanden ist. Die Lösung des Bayes-Theorems bzw. die Ermittlung der Posteriori-Dichte hängt dann wesentlich davon ab, wie die Priori-Dichte modelliert wird. Da Wahrscheinlichkeitsdichten, wie in (3.13) beispielhaft gezeigt, mittels Integralen formuliert werden, wird schnell ersichtlich, dass sich je nach Wahl von Priori-Dichte und Likelihood-Funktion komplexe Gleichungen ergeben, die sich nicht mehr geschlossen lösen lassen. In diesem Fall müssen Simulations-Algorithmen herangezogen werden, die eine numerische Approximation der Integrale und erst damit die Lösung des Bayes-Theorems ermöglichen. Häufig werden zu diesem Zweck Monte-Carlo Simulationen verwendet, die in Abschnitt 3.4 aufgegriffen werden. Für eine analytisch handhabbare Lösung der Integrale können alternativ *konjugierte Priori-Dichten* genutzt werden. Als konjugierte Dichte wird eine Priori-Dichte bezeichnet, die mit der Likelihood-Funktion multipliziert auf eine Posteriori-Dichte führt, die zu derselben Verteilungsfamilie wie die Priori-Dichte selbst gehört (VIERTL, 2003, S. 176), (TSCHIRK, 2014, S. 92). Wesentliche Vorteile sind im Gegensatz zu den nicht analytisch lösbaren *nichtkonjugierten Dichten* ein einfacher Zugang zu der Posteriori-Dichte in Form vereinfachter Berechnung sowie gute Approximationseigenschaften (GELMAN ET AL., 2004, S. 41 f.). Die Herleitung der Bayesschen Modelle in den Abschnitten 3.2 und 3.3 basiert auf der Verwendung konjugierter Priori-Dichten, sodass auf eine Beispielrechnung an dieser Stelle verzichtet werden kann. Im folgenden Abschnitt wird auf die Aspekte der Parameterschätzung sowie der Konfidenzregionen mittels Bayesscher Inferenz eingegangen.

3.1.3. Parameterschätzung und Konfidenzbereiche im Bayesschen Modell

Unabhängig von der Wahl konjugierter oder nichtkonjugierter Priori-Dichten ermöglicht die Auswertung einer Posteriori-Dichte die Schätzung der unbekannten Parameter sowie die Angabe von Bereichen, in denen die Parameter mit hoher Wahrscheinlichkeit liegen. Insofern entsprechen die Möglichkeiten der Bayesschen Inferenz denen der frequentistischen Ansätze, lediglich Herleitung und Interpretation weichen voneinander ab. In den folgenden Unterabschnitten werden die Bayessche Punktschätzung sowie die Herleitung von spezifischen Konfidenzregionen behandelt.

3.1.3.1. Bayessche Punktschätzung

KOCH (2000, S. 63) formuliert die Wahl der geeigneten Schätzung als Problem der Entscheidungstheorie: Demnach ist darüber zu entscheiden, ob eine Parameterschätzung zutrifft und ein Phänomen geeignet beschreibt. Die Formulierung einer Entscheidungsregel $\delta(\mathbf{y})$ wird als *Bayes-Strategie* bezeichnet. Entscheidungsregeln in Bayesschen Modellen geben an, wie das Ergebnis in Abhängigkeit von vorliegenden Daten $\mathbf{y}$ beurteilt wird: Hierfür werden die Kosten betrachtet, die eine gewählte, auslösende Entscheidungsregel $\delta(\mathbf{y})$ verursacht. Wird in Abhängigkeit von

$\boldsymbol{\theta}$ und $\delta(\mathbf{y})$ eine Kostenfunktion $L(\boldsymbol{\theta}, \delta(\mathbf{y}))$ eingeführt, können die posteriori zu erwartenden Kosten über die Posteriori-Dichte $p(\boldsymbol{\theta}|\mathbf{y})$ berechnet werden zu

$$E\{L(\boldsymbol{\theta}, \delta(\mathbf{y}))\} = \int_{\Theta} L(\boldsymbol{\theta}, \delta(\mathbf{y})) p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}, \quad (3.21)$$

wobei Θ den Parameterraum von $\boldsymbol{\theta}$ mit $\boldsymbol{\theta} \in \Theta$ beschreibt (KOCH, 2000, S. 64). Gleichung (3.21) wird alternativ auch als *Bayessches Risiko* bezeichnet (VIERTL, 2003, S. 183). Ziel der Bayes-Strategie ist es, die Entscheidungsregel $\delta(\mathbf{y})$ derart zu wählen, dass die posteriori zu erwartenden Kosten des Bayesschen Risikos nach (3.21) minimal werden. Während in der Fachliteratur zu diesem Zweck verschiedene Kostenfunktionen wie die Kostenfunktion der absoluten Fehler oder Null-Eins-Kosten gegeben sind (vgl. u. a. KOCH (2000), GELMAN ET AL. (2004), CHRISTENSEN (2011)), werden in der vorliegenden Arbeit ausschließlich quadratische Kostenfunktionen betrachtet.

Ziel einer quadratischen Kostenfunktion ist es, Schätzwerte für die Parameter zu ermitteln, deren Abweichungen zu den wahren Werten möglichst minimal sind. Zu diesem Zweck wird der Erwartungswert der quadrierten Differenz zwischen den Schätzwerten $\bar{\boldsymbol{\theta}}$ und den wahren Werten $\boldsymbol{\theta}$ als Kostenfunktion angesetzt, sodass sich

$$L(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) = (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}\bar{\boldsymbol{\theta}}}^{-1} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) \quad (3.22)$$

mit $\boldsymbol{\Sigma}_{\boldsymbol{\theta}\bar{\boldsymbol{\theta}}}^{-1}$ als VKM der Parameter ergibt. Wird der Erwartungswert für die Kostenfunktion gebildet, lässt sich dieser als Spezialfall von (3.21) zu

$$E\{L(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})\} = \int_{\Theta} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}\bar{\boldsymbol{\theta}}}^{-1} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \quad (3.23)$$

formulieren (KOCH, 2000, S. 66). Die Lösung von (3.23) stellt sich als Extremwertaufgabe mit der Forderung $\partial E\{L(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})\} / \partial \bar{\boldsymbol{\theta}} = 0$ dar. Als Lösung der Extremwertaufgabe ergibt sich

$$\bar{\boldsymbol{\theta}} = \int_{\Theta} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = E(\boldsymbol{\theta}|\mathbf{y}), \quad (3.24)$$

wobei an dieser Stelle für die Herleitung auf die Ausführungen von VIERTL (2003, S. 186) oder TSCHIRK (2014, S. 113 f.) verwiesen wird. Die Beziehung (3.24) wird generell als *Punktschätzer* der unbekannten Parameter und auf Grund des Ansatzes einer quadratischen Kostenfunktion explizit als *Bayes-Schätzer* bezeichnet. Der Bayes-Schätzer gibt somit an, welche Werte für $\bar{\boldsymbol{\theta}}$ posteriori zu erwarten sind, wenn die Beobachtungen $\mathbf{y}$ vorliegen. Wie in (3.24) ersichtlich, hängen Punktschätzer entscheidend von Priori-Dichte und Likelihood-Funktion ab. In Ergänzung zu der reinen Punktschätzung kann die Genauigkeit für $\bar{\boldsymbol{\theta}}$ angegeben werden mittels der VKM

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}\bar{\boldsymbol{\theta}}}(\boldsymbol{\theta}|\mathbf{y}) = E\{(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{y}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{y}))'\} = \int_{\Theta} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \quad (3.25)$$

(KOCH, 2000, S. 66). Neben der reinen Punktschätzung ist ebenso von Interesse, in welchen Bereichen die Parameter mit einer vorgegebenen Wahrscheinlichkeit liegen. Erst durch diese Angabe kann die Unsicherheit einer Schätzung im vorstehend definierten Sinne umfassend angegeben werden. Im folgenden Abschnitt werden daher Bayessche Konfidenzintervalle eingeführt.

3.1.3.2. Bayessche Intervallschätzung

Weder Parameterschätzung noch Varianz nach den im vorherigen Abschnitt erläuterten Gleichungen geben eine Aussage darüber, wie dicht ein Schätzwert tatsächlich an seinem wahren Wert liegt. In der klassischen Statistik werden Intervalle angegeben, die die unbekannten Parameter mit einer vorgegebenen Wahrscheinlichkeit enthalten; aus Bayesscher Sichtweise werden hingegen Intervalle formuliert, die die Parameter mit der gewählten Wahrscheinlichkeit überdecken (KUTTERER, 2002, S. 29). Entgegen der Bezeichnung der Konfidenzintervalle aus frequentistischer Sichtweise wird in Bayesschen Anwendungen dazu eine sogenannte *Region höchster Posteriori-Dichte* (englisch (eng.) *Highest Posterior Density Interval (HPDI)*) genutzt. Durch die Tatsache, dass die unbekannten Parameter

als Zufallsvariable mit einer zugeordneten Wahrscheinlichkeitsdichte interpretiert werden können, ergibt sich ein direkter, leicht verständlicher Zugang für die Konstruktion der Bayesschen Konfidenzintervalle. Basierend auf der Posteriori-Dichte $p(\boldsymbol{\theta}|\mathbf{y})$ lässt sich durch die Beziehung

$$P(\boldsymbol{\theta} \in \Theta_u|\mathbf{y}) = \int_{\Theta_u} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \quad (3.26)$$

die Wahrscheinlichkeit angeben, dass der Parametervektor $\boldsymbol{\theta}$ in dem Unterraum Θ_u des gesamten Parameterraums Θ mit $\Theta_u \subset \Theta$ liegt. Für eine Verwendung als Konfidenzregion wird der Unterraum Θ_u in der Folge derart festgelegt, dass er einen großen Teil der Wahrscheinlichkeit für maximale Dichten enthält, i. d. R. 95 % – oder alternativ formuliert, ein HPDI stellt einen Konfidenzbereich für $\boldsymbol{\theta}$ mit maximalen Werten der Posteriori-Dichte dar (VIERTL, 2003, S. 180). Ein HPDI für einen Bayes-Schätzer ergibt sich dann in Verbindung mit dem spezifischen Unterraum $\Theta_B \subset \Theta$ zu

$$P(\boldsymbol{\theta} \in \Theta_B|\mathbf{y}) = \int_{\Theta_B} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = 1 - \alpha \quad (3.27)$$

und der Festlegung von $1 - \alpha$ als Konfidenzniveau. Zusätzlich muss die Beziehung

$$p(\boldsymbol{\theta}_1|\mathbf{y}) \geq p(\boldsymbol{\theta}_2|\mathbf{y}) \quad \text{für} \quad \boldsymbol{\theta}_1 \in \Theta_B, \boldsymbol{\theta}_2 \notin \Theta_B \quad (3.28)$$

erfüllt sein, wobei mit $\boldsymbol{\theta}_1$ und $\boldsymbol{\theta}_2$ zwei sich gegenseitig ausschließende Parametervektoren bezeichnet werden. Ein HPDI fordert somit, dass die Dichte eines jeden Parameters innerhalb des Unterraums Θ_B gleich oder größer als die Dichte eines Parameters außerhalb des Unterraums ist (KOCH, 2000, S. 71 f.).

In praktischen Anwendungen ergibt sich, dass sich frequentistisch begründete Konfidenzintervalle und HPDIs oftmals sehr ähnlich sind. Insbesondere in Fällen, in denen eine Posteriori-Verteilung von der Stichprobe der Beobachtungen $\mathbf{y}$ dominiert wird, ergeben sich kaum Unterschiede. Entscheidend ist daher die Formulierung der Priori-Dichte, der ein erheblicher Einfluss beigemessen werden muss, um signifikant unterschiedliche Ergebnisse der Ansätze zu erhalten (TSCHIRK, 2014, S. 126 f.). Ebenso wie im Fall der im vorherigen Abschnitt formulierten Punktschätzung hängt die Angabe konkreter HPDIs von der Definition der Priori-Dichte und der Likelihood-Funktion ab. Konjugierte Priori-Dichten erlauben eine relativ einfache Ableitung der entsprechenden Schätzer und Intervalle, während die Lösungen für nichtkonjugierte Dichten oftmals nur über numerische Verfahren herbeigeführt werden können. Da in Abhängigkeit der angesetzten Wahrscheinlichkeitsdichten eine nicht zu begrenzende Vielzahl von konkreten Formulierungen für (3.24) sowie (3.27) möglich ist, wird an dieser Stelle lediglich auf die abstrakten Formulierungen eingegangen. Die für die Arbeit relevanten Ableitungen werden im Zuge der Modellformulierung erläutert. Im folgenden Abschnitt 3.2 wird zunächst die Bayessche Variante der multiplen Regressionsanalyse behandelt.

3.2. Lineares Modell der Bayesschen Regression für konjugierte Verteilungsdichten

Die Schätzung unbekannter Parameter und weiterer Momente nach den Ausführungen in Abschnitt 3.1.3 setzt voraus, dass die Beziehungen zwischen den Parametern und den Beobachtungen ebenso definiert werden wie die statistischen Eigenschaften der Beobachtungen. Während die Ausführungen in vorstehendem Abschnitt allgemein gehalten sind, werden sie sich im Folgenden konkret auf die Vergleichswertermittlung bezogen.

Die bislang abstrakt gehaltenen unbekannten Parameter $\boldsymbol{\theta}$ sind gemäß den bisherigen klassischen Ansätzen der Regression und Kollokation als Trendkoeffizienten $\boldsymbol{\beta}$ sowie als Varianzfaktor der Gewichtseinheit σ^2 zu interpretieren. Als Beobachtungen $\mathbf{y}$ werden entsprechend die realen (ggf. normierten) Kaufpreise bezeichnet, während zugehörige Posteriori-Schätzwerte $\bar{\mathbf{y}}$ Vergleichswerte symbolisieren. Für das zu formulierende Modell lässt sich mit diesen Definitionen unter Berücksichtigung des verallgemeinerten Bayes-Theorems (3.18) die Beziehung

$$\underbrace{p(\boldsymbol{\beta}, \sigma^2|\mathbf{y})}_{\text{Posteriori-Dichte}} \propto \underbrace{p(\boldsymbol{\beta}) \cdot p(\sigma^2)}_{\text{Priori-Dichten}} \cdot \underbrace{p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)}_{\text{Likelihood-Funktion}} \quad (3.29)$$

angeben, für die geeignete Priori-Dichten sowie eine Likelihood-Funktion gewählt werden müssen. Hauptaugenmerk liegt auf der Nutzung konjugierter Wahrscheinlichkeitsdichten, sodass für die Posteriori-Dichte eine analytisch handhabbare Lösung hergeleitet werden kann. Zu diesem Zweck erfolgt im nächsten Abschnitt zunächst eine Definition der Likelihood-Funktion. Um ein analytisch lösbares Modell zu erhalten, wird nachfolgend die Priori-Dichte geeignet bestimmt. In der zunächst rein theoretisch fundierten Modellentwicklung ist es dabei unerheblich, auf welchem Weg die Priori-Dichte generiert wird. Ausgehend vom Bayes-Theorem werden Priori-Dichte und Likelihood-Funktion kombiniert, um abschließend die Posteriori-Dichte sowie die resultierenden Schätzwerte abzuleiten.

3.2.1. Modelldefinition

Durch die Definition des Zusammenhangs von Parametern und Beobachtungen wird das Modell der Auswertung bestimmt. Grundsätzlich wird ein linearer Zusammenhang zwischen den Beobachtungen und den wertbeeinflussenden Merkmalen als Einflussgrößen unterstellt, was zur Anwendung linearer Modelle führt. Entsprechend sind die Parameter β als lineare Koeffizienten aufzufassen. Liegen nichtlineare Zusammenhänge vor, können diese entweder durch Linearisierung im Modell berücksichtigt oder alternativ nichtlineare Modelle angesetzt werden (KOCH, 2007, S. 85). Unter den genannten Annahmen kann das lineare Modell analog zu (2.11) und (2.12) formuliert werden mit

$$E(\mathbf{y}|\beta) = \mathbf{X}\beta, \quad \Sigma_{\mathbf{y}\mathbf{y}}(\mathbf{y}|\sigma^2) = \sigma^2 \mathbf{P}^{-1}, \quad (3.30)$$

wobei der Erwartungswert von $\mathbf{y}$ unter Bezugnahme auf die Ausführungen in Abschnitt 3.1.1 nunmehr als bedingt durch die Parameter β und die VKM der Beobachtungen bedingt durch den Varianzfaktor σ^2 beschrieben werden. Entsprechend müssen auch die Residuen bedingt durch die Parameter angegeben werden, sodass sich

$$E(\epsilon|\beta) = 0, \quad \Sigma_{\epsilon\epsilon}(\epsilon|\beta, \sigma^2) = \Sigma_{\mathbf{y}\mathbf{y}}(\mathbf{y}|\sigma^2) = \sigma^2 \mathbf{P}^{-1}, \quad (3.31)$$

ergibt. Der Varianzfaktor selbst wird als unbekannte Zufallsvariable angenommen. Lösungen von Modellen mit bekanntem Varianzfaktor sind für die hier dargestellten Anwendungen in der Verkehrswertermittlung von untergeordneter Bedeutung und werden entsprechend nicht weiter behandelt. Eine ausführliche Lösung findet sich bei KOCH (2007).

3.2.1.1. Likelihood-Funktion der Bayesschen Regression

Die Wahrscheinlichkeitsverteilung der Kaufpreise als Beobachtungen bestimmt die zu wählende Dichte für die Likelihood-Funktion $p(\mathbf{y}|\beta, \sigma^2)$. Wie bereits im Zusammenhang mit der Definition der klassischen Modelle in Abschnitt 2.3.2 (S. 43 ff.) erläutert, wird für die klassischen Modelle der Regression und Kollokation angenommen, dass die Beobachtungen einer Normalverteilung folgen. Auf Grund einer rechentechnisch günstigen Handhabung wird die Likelihood-Funktion der Beobachtungen $\mathbf{y}$ daher als *multivariate Normalverteilung (NV)* formuliert, die sich als mehrdimensionale Erweiterung der Normalverteilung angeben lässt:

$$\mathbf{y}|\beta, \sigma^2 \sim NV(\mathbf{X}\beta, \sigma^2 \mathbf{P}^{-1}). \quad (3.32)$$

Eine Begründung für die Wahl der Normalverteilung aus Sicht der Verkehrswertermittlung erfolgt im Zuge der Entwicklung der Auswertestrategie in Kapitel 4. Die Wahrscheinlichkeitsdichtefunktion für lineare Modelle in dem Bayes-Theorem nach (3.29) ist dann gegeben durch

$$p(\mathbf{y}|\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2} (\det(\mathbf{P}))^{-1/2}} e^{-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' \mathbf{P} (\mathbf{y} - \mathbf{X}\beta)} \quad (3.33)$$

(FAHRMEIR ET AL., 2009, S. 464). In dieser Gleichung wird die Annahme dargestellt, dass die Beobachtungen $\mathbf{y}$ sich als Linearkombination $\mathbf{X}\beta$ der unbekannten Parameter ergeben, unter der Bedingung, dass die Parameter β und σ^2 gegeben sind (KOCH, 2000, S. 85). Für eine Vereinfachung der Folgeberechnungen kann für die Varianz σ^2 der *Präzisionsparameter* $\tau = 1/\sigma^2$ eingeführt werden, sodass sich die Likelihood-Funktion umformulieren lässt zu

$$p(\mathbf{y}|\boldsymbol{\beta}, \tau) = (2\pi)^{-n/2} (\det \mathbf{P})^{1/2} \tau^{n/2} e^{-\frac{\tau}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{P} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})} \quad (3.34)$$

(CHRISTENSEN, 2011, S. 111), (KOCH, 2000, S. 111). Mit (3.34) ist die Likelihood-Funktion in (3.29) vollständig spezifiziert. Im folgenden Abschnitt werden ergänzend die Priori-Dichten für die Parameter und die Varianz angegeben. Dabei wird für die Priori-Dichte zunächst eine gemeinsame Wahrscheinlichkeitsdichte formuliert, auf deren Basis die jeweiligen Randverteilungen anzugeben sind.

3.2.1.2. Priori-Dichten der Bayesschen Regression

Für die Notation im weiteren Verlauf der Arbeit wird festgelegt, dass unterstrichene Variablen für die Kennzeichnung als Priori-Wissen genutzt werden, während durch einen Überstrich Posteriori-Variablen gekennzeichnet sind.

Im Fall *informativer Priori-Dichten* wird angenommen, dass für die unbekannten Parameter ein Erwartungswertvektor $E(\boldsymbol{\beta}) = \underline{\boldsymbol{\beta}}$ vorliegt und Vorinformation über die Unsicherheit der Parameter durch die VKM $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ gegeben ist. Für den unbekannten Varianzfaktor liegt im Fall der informativen Priori-Dichte Vorwissen durch $E(\sigma^2) = \underline{\sigma^2}$ sowie zu seiner Varianz $V(\sigma^2) = V_{\underline{\sigma^2}}$ vor. In Anbetracht der Vorinformation $\underline{\sigma^2}$ ergibt sich damit die VKM der Parameter zu $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \underline{\sigma^2} \mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}$. Aufgrund der geforderten Eigenschaften der konjugierten Dichten wird für die gemeinsame Priori-Verteilung der Parameter $\boldsymbol{\beta}$ und des unbekannten Varianzfaktors σ^2 eine *Normal-Gammaverteilung* (NG) gewählt,

$$\boldsymbol{\beta}, \tau \sim NG(\underline{\boldsymbol{\beta}}, \mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}, \underline{b}, \underline{p}), \quad (3.35)$$

wobei wiederum der Präzisionsparameter τ anstelle der Varianz angegeben wird. Die Wahrscheinlichkeitsdichtefunktion ergibt sich damit zu

$$p(\boldsymbol{\beta}, \tau | \underline{\boldsymbol{\beta}}, \mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}, \underline{b}, \underline{p}) = (2\pi)^{-n/2} \left(\det \mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}} \right)^{-1/2} \underline{b}^{\underline{p}} (\Gamma(\underline{p}))^{-1} \tau^{n/2 + \underline{p} - 1} e^{-\frac{\tau}{2} \left\{ 2\underline{b} + (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}})' \mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}}) \right\}} \quad (3.36)$$

für die Verteilungsparameter $\underline{b} > 0$, $\underline{p} > 0$, $0 < \tau < \infty$, $-\infty < \beta_i < \infty$ und in der $\Gamma(p)$ die Gamma-Funktion $\Gamma(p) = (p-1)!$ bezeichnet (KOCH, 2007, S. 61). Da die Parameter und der Varianzfaktor einzeln zu betrachten sind, können zusätzlich die Randverteilungen angegeben werden. Im Falle einer gemeinsamen Normal-Gammaverteilung lässt sich nach KOCH (2000, S. 56) die Randverteilung der Parameter $\boldsymbol{\beta}$ durch eine (*multivariate*) *t-Verteilung* (*t*) mit

$$\boldsymbol{\beta} \sim t(\boldsymbol{\beta} | \underline{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}\boldsymbol{\beta}}, \nu) \quad (3.37)$$

und ν als Parameter angeben, womit sich die Wahrscheinlichkeitsdichtefunktion entsprechend (FAHRMEIR ET AL., 2009, S. 467) zu

$$p(\boldsymbol{\beta} | \underline{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}\boldsymbol{\beta}}, \nu) = \frac{\nu^{\nu/2} \Gamma((k + \nu)/2) \left(\det \boldsymbol{\Sigma}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} \right)^{1/2}}{\pi^{k/2} \Gamma(\nu/2)} \left(\nu + (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}})' \boldsymbol{\Sigma}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}}) \right)^{-(k + \nu)/2} \quad (3.38)$$

ergibt. Die Randdichte des Präzisionsparameters τ lässt sich als *Gamma-Verteilung* (*G*) mit entsprechender Wahrscheinlichkeitsdichtefunktion formulieren:

$$\tau \sim G(\underline{b}, \underline{p}), \quad p(\tau | \underline{b}, \underline{p}) = \frac{\underline{b}^{\underline{p}}}{\Gamma(\underline{p})} \tau^{\underline{p} - 1} e^{-b\tau}, \quad (3.39)$$

wobei $\underline{b} > 0$, $\underline{p} > 0$, $0 < \tau < \infty$, $-\infty < \beta_i < \infty$ angesetzt wird (KOCH, 2000, S. 56). Da sich der Präzisionsparameter per Definition als Inverse des Varianzfaktors bestimmt, folgt für den Varianzfaktor aus (3.39) eine *Invertierte Gamma-Verteilung* (*IG*) mit

$$\sigma^2 \sim IG(\underline{b}, \underline{p}), \quad p(\sigma^2 | \underline{b}, \underline{p}) = \frac{\underline{b}^{\underline{p}}}{\Gamma(\underline{p})} \left(\frac{1}{\sigma^2} \right)^{\underline{p}+1} e^{-\underline{b}/\sigma^2} \quad (3.40)$$

für $\underline{b} > 0$, $\underline{p} > 0$, $0 < \sigma^2 < \infty$. Erwartungswert des Varianzfaktors und seine Varianz lassen sich dann berechnen durch

$$E(\sigma^2) = \frac{\underline{b}}{\underline{p} - 1} = \underline{\sigma}^2, \quad V(\sigma^2) = \frac{\underline{b}^2}{(\underline{p} - 1)^2 (\underline{p} - 2)} = V_{\underline{\sigma}^2}, \quad (3.41)$$

woraus nach hier nicht dargestellter Umformung

$$\underline{p} = \frac{(\underline{\sigma}^2)^2}{V_{\underline{\sigma}^2}} + 2, \quad \underline{b} = \left(\frac{(\underline{\sigma}^2)^2}{V_{\underline{\sigma}^2}} + 1 \right) \underline{\sigma}^2 \quad (3.42)$$

folgen (FAHRMEIR ET AL., 2009, S. 460 f.), (KOCH, 2007, S. 118). Deutlich wird, dass sich die Parameter der gemeinsamen Normal-Gamma-Verteilung der Priori-Dichte nach (3.35) direkt aus der Vorinformation bestimmen lassen. Die bislang dargestellten Wahrscheinlichkeitsdichten und Zusammenhänge beziehen sich ausschließlich auf den Fall vorhandenen Vorwissens. Werden hingegen *nichtinformative Priori-Dichten* betrachtet, das heißt (d. h.) es liegt kein Vorwissen über die Parameter sowie den Varianzfaktor vor, können die nichtinformativen Priori-Dichten

$$p(\beta) \propto c \quad \text{für } -\infty < \beta_i < \infty, i \in \{1, \dots, n\}, \quad (3.43)$$

$$p(\tau) \propto 1/\tau \quad \text{für } 0 < \tau < \infty \quad (3.44)$$

in das Bayes-Theorem eingeführt werden (KOCH, 2007, S. 111).

An dieser Stelle kann jeweils nur eine verkürzte Darstellung der Zusammenhänge gegeben werden. Für ein vertiefendes Studium sowie insbesondere die Herleitungen der Wahrscheinlichkeitsdichten wird auf die weiterführende Literatur verwiesen (vgl. u. a. GELMAN ET AL. (2004), KOCH (2007), CHRISTENSEN (2011)). Sowohl für den informativen als auch den nichtinformativen Fall kann mit den gegebenen Priori-Dichten sowie der Likelihood-Funktion nach (3.34) das Bayes-Theorem nach (3.18) gelöst werden. Die analytische Lösung wird im folgenden Abschnitt dargestellt.

3.2.2. Analytische Lösung im Modell mit unbekanntem Varianzfaktor

Werden die Priori-Dichte und die Likelihood-Funktion im Bayes-Theorem kombiniert, lässt sich die Posteriori-Dichte ableiten. Basierend auf der Posteriori-Dichte können dann Schätzwerte der unbekannten Parameter abgeleitet sowie Aussagen über deren Unsicherheit getroffen werden. Wie im vorherigen Abschnitt bereits deutlich wird, muss in der Ableitung der Posteriori-Dichte des vorliegenden Modells grundsätzlich zwischen Ansätzen mit informativem und nichtinformativem Vorwissen unterschieden werden. Es ergeben sich entsprechend zwei verschiedene Modelle, die als einer der zentralen Aspekte der Arbeit gegenübergestellt werden sollen. Im folgenden Abschnitt wird zunächst die analytische Lösung der Bayesschen Regression, basierend auf informativen Priori-Dichten, behandelt.

3.2.2.1. Informative Lösung der Bayesschen Regression

Für einen ersten Ansatz werden die mit der gemeinsamen Dichte nach (3.35) im vorherigen Abschnitt definierten informativen Priori-Dichten eingeführt. Wird die gemeinsame Wahrscheinlichkeitsdichtefunktion der Parameter und des Varianzfaktors nach (3.38) mit der Likelihood-Funktion nach (3.34) multipliziert, ergibt sich die gemeinsame Wahrscheinlichkeitsfunktion der Posteriori-Dichte zu

$$p(\beta, \tau | \mathbf{y}) \propto \tau^{u/2 + \underline{p} - 1} e^{\left\{ -\frac{\tau}{2} \left(2\underline{b} + (\beta - \underline{\beta})' \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} (\beta - \underline{\beta}) \right) \right\}} \cdot \tau^{n/2} e^{\left\{ -\frac{\tau}{2} ((\mathbf{y} - \mathbf{X}\beta)' \mathbf{P} (\mathbf{y} - \mathbf{X}\beta)) \right\}}, \quad (3.45)$$

$$p(\beta, \tau | \mathbf{y}) \propto \tau^{n/2 + \underline{p} + u/2 - 1} e^{\left\{ -\frac{\tau}{2} \left(2\underline{b} + (\beta - \underline{\beta})' \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} (\beta - \underline{\beta}) + (\mathbf{y} - \mathbf{X}\beta)' \mathbf{P} (\mathbf{y} - \mathbf{X}\beta) \right) \right\}}. \quad (3.46)$$

Aufgrund der Eigenschaft als konjugierte Dichte folgt die Posteriori-Dichte damit ebenso wie die eingeführte Priori-Dichte einer Normal-Gammaverteilung:

$$\beta, \tau \sim NG(\bar{\beta}, \mathbf{Q}_{\bar{\beta}\bar{\beta}}, \bar{b}, \bar{p}). \quad (3.47)$$

Auf eine explizite Herleitung der Lösung des vorstehenden Bayes-Theorems wird verzichtet. Nach KOCH (2007, S. 61) ergeben sich durch Umformung und Auflösung der Gleichung die einzelnen Parameter der Posteriori-Dichte (3.47), die für die Berechnung von Schätzwerten benötigt werden:

$$\bar{\beta} = \left(\mathbf{X}' \mathbf{P} \mathbf{X} + \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} \right) \left(\mathbf{X}' \mathbf{P} \mathbf{y} + \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} \underline{\beta} \right), \quad (3.48)$$

$$\mathbf{Q}_{\bar{\beta}\bar{\beta}} = \left(\mathbf{X}' \mathbf{P} \mathbf{X} + \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} \right)^{-1}, \quad (3.49)$$

$$\bar{b} = \left(2 \left[(\underline{\sigma}^2)^2 / V_{\underline{\sigma}^2} + 1 \right] \underline{\sigma}^2 + (\underline{\beta} - \bar{\beta})' \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} (\underline{\beta} - \bar{\beta}) + (\mathbf{y} - \mathbf{X}\bar{\beta})' \mathbf{P} (\mathbf{y} - \mathbf{X}\bar{\beta}) \right) / 2, \quad (3.50)$$

$$\bar{p} = \left(n + 2 (\underline{\sigma}^2)^2 / V_{\underline{\sigma}^2} + 4 \right) / 2. \quad (3.51)$$

Analog zu den Ausführungen für die Priori-Verteilung gilt für die Randverteilungen der Posteriori-Dichte, dass sich die Randverteilung der Trendkoeffizienten als multivariate t -Verteilung ergibt. Diese folgt entsprechend der Formulierung

$$\beta | \mathbf{y} \sim t(\bar{\beta}, \bar{b} \mathbf{Q}_{\bar{\beta}\bar{\beta}} / \bar{p}, 2\bar{p}) \quad \text{mit} \quad E(\beta | \mathbf{y}) = \bar{\beta}. \quad (3.52)$$

Für die Schätzung der Posteriori-Parameter der Trendkoeffizienten ergibt sich unter Ansatz einer quadratischen Kostenfunktion daraus ohne Herleitung

$$\bar{\beta} = \left(\mathbf{X}' \mathbf{P} \mathbf{X} + \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} \right) \left(\mathbf{X}' \mathbf{P} \mathbf{y} + \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} \underline{\beta} \right) \quad (3.53)$$

(KOCH, 2007, S. 119). Zu erkennen ist deutlich, wie das Vorwissen – repräsentiert durch die Kofaktormatrix $\mathbf{Q}_{\underline{\beta}\underline{\beta}}$ sowie die Priori-Trendkoeffizienten $\underline{\beta}$ – die geschätzten Posteriori-Trendkoeffizienten beeinflusst. Insbesondere die Kofaktormatrix $\mathbf{Q}_{\underline{\beta}\underline{\beta}}$ ist Träger der Unsicherheiten über die Priori-Parameter und pflanzt diese auf die aktuelle Schätzung fort.

Für die Auswirkungen der erweiterten Unsicherheitsbetrachtung muss zusätzlich eine Schätzung für den Posteriori-Varianzfaktor angegeben werden. Die Randverteilung des Varianzfaktors lässt sich analog zum Priori-Varianzfaktor aufgrund der Eigenschaft als konjugierte Dichte durch eine Invertierte Gamma-Verteilung formulieren: So ergibt sich für $\sigma^2 | \mathbf{y} \sim IG(\bar{b}, \bar{p})$ mit dem Erwartungswert $E(\sigma^2 | \mathbf{y}) = \bar{b} / (\bar{p} - 1)$ die Bayessche Schätzung mit

$$\begin{aligned} \bar{\sigma}^2 &= \left(n + 2 (\underline{\sigma}^2)^2 / V_{\underline{\sigma}^2} + 4 \right)^{-1} \\ &\cdot \left\{ 2 \left[(\underline{\sigma}^2)^2 / V_{\underline{\sigma}^2} + 1 \right] \underline{\sigma}^2 + (\underline{\beta} - \bar{\beta})' \mathbf{Q}_{\underline{\beta}\underline{\beta}}^{-1} (\underline{\beta} - \bar{\beta}) + (\mathbf{y} - \mathbf{X}\bar{\beta})' \mathbf{P} (\mathbf{y} - \mathbf{X}\bar{\beta}) \right\} \end{aligned} \quad (3.54)$$

sowie die Varianz des Varianzfaktors zu

$$V(\sigma^2 | \mathbf{y}) = V_{\bar{\sigma}^2} = \frac{2 (\bar{\sigma}^2)^2}{n + 2 (\underline{\sigma}^2)^2 / V_{\underline{\sigma}^2}}. \quad (3.55)$$

Die VKM der geschätzten Trendkoeffizienten lässt sich dann nach (KOCH, 2007, S. 120) angeben zu

$$\Sigma_{\beta\beta}(\beta|\mathbf{y}) = \Sigma_{\bar{\beta}\bar{\beta}} = \bar{\sigma}^2 \left(\mathbf{X}'\mathbf{P}\mathbf{X} + \mathbf{Q}_{\bar{\beta}\bar{\beta}}^{-1} \right)^{-1}. \quad (3.56)$$

Mit der Schätzung der Trendkoeffizienten sowie des Varianzfaktors ist die Bayessche Regression im Modell mit unbekanntem Varianzfaktor und informativer Priori-Dichte gelöst. Das Modell kann im folgenden zudem genutzt werden, um die ausgeglichenen Beobachtungen zu schätzen, die der Gleichung $\bar{\mathbf{y}} = \mathbf{X}\bar{\beta}$ folgen. Die Verteilung der geschätzten Beobachtungen folgt als lineare Transformation aus der Designmatrix $\mathbf{X}$ und der Parameter $\bar{\beta}$ ebenfalls einer multivariaten t -Verteilung der Form

$$\bar{\mathbf{y}}|\mathbf{y} \sim t(\mathbf{X}\bar{\beta}, \bar{b}(\mathbf{X}\mathbf{Q}_{\bar{\beta}\bar{\beta}}\mathbf{X}')/\bar{p}, 2\bar{p}) \quad (3.57)$$

(NIEMEIER, 2008, S. 141). Hieraus ergibt sich die VKM der Posteriori-Schätzwerte zu

$$\Sigma_{\mathbf{y}\mathbf{y}}(\bar{\mathbf{y}}|\mathbf{y}) = \Sigma_{\bar{\mathbf{y}}\bar{\mathbf{y}}} = \bar{\sigma}^2 (\mathbf{X}\mathbf{Q}_{\bar{\beta}\bar{\beta}}\mathbf{X}')^{-1} \quad (3.58)$$

Neben der reinen Schätzung der unbekannten Modellparameter steht häufig auch die Prädiktion neuer Schätzwerte im Mittelpunkt. Ziel des indirekten Vergleichswertverfahrens ist es, Vergleichswerte für Objekte vorherzusagen, die bislang noch nicht am Markt gehandelt wurden und von denen lediglich die wesentlichen wertbeeinflussenden Merkmale erfasst werden können. Basierend auf dem beschriebenen mathematischen Modell, welches die durchschnittlichen Verhältnisse eines gewählten Teilmarktes beschreibt, sind daher Prädiktionswerte abzuleiten, denen spezielle Prädiktionsdichten zugrunde liegen. Im Folgenden bezeichnet $\mathbf{y}_p$ einen beliebigen Prädiktionsvektor, in dem die zu prädizierenden Vergleichswerte von Wertermittlungsobjekten enthalten sind; die Designmatrix $\mathbf{X}_p$ enthält die zugehörigen wertbeeinflussenden Merkmale, die bekannt sind und entsprechend des Modells gesetzt werden. Die Formulierung einer Prädiktivdichte erfolgt dann auf Grundlage der Posteriori-Dichte $p(\beta, \sigma^2|\mathbf{y})$. Eine gemeinsame Wahrscheinlichkeitsdichtefunktion der zu prädizierenden Beobachtungen und der Parameter ist nach KOCH (1990, S. 49) gegeben durch

$$p(\mathbf{y}_p, \boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}_p|\boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}|\mathbf{y}), \quad (3.59)$$

wobei $\boldsymbol{\theta} = [\beta, \sigma^2]'$. Weiter wird angenommen, dass $\mathbf{y}_p$ derselben Verteilung folgt wie $\mathbf{y}$, im vorliegenden Fall einer multivariaten Normalverteilung. Unter diesen Voraussetzungen gibt KOCH (1990, S. 358) die Randdichte von $\mathbf{y}_p$ an mit

$$p(\mathbf{y}_p|\mathbf{y}) = \int_{\boldsymbol{\theta}} p(\mathbf{y}_p|\boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}) d\boldsymbol{\theta}. \quad (3.60)$$

Die Lösung des Integrals liefert unter Annahme der multivariaten Normalverteilung von $\mathbf{y}_p$ den Erwartungswert $E(\cdot)$ und die VKM der posteriori prädiktiven Dichte zu

$$\begin{aligned} E(\mathbf{y}_p|\sigma^2, \mathbf{y}) &= E(E(\mathbf{y}_p|\beta, \sigma^2, \mathbf{y})|\sigma^2, \mathbf{y}) \\ &= E(\mathbf{X}_p\beta|\sigma^2, \mathbf{y}) \\ &= \mathbf{X}_p\bar{\beta} \end{aligned} \quad (3.61)$$

$$\begin{aligned} E(\Sigma_{\mathbf{y}_p\mathbf{y}_p}|\sigma^2, \mathbf{y}) &= E(V(\mathbf{y}_p|\beta, \sigma^2, \mathbf{y})|\sigma^2, \mathbf{y}) + V(E(\mathbf{y}_p|\beta, \sigma^2, \mathbf{y})|\sigma^2, \mathbf{y}) \\ &= E(\sigma^2\mathbf{I}|\sigma^2, \mathbf{y}) + V(\mathbf{X}_p\beta|\sigma^2, \mathbf{y}) \\ &= \sigma^2(\mathbf{I} + \mathbf{X}_p\mathbf{Q}_{\bar{\beta}\bar{\beta}}\mathbf{X}_p') \end{aligned} \quad (3.62)$$

(GELMAN ET AL., 2004, S. 357 f.). Mit der prädiktiven Dichte wird das entwickelte Bayessche Regressionsmodell vollständig angegeben. Im folgenden Abschnitt wird die Lösung unter Verwendung nichtinformativer Priori-Dichten aufgezeigt.

3.2.2.2. Nichtinformative Lösung der Bayesschen Regression

Für die nichtinformative Lösung werden die Priori-Dichten genutzt, die mit (3.96) für die Trendkoeffizienten sowie mit (3.97) für den Präzisionsparameter gegeben sind. Werden die Priori-Dichten im Bayes-Theorem mit der Likelihood-Funktion (3.34) zusammengeführt, lässt sich die Posteriori-Dichte ohne Herleitung formulieren zu

$$p(\beta, \tau | \mathbf{y}) \propto \tau^{n/2-1} e^{-\frac{\tau}{2} (\mathbf{y} - \mathbf{X}\beta)' \mathbf{P} (\mathbf{y} - \mathbf{X}\beta)}. \quad (3.63)$$

Nach Umformung lässt sich zeigen, dass die Posteriori-Dichte einer Normal-Gammaverteilung mit

$$\beta, \tau | \mathbf{y} \sim NG \left(\bar{\beta}, (\mathbf{X}' \mathbf{P} \mathbf{X})^{-1}, (n-u) \bar{\sigma}^2 / 2, (n-u) / 2 \right). \quad (3.64)$$

folgt und in der mit n die Anzahl der Beobachtungen und mit u die Anzahl der unbekannten Parameter bezeichnet werden (KOCH, 2007, S. 111). Wie im vorstehenden Fall der informativen Priori-Dichten lässt sich die Randverteilung der Parameter β durch eine multivariate t -Verteilung mit

$$\beta | \mathbf{y} \sim t \left(\bar{\beta}, (\mathbf{X}' \mathbf{P} \mathbf{X})^{-1}, n-u \right). \quad (3.65)$$

bestimmen, deren Erwartungswert mit $E(\beta | \mathbf{y}) = \bar{\beta}$ gegeben ist. Unter Berücksichtigung einer quadratischen Kostenfunktion ergibt sich gemäß (3.24) daher die nichtinformative Bayes-Schätzung der Parameter mit

$$\bar{\beta} = (\mathbf{X}' \mathbf{P} \mathbf{X})^{-1} \mathbf{X}' \mathbf{P} \mathbf{y} \quad (3.66)$$

und die Schätzung der VKM der Parameter mit

$$\Sigma_{\beta\beta}(\beta | \mathbf{y}) = \Sigma_{\bar{\beta}\bar{\beta}} = \frac{n-u}{n-u-2} \bar{\sigma}^2 (\mathbf{X}' \mathbf{P} \mathbf{X})^{-1}. \quad (3.67)$$

Diese Lösung entspricht der Schätzung nach dem klassischen Ansatz der Regressionsanalyse, was durch die Modellformulierung (3.30) sowie durch die Nichtbeachtung von Priori-Wissen zu erwarten ist. Für die VKM ergibt sich ein Vorfaktor, der jedoch für große Werte von $(n-u)$ gegen 1 strebt und lediglich bei kleineren Stichprobenumfängen eine Veränderung gegenüber der klassischen Schätzung bewirkt.

Für den Präzisionsparameter lässt sich die Randverteilung wie bereits beschrieben als Gamma-Verteilung angeben, sodass für den Posteriori-Varianzfaktor eine Invertierte Gamma-Verteilung mit

$$\sigma^2 | \mathbf{y} \sim IG \left((n-u) \bar{\sigma}^2 / 2, (n-u) / 2 \right), \quad (3.68)$$

folgt, deren Erwartungswert mit $E(\sigma^2 | \mathbf{y}) = ((n-u) / (n-u-2)) \bar{\sigma}^2$ gegeben ist. Für die Bayes-Schätzung des Varianzfaktors ergibt sich daraus

$$\bar{\sigma}^2 = \frac{(n-u)}{(n-u-2)} \frac{(\mathbf{y} - \mathbf{X}\bar{\beta})' \mathbf{P} (\mathbf{y} - \mathbf{X}\bar{\beta})}{n-u}, \quad (3.69)$$

welche sich von der klassischen Lösung nur durch einen Vorfaktor unterscheidet. Die Varianz des Varianzfaktors wird abschließend mit

$$V(\sigma^2 | \mathbf{y}) = V_{\bar{\sigma}^2} = \frac{2(n-u)^2 (\bar{\sigma}^2)^2}{(n-u-2)(n-u-4)}. \quad (3.70)$$

angegeben (KOCH, 2007, S. 112). Die VKM der geschätzten Beobachtungen folgt wie im klassischen Fall der multiplen linearen Regressionsanalyse als lineare Transformation aus der Designmatrix $\mathbf{X}$ und der Parameter $\bar{\beta}$ zu

$$\Sigma_{\mathbf{y}\mathbf{y}}(\bar{\mathbf{y}} | \mathbf{y}) = \Sigma_{\bar{\mathbf{y}}\bar{\mathbf{y}}} = \bar{\sigma}^2 \left(\mathbf{X} (\mathbf{X}' \mathbf{P} \mathbf{X})^{-1} \mathbf{X}' \right)^{-1} \quad (3.71)$$

(NIEMEIER, 2008, S. 141).

Als prädiktive Dichten gelten weiterhin die Ausführungen zu (3.61) und (3.62) unter Berücksichtigung der im nichtinformativen Fall abweichenden Posteriori-Kofaktormatrix der Parameter mit $\mathbf{Q}_{\bar{\beta}\bar{\beta}} = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}$.

3.2.3. Konfidenzregion der Parameter und Schätzwerte

Für eine Beurteilung der Unsicherheiten der Modellparameter sind neben den Informationen aus der VKM der geschätzten Parameter die HPDIs nach der in Abschnitt 3.1.3.2 (S. 56 folgende (f.)) erläuterten Bayesschen Sichtweise von Interesse. Wird (3.27) im Ansatz der Regressionsanalyse formuliert, ergibt sich die Beziehung

$$P(\beta \in B|\mathbf{y}) = \int_B p(\beta|\mathbf{y}) d\beta = 1 - \alpha, \quad (3.72)$$

in der B den Parameterraum für β bezeichnet. Um eine Beurteilung der HPDIs der einzelnen Parameter zu ermöglichen, muss vorstehende Gleichung für einzelne Parameter des Parametervektors β umformuliert werden:

$$P(\beta_i \in B|\mathbf{y}) = \int_B p(\beta_i|\mathbf{y}) d\beta = 1 - \alpha \quad \text{für } i \in \{1, \dots, u\}. \quad (3.73)$$

(ALKHATIB & WEITKAMP, 2012, S. 100). Die den Parametern als Randverteilung zugrunde liegende multivariate t -Verteilung nach (3.52) ist als die multivariate Verallgemeinerung einer t -Verteilung zu sehen. Dieser Zusammenhang hat zur Folge, dass eine beliebige Zufallsvariable, die durch eine verallgemeinerte t -Verteilung beschrieben wird, durch Standardisierung in eine t -Verteilung transformiert werden kann. Die t -Verteilung stellt sich damit als Spezialfall der multivariaten t -Verteilung dar, da in (3.38) $k = 1$, $\beta = \beta_i$, $\bar{\beta} = \bar{\beta}_i$ und $(\mathbf{Q}_{\bar{\beta}\bar{\beta}})_{ii} = f$ mit $i \in \{1, \dots, u\}$ (unter Berücksichtigung der Substitution von Priori- durch Posteriori-Größen) gesetzt werden kann, sodass sich diese als

$$p(\beta_i|\bar{\beta}_i, 1/f, \nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi}\Gamma(\nu/2)} \left(\frac{f}{\nu}\right)^{1/2} \left(1 + \frac{f}{\nu}(\beta_i - \bar{\beta}_i)^2\right)^{-(\nu+1)/2} \quad (3.74)$$

darstellen lässt und im vorliegenden Fall $\nu = 2\bar{p}$ ist (KOCH, 2000, S. 55). Die Zufallsvariable β_i folgt damit der Dichte $\beta_i \sim t(\bar{\beta}_i, 1/f, \nu)$. Durch Standardisierung wird die Variable β_i nach β_i^t durch

$$\begin{aligned} \beta_i^t &= \sqrt{f}(\beta_i - \bar{\beta}_i) \\ \text{mit } \sqrt{f}(\beta_i - \bar{\beta}_i) &\sim t(\nu) \end{aligned} \quad (3.75)$$

transformiert. Mit der Beziehung $\partial\beta_i/\partial\beta_i^t = 1/\sqrt{f}$ ergibt sich daraus nach KOCH (1999, S. 93) die Dichte der Zufallsvariablen β_i^t :

$$p(\beta_i^t|\nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi}\Gamma(\nu/2)} \left(1 + \frac{(\beta_i^t)^2}{\nu}\right)^{-(\nu+1)/2}. \quad (3.76)$$

Durch Vergleich mit der Dichte einer t -Verteilung wird deutlich, dass das Ergebnis β_i^t der transformierten Zufallsvariablen der t -Verteilung mit dem Parameter ν folgt mit $\beta_i^t \sim t(\nu)$ (KOCH, 2000, S. 55). Ein HPDI des Parameters β_i lässt sich dann angeben mit

$$P\left(-t_{1-\alpha;\nu} < (\beta_i - \bar{\beta}_i) \sqrt{f} < t_{1-\alpha;\nu}\right) = 1 - \alpha, \quad (3.77)$$

in der $t_{1-\alpha;\nu}$ das Quantil der t -Verteilung bezeichnet. Durch Auflösen lässt sich schließlich die Konfidenzregion als HPDI für eine einzelne Komponente des Parametervektors ableiten:

$$P\left(\bar{\beta}_i - t_{1-\alpha;\nu}/\sqrt{f} < \beta_i < \bar{\beta}_i + t_{1-\alpha;\nu}/\sqrt{f}\right) = 1 - \alpha. \quad (3.78)$$

Für eine umfassende Beurteilung der Auswirkungen von Unsicherheiten auf das Schätzergebnis können abschließend die HPDIs der ausgeglichenen Beobachtungen $\bar{\mathbf{y}}$ sowie der prädizierten Werte $\bar{\mathbf{y}}_p$ angegeben werden. Der

Ansatz lässt sich insofern einfach ableiten, als das die geschätzten Werte wie beschrieben ebenfalls einer multivariaten t -Verteilung folgen. Für einzelne ausgeglichene Beobachtungen $\bar{y}_j$ mit $j \in \{1, \dots, n\}$ ergibt sich daher in Konsequenz zu den bisherigen Ausführungen

$$P\left(\bar{y}_j - t_{1-\alpha;\nu}/\sqrt{f} < y_j < \bar{y}_j + t_{1-\alpha;\nu}/\sqrt{f}\right) = 1 - \alpha \quad (3.79)$$

$$\text{mit } 1/f = \bar{b} \left(\mathbf{Q}_{\bar{\mathbf{y}}\bar{\mathbf{y}}} \right)_{jj} / \bar{p}, \quad (3.80)$$

sowie für die prädizierten Werte $\bar{y}_{p,k}$ mit $k \in \{1, \dots, v\}$ und v als Anzahl der Prädiktionsobjekte

$$P\left(\bar{y}_{p,k} - t_{1-\alpha;\nu}/\sqrt{f} < y_{p,k} < \bar{y}_{p,k} + t_{1-\alpha;\nu}/\sqrt{f}\right) = 1 - \alpha \quad (3.81)$$

$$\text{mit } 1/f = \bar{b} \left(\mathbf{Q}_{\bar{\mathbf{y}}_p \bar{\mathbf{y}}_p} \right)_{kk} / \bar{p}. \quad (3.82)$$

Die in diesem Abschnitt erläuterten theoretischen Grundlagen der Bayesschen Regression sind als erweiterter Ansatz zum klassischen Schätzprinzip aufzufassen. Im folgenden Abschnitt wird zusätzlich die einfache Kollokation um den Bayesschen Ansatz erweitert.

3.3. Lineares Modell der Bayesschen Kollokation für konjugierte Verteilungsdichten

In Abschnitt 2.3.2.2 (S. 45 ff.) wird die einfache Kollokation als Modellerweiterung zur klassischen linearen Regressionsanalyse eingeführt; ebenso kann die Bayessche Kollokation aus dem linearen Modell der Bayesschen Regression entwickelt werden. Alternativ wird das Modell in der Fachliteratur auch als gemischtes Modell oder Modell der Prädiktion und Filterung bezeichnet; für eine konsequente Terminologie in der vorliegenden Arbeit wird der Begriff der Bayesschen Kollokation beibehalten. Die Modellformulierung wird insoweit vorgenommen, wie sie i. S. der Forschungsziele relevant ist. Der erläuterte Grundgedanke der einfachen Kollokation mit den Einzelschritten der Ausgleichung, Filterung und Prädiktion kann in Bezug auf die Bayessche Kollokation analog beibehalten werden. Im Gegensatz zum klassischen Ansatz werden die unbekannten Parameter wie auch im Fall der Bayesschen Regression als Zufallsvariable behandelt, denen eine Wahrscheinlichkeitsverteilung zugeordnet werden kann. Ebenso wird die Signalkomponente nunmehr mit in die Priori-Dichte eingeführt:

$$\underbrace{p(\boldsymbol{\beta}, \mathbf{s}, \sigma^2 | \mathbf{y})}_{\text{Posteriori-Dichte}} \propto \underbrace{p(\boldsymbol{\beta}) \cdot p(\mathbf{s}) \cdot p(\sigma^2)}_{\text{Priori-Dichten}} \cdot \underbrace{p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{s}, \sigma^2)}_{\text{Likelihood-Funktion}}. \quad (3.83)$$

Die Posteriori-Dichte enthält unter diesen Voraussetzungen die gesamten Informationen, die für die Bayessche Inferenz benötigt werden (KOCH, 1994, S. 464). Für eine einfache Zugänglichkeit der Bayesschen Kollokation werden wiederum konjugierte Wahrscheinlichkeitsdichten genutzt.

3.3.1. Modelldefinition

Die Modelldefinition der Bayesschen Kollokation basiert in ihrer theoretischen Begründung zunächst auf einem gemischten Modell der Prädiktion und Filterung nach (2.13). Wie auch im Fall der einfachen Kollokation kann jedoch die Vereinfachung $\mathbf{s} = \mathbf{Z}\boldsymbol{\gamma}$ getroffen werden, falls durch die Beobachtungen $\mathbf{y}$ die Signale direkt messbar sind. Während insbesondere in KOCH (1994, S. 466 ff.) sowie KOCH (2007, S. 129 ff.) ausführliche Herleitung des ursprünglichen, gemischten Modells zu finden sind, wird an dieser Stelle direkt auf das vereinfachte Modell mit $\mathbf{s} = \mathbf{Z}\boldsymbol{\gamma}$ eingegangen. Wird vergleichbar zu (3.30) zunächst ein allgemeines lineares Modell der Form

$$E(\mathbf{y} | \boldsymbol{\beta}^*) = \mathbf{y} + \mathbf{v} = \mathbf{X}^* \boldsymbol{\beta}^* \quad (3.84)$$

$$\text{mit } \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}}(\mathbf{y} | \sigma^2) = \boldsymbol{\Sigma}_{\mathbf{v}\mathbf{v}}(\mathbf{v} | \boldsymbol{\beta}^*, \sigma^2) = \sigma^2 \mathbf{P}^{*-1} \quad (3.85)$$

eingeführt, zeigt sich durch eine geeignete Substitution mit

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{X} & \mathbf{I} \\ [n \times u] & [n \times n] \end{bmatrix}, \quad \boldsymbol{\beta}^* = \begin{bmatrix} \boldsymbol{\beta} & \mathbf{s} \\ [u \times 1] & [n \times 1] \end{bmatrix}', \quad \mathbf{P}^{*-1} = \mathbf{H}^{-1} = (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1}, \quad (3.86)$$

dass sich das lineare Modell der Bayesschen Kollokation vergleichbar zum Modell der einfachen Kollokation (2.17) durch

$$E(\mathbf{y}|\boldsymbol{\beta}, \mathbf{s}) = \mathbf{y} + \mathbf{v} = \mathbf{X}\boldsymbol{\beta} + \mathbf{s} \quad (3.87)$$

$$\text{mit } E(\mathbf{v}|\boldsymbol{\beta}, \mathbf{s}) = \mathbf{0}, \quad \boldsymbol{\Sigma}_{vv}(\mathbf{v}|\boldsymbol{\beta}, \mathbf{s}, \sigma^2) = \boldsymbol{\Sigma}_{yy}(\mathbf{y}|\sigma^2) = \sigma^2 \mathbf{H}^{-1} = \sigma^2 (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \quad (3.88)$$

darstellen lässt. In dem so definierten Modell bezeichnet wie im Fall der einfachen Kollokation $\mathbf{X}\boldsymbol{\beta}$ den Trendanteil, während der Vektor $\mathbf{s}$ die Signalkomponenten beschreibt, die den geschätzten Beobachtungen des Trends hinzuzuaddieren sind. Wie aus den Gleichungen ersichtlich, werden die Beobachtungen $\mathbf{y}$ als bedingt durch die Parameter $\boldsymbol{\beta}$ sowie das Signal $\mathbf{s}$ angegeben, während das Rauschen im stochastischen Modell zusätzlich durch den unbekannten Varianzfaktor σ^2 bedingt wird. Das Modell nach (3.87) bildet die Grundlage für die Formulierung der Beziehungen von Beobachtungen, Parametern und Signal durch Wahrscheinlichkeitsdichten. Für die Anwendung des Bayes-Theorems nach (3.83) wird im folgenden Abschnitt zunächst die Likelihood-Funktion definiert.

3.3.1.1. Likelihood-Funktion der Bayesschen Kollokation

Die Wahrscheinlichkeitsverteilung der Kaufpreise als Beobachtungen $\mathbf{y}$ wird wie in den bisherigen drei Modellansätzen als normalverteilt angesetzt. Die Dichte der Likelihood-Funktion $p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{s}, \sigma^2)$ ergibt sich daher unter der Annahme einer multivariaten Normalverteilung mit

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{s}, \sigma^2 \sim NV(\mathbf{X}^*\boldsymbol{\beta}^*, \sigma^2 \mathbf{P}^{*-1}). \quad (3.89)$$

Die Wahrscheinlichkeitsdichtefunktion für Beobachtungen im linearen Modell der Bayesschen Kollokation lässt sich damit als

$$p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{s}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2} (\det(\mathbf{P}^*))^{-1/2}} e^{-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}^*\boldsymbol{\beta}^*)' \mathbf{P}^* (\mathbf{y} - \mathbf{X}^*\boldsymbol{\beta}^*)} \quad (3.90)$$

angeben. Ebenso wie im Fall der Bayesschen Regression soll weiterhin der Präzisionsparameter $\tau = 1/\sigma^2$ eingeführt werden, um die Folgeberechnungen zu vereinfachen. Die Wahrscheinlichkeitsdichtefunktion wird damit entsprechend modifiziert zu

$$p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{s}, \tau) = (2\pi)^{-n/2} (\det \mathbf{P}^*)^{1/2} \tau^{n/2} e^{-\frac{\tau}{2} (\mathbf{y} - \mathbf{X}^*\boldsymbol{\beta}^*)' \mathbf{P}^* (\mathbf{y} - \mathbf{X}^*\boldsymbol{\beta}^*)}, \quad (3.91)$$

womit abschließend die Likelihood-Funktion für die Verwendung im Bayes-Theorem nach (3.83) gegeben ist. In folgendem Abschnitt werden die Priori-Dichten der informativen und nichtinformativen Lösungen inklusive derer Randverteilungen eingeführt.

3.3.1.2. Priori-Dichten der Bayesschen Kollokation

Wie bereits beschrieben, beruht das Modell der Bayesschen Kollokation auf konjugierten Dichten, sodass eine analytische Lösung ermöglicht wird. Als Vorinformation für den Parametervektor $\boldsymbol{\beta}^*$ sowie die Varianz σ^2 können dabei wie im Fall der Bayesschen Regression deren Erwartungswerte und das zugehörige Dispersionsmodell eingeführt werden:

$$E(\boldsymbol{\beta}^*|\mathbf{y}) = E\left(\begin{bmatrix} \boldsymbol{\beta} & \mathbf{s} \end{bmatrix}'\right) = \underline{\boldsymbol{\beta}}^* = \begin{bmatrix} \underline{\boldsymbol{\beta}} & \mathbf{0} \end{bmatrix}', \quad \boldsymbol{\Sigma}_{\boldsymbol{\beta}^* \boldsymbol{\beta}^*} = \underline{\sigma}^2 \mathbf{Q}_{\boldsymbol{\beta}^* \boldsymbol{\beta}^*} = \underline{\sigma}^2 \begin{bmatrix} \mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{ss} \end{bmatrix}, \quad (3.92)$$

$$E(\sigma^2) = \underline{\sigma}^2, \quad V(\sigma^2) = V_{\underline{\sigma}^2}. \quad (3.93)$$

Die Vorinformation für den Signalvektor $\mathbf{s}$ wird mit dem Erwartungswert $E(\mathbf{s}) = \mathbf{0}$ eingeführt, da davon ausgegangen wird, dass der Trendanteil zutreffend formuliert ist und die Beobachtungen hinreichend erklärt. Die Varianzen und Kovarianzen in $\Sigma_{\mathbf{ss}} = \sigma^2 \mathbf{Q}_{\mathbf{ss}}$ werden hingegen als gegeben eingeführt (KOCH, 2007, S. 130). Insbesondere im Hinblick auf die Anwendung der Bayesschen Kollokation als Vergleichswertverfahren stellt diese Annahme eine wichtige Voraussetzung dar: Im Vorgriff auf die Erläuterungen in Kapitel 4 können die Signalkomponenten von Wertermittlungsobjekten ausschließlich aus dem aktuell zu untersuchenden Datensatz abgeleitet werden. Die grundlegende Form der Autokovarianzfunktion ist dabei bekannt, muss jedoch mit den aktuellen Werten berechnet werden. Die Vorinformation bezieht sich im Fall der Signalkomponente insofern stets auf einen ausgewerteten aktuellen Datensatz, sodass $\Sigma_{\mathbf{ss}}$ als vor der Auswertung gegeben angenommen wird und so die Vorinformation resultierend aus der Untersuchung wertbeeinflussender Merkmale repräsentiert. Für eine Kennzeichnung dieser Besonderheit wird die Vorinformation des Signals nicht wie in der bisherigen Notation mit einem Unterstrich für Priori-Wissen gekennzeichnet, sodass in den folgenden Ausführungen allgemein $\mathbf{Q}_{\underline{\mathbf{ss}}} = \mathbf{Q}_{\mathbf{ss}}$ gesetzt wird.

Für die gemeinsame *informative Priori-Dichte* des Parametervektors β^* und des Varianzfaktors σ^2 ergibt sich nach dem Prinzip der konjugierten Dichten eine Normal-Gamma-Verteilung der Form

$$\beta^*, \tau \sim NG(\underline{\beta}^*, \mathbf{Q}_{\underline{\beta}^* \underline{\beta}^*}, \underline{b}, \underline{p}), \quad (3.94)$$

wobei wiederum der Präzisionsparameter τ anstelle der Varianz angegeben wird. Die Wahrscheinlichkeitsdichtefunktion lässt sich daraus mit

$$p(\beta^*, \tau | \underline{\beta}, \mathbf{Q}_{\underline{\beta}^* \underline{\beta}^*}, \underline{b}, \underline{p}) = (2\pi)^{-n/2} \left(\det \mathbf{Q}_{\underline{\beta}^* \underline{\beta}^*} \right)^{-1/2} \underline{b}^{\underline{p}} (\Gamma(\underline{p}))^{-1} \tau^{n/2 + \underline{p} - 1} e^{-\frac{\tau}{2} \left\{ 2\underline{b} + (\beta^* - \underline{\beta})' \mathbf{Q}_{\underline{\beta}^* \underline{\beta}^*}^{-1} (\beta^* - \underline{\beta}) \right\}} \quad (3.95)$$

für die Verteilungsparameter $\underline{b} > 0$, $\underline{p} > 0$, $0 < \tau < \infty$, $-\infty < \beta_i^* < \infty$ angeben (KOCH, 2007, S. 61). Aus der gemeinsamen Priori-Dichte resultiert als Randverteilung für β^* eine multivariate t -Verteilung nach (3.37), für den Präzisionsparameter eine Gamma-Verteilung nach (3.39) sowie für die Varianz eine Inverse Gamma-Verteilung nach (3.40). Auf eine Darstellung der Wahrscheinlichkeitsdichten der Randverteilungen und eine Angabe der Berechnung der Verteilungsparameter kann an dieser Stelle verzichtet werden, da sie sich analog zu den Gleichungen (3.37) bis (3.42) unter der Berücksichtigung der Substitution $\beta = \beta^*$, $\underline{\beta} = \underline{\beta}^*$ sowie $\mathbf{Q}_{\underline{\beta}\underline{\beta}} = \mathbf{Q}_{\underline{\beta}^* \underline{\beta}^*}$ ergeben. Für die Lösung basierend auf *nichtinformativen Priori-Dichten* gelten entsprechend die Beziehungen

$$p(\beta^*) \propto c \quad \text{für } -\infty < \beta_i^* < \infty, i \in \{1, \dots, n\}, \quad (3.96)$$

$$p(\tau) \propto 1/\tau \quad \text{für } 0 < \tau < \infty. \quad (3.97)$$

Mit den informativen und nichtinformativen Priori-Dichten kann in Kombination mit der definierten Likelihood-Funktion das Bayes-Theorem (3.83) gelöst werden. Im folgenden Abschnitt werden die analytischen Lösungen der jeweiligen Modellansätze aufgezeigt.

3.3.2. Analytische Lösung im Modell mit unbekanntem Varianzfaktor

Die Lösung des Bayes-Theorems mit den vorgestellten Priori-Dichten und der Likelihood-Funktion erlaubt wie im Fall der Bayesschen Regression die Bestimmung von Posteriori-Schätzwerten der unbekannten Parameter aus der Posteriori-Dichte sowie eine Angabe der Unsicherheitsmaße in Form von HPDIs. Als wesentliche Erweiterung durch die Bayessche Kollokation wird neben der Schätzung der unbekannten Trendkoeffizienten $\bar{\beta}$, des Varianzfaktors $\bar{\sigma}^2$ und der ausgeglichenen Schätzwerte $\bar{\mathbf{y}}$ auch die Schätzung der Signalkomponente $\bar{\mathbf{s}}$ ermöglicht.

3.3.2.1. Informative Lösung der Bayesschen Kollokation

Werden im Bayes-Theorem die genannten Priori-Dichten und die Likelihood-Funktion eingesetzt, ergibt sich für die Posteriori-Dichte ohne Herleitung

$$p(\beta^*, \tau | \mathbf{y}) \propto \tau^{n/2 + \underline{p} + u/2 - 1} \cdot e^{\left\{ -\frac{\tau}{2} \left(2\underline{b} + (\beta^* - \underline{\beta})' \mathbf{Q}_{\underline{\beta}^* \underline{\beta}^*}^{-1} (\beta^* - \underline{\beta}) + (\mathbf{y} - \mathbf{X}^* \beta^*)' \mathbf{P}^* (\mathbf{y} - \mathbf{X}^* \beta^*) \right) \right\}}. \quad (3.98)$$

Wie aus vorstehender Gleichung ersichtlich und durch die Nutzung konjugierter Dichten zu schlussfolgern, resultiert als Posteriori-Dichte wiederum eine Normal-Gamma-Verteilung der Form

$$\left| \begin{array}{c} \beta \\ \mathbf{s} \end{array} \right|, \tau | \mathbf{y} \sim NG \left(\left| \begin{array}{c} \bar{\beta} \\ \bar{\mathbf{s}} \end{array} \right|, \mathbf{Q}_{\bar{\beta}^* \bar{\beta}^*}, \bar{b}, \bar{p} \right), \quad (3.99)$$

in der neben den Posteriori-Parametern $\bar{\beta}$ und $\bar{\mathbf{s}}$ sowie den Posteriori-Verteilungsparametern der Normal-Gamma-Verteilung $\bar{b}$ und $\bar{p}$ die Matrix $\mathbf{Q}_{\bar{\beta}^* \bar{\beta}^*}$ die gemeinsame Posteriori-Kofaktormatrix von $\bar{\beta}$ und $\bar{\mathbf{s}}$ bezeichnet. Da die Herleitung der Berechnung der einzelnen Verteilungsparameter nicht im Vordergrund steht, wird zur Lösung des Problems auf KOCH (2007, S. 131 f.) unter Berücksichtigung von $\mathbf{Z}\gamma = \mathbf{s}$ verwiesen, sodass an dieser Stelle lediglich die Endergebnisse angegeben werden. Diese ergeben sich nach KOCH (2007, S. 132) im Einzelnen zu:

$$\left| \begin{array}{c} \bar{\beta} \\ \bar{\mathbf{s}} \end{array} \right| = \mathbf{Q}_{\bar{\beta}^* \bar{\beta}^*} \left| \begin{array}{c} \mathbf{X}' \mathbf{Q}_{vv}^{-1} \mathbf{y} + \mathbf{Q}_{\underline{\beta} \underline{\beta}}^{-1} \underline{\beta} \\ \mathbf{I}' \mathbf{Q}_{vv}^{-1} \mathbf{y} \end{array} \right|, \quad (3.100)$$

$$\mathbf{Q}_{\bar{\beta}^* \bar{\beta}^*} = \left| \begin{array}{cc} \mathbf{X}' \mathbf{Q}_{vv}^{-1} \mathbf{X} + \mathbf{Q}_{\underline{\beta} \underline{\beta}}^{-1} & \mathbf{X}' \mathbf{Q}_{vv}^{-1} \mathbf{I} \\ \mathbf{I}' \mathbf{Q}_{vv}^{-1} \mathbf{X} & \mathbf{I}' \mathbf{Q}_{vv}^{-1} \mathbf{I} + \mathbf{Q}_{ss}^{-1} \end{array} \right|^{-1}, \quad (3.101)$$

$$\begin{aligned} \bar{b} = & \left(2 \left[(\sigma^2)^2 / V_{\underline{\sigma}^2} + 1 \right] \underline{\sigma}^2 + (\underline{\beta} - \bar{\beta})' \mathbf{Q}_{\underline{\beta} \underline{\beta}}^{-1} (\underline{\beta} - \bar{\beta}) \right. \\ & + (\mathbf{y} - \mathbf{X} \bar{\beta}) (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{Q}_{ss} (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} (\mathbf{y} - \mathbf{X} \bar{\beta}) \\ & \left. + (\mathbf{y} - \mathbf{X} \bar{\beta} - \bar{\mathbf{s}})' \mathbf{Q}_{vv}^{-1} (\mathbf{y} - \mathbf{X} \bar{\beta} - \bar{\mathbf{s}}) \right) / 2, \end{aligned} \quad (3.102)$$

$$\bar{p} = \left(n + 2 (\sigma^2)^2 / V_{\underline{\sigma}^2} + 4 \right) / 2. \quad (3.103)$$

Werden aus der angegebenen gemeinsamen Posteriori-Dichte die Randverteilungen gebildet, folgt der Vektor β^* einer multivariaten t -Verteilung, die mit

$$\left| \begin{array}{c} \beta \\ \mathbf{s} \end{array} \right| | \mathbf{y} \sim t \left(\left| \begin{array}{c} \bar{\beta} \\ \bar{\mathbf{s}} \end{array} \right|, \bar{b} \mathbf{Q}_{\bar{\beta}^* \bar{\beta}^*} / \bar{p}, 2\bar{p} \right) \quad (3.104)$$

angegeben werden kann. Die vorstehende Posteriori-Dichte lässt sich weiterhin zerlegen in die für die Einzelkomponenten β und $\mathbf{s}$ jeweilig gültigen Randverteilungen:

$$\beta | \mathbf{y} \sim t \left(\bar{\beta}, \bar{b} \left(\mathbf{X}' (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{X} + \mathbf{Q}_{\underline{\beta} \underline{\beta}}^{-1} \right)^{-1} / \bar{p}, 2\bar{p} \right) \quad (3.105)$$

$$\mathbf{s} | \mathbf{y} \sim t \left(\bar{\mathbf{s}}, \bar{b} \left(\mathbf{Q}_{ss} - \mathbf{Q}_{ss} \left(\mathbf{Q}_{ss} + \mathbf{Q}_{vv} + \mathbf{X} \mathbf{Q}_{\underline{\beta} \underline{\beta}} \mathbf{X}' \right)^{-1} \mathbf{Q}_{ss} \right) / \bar{p}, 2\bar{p} \right). \quad (3.106)$$

Aus den multivariaten t -Verteilungen können unter dem Ansatz einer quadratischen Kostenfunktion direkt Schätzwerte für die unbekannten Parameter und das Signal abgeleitet werden (KOCH, 2007, S. 135 f.):

$$\bar{\beta} = \left(\mathbf{X}' (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{X} + \mathbf{Q}_{\underline{\beta} \underline{\beta}}^{-1} \right)^{-1} \left(\mathbf{X}' (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{y} + \mathbf{Q}_{\underline{\beta} \underline{\beta}}^{-1} \underline{\beta} \right) \quad (3.107)$$

$$\bar{\mathbf{s}} = \mathbf{Q}_{ss} (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} (\mathbf{y} - \mathbf{X} \bar{\beta}). \quad (3.108)$$

Anhand der Gleichungen wird ersichtlich, dass die Schätzung der Posteriori-Trendkoeffizienten im Bayesschen Ansatz durch die Berücksichtigung der Priori-Trendkoeffizienten und -VKM direkt beeinflusst wird. Die Modellunsicherheit ergibt sich primär aus dem Trendanteil. Da sich die Signalkomponente wie beschrieben zwingend aus

einem aktuellen Datensatz ergibt, bezieht sich das Priori-Wissen hier nicht auf vergangene Auswertensätze. Die Signalkomponente wird daher maßgeblich durch Priori-Information über die Trendkoeffizienten beeinflusst.

Die Schätzung des Varianzfaktors ergibt sich aus der Randverteilung des Präzisionsparameters

$$\tau|\mathbf{y} \sim G(\bar{b}, \bar{p}) \implies \sigma^2|\mathbf{y} \sim IG(\bar{b}, \bar{p}), \quad (3.109)$$

sodass sich die Schätzung des Varianzfaktors und dessen Varianz zu

$$\bar{\sigma}^2 = \frac{\bar{b}}{\bar{p} - 1}, \quad V(\sigma^2|\mathbf{y}) = \frac{2(\bar{\sigma}^2)^2}{n + 2(\bar{\sigma}^2)^2/V_{\sigma^2}} \quad (3.110)$$

ergibt (KOCH, 2007, S. 134 f.). Aus Gründen der Übersichtlichkeit wird an dieser Stelle auf die Substitution von $\bar{b}$ und $\bar{p}$ aus (3.102) bzw. (3.103) verzichtet. Die VKM der Parameter und des Signals können mit Hilfe der geschätzten Varianz abgeleitet werden:

$$\Sigma_{\beta\beta}(\beta|\mathbf{y}) = \Sigma_{\bar{\beta}\bar{\beta}} = \bar{\sigma}^2 \left(\mathbf{X}'(\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{X} + \mathbf{Q}_{\beta\beta}^{-1} \right)^{-1}, \quad (3.111)$$

$$\Sigma_{ss}(s|\mathbf{y}) = \Sigma_{\bar{s}\bar{s}} = \bar{\sigma}^2 \left(\mathbf{Q}_{ss} - \mathbf{Q}_{ss} \left(\mathbf{Q}_{ss} + \mathbf{Q}_{vv} + \mathbf{X} \mathbf{Q}_{\beta\beta} \mathbf{X}' \right)^{-1} \mathbf{Q}_{ss} \right). \quad (3.112)$$

Mit der dargestellten Lösung des Bayesschen Kollokationsmodells werden die Auswerteschritte der Ausgleichung und Filterung ermöglicht. Die geschätzten Beobachtungen lassen sich demnach mit $\bar{\mathbf{y}} = \mathbf{X}\bar{\beta} + \bar{s}$ ermitteln. Die Verteilung der Posteriori-Schätzwerte $\bar{\mathbf{y}}$ folgt nach KOCH (1994, S. 468) als lineare Transformation der Parameter und des Signals einer multivariaten t -Verteilung mit

$$\left| \begin{array}{c} \mathbf{X} \quad \mathbf{I} \\ \hline \beta \\ s \end{array} \right| \mathbf{y} \sim t \left(\left| \begin{array}{c} \mathbf{X} \quad \mathbf{I} \\ \hline \bar{\beta} \\ \bar{s} \end{array} \right|, \bar{b}/\bar{p} \left| \begin{array}{c} \mathbf{X} \quad \mathbf{I} \\ \hline \mathbf{Q}_{\beta\beta}^{-1} \mathbf{X}' \\ \mathbf{I}' \end{array} \right|, 2\bar{p} \right), \quad (3.113)$$

woraus sich die VKM der ausgeglichenen und gefilterten Beobachtung mit

$$\Sigma_{\mathbf{y}\mathbf{y}}(\bar{\mathbf{y}}|\mathbf{y}) = \Sigma_{\bar{\mathbf{y}}\bar{\mathbf{y}}} = \bar{\sigma}^2 \left(\left| \begin{array}{c} \mathbf{X} \quad \mathbf{I} \\ \hline \mathbf{Q}_{\beta\beta}^{-1} \mathbf{X}' \\ \mathbf{I}' \end{array} \right| \right) \quad (3.114)$$

ergibt. Aus Gründen der Lesbarkeit wird auf eine Ausmultiplikation und Inversion der Terme verzichtet. Das vorgestellte Modell erlaubt im dritten Auswerteschritt der Bayesschen Kollokation die Prädiktion von neuen Werten $\mathbf{y}_p$, für die keine Beobachtungen, sondern lediglich die wertbeeinflussenden Merkmale in Form der Designmatrix $\mathbf{X}_p$ vorliegen. Die Schätzwerte lassen sich über die Beziehung $\bar{\mathbf{y}}_p = \mathbf{X}_p \bar{\beta} + \bar{s}_p$ angeben, in der $\bar{s}_p$ die prädizierte Signalkomponente angibt. Die Randverteilung des prädizierten Signals lässt sich analog zum Signal nach (3.106) mittels einer multivariaten t -Verteilung bestimmen (KOCH, 1994, S. 469)

$$s_p|\mathbf{y} \sim t \left(\bar{s}_p, \bar{b} \left(\mathbf{Q}_{s_p s_p} - \mathbf{Q}_{s_p s} \left(\mathbf{Q}_{ss} + \mathbf{Q}_{vv} + \mathbf{X} \mathbf{Q}_{\beta\beta} \mathbf{X}' \right)^{-1} \mathbf{Q}_{ss_p} \right) / \bar{p}, 2\bar{p} \right), \quad (3.115)$$

woraus sich die Schätzung des prädizierten Signals ergibt mit

$$\bar{s}_p = \mathbf{Q}_{s_p s} (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} (\mathbf{y} - \mathbf{X}\bar{\beta}), \quad (3.116)$$

$$\Sigma_{s_p s_p}(s_p|\mathbf{y}) = \Sigma_{\bar{s}_p \bar{s}_p} = \bar{\sigma}^2 \left(\mathbf{Q}_{s_p s_p} - \mathbf{Q}_{s_p s} \left(\mathbf{Q}_{ss} + \mathbf{Q}_{vv} + \mathbf{X} \mathbf{Q}_{\beta\beta} \mathbf{X}' \right)^{-1} \mathbf{Q}_{ss_p} \right). \quad (3.117)$$

Der Erwartungswert sowie die VKM der prädizierten Beobachtungen ist weiterhin unter Beachtung des Prinzips der prädiktiven Dichten nach

$$\begin{aligned} E(\mathbf{y}_p | \sigma^2, \mathbf{y}) &= E(E(\mathbf{y}_p | \boldsymbol{\beta}, \mathbf{s}, \sigma^2, \mathbf{y}) | \sigma^2, \mathbf{y}) = E(\mathbf{X}_p \boldsymbol{\beta} + \mathbf{s}_p | \sigma^2) \\ &= \mathbf{X}_p \bar{\boldsymbol{\beta}} + \mathbf{s}_p, \end{aligned} \quad (3.118)$$

$$\begin{aligned} E(\boldsymbol{\Sigma}_{\mathbf{y}_p \mathbf{y}_p} | \sigma^2, \mathbf{y}) &= E(V(\mathbf{y}_p | \boldsymbol{\beta}, \mathbf{s}_p, \sigma^2, \mathbf{y}) | \sigma^2, \mathbf{y}) + V(E(\mathbf{y}_p | \boldsymbol{\beta}, \mathbf{s}_p, \sigma^2, \mathbf{y}) | \sigma^2, \mathbf{y}) \\ &= E(\sigma^2 \mathbf{H}^{-1} | \sigma^2, \mathbf{y}) + V(\mathbf{X}_p \boldsymbol{\beta} + \mathbf{s}_p | \sigma^2, \mathbf{y}) \\ &= \sigma^2 \left(\mathbf{H}^{-1} + \left(\mathbf{Q}_{\mathbf{s}_p \mathbf{s}_p} + \mathbf{X}_p \mathbf{Q}_{\bar{\boldsymbol{\beta}} \bar{\boldsymbol{\beta}}} \mathbf{X}_p' \right) \right), \end{aligned} \quad (3.119)$$

$$\boldsymbol{\Sigma}_{\mathbf{y}_p \mathbf{y}_p}(\bar{\mathbf{y}}_p | \mathbf{y}) = \boldsymbol{\Sigma}_{\bar{\mathbf{y}}_p \bar{\mathbf{y}}_p} = \bar{\sigma}^2 \left\{ \left(\mathbf{Q}_{\mathbf{s}_p \mathbf{s}_p} + \mathbf{Q}_{\mathbf{v}_p \mathbf{v}_p} \right)^{-1} + \left(\mathbf{Q}_{\bar{\mathbf{s}}_p \bar{\mathbf{s}}_p} + \mathbf{X}_p \mathbf{Q}_{\bar{\boldsymbol{\beta}} \bar{\boldsymbol{\beta}}} \mathbf{X}_p' \right) \right\} \quad (3.120)$$

anzugeben, wobei $\mathbf{Q}_{\mathbf{v}_p \mathbf{v}_p}$ die Kofaktormatrix des Rauschanteils der Prädiktionsobjekte mit der Dimension $n_p \times n_p$ symbolisiert (KOCH, 2007, S. 136). Die Bayessche Kollokation ist damit unter der Voraussetzung informativer Priori-Dichten vollständig angegeben, mit den dargestellten Gleichungen können die Aufgaben der Ausgleichung, Filterung und Prädiktion gelöst werden. Im folgenden Abschnitt wird ergänzend die Lösung für das Modell basierend auf nichtinformativen Priori-Dichten angegeben.

3.3.2.2. Nichtinformativ Lösung der Bayesschen Kollokation

Die nichtinformativ Lösung der Bayesschen Kollokation ist gleichbedeutend damit, dass keine Vorinformation über die Parameter $\boldsymbol{\beta}$ vorliegt. Das Priori-Wissen wird damit zu $E(\boldsymbol{\beta}) = \underline{\boldsymbol{\beta}} = \mathbf{0}$ sowie $\mathbf{Q}_{\bar{\boldsymbol{\beta}} \bar{\boldsymbol{\beta}}} = \mathbf{0}$ angenommen und muss in den entsprechenden Gleichungen substituiert werden. Für die gemeinsame Posteriori-Dichte lässt sich daher eine Normal-Gamma-Verteilung ansetzen, deren Verteilungsparameter analog zur informativ basierten Vorgehensweise mit

$$\left| \begin{array}{c} \boldsymbol{\beta} \\ \mathbf{s} \end{array} \right|, \tau | \mathbf{y} \sim NG \left(\left| \begin{array}{c} \bar{\boldsymbol{\beta}} \\ \bar{\mathbf{s}} \end{array} \right|, \mathbf{Q}_{\bar{\boldsymbol{\beta}}^* \bar{\boldsymbol{\beta}}^*}, \bar{b}, \bar{p} \right), \quad (3.121)$$

angegeben werden. Im Vergleich zur informativen Lösung entfallen im nichtinformativen Ansatz für die Bestimmung der einzelnen Verteilungsparameter nach (3.100) bis (3.103) sämtliche Terme, die das Priori-Wissen repräsentieren. Für die nichtinformativ Lösung ergeben sich dann

$$\left| \begin{array}{c} \bar{\boldsymbol{\beta}} \\ \bar{\mathbf{s}} \end{array} \right| = \mathbf{Q}_{\bar{\boldsymbol{\beta}}^* \bar{\boldsymbol{\beta}}^*} \left| \begin{array}{c} \mathbf{X}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{y} \\ \mathbf{I}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{y} \end{array} \right| \quad (3.122)$$

$$\mathbf{Q}_{\bar{\boldsymbol{\beta}} \bar{\boldsymbol{\beta}}} = \left| \begin{array}{cc} \mathbf{X}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{X} & \mathbf{X}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{I} \\ \mathbf{I}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{X} & \mathbf{I}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{I} + \mathbf{Q}_{\mathbf{s} \mathbf{s}}^{-1} \end{array} \right|^{-1} \quad (3.123)$$

mit den Verteilungsparametern der Gamma-Verteilung

$$\begin{aligned} \bar{b} &= \left((\mathbf{y} - \mathbf{X} \bar{\boldsymbol{\beta}}) (\mathbf{Q}_{\mathbf{s} \mathbf{s}} + \mathbf{Q}_{\mathbf{v} \mathbf{v}})^{-1} \mathbf{Q}_{\mathbf{s} \mathbf{s}} (\mathbf{Q}_{\mathbf{s} \mathbf{s}} + \mathbf{Q}_{\mathbf{v} \mathbf{v}})^{-1} (\mathbf{y} - \mathbf{X} \bar{\boldsymbol{\beta}}) \right. \\ &\quad \left. + (\mathbf{y} - \mathbf{X} \bar{\boldsymbol{\beta}} - \bar{\mathbf{s}})' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} (\mathbf{y} - \mathbf{X} \bar{\boldsymbol{\beta}} - \bar{\mathbf{s}}) \right) / 2 \end{aligned} \quad (3.124)$$

$$\bar{p} = (n + 4) / 2. \quad (3.125)$$

Eine gemeinsame Verteilung von Parametern und Signal ist unter der Annahme der multivariaten t -Verteilung anzusetzen, deren Randverteilungen nach KOCH (2007, S. 136) direkt anzugeben sind mit

$$\boldsymbol{\beta} | \mathbf{y} \sim t \left(\bar{\boldsymbol{\beta}}, \bar{b} / \bar{p} \left(\mathbf{X}' (\mathbf{Q}_{\mathbf{s} \mathbf{s}} + \mathbf{Q}_{\mathbf{v} \mathbf{v}})^{-1} \mathbf{X} \right), 2\bar{p} \right) \quad (3.126)$$

$$\mathbf{s} | \mathbf{y} \sim t \left(\bar{\mathbf{s}}, \bar{b} / \bar{p} \underbrace{\left(\mathbf{A} + \mathbf{A} \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{X} - \mathbf{X}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{A} \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{Q}_{\mathbf{v} \mathbf{v}}^{-1} \mathbf{A} \right)}_{\mathbf{M}}, 2\bar{p} \right). \quad (3.127)$$

Eine Vereinfachung der Darstellung wird durch die Beziehung $\mathbf{A} = \left(\mathbf{Q}_{ss} - \mathbf{Q}_{ss} (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{Q}_{ss} \right)$ genutzt. Die Schätzwerte der Parameter und des Signals folgen daraus unter Ansatz der quadratischen Kostenfunktion direkt zu

$$\bar{\beta} = \left(\mathbf{X}' (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{X} \right)^{-1} \left(\mathbf{X}' (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{y} \right)^{-1}, \quad (3.128)$$

$$\bar{s} = \mathbf{Q}_{ss} (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} (\mathbf{y} - \mathbf{X} \bar{\beta}). \quad (3.129)$$

Die Randverteilung des Präzisionsparameters folgt wie im informativen Fall der Gamma-Verteilung mit $\tau|\mathbf{y} \sim G(\bar{b}, \bar{p})$, sodass sich ein Schätzwert für den Varianzfaktor und dessen Varianz über die Invertierte Gamma-Verteilung ergibt mit

$$\sigma^2|\mathbf{y} \sim IG(\bar{b}, \bar{p}), \quad (3.130)$$

$$\bar{\sigma}^2 = \frac{\bar{b}}{\bar{p} - 1}, \quad V(\sigma^2|\mathbf{y}) = \frac{2(\bar{\sigma}^2)^2}{n} \quad (3.131)$$

und die Variablen $\bar{b}$ und $\bar{p}$ durch die entsprechenden Ausdrücke (3.124) bzw. (3.125) zu substituieren sind. Mit Hilfe des geschätzten Varianzfaktors lassen sich in der Folge aus den Randdichten die VKM der Parameter und der Signalkomponenten ableiten:

$$\Sigma_{\beta\beta}(\beta|\mathbf{y}) = \Sigma_{\bar{\beta}\bar{\beta}} = \bar{\sigma}^2 \left(\mathbf{X}' (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{X} \right)^{-1}, \quad (3.132)$$

$$\Sigma_{ss}(\mathbf{s}|\mathbf{y}) = \Sigma_{\bar{s}\bar{s}} = \bar{\sigma}^2 \mathbf{M}. \quad (3.133)$$

Analog zu den Ausführungen im vorherigen Abschnitt ermöglichen die Schritte der Ausgleichung und Filterung im nichtinformativen Bayesschen Kollokationsmodell durch die funktionale Gleichung $\bar{\mathbf{y}} = \mathbf{X} \bar{\beta} + \bar{\mathbf{s}}$ die Schätzung ausgeglichener Beobachtungen. Somit folgt die Verteilung der geschätzten Beobachtung wiederum als lineare Transformation einer multivariaten t -Verteilung nach (3.114) und einer daraus folgenden VKM nach (3.115), wobei die entsprechenden Verteilungsparameter durch die nichtinformativen Ansätze (3.122) bis (3.125) zu substituieren sind.

In Hinblick auf den Schritt der Prädiktion im nichtinformativen Bayesschen Kollokationsmodell muss zunächst die prädizierte Signalkomponente $\bar{s}_p$ bestimmt werden, um die der Trendanteil verbessert wird. Die Randverteilung lässt sich nach KOCH (2007, S. 136) angeben als

$$s_p|\mathbf{y} \sim t\left(\bar{s}_p, \bar{b}/\bar{p} \underbrace{\left(\mathbf{B} + \mathbf{B} \mathbf{Q}_{vv}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{Q}_{vv}^{-1} \mathbf{X} - \mathbf{X}' \mathbf{Q}_{vv}^{-1} \mathbf{B} \mathbf{Q}_{vv}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{Q}_{vv}^{-1} \mathbf{B} \right)}_N, 2\bar{p}\right), \quad (3.134)$$

wobei als Vereinfachung die Beziehung $\mathbf{B} = \left(\mathbf{Q}_{s_p s_p} - \mathbf{Q}_{s_p s} (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{Q}_{ss s_p} \right)$ eingeführt wird. Als Schätzwert für die prädizierte Signalkomponente sowie der zugehörigen VKM ergibt sich daraus nach KOCH (2007, S. 136)

$$\bar{s}_p = \mathbf{Q}_{s_p s} (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} (\mathbf{y} - \mathbf{X} \bar{\beta}), \quad (3.135)$$

$$\Sigma_{s_p s_p}(s_p|\mathbf{y}) = \Sigma_{\bar{s}_p \bar{s}_p} = \bar{\sigma}^2 \mathbf{N}. \quad (3.136)$$

Für die prädiktive Dichte von $\bar{\mathbf{y}}_p$ gelten unter der Voraussetzung einer linearen Transformation die Ausführungen zu (3.120) unter Berücksichtigung der im nichtinformativen Fall abweichenden Posteriori-Kofaktormatrix der Parameter mit $\mathbf{Q}_{\bar{\beta}\bar{\beta}} = \left(\mathbf{X}' (\mathbf{Q}_{ss} + \mathbf{Q}_{vv})^{-1} \mathbf{X} \right)^{-1}$. Ebenso wie die informative Lösung ist die nichtinformativ Lösung der Bayesschen Kollokation damit vollständig angegeben. Im folgenden Abschnitt werden ergänzend die Definitionen der Konfidenzregionen für die Modellparameter und Posteriori-Schätzwerte der Beobachtungen gegeben.

3.3.3. Konfidenzregionen der Parameter und Schätzwerte

Die Formulierung eines HPDIs, welches für die Parameter der Bayesschen Regression bereits durch die Erläuterungen in Abschnitt 3.2.3 (S. 64 f.) eingeführt ist, gilt aufgrund der Annahme der multivariaten t -Verteilung als Randverteilung für β äquivalent für die Modelle der Bayesschen Kollokation. Wie erläutert, lassen sich für die Einzelkomponenten $\bar{\beta}_i$ mit $i \in \{1, \dots, u\}$ des Vektors der Trendkoeffizienten verallgemeinerte t -Verteilungen bilden, sodass sich für die Konfidenzregionen der Posteriori-Parameter nach der Bayesschen Schätzung

$$P\left(\bar{\beta}_i - t_{1-\alpha;\nu}/\sqrt{f} < \beta_i < \bar{\beta}_i + t_{1-\alpha;\nu}/\sqrt{f}\right) = 1 - \alpha \quad (3.137)$$

$$\text{mit } 1/f = \bar{b}(\mathbf{Q}_{\bar{\beta}\bar{\beta}})_{ii} / \bar{p} \quad (3.138)$$

ergibt. Da neben den eigentlichen Modellparametern auch die Signalkomponenten in einem Kollokationsmodell von besonderem Interesse sind, erlauben die HPDIs der Signalkomponenten eine erweiterte Beurteilung der Unsicherheitsmodellierung. Wie aus den bisherigen Ausführungen zu schließen ist, wird die Unsicherheit eines Kollokationsmodells neben der rein funktionalen Trendmodellierung insbesondere auch von der stochastisch formulierten Signalkomponente beeinflusst, sodass sich an dieser Stelle gegenüber der (Bayesschen) Regression eine zweite Unsicherheitskomponente ergibt. Diese wirkt sich direkt auf die Unsicherheit der Schätzung ausgeglichener Beobachtungen und neuer Prädiktionswerte aus.

Da sich, wie im Zusammenhang mit (3.105) und (3.106) beschrieben, die gemeinsame Posteriori-Dichte von Parametern β und Signal s (bzw. s_p) in jeweilige multivariate t -Verteilungen als Randverteilungen zerlegen lässt, gelten für die Herleitung der HPDIs der Signalkomponenten dieselben Grundsätze, die für die Einzelkomponenten $\bar{\beta}_i$ des Vektors der Trendkoeffizienten zur Herleitung über die verallgemeinerte t -Verteilung führen. Die Beschreibungen an genannter Stelle können daher direkt auf eine beliebige Einzelkomponente $\bar{s}_j$ mit $j \in \{1, \dots, n\}$ übertragen werden. Während sich eine Herleitung der Beziehungen insbesondere (KOCH, 2007) entnehmen lässt, kann für das HPDI des Signals direkt

$$P\left(\bar{s}_j - t_{1-\alpha;\nu}/\sqrt{f} < s_j < \bar{s}_j + t_{1-\alpha;\nu}/\sqrt{f}\right) = 1 - \alpha \quad (3.139)$$

$$\text{mit } 1/f = \bar{b}(\mathbf{Q}_{ss})_{jj} / \bar{p} \quad (3.140)$$

angegeben werden. Für eine Einzelkomponente des posteriori prädizierten Signals $\bar{s}_{p,k}$ mit $k \in \{1, \dots, v\}$ und v als Anzahl der Prädiktionsobjekte ergibt sich entsprechend

$$P\left(\bar{s}_{p,k} - t_{1-\alpha;\nu}/\sqrt{f} < y_{p,k} < \bar{s}_{p,k} + t_{1-\alpha;\nu}/\sqrt{f}\right) = 1 - \alpha \quad (3.141)$$

$$\text{mit } 1/f = \bar{b}(\mathbf{Q}_{s_p s_p})_{kk} / \bar{p}. \quad (3.142)$$

Für eine umfassende Beurteilung der Auswirkungen von Unsicherheiten auf das Schätzergebnis können abschließend die HPDIs der ausgeglichenen Beobachtungen $\bar{y}$ sowie der prädizierten Werte $\bar{y}_p$ angegeben werden. Der Ansatz lässt sich insofern einfach angeben, als dass die geschätzten Werte ebenfalls einer multivariaten t -Verteilung folgen.

Für einzelne Posteriori-Schätzwerte der Beobachtungen $\bar{y}_j$ mit $j \in \{1, \dots, n\}$ ergibt sich daher in Konsequenz zu den bisherigen Ausführungen

$$P\left(\bar{y}_j - t_{1-\alpha;\nu}/\sqrt{f} < y_j < \bar{y}_j + t_{1-\alpha;\nu}/\sqrt{f}\right) = 1 - \alpha \quad (3.143)$$

$$\text{mit } 1/f = \bar{b}(\mathbf{Q}_{\bar{y}\bar{y}})_{jj} / \bar{p}, \quad (3.144)$$

sowie für die prädizierten Werte $\bar{y}_{p,k} = \bar{y}_{p,k}$ mit $k \in \{1, \dots, v\}$

$$P\left(\bar{y}_{p,k} - t_{1-\alpha;\nu}/\sqrt{f} < s_{p,k} < \bar{s}_{p,k} + t_{1-\alpha;\nu}/\sqrt{f}\right) = 1 - \alpha \quad (3.145)$$

$$\text{mit } 1/f = \bar{b}(\mathbf{Q}_{\bar{y}_p \bar{y}_p})_{kk} / \bar{p}. \quad (3.146)$$

Die bis zu diesem Punkt vorgestellten Modellformulierungen basieren ausschließlich auf konjugierten Dichten, sodass eine analytische Lösung angegeben werden kann. Insbesondere im Hinblick auf die Aufgaben der Verkehrswertermittlung ist eine leicht zugängliche Lösung von Vorteil, wie in den folgenden Kapitel gezeigt wird. Wie bereits angedeutet, stellt eine analytische Lösung je nach Wahl von Priori-Dichten und Likelihood-Funktion jedoch nicht zwingend den Regelfall dar, sodass der gezeigte Lösungsweg u. U. nicht anwendbar ist. Insbesondere in praktischen Untersuchungen kann sich herausstellen, dass von der Annahme konjugierter Verteilungsfamilien abgewichen werden muss, um ein reales Phänomen adäquat zu beschreiben. Die Lösung solcher Modellformulierungen muss dann über Simulations-Methoden erfolgen, die im folgenden Abschnitt kurz beleuchtet werden sollen.

3.4. Numerische Lösung für nicht konjugierte Verteilungsdichten

Anhand der Erläuterungen zum verallgemeinerten Bayes-Theorem in Abschnitt 3.1.2.2 (S. 54 ff.) wird deutlich, dass für die Ermittlung der Posteriori-Dichte die resultierenden Integrale aus Priori-Dichte und Likelihood-Funktion zu lösen sind. Während die Verwendung konjugierter Dichten sich hier aus genannten Gründen als Vorteil erweist, ist für nicht konjugierte Dichten eine numerische Approximation erforderlich, die die Generierung von Zufallszahlen aus einer analytisch nicht zugänglichen Posteriori-Dichte erlaubt. Auf diese Weise können die Verteilungsmomente der Posteriori-Dichte, wie bspw. Erwartungswert und Varianz, sowie Quantile oder die Dichte der Verteilung selbst geschätzt werden (FAHRMEIR ET AL., 2009, S. 482), (BANDEMER, 1997, S. 119).

Als numerische Approximationsmethoden haben sich insbesondere Monte-Carlo Verfahren durchgesetzt, die sich speziell für die Lösung komplexer Dichten mit einer hohen Anzahl von Dimensionen eignen. Monte-Carlo Verfahren beruhen auf dem Prinzip der Monte-Carlo Integration, die in Abschnitt 3.4.1 kurz vorgestellt werden soll. Aufbauend auf dem Grundgedanken wird als Erweiterung in Abschnitt 3.4.2 auf die Monte-Carlo-Markov-Chain (MCMC) Methoden eingegangen, die eine Generierung von Zufallswerten aus einer beliebigen Posteriori-Dichte erlauben.

3.4.1. Monte-Carlo Integration

Das Grundprinzip der Monte-Carlo Integration basiert auf der Generierung von Zufallszahlen, die in ihrer Summation ein beliebig komplexes Integral approximieren können: Wird mit ζ eine allgemeine Zufallsvariable mit der Wahrscheinlichkeitsdichte $p(z)$, $z \in \mathbb{R}$ im Integrationsraum $Z \subset \mathbb{R}$ bezeichnet, lässt sich der Erwartungswert einer beliebigen Funktion $f(\zeta)$ über $p(z)$ durch das Integral

$$E(f(\zeta)) = \int_Z f(z) p(z) dz \quad (3.147)$$

darstellen. Wird auf dieser Basis eine hohe Anzahl N von gleichverteilten und voneinander unabhängigen (eng. *identically and independently distributed (i. i. d.)*) Zufallsvariablen $s_z^1, \dots, s_z^N$ aus der Wahrscheinlichkeitsdichte $p(z)$ generiert, lässt sich für den Erwartungswert der Funktion ein Schätzwert mit

$$\bar{E}(f(\zeta)) = \frac{1}{N} \sum_{i=1}^N f(s_z^i) \quad (3.148)$$

angeben. Durch diese Formulierung kann die analytische Integration durch eine finite Summe von Zufallszahlen ersetzt werden. Ein entsprechend multivariater Ansatz lässt sich durch Übergang auf einen m -dimensionalen Vektor ζ und einer n -dimensionalen Funktion $\mathbf{f}(\zeta)$ erreichen, dessen geschätzter Erwartungswert mit

$$\bar{E}(\mathbf{f}(\zeta)) = \frac{1}{N} \sum_{i=1}^N \mathbf{f}(s_z^i) \quad (3.149)$$

angegeben werden kann und s_z^i i. i. d. multivariate Zufallszahlen mit der gemeinsamen Dichte $p(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^m$ bezeichnen. Vorstehende Gleichung ermöglicht durch die Summation der einzelnen Dichtewerte somit (unter der Voraussetzung, dass N gegen unendlich strebt) die Approximation des geschätzten Erwartungswertes des entspre-

chenden Integrals. Analog zur Schätzung von Erwartungswerten lässt sich auch die Schätzung der Varianz als zweites zentrales Moment angeben:

$$\bar{E}(\mathbf{f}(\zeta)) = \frac{1}{N} \sum_{i=1}^N (\mathbf{s}_f^i - E(\mathbf{f}(\zeta))) (\mathbf{s}_f^i - E(\mathbf{f}(\zeta)))', \quad (3.150)$$

wobei die Transformation $\mathbf{s}_f^i = \mathbf{f}(\mathbf{s}_f^i)$ gilt (ALKHATIB, 2008, S. 17 f.).

Neben der Ermittlung von Kovarianzen ermöglicht die Monte-Carlo Integration auch die Bestimmung von höheren Momenten der zu Grunde liegenden Wahrscheinlichkeitsdichte, die an dieser Stelle jedoch nicht weiter vertieft werden sollen. Die Monte-Carlo Integration soll an dieser Stelle lediglich in ihrem Grundgedanken vorgestellt werden, vertiefende Darstellungen finden sich neben den genannten Quellen insbesondere auch bei GELMAN ET AL. (2004), CHRISTENSEN (2011) oder LEE (2012).

Als wesentlichster Schritt in der Anwendung der Monte-Carlo Integration gilt die Generierung von Zufallszahlen aus einer zutreffenden Dichte $p(\mathbf{z})$. Während KOCH (2000, S. 183 ff.) und ALKHATIB (2008, S. 15 ff.) verschiedene Verfahren zu deren Generierung aufzeigen, lassen sich bei komplexeren Wahrscheinlichkeitsdichten nur schwer zutreffende Algorithmen mit zufriedenstellendem Ergebnis angeben. Insbesondere für Posteriori-Dichten, deren Verteilungsdichte auf Grund nichtkonjugierter Priori-Dichte und Likelihood-Funktion nicht bekannt ist, müssen als Weiterentwicklung zur Monte-Carlo Integration MCMC-Methoden angewendet werden. Das Prinzip von sogenannten Markov-Ketten und den darauf basierenden MCMC-Methoden wird im folgenden Abschnitt thematisiert.

3.4.2. Monte-Carlo-Markov-Chain-Methoden zur numerischen Approximation

Als *Markov-Ketten* werden stochastische Prozesse bezeichnet, die ein Modell für ein System sequenzieller, stochastisch abhängiger Zufallsfolgen beschreiben. Der Startpunkt einer Markovkette kann im Falle eines unbekannten Parameters allgemein mit $\theta^{(1)}$ bezeichnet werden. Unter Verwendung eines sogenannten Übergangskerns $p(\theta^{(t)}|\theta^{(t-1)})$ kann der durch $\theta^{(1)}$ bedingte Zustand $\theta^{(2)}$ des beschriebenen Systems als zweite Iteration ermittelt werden. Auf diese Weise entsteht eine Kette von einzelnen Zuständen, bei denen der jeweils vorherige Zustand auf den aktuellen Zustand übertragen und bis zu einer festgelegten Anzahl T von Iterationen fortgeführt wird ($\theta^{(1)} \rightarrow \theta^{(2)} \rightarrow \dots \rightarrow \theta^{(T)}$) (LEE, 2012, S. 282). Der Prozessablauf einer Markov-Kette lässt sich daher zusammenfassen zu

1. Setzen von $t = 1$.
2. Generierung einen beliebigen Initialwert u mit $\theta^{(t)} = u$.
3. Durchlauf des iterativen Prozesses bis $t = T$ mit
 - $t = t + 1$,
 - Generierung eines neuen Zufallswertes u mittels des Übergangskerns $p(\theta^{(t)}|\theta^{(t-1)})$,
 - $\theta^{(t)} = u$.

Da ein aktueller Zustand innerhalb des Prozesses ausschließlich von seinem vorherigen abhängt und Vergangenheit sowie Zukunft keine Rolle spielen, werden Markov-Ketten auch als „gedächtnislos“ bezeichnet. Wird eine Kette mit einem beliebigen Startwert initialisiert, wird sie sich daher zunächst im Zustandsraum um diesen Wert bewegen. Werden weiterhin eine Vielzahl von Ketten mit dem gleichen Übergangskern aber mit unterschiedlichen Werten für u initialisiert, bewegen sie sich dementsprechend zunächst ebenfalls um den jeweiligen Startwert. Diese Phase wird als *Burn-In* bezeichnet (FAHRMEIR ET AL., 2009, S. 482). Als charakteristische Eigenschaft von Markov-Ketten ergibt sich nach der Burn-In-Phase auf Grund der ausschließlichen Berücksichtigung des Zustands $(t - 1)$ für die Bestimmung des aktuellen Zustands t , dass der Startwert keinen Einfluss mehr auf den aktuellen Zustand hat und dementsprechend gegen die Zielverteilung konvergiert. Für die Markov-Ketten mit unterschiedlichen Startwerten bedeutet dies, dass sie sich unabhängig voneinander der gemeinsamen *stationären Verteilung* des Übergangskerns annähern, was als *Steady-State-Phase* bezeichnet wird. Das Grundprinzip von Markov-Ketten wird an dieser Stelle nicht vertieft, eine ausführliche Diskussion insbesondere der Eigenschaften von Markov-Ketten

findet sich u. a. in LAUER (2004, S. 25 ff.). Zusammenfassend beruht die Monte-Carlo Integration auf dem Prinzip, aus einer Wahrscheinlichkeitsdichte eine unabhängige Stichprobe zu ziehen, um auf dieser Basis Schätzungen zentraler Verteilungsmomente zu ermöglichen. Im Vergleich können Markov-Ketten hingegen genutzt werden, um aufgrund ihrer Definition als stochastische sequenzielle Prozesse Zustände von interessierenden Zufallsparametern aus einer unbekannten, stationären Verteilung zu generieren. Wird der Grundgedanke auf Bayessche Anwendungen übertragen, kann somit unter der Voraussetzung einer hohen Anzahl von Iterationen des Übergangskerns eine Konvergenz gegen die gewünschte Dichte erreicht werden, die die Ziehung einer Stichprobe aus der Posteriori-Verteilung erlaubt.

Werden Monte-Carlo Methoden und Markov-Ketten zu einem Verfahren kombiniert, ergeben sich MCMC-Methoden, die die wesentlichen Eigenschaften der Einzelmethode vereinen. Ziel von MCMC-Methoden ist es, eine Markov-Kette zu formulieren, deren stationäre Verteilung der Posteriori-Dichte entspricht, wobei letztere dann als *Zielverteilung* bezeichnet wird. Von entscheidender Bedeutung ist daher die Formulierung des Übergangskerns, der nach der Burn-In-Phase die Zielverteilung bestmöglich approximieren soll. In der Fachliteratur lassen sich eine Vielzahl von Algorithmen finden, die unter dem Begriff der MCMC-Methoden zu subsumieren sind. Zu den bekanntesten zählt dabei der *Metropolis-Hastings-Algorithmus*, der von METROPOLIS ET AL. (1953) veröffentlicht und von HASTINGS (1970) erweitert wurde. Da der Metropolis-Hastings-Algorithmus als Basis-Algorithmus angesehen werden kann, von dem sich alle weiteren ableiten lassen, soll er im Folgenden kurz für die Anwendung in der Bayes-Theorie beschrieben werden.

Prinzip des Metropolis-Hastings-Algorithmus ist es, die Erzeugung eines nachfolgenden Zustands eines Parametervektors θ einer Markov-Kette von der Zielverteilung unabhängig zu machen und in einem zweiten Schritt den erzeugten Zustand entweder zu akzeptieren oder zu verwerfen (LAUER, 2004, S. 28). Nach der Festlegung von Startwerten $\theta^{(0)}$ für einen allgemein formulierten Vektor unbekannter Parameter wird die Anzahl der Iterationen mit T festgelegt und $t = 1$ gesetzt. Da die interessierende Posteriori-Dichte $p(\theta|\mathbf{y})$ nicht bekannt oder zugänglich ist, werden in jeder Iteration Zufallswerte θ^* aus einer *Vorschlagsdichte* q gezogen. Gemäß der Definition des Übergangskerns hängt die Vorschlagsdichte vom Zustand $\theta^{(t-1)}$ ab, sodass $q = q(\theta^*|\theta^{(t-1)})$ gilt. Eine Vorschlagsdichte kann dabei beliebig gewählt werden, sollte jedoch eine relativ einfache Generierung von Zufallszahlen erlauben. Da sich Vorschlagsdichte und Posteriori-Dichte nicht entsprechen, werden nicht alle vorgeschlagenen Zufallszahlen θ^* als neuer Zustand t^* des Systems akzeptiert, sondern lediglich mit einer gewissen Wahrscheinlichkeit α . Die Akzeptanzwahrscheinlichkeit lässt sich berechnen aus dem Quotienten der Posteriori-Dichte und der Vorschlagsdichte, welche am aktuellen Zustand $\theta^{(t-1)}$ ausgewertet wird:

$$\alpha(\theta^*|\theta^{(t-1)}) = \min \left\{ \frac{p(\theta^*|\mathbf{y}) q(\theta^{(t-1)}|\theta^*)}{p(\theta^{(t-1)}|\mathbf{y}) q(\theta^*|\theta^{(t-1)})}, 1 \right\}. \quad (3.151)$$

Für den Fall, dass der vorgeschlagene Parametervektor θ^* nicht akzeptiert wird, ist $t = (t - 1)$ zu setzen. Da die Dichte $p(\theta|\mathbf{y})$ offensichtlich in $\alpha(\theta^*|\theta^{(t-1)})$ ausschließlich im Verhältnis $p(\theta^*|\mathbf{y})/p(\theta^{(t-1)}|\mathbf{y})$ eingeht, brauchen alle konstanten Ausdrücke in der Posteriori-Dichte $p(\theta|\mathbf{y})$ nicht berücksichtigt zu werden, was sich insbesondere auf die Normierungskonstante im Bayes-Theorem bezieht. Nach der Burn-In-Phase und einer gewissen Konvergenzzeit t_0 können die erhaltenen Zufallszahlen $\theta^{(t_0+1)}, \dots, \theta^{(T)}$ als Stichprobe der Posteriori-Dichte angesehen werden (FAHRMEIR ET AL., 2009, S. 483 f.).

Wie dargestellt, ermöglichen MCMC-Methoden die Generierung von Zufallsstichproben aus einer Posteriori-Dichte und erlauben die Lösung des Bayes-Theorems für den Fall nicht konjugierter Dichten. In Erweiterung des dargestellten Metropolis-Hastings-Algorithmus lassen sich weitere Algorithmen formulieren, wie bspw. der Gibbs-Sampler, der sich für hochdimensionale Parametervektoren eignet und als Spezialfall aus dem Metropolis-Hastings-Algorithmus ableiten lässt. Weiterhin wird auf die Entwicklung hybrider Algorithmen hingewiesen. Auf Grund der Nutzung konjugierter Verteilungsmodelle bleibt die eingehende Diskussion der MCMC-Methoden der weiterführenden Literatur vorbehalten. Anwendungsbezogene Problemstellungen lassen sich bspw. LAUER (2004) und ALKHATIB (2008) entnehmen, Grundlagen der Monte-Carlo Theorie finden sich u. a. in DOUCET & SMITH (2001).

4. Entwicklung einer Auswertestrategie für eine Bayessche Vergleichswertermittlung

Durch die Erläuterungen des Kapitels 2 wird deutlich, dass sich Aufgaben der Verkehrswertermittlung in einem Spannungsfeld aus normativen Anforderungen, stark differenzierten Teilmärkten und einer zumeist heterogenen bis lückenhaften Datengrundlage bewegen. Dementsprechend hohe Anforderungen sind an die Auswertemethoden inklusive der Behandlung von Unsicherheiten zu stellen, mit denen Marktanalysen durchgeführt werden. Die hier betrachteten Auswertemethoden, basierend auf der Bayes-Strategie, wurden theoretisch ausführlich in Kapitel 3 behandelt, bislang jedoch losgelöst von einem anwendungsbezogenen Kontext. In dem folgenden Kapitel soll die Verknüpfung der beiden Themenfelder erfolgen: Im Fokus steht die Entwicklung einer Auswertestrategie, die die Voraussetzungen und Ansprüche der jeweiligen theoretischen Grundlagen vereint und deren praktische Anwendung die definierten Forschungsziele erreicht.

In Abschnitt 4.1 werden zunächst allgemeine Voraussetzungen und die schematische Formulierung der Auswertestrategie behandelt. Insbesondere stehen spezifische Voraussetzungen der Verkehrswertermittlung an eine mathematisch-statistisch fundierte Auswertestrategie sowie an eine geeignete Datengrundlage im Mittelpunkt. Differenziert werden Annahmen über die Verteilung der Beobachtungen und der Modellparameter. Ergänzt werden die Ausführungen durch allgemeine Annahmen, die den Vergleichswertmodellen zugrunde liegen und sich vornehmlich auf Eigenschaften der Einflussgrößen und den Einfluss zeitlicher Entwicklungen auf dem Immobilienmarkt beziehen. Der Abschnitt schließt mit einer ersten schematischen Übersicht der Auswertestrategie. Aufbauend auf dem entwickelten Schema werden in Abschnitt 4.2 die ersten beiden Schritte der Auswertestrategie erläutert. Diese umfassen zum einen die Definition des Modellrahmens, der von der untersuchten Fragestellung abhängt. Im Anschluss wird die Aufbereitung einer gegebenen Stichprobe als Datengrundlage für die Anwendung der Bayesschen Modelle aufgezeigt. Abschnitt 4.3 ist der eigentlichen Auswertung im Rahmen der Bayesschen Modelle gewidmet. Eingehend erläutert werden in diesem Zuge der Ablauf der Bayesschen Schätzung, die Modellwahl über Hypothesentests sowie die Rekursion als Mittel für die Fortpflanzung von Schätzungsergebnissen und Unsicherheiten.

4.1. Voraussetzungen und Formulierung der Auswertestrategie

Werden die jeweils theoretischen Grundlagen der Kapitel 2 und 3 in einem gemeinsamen Kontext betrachtet, lassen sich zwei Anforderungen an die Nutzung Bayesscher Modelle allgemein in der Verkehrswertermittlung identifizieren:

1. Grundstücksmarktanalysen liegen i. d. R. aufgrund der erläuterten Eigenschaften von Immobilien stark heterogene Daten zu Grunde. Eine zusätzliche Diversifizierung ergibt sich durch die Aufsplitterung des Immobilienmarktes in zahlreiche Teilmärkte. Unter der generellen Voraussetzung, dass sich Grundstücksmarktanalysen stets auf einen durch Sachverstand sorgfältig abgegrenzten Teilmarkt beziehen lassen, verbleibt stets das Problem der Heterogenität innerhalb der gewählten Stichprobe. Die erste Anforderung ergibt sich daher bezüglich des zu wählenden Modells: Dieses muss zum einen geeignet sein, die Forschungsziele zu erreichen, zum anderen die Datenheterogenität berücksichtigen zu können.
2. Die zweite Anforderung bezieht sich auf die mathematisch-statistische Sichtweise: Modelle, wie sie in Kapitel 3 konkret beschrieben sind, lassen sich lediglich dann aufstellen, wenn grundlegende Modellannahmen getroffen werden. Diese beziehen sich insbesondere auf die Eingangsgrößen sowie die Modellparameter. Weitere, allgemeine Annahmen über anwendungsbezogene Gegebenheiten (hier: Rahmenbedingungen durch den Immobilienmarkt) ergänzen die mathematisch-statistischen Anforderungen.

Wie aus den Einzelaspekten deutlich wird, ergibt sich eine enge Wechselwirkung von fachlichen Voraussetzungen der Verkehrswertermittlung und theoretischen Erfordernissen der statistischen Auswertung. Als Konsequenz ist eine

Auswertestrategie für Anwendungen in der Verkehrswertermittlung zu entwickeln, die sämtlichen Anforderungen genügt. In Abschnitt 4.1.1 werden die Anforderungen aus Sicht der Verkehrswertermittlung erläutert. In Abschnitt 4.1.2 stehen die allgemeine Verteilungs- und Modellannahmen in Bezug auf die Nutzung Bayesscher Modelle in der Verkehrswertermittlung im Mittelpunkt. Abschließend werden die Einzelbetrachtungen in Abschnitt 4.1.3 zu einem Schema der Auswertestrategie verknüpft.

4.1.1. Voraussetzungen aus Sicht der Verkehrswertermittlung

Die Auswertestrategie basiert auf Vorüberlegungen, die Gegenstand der Ausführungen im folgenden Unterabschnitt sind. Aufbauend sind ebenso Anforderungen an die Datengrundlage zu stellen, die den zentralen Baustein der Verfahrens- und Ergebnisunsicherheiten im Sinne des Abschnitts 2.2 (S. 31 ff.) darstellen.

4.1.1.1. Auswertestrategie

Statistische Methoden haben sich seit Mitte der 1970iger Jahre als Standard in der Analyse des Grundstücksmarktes und in Aufgaben der Verkehrswertermittlung entwickelt. Neben der rein deskriptiven Statistik erfordern komplexere Verfahren wie multivariate Analysemethoden ein umfangreiches Vorgehen bezüglich der Aufbereitung und Handhabung der Wertermittlungsdaten. Die in Abschnitt 2.1.2 (S. 26 ff.) aufgeführten Bestimmungen über die Ermittlung des Verkehrswertes setzen die Auswertung von Kaufpreisen voraus. Dem Verkehrswert selbst wird unterstellt, dass er den Preis reflektiert, der im gewöhnlichen Geschäftsverkehr für ein Objekt gezahlt werden würde. Zwingende Voraussetzung ist, dass bekannt ist, wie sich die Gesamtheit der am Grundstücksmarkt Beteiligten verhält und welche Preise sie für konkrete Objekte zu zahlen bereit sind (GERARDY ET AL., 2014, S. 2.2.1/1). Die Erkenntnisse können in diesem Hinblick aus den Erfassungsdaten gewonnen werden: Das Rohmaterial der statistischen Untersuchungen besteht aus wertermittlungsrelevanten Einzelwerten, zwischen denen Beziehungen bestehen, welche durch einen funktionalen bzw. stochastischen Zusammenhang ausgedrückt werden können. Nach GERARDY ET AL. (2014, S. 2.2.1/1) sind auch Kaufpreise zu diesen Einzelwerten zu zählen. Zwischen den Kaufpreisen besteht eine stochastische Abhängigkeit, da sie in ihrer Grundgesamtheit den Grundstücksmarkt repräsentieren. Aus den Kaufpreisen lassen sich auf diesem Weg Werte und Beziehungen ableiten, welche ein umfassendes Bild des Kaufverhaltens geben und die Beurteilung des gewöhnlichen Geschäftsverkehrs ermöglichen. Im Zuge einer Auswertestrategie für die Verkehrswertermittlung sollen aufgabenspezifisch die Voraussetzungen geschaffen werden, ein zutreffendes und objektives Modell des Grundstücksmarktes zu erstellen, welches die Zusammenhänge adäquat widerspiegelt.

Eine aufgabenspezifische Modellformulierung erfordert im Vorfeld einer komplexeren statistischen Untersuchung zunächst die Prüfung, ob das angestrebte Ziel der Verkehrswertermittlung den Einsatz multivariater Analyseverfahren erfordert. Unabhängig von dem hier im Fokus stehenden Vergleichswertverfahren sind statistische Methoden zur Verkehrswertermittlung nicht zwingend notwendig oder möglich und dürfen keinen reinen Selbstzweck erfüllen. Werden die Erläuterungen des Abschnitts 2.3.1 (S. 40 ff.) herangezogen, ist die praktische Anwendung von Wertermittlungsverfahren abhängig von der Objektart sowie der generellen Verfügbarkeit von Daten. Die hier thematisierten Analyseverfahren werden vornehmlich für die folgenden Aufgabenbereiche angewendet:

1. (Indirektes) Vergleichswertverfahren,
2. Ermittlung von Bodenrichtwerten,
3. Ableitung sonstiger für die Wertermittlung erforderlicher Daten (Indexreihen, Umrechnungskoeffizienten, Vergleichsfaktoren für bebaute Grundstücke, Sachwertfaktoren, Liegenschaftszinssätze).

Verfolgt die Analyse des Grundstücksmarktes eines der vorgenannten Ziele, ist im Rahmen von Vorüberlegungen die geeignete Auswertemethode zu wählen. Eine Klassifizierung der gängigen Methoden wird in ZIEGENBEIN (1977, S. 13 ff.) gegeben und orientiert sich an der Fragestellung, ob mittels eines Analyseverfahrens die Beziehungen einer Reihe von Variablen untereinander untersucht werden sollen (Interdependenz-Analyse) oder ob die Abhängigkeit einer Variablen von mehreren weiteren Variablen im Fokus steht (Dependenz-Analyse). Basierend auf einer

Einteilung nach KENDALL & STUART (1969) listet Ziegenbein die für Wertermittlungsfragestellungen relevanten Methoden auf, die in Tabelle 4.1 um die Bayesschen Ansätze ergänzt werden.

Tabelle 4.1.: Multivariate statistische Verfahren in der Verkehrswertermittlung

Analyse der ...	
... Dependenz	... Interdependenz
Multiple lineare Regression	Faktorenanalyse
Einfache Kollokation	Clusteranalyse
Varianzanalyse	Korrelationsanalyse
Bayessche Regression	
Bayessche Kollokation	

Die folgenden Ausführungen beziehen sich ausschließlich auf das Vergleichswertverfahren: Eine Auswertestrategie wird konkret für die Schätzung von Vergleichswerten bzw. -faktoren aufgezeigt. Der Ablauf der Auswertestrategie lässt sich jedoch im Grundsatz auch auf die übrigen aufgelisteten Wertermittlungsaufgaben übertragen; eine Diskussion der Übertragungsmöglichkeiten wird in Kapitel 6 aufgegriffen.

Nach ZIEGENBEIN (1977, S. 46 ff.) ist eine wesentliche Voraussetzung für die Anwendung der Verfahren, dass sich eine Gesetzmäßigkeit erkennen lässt, mit der sich Preise am Grundstücksmarkt bilden und sich so kausale Zusammenhänge zwischen Kaufpreisen und wertrelevanten Merkmalen erkennen lassen. Insbesondere bezüglich der in Abschnitt 2.1.1 (S. 23 ff.) thematisierten Heterogenität am Grundstücksmarkt ist diese Voraussetzung nicht sofort erkennbar und in der Realität nicht zwingend gegeben. Werden jedoch ergänzend die normativen Vorgaben der Verkehrswertdefinition des § 194 BauGB herangezogen, zeigen sich die Bedingungen der Preisbildung als stark eingeschränkt: Wirtschaftliches Verhalten der Akteure am Grundstücksmarkt vorausgesetzt, kann für Kaufpreise unterstellt werden, dass sie sich unter rationaler Beurteilung und Bewertung der jeweiligen Grundstücksmerkmale bilden. Der unmittelbare Zusammenhang zwischen den Kaufpreisen und den zugeordneten wesentlichen Grundstücksmerkmalen als Einzelwerten ist dabei von besonderer Bedeutung, da multivariate Analyseverfahren ausschließlich auf dem angenommenen Zusammenhang beruhen.

Um gewährleisten zu können, dass eine zutreffende und objektive Modellformulierung der Problemstellung erfolgt, müssen die Daten sorgfältig auf ihre Verwendbarkeit geprüft werden. Zum einen sind Anforderungen im Hinblick auf die Kriterien des gewöhnlichen Geschäftsverkehrs, zum anderen betreffend statistischer Voraussetzungen zu formulieren. Werden die jeweiligen Vorgaben nicht erfüllt, ergibt sich als Konsequenz ein nicht korrektes Modell des Grundstücksmarktes, welches i. d. R. unzutreffend oder falsch, im besten Fall höchstens bedingt aussagefähig ist. Um der Heterogenität der Datengrundlage bei der Verkehrswertermittlung gerecht zu werden, sollte eine direkte Anwendung der Modelle nach Kapitel 3 ohne weitere Datenaufbereitung vermieden werden. Zudem ist keine noch so sorgfältige Modellformulierung in der Lage, das Fachwissen zu ersetzen: Zielsetzung, Modellwahl, Beurteilung der Eingangsdaten sowie Interpretation der Ergebnisse können ausschließlich durch Sachverstand angemessen gewürdigt werden. Eine Auswertestrategie, wie sie nachfolgend in Abschnitt 4.1.3 vorgestellt und im Detail in den Abschnitten 4.2 und 4.3 beschrieben wird, ermöglicht die Berücksichtigung der praktischen Belange der Verkehrswertermittlung, insbesondere der Anforderungen im Vergleichswertverfahren, die durch die VW-RL definiert sind.

Wie durch die Ausführungen verdeutlicht, ist ein zentraler Aspekt der Auswertestrategie der verwendeten Datengrundlage zuzuschreiben. Die Anforderungen und Grundlagen werden nachfolgend erläutert.

4.1.1.2. Datengrundlage für das indirekte Vergleichswertverfahren

Elementare Voraussetzung für eine zutreffende Wertermittlung mittels statistischer Analysen ist die verwendete Datengrundlage: Wie in Abschnitt 2.2.1 (S. 31 ff.) thematisiert, sind Unsicherheiten im Bewertungsprozess zum einen auf die Daten zurückzuführen, zum anderen auf das verwendete Modell selbst. Voraussetzung für weitere Auswertungen ist zugleich die Verfügbarkeit von Marktdaten, da sich Analysen im Vergleichswertverfahren ausschließlich auf die mit einem Wertermittlungsobjekt vergleichbaren Kauffalldaten stützen. An die Daten wird der Anspruch gestellt, dass sie direkt aus dem Markt stammen: Dies bezieht sich nicht nur ausschließlich auf die Erhebung von Kaufpreisen als direktem Indikator für die Preisentwicklung, vielmehr sind auch die wesentlichen wertbeeinflussenden

den Merkmale zu berücksichtigen, die den Kaufpreisen zugeordnet werden können. Welche Merkmale als relevant beurteilt werden, ist in erster Linie sachverständigem Ermessen und Erfahrung in der Analyse des Grundstücksmarktes zuzuordnen. Ein allgemeiner Überblick über potenzielle Größen ist bereits in Tabelle 2.1 (S. 41) aufgeführt. Eine Konkretisierung kann stets nur im Hinblick auf den betrachteten sachlichen Teilmarkt erfolgen.

Als wichtigste Datengrundlage im Vergleichswertverfahren, die auch für die vorliegende Arbeit genutzt wird, dient regelmäßig eine Kaufpreissammlung nach § 195 BauGB, die durch den jeweils zuständigen Gutachterausschuss für seinen gesamten Amtsbezirk zu führen ist. Erst durch die Einrichtung einer Kaufpreissammlung werden die Voraussetzungen für die weitere Analyse des Grundstücksmarktes geschaffen. Die grundlegenden Kauffalldaten stammen aus Kaufverträgen, die gemäß § 195 BauGB durch jede eine Transaktion von Grundstücken beurkundende Stelle in Kopie an den Gutachterausschuss zu übermitteln ist. Die Entscheidung, welche Grundstücksmerkmale als wertrelevante Einflussgrößen standardmäßig in einer Kaufpreissammlung zu erfassen sind, obliegt den für das Gutachterausschusswesen verantwortlichen Stellen der Länder. In Niedersachsen wurden durch die zuständige Niedersächsische Vermessungs- und Katasterverwaltung (VKV) umfangreiche Vorgaben für die Führung und Auswertung der AKS in Form von Elementkatalogen herausgegeben (vgl. LGLN (2015), MI (2015)). In Abhängigkeit einer Kategorisierung der sachlichen Teilmärkte (bebaute Grundstücke, unbebaute Flächen sowie Eigentumswohnungen) wird eine Mindestzahl an Elementen gegeben, die zwingend in der AKS zu belegen ist, um eine standardisierte Auswertung zu ermöglichen (Muss-Elemente). Diese beziehen sich sowohl auf sogenannte *Ordnungsmerkmale*, die insbesondere die räumliche Lage der einzelnen Kauffälle beschreiben (bspw. Georeferenzierung, Flurstücksangaben) als auch auf die *Vertragsmerkmale*, die die spezifischen Eigenarten des einzelnen Kauffalls näher beschreiben (vgl. Tabelle 2.1, S. 41).

Neben den Daten, die dem Kaufvertrag entnommen werden können, steht es den Gutachterausschüssen frei, ergänzende Informationen zum jeweiligen Objekt zu erheben, die als wertrelevant angesehen werden, und in der Kaufpreissammlung zu speichern. So sind trotz des umfangreich formulierten Katalogs mit obligatorischen Erfassungselementen auch Erweiterungen in der Erhebung denkbar und können sich bspw. auf sozio-ökonomische Faktoren wie Bevölkerungsentwicklung und -dichte sowie allgemeine Geodaten beziehen. Einen Einblick in die Möglichkeiten gibt MÜRLE (2006); die Einflüsse zusätzlicher Merkmale auf den Kaufpreis sind jedoch bis heute nicht tiefer gehend analysiert oder als Standardelemente in der AKS implementiert. Die Wertrelevanz zusätzlicher Einflussgrößen in Form von Geodaten wird u. a. in ZADDACH & ALKHATIB (2013a), ZADDACH & ALKHATIB (2013b), ZADDACH & ALKHATIB (2014a) sowie ZADDACH & ALKHATIB (2014b) für unterschiedliche Modellansätze des Vergleichswertverfahrens erfolgreich genutzt. Die standardisierte Übernahme zusätzlicher Daten bleibt künftigen Forschungs- und Entwicklungsfragen überlassen. Aus Sicht der Auswertestrategie liegt der Fokus auf der reinen Entwicklung der methodischen Vorgehensweise. Für die Modelle des indirekten Vergleichswertverfahrens wird daher in Bezug auf die Datengrundlage der Anspruch abgeleitet, eine adäquate Approximation der Realität durch diejenigen Determinanten zu erreichen, die standardmäßig verfügbar sind oder ohne großen Aufwand erhoben werden können. Als wesentliche Datengrundlage dient damit die AKS, auf der die praktischen Auswertungen in Kapitel 5 beruhen.

Im Hinblick auf die Nutzung der Kauffalldaten im Rahmen der Auswertestrategie sind darüber hinaus Anforderungen an die Daten zu formulieren und zu prüfen: Im Einzelnen sind dies die Definition einer geeigneten Grundgesamtheit, die Prüfung auf Vollständigkeit der Datensätze, die Elimination von Ausreißern in den Daten sowie die Berücksichtigung von Korrelationen in der Datengrundlage. Detailliert wird auf die Punkte im Zuge der Beschreibung der Auswertestrategie eingegangen. Neben den Anforderungen aus der Sicht der Verkehrswertermittlung an Auswertestrategie und Datengrundlage sind für die mathematisch-statistische Modellbildung zusätzliche Verteilungs- und Modellannahmen zu treffen, die im folgenden Abschnitt vorgestellt werden.

4.1.2. Allgemeine Verteilungs- und Modellannahmen

Bislang wurden ausschließlich Anforderungen aus Sicht der Verkehrswertermittlung formuliert. Werden diese mit den Erkenntnissen aus Kapitel 3 über die Theorie der Bayes-Strategie und der exemplarischen Formulierung von Bayesschen Regressions- und Kollokationsmodellen verknüpft, sind allgemeine Verteilungs- und Modellannahmen notwendig, um die angestrebte Lösung in konjugierten linearen Modellen zu ermöglichen. Die Verteilungsannahmen

beziehen sich auf die Verteilung der Beobachtungen sowie der unbekannten Parameter und werden in den folgenden Unterabschnitten dargestellt. Darauf folgend wird auf ergänzende Modellannahmen eingegangen. Die Erläuterungen beziehen sich dabei gleichermaßen auf das Modell der Bayesschen Regression und der Bayesschen Kollokation.

4.1.2.1. Verteilung der Beobachtungen

Als Beobachtung i. S. einer Dependenz-Analyse nach Tabelle 4.1 (S. 79) ist die Zielgröße $\mathbf{y}$ aufzufassen, die geeignet definiert werden muss. Eine exemplarische Auflistung über allgemein mögliche Faktoren als Zielgrößen in der Verkehrswertermittlung gibt Tabelle 4.2. Die aufgelisteten Beispiele sind nicht abschließend und können im Rahmen einer Aufgabendefinition geeignet festgelegt werden.

Tabelle 4.2.: Exemplarische Zielgrößen in der Verkehrswertermittlung

Anwendung	Zielgröße	Einheit
Schätzung von Vergleichswerten	Kaufpreis	[€]
Schätzung von Vergleichsfaktoren	Kaufpreis/Wohnfläche	[€/m ²]
Bodenrichtwertermittlung	Kaufpreis/Grundstücksfläche	[€/m ²]
Schätzung von Marktanpassungsfaktoren	Kaufpreis/Sachwert	[-]

Wird explizit das Vergleichswertverfahren betrachtet, ergibt sich die Zielgröße i. d. R. aus dem Kaufpreis. Je nach Aufgabenstellung kann dies der Gesamtkaufpreis eines Objekts sein oder ein Faktor, der durch Normierung aus diesem hervorgeht. Voraussetzung für die Normierung ist lediglich, dass die Größen Kaufpreis und Normierungsfaktor einen Zusammenhang aufweisen (ZIEGENBEIN, 1977, S. 49). Die grundsätzliche Normierung wurde von ZIEGENBEIN (1977, S. 49 f.) vorgeschlagen und hat sich seither in praktischen Anwendungen etabliert. In den weiteren Ausführungen zur Auswertestrategie wird daher die Annahme $\mathbf{y} = (\text{Kaufpreis}/\text{Normierungsfaktor})$ getroffen. Eine konkrete inhaltliche Auslegung erfolgt in Kapitel 5.

Wie bereits im Zusammenhang mit der Definition der Modelle in Abschnitt 2.3.2 (S. 43 ff.) erläutert, wird für die klassischen Modelle der Regression und Kollokation angenommen, dass die Beobachtungen $\mathbf{y}$ einer Normalverteilung folgen. Aus Sicht der Verkehrswertermittlung lässt sich dies nach ZIEGENBEIN (1977, S. 3) und UHDE (1982, S. 25) insbesondere damit begründen, dass nur Preise, die im gewöhnlichen Geschäftsverkehr ohne Beeinflussung durch ungewöhnliche oder persönliche Verhältnisse erzielt werden, zum direkten oder indirekten Preisvergleich zugelassen sind. Wird die rein statistische Sichtweise hinzugezogen, lässt sich eine Begründung über den zentralen Grenzwertsatz finden: Demzufolge folgt die Summe einer gegen unendlich strebenden Zahl von unabhängigen Zufallsvariablen asymptotisch einer stabilen Verteilung; bei endlicher und positiver Varianz ist dies i. d. R. die Normalverteilung. In Anwendungen der Verkehrswertermittlung hat sich daher die Annahme etabliert, dass Kaufpreise bei hinreichender Abgrenzung des betrachteten Teilmarktes generell der Normalverteilung folgen, während von der Verteilungsannahme abweichende Ausreißer als Kaufpreise mit Besonderheiten gelten, die erkannt und eliminiert werden müssen. Eine entsprechende Aufbereitung der Datengrundlage wird mit Hilfe einer Ausreißerelimination vorgenommen und in den Abschnitten 4.2 und 4.3 beschrieben.

Durch die sachlich begründete Annahme einer Normalverteilung für $\mathbf{y}$ lassen sich die Likelihood-Funktionen der Bayesschen Regression und Bayesschen Kollokation gemäß der Beziehung (3.32) bzw. (3.89) angeben, unabhängig von informativen oder nichtinformativen Ansätzen. Neben der Likelihood-Funktion sind zusätzlich Annahmen über die Verteilung der Modellparameter zu treffen, die im folgenden Abschnitt benannt werden.

4.1.2.2. Verteilung der Modellparameter

Für die Bayesschen Modelle wird ein linearer Zusammenhang zwischen den Beobachtungen und den wertbeeinflussenden Merkmalen unterstellt, was zu linearen Modellen führt. Um zusätzlich die angestrebte analytische Lösung durch die Nutzung konjugierter linearer Modelle zu ermöglichen, sind geeignete Annahmen über die Verteilung der Modellparameter zu treffen: Diese umfassen nach den Ausführungen in Kapitel 3 die Schätzwerte für die Koeffizienten β , die unbekannte Varianz σ^2 sowie im Bayesschen Kollokationsmodell zusätzlich die Signalkomponente $\mathbf{s}$.

Wie in den Abschnitten 3.2 und 3.3 (S. 57 ff.) ausführlich dargelegt, implizieren die Annahme der Normalverteilung für die Beobachtungen $\mathbf{y}$ sowie für die Residuen nach (3.31) in Verbindung mit der Forderung nach einem konjugierten Modell bereits die Randverteilungen der genannten Modellparameter. Generell unterschieden werden muss zwischen der informativen Lösung und der nichtinformativen Lösung, da unterschiedliche Verteilungsansätze zu nutzen sind.

Werden zunächst die zu schätzenden *Trendkoeffizienten* im informativen Ansatz betrachtet, stellt sich in einem linearen Modell die Anforderung, dass sich die Beobachtungen $\mathbf{y}$ als Linearkombination von $\boldsymbol{\beta}$ ergeben. In Kombination mit dem Anspruch eines konjugierten Modells wird die Priori-Randverteilung der unbekannten Parameter gemäß (3.37) als multivariate t-Verteilung in die Auswertung eingeführt, woraus sich nach Systematik der Bayes-Theorie (und unter Beachtung der Verteilung der übrigen Modellparameter) auch zwingend eine multivariate t-Verteilung der Posteriori-Schätzung ergibt. Durch die Verteilungsparameter kann direkt die Unsicherheit der unbekannten Trendkoeffizienten ausgedrückt werden. Unbeachtet bleibt an dieser Stelle die tatsächliche Verteilung der Einflussgrößen, die sich in Anwendungen der Verkehrswertermittlung in den überwiegenden Fällen als schiefsymmetrisch darstellt. Während gängige Anwendungen i. d. R. durch Transformationsansätze schiefsymmetrisch verteilte Einflussgrößen in eine symmetrische Verteilung überführen, wird in der folgenden Auswertestrategie auf diesen Zwischenschritt verzichtet. Die Unsicherheit der Daten spiegelt sich unter anderem in deren Verteilung wider und soll möglichst originär in den Modellansatz übernommen werden, um eine realitätsnahe Beschreibung der Unsicherheiten zu ermöglichen. Es wird unterstellt, dass sich Unsicherheiten u. a. in der Streuung der Modellparameter niederschlagen. Lineare Modelle setzen lediglich Linearität in den unbekannten Parametern voraus, sodass diese Voraussetzung der Modellformulierung nicht entgegensteht. Im Fall einer nichtinformativen Lösung des Bayesschen Ansatzes ist Vorwissen über die unbekannten Koeffizienten nicht vorhanden, was bedeutet, dass alle Elemente $\boldsymbol{\beta}$ im zugehörigen Parameterraum B gleich wahrscheinlich sind. Mathematisch lässt sich dies durch eine Priori-Gleichverteilung ausdrücken (vgl. (3.43) und (3.96)), die wiederum zu einer multivariaten t-Verteilung als Posteriori-Lösung führt.

Für die Schätzung der *Varianz* als weiterem Modellparameter wird für die informative Lösung – wie in (3.39) gezeigt – eine Gamma-Verteilung formuliert. Diese bezieht sich auf den Gewichtsparameter τ , der die Varianz σ^2 substituiert. Die Gamma-Verteilung wird als Randverteilung gewählt, um die gemeinsame Dichte von $\boldsymbol{\beta}$ und τ als Normal-Gamma-Verteilung modellieren zu können und um so die Voraussetzung für ein konjugiertes Modell zu schaffen. Im nichtinformativen Fall kann entsprechend eine Gleichverteilung für die Priori-Information angesetzt werden (vgl. (3.44 und 3.97)), sodass sich für die Posteriori-Randverteilung eine Gamma-Verteilung ergibt. Sowohl der informative wie auch der nichtinformativ Ansatz führen dann mit $\boldsymbol{\beta}$ zu einer gemeinsamen Normal-Gamma-Verteilung als Posteriori-Dichte.

Die *Signalkomponente* $\mathbf{s}$ der Bayesschen Kollokation spiegelt den unsystematischen, aber regelmäßigen Anteil wider, der nach Trendabzug in den Daten verbleibt. Wie durch die Herleitung in (2.13) bis (2.16) gezeigt, lässt sie sich mathematisch durch Umformulierung eines gemischten Modells ausdrücken. Da sich das Signal aus Autokorrelationen in den Residuen nach Trendabspaltung ergibt, für die nach (2.13) und (2.16) $E(\boldsymbol{\epsilon}) = E(\mathbf{s}) = E(\mathbf{u}) = \mathbf{0}$ gilt, ist die Randverteilung äquivalent zur Verteilung von $\boldsymbol{\beta}$ als multivariate t-Verteilung zu formulieren. Zu beachten ist, dass im Hinblick auf das Signal die Varianzen und Kovarianzen in $\boldsymbol{\Sigma}_{ss} = \sigma^2 \mathbf{Q}_{ss}$ wie erläutert als gegeben eingeführt werden (KOCH, 2007, S. 130). Als Posteriori-Dichte ergibt sich daher sowohl im informativen wie auch im nichtinformativen Fall zusammen mit $\boldsymbol{\beta}$ eine gemeinsame multivariate t-Verteilung, unter Berücksichtigung der unbekannten Varianz wie im Fall der Bayesschen Regression eine Normal-Gamma-Verteilung.

Die vorstehend beschriebenen Ansätze für die Wahrscheinlichkeitsdichten sind explizite Annahmen für die Auswertestrategie des Vergleichswertverfahrens. Alternativ kann ein Ansatz abweichender Wahrscheinlichkeitsdichten erfolgen, die sich nicht mehr geschlossen lösen lassen. In diesen Fällen ist ein Lösungsweg über Monte-Carlo Simulationen nach den Ausführungen in Abschnitt 3.4 (S. 73 ff.) zu wählen. Abweichende Verteilungsannahmen sowie die Untersuchung von deren Auswirkungen auf die Unsicherheitsmodellierung sind auf diese Weise zwar möglich, stehen jedoch nicht im Fokus der Arbeit. Tiefergehende Untersuchungen bleiben künftigen Forschungsfragen vorbehalten. Neben den Annahmen über die Verteilung der Beobachtungen und Modellparameter sind allgemeine Rahmenbedingungen für die Modellformulierung zu berücksichtigen. Diese werden im folgenden Abschnitt zusammengefasst.

4.1.2.3. Allgemeine Modellannahmen

Neben der rein mathematisch-statistischen Sichtweise stellen gleichermaßen Vorgaben der Verkehrswertermittlung Anforderungen an die Modellformulierung. Diese beziehen sich insbesondere auf die Auswahl der Einflussgrößen sowie auf Annahmen über die Stabilität der Zielfunktion hinsichtlich zeitlicher Entwicklungen am Grundstücksmarkt.

Für die in den Modellansatz eingebrachten *Einflussgrößen* wird vorausgesetzt, dass sie ein hinreichend genaues Abbild des Grundstückmarktes ermöglichen. Die Einflussgrößen unterliegen damit der Annahme, dass sie jeweils einen hohen Beitrag zur Erklärung der Zielgröße liefern. Die Auswahl eines geeigneten Datensatzes obliegt so dem Sachverstand, anhand dessen die marktüblichen Mechanismen beurteilt und auf einzelne wertbeeinflussende Merkmale übertragen werden müssen. In der Theorie ist eine Einschränkung wertbeeinflussender Merkmale nur schwer bzw. nicht möglich. Anzahl und Art der Einflussgrößen im Modell wird hingegen unterstellt, dass sie ausreichend sind, um den betrachteten räumlichen und sachlichen Teilmarkt adäquat zu beschreiben. Als Datengrundlage dienen die Erfassungselemente der AKS, die die wesentlichen Merkmale erfasster Kauffälle enthält. Die Ausweitung auf zusätzliche Einflussgrößen ist grundsätzlich denkbar, bedingt jedoch i. d. R. umfangreiche Ergänzungserhebungen. Der Fokus und das Ziel liegen insbesondere auf der Nutzung einiger weniger Merkmale, die dennoch eine zutreffende Schätzung einer einfach gehaltenen Trendfunktion ermöglichen und von denen angenommen wird, dass sie den betrachteten Teilmarkt ausreichend beschreiben.

Eine weitere Annahme des Modells bezieht sich auf *zeitliche Entwicklungen am Grundstücksmarkt*. Wie bereits HAWERK (1992, S. 57) feststellt, lassen sich lang-, mittel- und kurzfristige Änderungen wahrnehmen, die sich im Kaufverhalten der Marktteilnehmer ausdrücken. Durch die Veränderungen werden die Angebots- und Nachfragesituationen in den einzelnen Teilmärkten beeinflusst und sind entsprechend in der Modellbildung zu berücksichtigen. Veränderungen sind stets als Funktion der Zeit aufzufassen und lassen sich in stetige und nichtstetige Einflüsse differenzieren. *Stetige Veränderungen* stellen sich als ein langfristiger, allmählicher Prozess dar. Als Beispiel nennt HAWERK (1992, S. 57 f.) den Teilmarkt unbebauter Einfamilienhausgrundstücke: Bedingt durch kontinuierlich steigende Bau- und Grundstückspreise bevorzugen viele Kaufinteressenten kleinere Grundstücke. Die erhöhte Nachfrage spiegelt sich dann im Rahmen einer Modellbildung in der Entwicklung des Trendkoeffizienten für die Grundstücksfläche wieder, welcher einer längerfristigen Änderung unterliegt. Ebenso kann beobachtet werden, dass bei einem Überangebot von Grundstücken – den marktwirtschaftlichen Grundsätzen folgend – das mittlere Preisniveau sinkt und als Folge Grundstücke in schlechten Lagen kaum mehr veräußerbar sind. Die Nachfrage konzentriert sich dann auf gute Lagen, in denen zwangsläufig eine Mangelsituation mit der Folge eines allgemeinen Preisanstiegs entsteht. In diesem Fall würde sich eine langfristige Änderung in einem Trendkoeffizienten der Lage ergeben. Entgegen der stetigen Veränderungen können sich kurz- bis mittelfristige, *nichtstetige Veränderungen* insbesondere durch neue steuerliche Abschreibungsmöglichkeiten, Änderungen bei Kreditzinsen oder der Ausweisung neuer Baugebiete zugeordnet werden. Diese Einflüsse können u. U. eine sofortige Auswirkungen auf den betrachteten Teilmarkt bewirken. Ebenso zu nennen sind als aktuelles Beispiel die Einflüsse der Immobilien- und Finanzkrise von 2008, die unmittelbare Schwankungen auf dem Immobilienmarkt auslösen und deren Wirkung sich kurzfristig auf die Preisbildung niederschlagen können.

Die Folge von zeitlichen Veränderungen am Immobilienmarkt ist, dass mathematisch-statistische Modelle nicht als konstant angesehen werden können und sich so die Frage nach einem geeigneten Auswertezeitraum stellt. Für den Fall zeitlich zu umfangreich definierter Stichproben sind sogenannte – insbesondere durch kurz- und mittelfristige Veränderungen bedingte – Strukturbrüche in den Koeffizienten eines Trendmodells zu erwarten. Dies bedeutet, dass die Koeffizienten über den Auswertezeitraum nicht als konstant angesehen werden können und einen falschen Sachverhalt wiedergeben. UHDE (1982, S. 32 ff.) sowie HAWERK (1992, S. 58 ff.) geben für diesen Fall (frequentistisch basierte) Teststatistiken vor, mittels derer Strukturbrüche aufgedeckt werden können. In praktischen Anwendungen der Verkehrswertermittlung werden aus diesem Grund bislang in einem regelmäßigen Abstand neue Modelle aufgestellt und Funktionen geschätzt. Für die einzelnen Auswertepochen wird Konstanz in der Entwicklung der allgemeinen Markteinflüsse und damit den Auswirkungen der Einflussgrößen auf die Zielgröße unterstellt. Klassische Regressionsansätze von Gutachterausschüssen beziehen sich i. d. R. auf eine Auswertepoche von einem Jahr, werden jedoch bei einer zu geringen Anzahl von vorliegenden Kauffällen auch über zwei oder mehr Jahre

ausgewertet. Die gängigen Modelle sind damit als *statische* Momentaufnahmen einzustufen, die einen kontinuierlichen, dynamischen Prozess beschreiben. HAWERK (1992, S. 59) bezeichnet diese Art der Modellierung in Analogie zu Anwendungen der Ingenieurgeodäsie als *quasistatisch*. Eine Lösung wird in seiner Arbeit durch die alternative dynamische Modellformulierung des Kalman-Filters als zeitabhängige Regressionsfunktion gegeben, welches sich nach den Ansätzen durch KOCH (2007, S. 107 ff.) anwendungsunabhängig auch als Bayessches Kalman-Filter entwickeln lässt. Der Fokus der Arbeit liegt allerdings auf einer quasistatischen Modellformulierung, die durch die Integration von Priori-Information aus den vorherigen AuswerteePOCHen ergänzt wird. Für das Bayessche Modell wird damit wie in klassischen Anwendungen angenommen, dass sich innerhalb einer einzelnen AuswerteePOCHe keine Änderungen in den Rahmenbedingungen bzw. dem Einfluss der wertrelevanten Merkmale ergeben, was eng mit der bereits angeführten Annahme über Art und Anzahl der Einflussgrößen verknüpft ist.

In Hinsicht auf die Auswertestrategie ist zu berücksichtigen, dass die Bayesschen Modelle im Zuge einer rekursiven Anwendung genutzt werden sollen. Dies setzt voraus, dass die Posteriori-Information einer AuswerteePOCHe als Priori-Information für die Folgeepoche genutzt werden kann. Als Bedingung wird daher formuliert, dass sich in zwei aufeinander folgenden Epochen $(t - 1)$ und t keine Änderungen in den Einflussgrößen des Modells ergeben. Erweist sich eine Einflussgröße in der Epoche t als nicht mehr signifikant oder ist eine bislang im Modell nicht enthaltene Einflussgröße mit Wertrelevanz festzustellen, kann der Posteriori-Datensatz $(t - 1)$ nicht mehr als Priori-Information für die Epoche t dienen. Eine Priori-Information für die Epoche t muss in diesen Fällen komplett neu generiert werden. Signifikante Veränderungen sind durch eine Bayessche Teststatistik festzustellen. Da die hier vorgestellte Auswertestrategie eine rekursive Anwendung für die Fortpflanzung von Unsicherheiten vorsieht, ist für das Modell die Annahme von Stabilität der Entwicklung innerhalb einzelner AuswerteePOCHen zu treffen. Die dynamische Entwicklung des Immobilienmarktes spiegelt sich dann ausschließlich in der Änderung der Trendkoeffizienten zwischen einzelnen AuswerteePOCHen wider, sodass ein Übergang von einer statischen Momentaufnahme auf eine dynamische Sichtweise in Form von Aufdatierung der Kauffallinformationen sowie Fortpflanzung der Unsicherheiten gegeben ist. Durch diesen Ansatz wird neben der geforderten realitätsnäheren Beschreibung von Unsicherheiten im Bayesschen Modell der wesentliche Mehrwert gegenüber den klassischen Ansätzen begründet.

Die allgemeinen Modellannahmen sind insbesondere darauf ausgelegt, die langfristigen, stetigen Veränderungen am Immobilienmarkt abzubilden. Zu diesem Zweck soll eine grundlegende Funktion über möglichst viele AuswerteePOCHen rekursiv angewendet werden können. Gleichwohl setzt dies voraus, dass die Modellannahmen in definierten Abständen auf ihre Gültigkeit überprüft werden. Für die hier vorgestellte Auswertestrategie werden AuswerteePOCHen mit Kauffalldaten von jeweils einem Jahr festgelegt, was somit der gängigen Anwendung in der Praxis entspricht. Es wird weiterhin unterstellt, dass die erläuterten kurz- und mittelfristige Änderungen durch entsprechend erhöhte Unsicherheiten im Modell berücksichtigt werden.

Zur weiteren Modellspezifikation werden zusätzlich Annahmen über die Residuen getroffen, die auch in abweichenden Anwendungsfällen genutzt werden (FAHRMEIR ET AL., 2009, S. 61 ff.):

1. Die erste Annahme bezieht sich auf den *Erwartungswert des Residuenvektors*, der im Mittel Null ergeben muss, d. h. $E(\epsilon_i) = 0$, woraus folgt, dass $E(\epsilon) = \mathbf{0}$.
2. Die zweite Annahme bezieht sich auf die *Varianz und Korrelation der Residuen*. Die Varianz wird als homoskedastisch für alle Beobachtungen angenommen, d. h. $\text{var}(\epsilon_i) = \sigma^2$. Zusätzlich wird angenommen, dass zwischen den einzelnen Residuen keine Korrelationen vorhanden sind, sodass für die VKM der Residuen $\Sigma_{yy} = \Sigma_{ee} = E(\epsilon\epsilon') = \sigma^2 P^{-1} = \sigma^2 \mathbf{I}$ mit $\mathbf{I}$ als Einheitsmatrix angenommen werden kann.
3. Zusätzlich wird in einer weiteren Annahmen von *normalverteilten Residuen* ausgegangen. In Verbindung mit den beiden vorgenannten Annahmen ergibt sich $\epsilon_i \sim N(0, \sigma^2)$ bzw. $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.
4. Die abschließende Annahme ist losgelöst von der Betrachtung der Residuen und bezieht sich auf das funktionale Modell. Hier wird für die Designmatrix $\mathbf{X}$ *voller Spaltenrang*, d. h. $\text{rg}(\mathbf{X}) = m + 1 = u$, unterstellt. Die notwendige Bedingung für die lineare Unabhängigkeit der Spalten der Designmatrix ist eine Anzahl n von Vergleichsfällen größer oder gleich der Anzahl u der Parameter β . Ist die Annahme verletzt, ergibt sich eine der Spalten aus einer Linearkombination einer weiteren, was einen gleichen Informationsgehalt induziert und

als Multikollinearität bezeichnet wird. Als Folge ergibt sich eine nicht eindeutige Schätzung der unbekannten Modellparameter β .

Da ein Kollokationsansatz die Aufspaltung der nach Trendabzug verbleibenden Residuen in einen systematischen unregelmäßigen und einen unsystematischen Anteil bezweckt, kann sich eine sinnvolle Anwendung nur bei vorhandener Autokorrelation der Residuen ergeben. Demgegenüber besteht das Ziel der Regressionsanalyse darin, die Modellspezifikationen derart vorzunehmen, dass Korrelationen zwischen den einzelnen Größen nicht auftreten und die Residuen damit voneinander unabhängig sind. Nach UHDE (1982, S. 43) kann für die Lösung des Problems die Komplexität der Zusammenhänge in der Verkehrswertermittlung angeführt werden. Demnach nehmen mit wachsender Verschiedenartigkeit der zu analysierenden Kauffälle auch die Anzahl der wertbeeinflussenden Merkmale zu, die

- a) lediglich mit hohem Aufwand erhoben werden können (z. B. Lageeinflüsse),
- b) in Form numerisch erfassbarer oder messbarer Daten nicht vorliegen (z. B. psychologisch bedingte Wertvorstellungen),
- c) in keinem sinnvoll funktional formulierbaren Zusammenhang mit dem Kaufpreis stehen oder nur unvollkommen beschrieben werden können (z. B. Einfluss des Kaufzeitpunktes).

Trotz der einschränkenden Gegebenheiten lässt sich nach UHDE (1982, S. 44) ein marktgerechtes Modell entwickeln, welches jedoch grundsätzlich mit den Schwierigkeiten komplexer Datenanalyse verbunden ist – insbesondere die funktionale Spezifikation des Modells ist mit zunehmender Anzahl der Einflussgrößen nur mit hohem Aufwand realisierbar. Als problematisch stellt sich in diesem Kontext die Multikollinearität im funktionalen Modell dar, da die theoretische Berücksichtigung sämtlicher wertbeeinflussender Merkmale in hohen Korrelationen zwischen den Einflussgrößen resultiert. Zum einen führt dies einen kaum nennenswerten Informationszugewinn mit sich, zum anderen ist die Lösung des Ausgleichsalgorithmus aufgrund des Rangabfalls der Designmatrix nicht möglich. Praktische Erfahrungen haben gezeigt, dass Regressions- und Kollokationsmodelle in der Verkehrswertermittlung meist auf einigen wenigen bestimmenden Einflussgrößen beruhen, die den wesentlichen Erklärungsbeitrag zur Zielgröße liefern. Weitere Einflussgrößen im Modellansatz können das Modell lediglich verfeinern, liefern aber keinen weiteren signifikanten Mehrwert (vgl. u. a. ZIEGENBEIN (1977), ZIEGENBEIN & HAWERK (1978), UHDE (1982)). Aufgrund der Überlegungen in Verbindung mit der rein objektiven Vergleichbarkeit von Objekten, die einer geeigneten Grundgesamtheit entstammen, lässt sich unterstellen, dass sich die in einem abgegrenzten räumlichen Bereich liegenden Kauffälle in Bezug auf die wertbeeinflussenden Merkmale zumindest approximativ ähnlich sind und sich dieses Verhalten auch in einem ähnlichen Verhalten der Residuen nach der Trendabspaltung widerspiegelt (UHDE, 1982, S. 44).

Inwiefern die Modellannahmen gerechtfertigt sind, ist lediglich durch praktische Untersuchungen festzustellen. Im folgenden Abschnitt wird zunächst das allgemeine Schema der Auswertestrategie für das indirekte Vergleichsverfahren im Bayesschen Ansatz erläutert.

4.1.3. Schema der Auswertestrategie

Anhand der vorstehenden Ausführungen wird deutlich, dass eine Auswertestrategie eine Reihe von Vorgaben zu erfüllen hat. Ziel ist es dabei, in Bayesscher Formulierung Schätzmodelle zu geben, die den erläuterten Anforderungen gerecht werden und dabei die getroffenen allgemeinen Modellannahmen berücksichtigen. ZIEGENBEIN (1977, S. 46) verweist darauf, dass ein Ablauf zu entwickeln ist, der sich als Zusammenspiel aus der mathematischen Statistik, der Verkehrswertermittlung sowie der durch die mathematischen Modelle gegebenen Voraussetzungen und Möglichkeiten ergibt. Die Anforderungen der einzelnen Aspekte sind in den vorherigen Kapiteln und Abschnitten eingehend diskutiert worden; in ihrer Gesamtheit sind die Wechselwirkungen in der Auswertestrategie jeweils zu berücksichtigen. Da a priori unbekannt ist, welche Einflussgrößen letztlich die Zielgröße optimal beschreiben, stellt sich die Auswertestrategie zudem als ein iterativer Prozess dar, der durchlaufen wird, bis sämtliche Anforderungen erfüllt und somit die als optimal zu bezeichnenden Funktionen gefunden sind. In Abbildung 4.1 wird das grundlegende Schema der Auswertestrategie gezeigt.

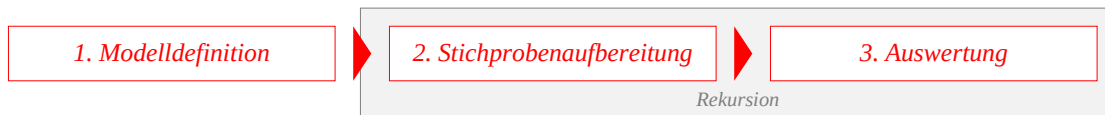


Abbildung 4.1: Allgemeiner Ablauf der Auswertestrategie

Demnach können drei aufeinander aufbauende Schritte identifiziert werden. Neben den generellen Vorüberlegungen über die Erfordernisse multivariater Analysemethoden für das angestrebte Ziel einer Verkehrswertermittlung erfolgt in Schritt 1, der *Modelldefinition*, die Formulierung der konkreten Aufgabe mit den wesentlichen Rahmenbedingungen. Hierbei spielt insbesondere die Eingrenzung der räumlichen und sachlichen Teilmärkte eine bedeutende Rolle. Ebenso müssen die Zielgröße und die potenziell wertrelevanten Merkmale benannt werden. Eine detaillierte Beschreibung der Modelldefinition erfolgt in Abschnitt 4.2.1.

Aufbauend auf der Festlegung der Rahmenbedingungen folgt Schritt 2 mit der *Stichprobenaufbereitung*, die sich in die eigentliche Stichprobenbildung und die Aufbereitung der erhobenen Daten untergliedern lässt. Wesentliches Ziel ist es, Eingangsdaten für die Analysemethoden aufzubereiten, die den diskutierten Erfordernissen der Verkehrswertermittlung – insbesondere im Hinblick auf den gewöhnlichen Geschäftsverkehr sowie auf den vermuteten Erklärungsbeitrag – gerecht werden. Dieser Schritt lässt sich somit als allgemeine Vorprüfung auffassen, in der die Daten einer gegebenen Stichprobe auf Auffälligkeiten untersucht werden. Resultat der (iterativen) Aufbereitung ist eine Stichprobe, von der angenommen werden kann, dass sie die Grundgesamtheit des betrachteten Teilmarkts hinreichend repräsentiert. Weitergehende Erläuterungen sind dem Abschnitt 4.2.2 zu entnehmen.

Im abschließenden Schritt 3 der Strategie nach Abbildung 4.1 erfolgt die eigentliche *Auswertung* der Stichprobe in den Modellen der Bayesschen Regression und der Bayesschen Kollokation. Kern ist hierbei die empirisch fundierte Generierung von Priori-Informationen aus zurückliegenden Epochen sowie die Kombination mit einem Algorithmus für eine optimale Modellwahl. Neben der Formulierung einer iterativen Ausreißersuche und Hypothesentests ist im Rahmen der Bayesschen Kollokation insbesondere die Ableitung von Autokovarianzfunktionen von Interesse. Der Themenkomplex der Auswertung wird in den Abschnitten 4.3.1 und 4.3.2 eingehend behandelt.

Die Schritte der Modelldefinition, Stichprobenaufbereitung und Auswertung ermöglichen die Analyse des Grundstücksmarktes bezüglich der definierten Zielsetzung. Da neben der rein Bayesschen Modellformulierung und -auswertung insbesondere auch die Fortpflanzung von Unsicherheiten auf nachfolgende Auswertepochen im Fokus steht, hat die Auswertestrategie ebenfalls den Anspruch einer *rekursiven Anwendbarkeit* zu erfüllen. Dieser ist zu stellen, da erst auf eine solche Weise die Vorteile der Nutzung von Priori-Informationen in Bayesschen Modellen im Vergleich zu klassisch frequentistischen Ansätzen deutlich werden. Aus diesem Grund sind gemäß Abbildung 4.1 die Schritte der Stichprobenaufbereitung und der Auswertung rekursiv nutzbar. Der Schritt der Modelldefinition hingegen erfolgt lediglich einmalig, da er sich als Definition der allgemeinen Rahmenbedingungen für die Auswertung auffassen lässt und keiner Änderung im Zeitablauf unterlegen ist. Die Rekursion wird in Abschnitt 4.3.3 aufgegriffen.

Der durch die Auswertestrategie entwickelte Prozess soll die Schätzung von Bayesschen Modellen ermöglichen, die sämtlichen Anforderungen genügen. Die in den folgenden Abschnitten detailliert erläuterten Auswerteschritte erlauben es, für eine gegebene Stichprobe von Daten Schätzfunktionen zu ermitteln, mit deren Hilfe Unterschiede zwischen den Realisierungen einer Zielgröße optimal durch die Einflussgrößen beschrieben und die den Eigenarten der immobilienpezifischen Datengrundlage gerecht werden.

4.2. Vorbereitung der Bayesschen Auswertung

Wie in Abbildung 2.3 (S. 32) schematisch gezeigt, kommt der Modelldefinition die Aufgabe zu, die allgemeinen Rahmenbedingungen festzulegen: die Definition der betrachteten Grundgesamtheit und die darauf aufbauende Definition der relevanten Daten sind die entscheidenden Schritte für die weitere Auswertung.

In Abschnitt 4.2.1 wird vertiefend auf die Einzelaspekte eingegangen. Gemäß der einführenden Erläuterungen des Kapitels sind die erhobenen Daten nicht direkt für die Bayesschen Modelle nutzbar, da zusätzliche Voraussetzungen zu erfüllen sind. Die Erhebung einer Stichprobe sowie deren Aufbereitung wird in Abschnitt 4.2.2 behandelt.

4.2.1. Modelldefinition

Die Modelldefinition markiert den ersten Schritt in der Auswertestrategie und ist individuell für jede Verkehrswertermittlung durchzuführen. Die wesentlichen Faktoren der Aufgabendefinition werden in Abbildung 4.2 als Erweiterung zu Abbildung 4.1 dargestellt. Demnach sind zwei aufeinander aufbauende Aspekte zu behandeln: Die Definition der Grundgesamtheit sowie die Definition der relevanten Daten für die Auswertung.

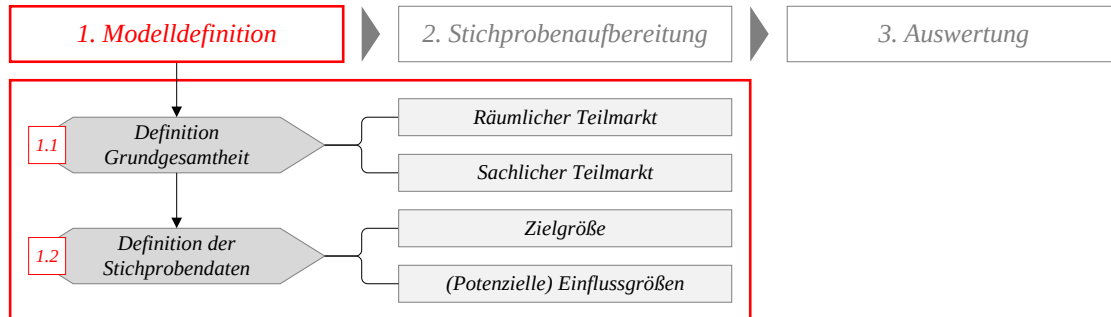


Abbildung 4.2: Modelldefinition

Schritt 1.1 nach Abbildung 4.2 beinhaltet die *Definition der Grundgesamtheit* anhand der spezifischen, im Vorfeld umrissenen Aufgabenstellung. Nach einer allgemeingültigen Definition von SCHLITGEN (1998, S. 1 ff.) handelt es sich bei der Grundgesamtheit um die Menge aller potenziellen Untersuchungsobjekte, die einem zu untersuchenden Phänomen zuzurechnen sind. Generell kann die Menge endliche oder unendlich viele Elemente enthalten und richtet sich nach der Art des Phänomens. Wird dieses, wie im Fall der Verkehrswertermittlung, durch Kaufpreise als Zufallsvariablen definiert, ist die Grundgesamtheit i. d. R. eine unendliche Menge: Wird keine weitere Einschränkung vorgenommen, kann eine umfassende Grundgesamtheit sämtliche Objekte enthalten, die in Abschnitt 2.1.1 (S. 23 ff.) allgemeingültig definiert werden. Eine diese theoretische Grundgesamtheit repräsentierende Auswertung ist nicht zielorientiert und in der Praxis zudem nicht realisierbar. Aus diesem Grund werden Grundgesamtheiten in der Verkehrswertermittlung durch die Identifikationskriterien

- räumlicher Bezug (global, lokal),
- sachlicher Bezug (Objekte nach Nutzungsarten) sowie
- zeitlicher Bezug (Auswertintervall)

eingeschränkt und umfassen damit lediglich einzelne Teilmärkte. Für jeden räumlichen und sachlichen Teilmarkt wird eine eigene Grundgesamtheit gebildet, die jeweils für sich wiederum theoretisch eine unendliche Menge von Elementen enthält. Für die weitere Auswertung kann basierend auf dieser Abgrenzung dann eine geeignete und insbesondere repräsentative Stichprobe der Grundgesamtheit gewählt werden.

Ist der gewünschte Untersuchungsbereich mittels der Identifikationskriterien klar umrissen, kann die *Definition relevanter Daten* in Schritt 1.2 nach Abbildung 4.2 erfolgen. Für die Dependenz-Analyse ist zunächst eine geeignete Zielgröße zu definieren, der nachfolgend die erklärenden Variablen hinzugefügt werden. Hierbei handelt es sich um diejenigen Einflussgrößen, denen nach sachlogischen Erwägungen und im Hinblick auf die festgelegten Identifikationskriterien ein Erklärungsbeitrag zur Zielgröße (Wertrelevanz) unterstellt wird (vgl. Abschnitt 2.2.1, S. 31 ff.). Auf die Möglichkeit der Normierung der Zielgröße in Abhängigkeit der gewählten Einflussgrößen wurde hingewiesen; sie stellt im Zuge der Auswertestrategie jedoch keine generelle Erfordernis dar. Mittels Methoden der deskriptiven Statistik wie bspw. Mittelwert, Median sowie Minima und Maxima lassen sich erste Aussagen über die Modellgrenzen treffen.

Sind die Ziel- und potenziellen Einflussgrößen auf Grundlage der Aufgabendefinition und der Festlegung des räumlichen und sachlichen Teilmarktes festgelegt, ist der Schritt der Modelldefinition abgeschlossen. Im zweiten Schritt erfolgt die Aufbereitung der Stichprobe. Die einzelnen zu beachtenden Aspekte werden im folgenden Abschnitt behandelt.

4.2.2. Aufbereitung einer Stichprobe

Im Schritt der Modelldefinition werden neben grundsätzlichen Überlegungen Anforderungen an die Stichprobe für eine konkrete Wertermittlungsaufgabe formuliert. Als zweite Komponente der Auswertestrategie erfolgt die Erhebung und Aufbereitung der eigentlichen Daten. In Abbildung 4.3 ist der Ablauf mit den wesentlichen Kernpunkten schematisch dargestellt, wie er in den folgenden Abschnitten beschrieben wird.

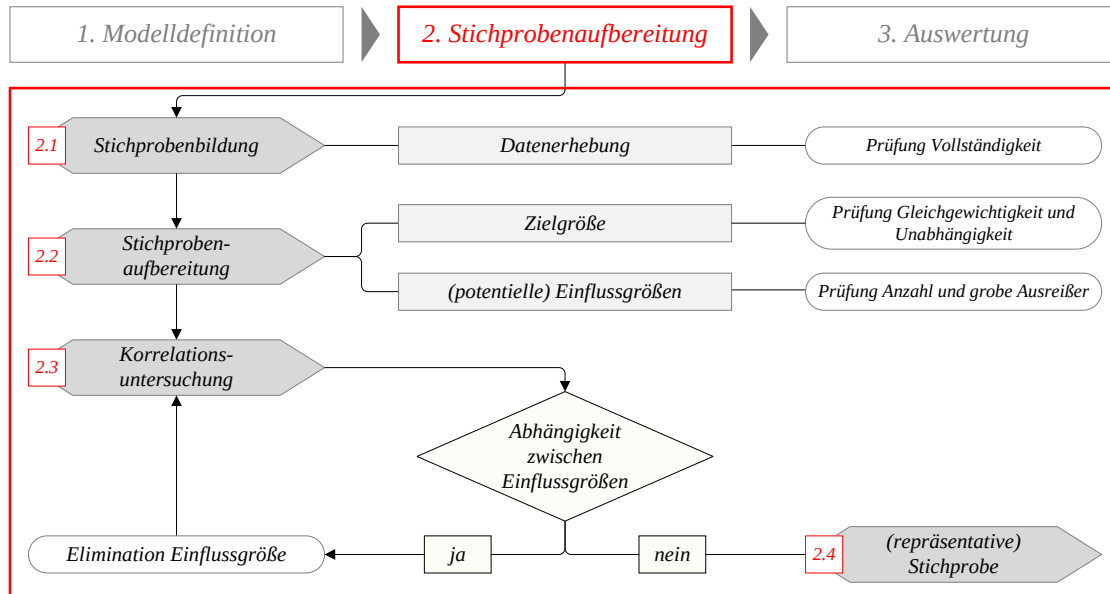


Abbildung 4.3: Ablauf der Stichprobenaufbereitung

4.2.2.1. Stichprobenbildung und -aufbereitung

Für die weitere Aufbereitung ist in Schritt 2.1 nach Abbildung 4.3 zunächst die eigentliche *Stichprobenbildung* erforderlich. In diesem Zuge erfolgt die Datenerhebung gemäß der Definition der Grundgesamtheit, die zu einer möglichst repräsentativen Stichprobe führen soll. Als wesentliche Datenquelle dient regelmäßig die AKS, die, wie erläutert, sämtliche Kauffalldaten in dem Zuständigkeitsbereich eines jeweiligen Gutachterausschusses enthält. Sind die relevanten Daten der AKS selektiert und ggf. um die genannten zusätzlichen Erhebungen ergänzt, folgt die Prüfung auf Vollständigkeit: Diese stellt sicher, dass die aus der AKS selektierten Kauffälle in den relevanten Grundstücksmerkmalen vollständig mit numerischen Werten belegt sind. Im Fall nicht erhobener Merkmale einzelner Kauffälle ergeben sich Datenlücken in der Designmatrix. Zwar existieren in klassisch ingenieurgeodätischen Auswertemethoden eine Reihe von Verfahren, mit deren Hilfe Datenlücken durch synthetische Daten oder Interpolationen geschlossen werden können, diese lassen sich jedoch als unmittelbare Konsequenz der in Abschnitt 2.1.1 (S. 23 ff.) erläuterten Eigenschaften von Immobilien nicht auf Anwendungen in der Verkehrswertermittlung übertragen: Rein mathematisch-statistisch begründete Interpolationen fehlender Einflussgrößen führen durch die Individualität von Immobilien grundsätzlich zu grob falschen Ergebnissen. Einzelne Datenlücken lassen sich ausschließlich durch Nacherhebungen der entsprechenden Daten am konkreten Objekt beheben, daher in den hier durchgeführten Untersuchungen in den Einflussgrößen nicht vollständig erfasste Kauffälle eliminiert und vom weiteren Auswerteprozess ausgeschlossen.

Ergebnis der Stichprobenbildung ist ein Datensatz, der neben der Zielgröße die zugehörigen potenziell wertbeeinflussenden Merkmale als Einflussgrößen enthält. Nach Schritt 2.2 nach Abbildung 4.3 müssen für die weitere Auswertung Ziel- und Einflussgrößen im Rahmen der *Stichprobenaufbereitung* auf die Erfüllung von allgemeinen, wertermittlungsrelevanten Voraussetzungen überprüft werden. Für die Formulierung der Likelihood-Funktion im konjugierten Modell wird bzgl. der Verteilung der Kaufpreise von einer n -dimensional normalverteilten Grundgesamtheit ausgegangen, in der n die Anzahl der einzelnen Stichprobenelemente y_i des Beobachtungsvektors $\mathbf{y}$ bezeichnet. Für die Stichprobenelemente y_i wird damit gefordert, dass sich ihre Randverteilungen wiederum als

eindimensional normalverteilt ergeben, sodass $y_i \sim N(\tilde{y}_i, \sigma_i^2)$, $\forall i \in \{1, \dots, n\}$. In diesem Fall stellt jede Beobachtung y_i eine Zufallsvariable dar, die der Menge hinzuzurechnen ist, die sich aus allen möglichen Kaufpreisen für das betrachtete Objekt i ergeben kann und daher aus den Kombinationen einer potenziell unendlichen Anzahl verschiedener Verkäufer und Käufer resultiert. Sämtliche auf diese Weise denkbaren Kaufpreise bilden eine objektbezogene Grundgesamtheit, wobei für das Zustandekommen der einzelnen Preise grundsätzlich der gewöhnliche Geschäftsverkehr unterstellt wird. Unter Berücksichtigung der Theorie des zentralen Grenzwertsatzes liegt die Zufallsvariable y_i dann als jeweils ein Stichprobenereignis einer normalverteilten Grundgesamtheit vor. Für die Auswertung einer gegebenen Stichprobe bedeutet dies, dass a priori eine multivariate Normalverteilung der Zielgröße angenommen wird. Um diese Vorgabe zumindest approximativ zu erfüllen, werden die Eingangsdaten auf Gleichgewichtigkeit und Unabhängigkeit der Kaufpreise geprüft.

Gleichgewichtigkeit kann unterstellt werden, wenn die Beobachtung mit gleichen Varianzen in das Verfahren eingehen (Homoskedastizität). Im Fall vorliegender Heteroskedastizität ist mit einer Verzerrung der Schätzergebnisse zu rechnen. Ein pragmatischer Untersuchungsansatz, der auch in Anwendungen der AKS genutzt wird, beruht auf dem Verhältnis der Kaufpreise: Das Verhältnis zwischen niedrigstem und höchstem Kaufpreis wird angegeben mit $y_{max} : y_{min} < 5$. Wird das Verhältnis überschritten, ist in Anwendungen der Verkehrswertermittlung von einer schiefsymmetrischen Verteilung der Daten auszugehen. Diese beruht zumeist auf einigen wenigen Ausreißern, die dafür sorgen, dass die Verteilung nicht mehr den gewöhnlichen Geschäftsverkehr – repräsentiert durch die Annahme einer Normalverteilung – abbildet. Da sich eine Verzerrung der erwarteten Normalverteilung ergibt, weisen die Ausreißer ein höheres Gewicht auf die Schätzungsergebnisse auf. Die Beobachtungen sind dann nicht mehr als gleichgewichtig im Ansatz zu beurteilen. Neben dieser ersten Kontrolle mittels deskriptiver Statistik erfolgt eine pragmatische Beurteilung der Symmetrie der Verteilung insbesondere über die Berechnung des 3. Verteilungsmomentes, die als Schiefe (v_3) bezeichnet wird und deren Erwartungswerte im Fall der Normalverteilung $E(v_3) = 0$ ist. Die empirische Schiefe v_3 sowie deren Standardabweichung σ_{v_3} berechnet sich nach PELZER (1976, S. 132 f.) zu

$$v_3 = -\frac{1}{n \sigma^3} \sum_{i=1}^n \epsilon_i^3, \quad \sigma_{v_3} \approx \sqrt{\frac{6}{n}}. \quad (4.1)$$

Liegt die Schiefe innerhalb einer nach pragmatischen Gesichtspunkten gewählten Schranke von $\pm 0,30$ und ergeben sich zudem keine Auffälligkeiten in der visuellen Kontrolle des zugehörigen Histogramms des Beobachtungsvektors, wird eine multivariat normalverteilte Stichprobe angenommen (LGLN, 2015, S. 338). Liegt die Schiefe hingegen außerhalb der genannten Grenzen, sind Ausreißer zu vermuten: Weichen Kaufpreise auf diese Weise von der Masse der übrigen Stichprobenelemente ab, ist davon auszugehen, dass diese nicht unter den Restriktionen des gewöhnlichen Geschäftsverkehrs entstanden sind. Sie müssen als Folge aus der Stichprobe entfernt werden. In der Praxis hat sich für die Bestimmung von Grenzwerten eine $k \cdot \sigma$ -Regel etabliert, in der ein geeigneter Faktor k mit der Standardabweichung der Stichprobe multipliziert wird und Beobachtungen außerhalb des Grenzwerts eliminiert werden. Anwendungen im Rahmen der AKS beruhen auf dem Faktor $k = 2,5$, sodass dieser Wert auch für die Auswertestrategie übernommen wird. An dieser Stelle wird auf eine tiefergehende statistische Behandlung verzichtet, da es sich i. d. R. um grobe Ausreißer in der Stichprobe handelt. Ob ein Kaufpreis tatsächlich als grober Ausreißer identifiziert werden kann, der der Stichprobenabgrenzung des räumlichen und sachlichen Teilmarktes nicht entspricht, hängt entgegen strenger Testverfahren wie bspw. in Auswertemethoden der Ingenieurgeodäsie von gutachterlichem Sachverstand ab: Das Vorliegen eines Ausreißers lässt sich nicht zwangsläufig pauschal mit Mitteln der Statistik beurteilen, sondern muss im Einzelfall interpretiert werden. Auf die generelle Ausreißersuche im Zuge des Bayesschen Auswertalgorithmus wird in Abschnitt 4.3.1 näher eingegangen.

Die *Unabhängigkeit* der Beobachtungen in $\mathbf{y}$ wird pragmatisch durch die Beschränkung der Kauffälle eines einzelnen Verkäufers zum selben Zeitpunkt erreicht, was im Rahmen der Datenselektion zu berücksichtigen ist. Als Grenze hat sich in der Praxis die Reduktion auf eine Anzahl von $\sqrt{n_v}$ bewährt, wobei n_v die Anzahl der Kauffälle bezeichnet, die auf einen einzelnen Verkäufer zurückzuführen sind (ZIEGENBEIN, 1995, S. 324).

Ist die Zielgröße durch die beschriebenen Maßnahmen aufbereitet, erfolgt die Untersuchung der vermuteten Einflussgrößen. Diese sind zunächst dahingehend zu prüfen, ob die *Anzahl der Kauffälle* – in Abhängigkeit vom Datentyp – zum einen für eine Auswertung ausreichen und zum anderen, ob sie so gleichmäßig über den Aussagen-

bereich verteilt sind, dass gesicherte statistische Aussagen möglich sind. Durch ZIEGENBEIN (1986), ZIEGENBEIN (1995) sowie ZIEGENBEIN (1999) wurden auf Basis umfangreicher Untersuchungen Vorgaben veröffentlicht, die bis heute als Standard in der automatisierten Auswertung der AKS hinterlegt und in Niedersachsen durch die *Hinweise zur Auswertung* der AKS für Gutachterausschüsse zur Anwendung vorgeschrieben sind (vgl. LGLN (2015)). Für die Vorgehensweise in der Bayesschen Auswertestrategie können diese Vorgaben – teilweise mit Modifikation – übernommen werden:

- a) Stetige Realisierungen von Einflussgrößen werden aus dem Datensatz entfernt, wenn sie um mehr als das vierfache der Standardabweichung (4σ) vom Mittel abweichen. Liegen größere Abweichungen einzelner Größen vor, kann von Inhomogenitäten in der Stichprobe ausgegangen werden, sodass eine allgemeine Repräsentativität für den gewählten räumlichen und sachlichen Teilmarkt nicht gegeben ist. Wie auch im Fall der Zielgröße handelt es sich hier um eine Prüfung auf grobe Ausreißer, die grundsätzlich dem Sachverstand unterliegt; auf die Erforderlichkeit der manuellen Prüfung der einzelnen Kauffälle wurde hingewiesen.
- b) Diskrete Realisierungen von Einflussgrößen werden aus dem Datensatz entfernt, wenn bei insgesamt d vorhandenen Ausprägungen ein Stichprobenumfang von $n/3d$ unterschritten wird. Das Entfernen von Datensätzen ist gleichbedeutend mit einer neuen Datenselektion unter verringerten Bereichsgrenzen.
- c) Dummy-Variablen mit der Belegung 0/1 müssen eine Mindestbelegung von 10 % pro Ausprägung aufweisen, da andernfalls der Anteil zu gering ist, um eine gesicherte Aussage treffen zu können. Wird die prozentuale Mindestbelegung nicht erreicht, wird die Einflussgröße aus dem Datensatz entfernt.

Während in den klassischen Auswertansätzen eine zusätzliche Vorgabe existiert, nach der pro geschätzter Einflussgröße im Regressionsmodell eine Anzahl von mindestens 15 Kauffällen vorliegen muss, soll diese Einschränkung durch die Einführung von Vorwissen im Bayesschen Modell aufgehoben werden. Eine erweiterte Anwendbarkeit des Ansatzes durch die Aufhebung ist zu erwarten und wird in Kapitel 7 aufgegriffen.

Durch die beschriebene Vorgehensweise ist die Stichprobe der Ziel- und Einflussgrößen aufbereitet. Im Anschluss folgt der Schritt der Beurteilung korrelativer Beziehungen der Größen, der die Stichprobenselektion abschließt und im nachstehenden Abschnitt erläutert wird.

4.2.2.2. Korrelationsuntersuchung

Die aufbereiteten Stichprobendaten beruhen bislang auf rein sachverständigem Ermessen. Es werden lediglich diejenigen potenziellen Einflussgrößen erhoben, die in dem gewählten räumlichen und sachlichen Teilmarkt nach Erfahrungswerten die Zielgröße nachhaltig beeinflussen. Neben den unterstellten Kausalzusammenhängen kann durch die Analyse der *Korrelationen* auch numerisch geprüft werden, inwieweit die erhobenen Daten voneinander abhängen und ob sich redundante Einflussgrößen im Modellansatz finden. Dabei ist zwischen zwei Arten der Abhängigkeit zu unterscheiden: Zum einen sollen zwischen der Zielgröße auf der einen und den Einflussgrößen auf der anderen Seite Korrelationen vorliegen, was auf einen Erklärungsbeitrag der Einflussgröße in dem zu formulierenden Modell hindeutet. Die zweite Art der Abhängigkeit ist zwischen den Einflussgrößen selbst zu untersuchen und zielt auf die Unabhängigkeit der einzelnen Merkmale ab. Im Gegensatz zu den Korrelationen zur Zielgröße sollen in diesem Fall möglichst geringe Korrelationen zwischen den einzelnen Einflussgrößen vorliegen, da sie andernfalls einen gleichgerichteten Einfluss auf die Zielgröße bewirken und zum Problem der Multikollinearität führen: Ergibt sich im ungünstigsten Fall eine Einflussgröße als Linearkombination einer oder mehrerer weiterer Einflussgrößen, erfährt die Designmatrix einen Rangabfall, der eine Inversion im Ausgleichsalgorithmus zur Schätzung der unbekannten Parameter verhindert oder zu numerischer Instabilität führt. Aber auch hohe Korrelationen, die noch eine Lösung ermöglichen, führen unter Umständen zu einer erschwerten Interpretierbarkeit der einzelnen Parameter. Gemäß Schritt 2.3 nach Abbildung 4.3 werden in der Auswertestrategie lediglich die Korrelationen zwischen den Einflussgrößen betrachtet; der Beitrag der Einflussgrößen zur Erklärung der Zielgröße wird im Zuge der Bayesschen Hypothesentests geprüft und kann daher an dieser Stelle vernachlässigt werden. Für die Untersuchung der Einflussgrößen wird der pragmatische Ansatz der AKS übernommen, der für Korrelation von $\rho > |0.7|$ die Entfernung einer der betroffenen Einflussgrößen aus dem Auswertansatz fordert. Liegen die Korrelationen zwischen

$|0.3| < \rho < |0.7|$ können die Einflussgrößen zwar im Ansatz verbleiben, sollten jedoch nicht einzeln interpretiert werden (ZIEGENBEIN, 1995, S. 326). Auf mögliche Korrelationen ist bereits im Schritt der Stichprobenaufbereitung zu achten: Einflussgrößen, von denen bekannt ist, dass sie sich ähneln oder ggf. sogar auseinander berechnen lassen, sind nach sachverständigem Ermessen zu identifizieren, und es ist nur eine Größe in den Modellansatz aufzunehmen. Als klassisches Beispiel sei die Grundflächenzahl (GRZ) genannt, die sich rechnerisch aus der Geschossflächenzahl (GFZ) ableiten lässt. Eine Aufnahme beider Größen in den Modellansatz wäre als nicht sinnvoll zu beurteilen, sodass eine Entscheidung getroffen werden muss, welche der Einflussgrößen eingeführt werden soll.

Für die Berechnung der Korrelationen muss nach der Skalierung der Daten getrennt werden. Für stetig verteilte Zufallsvariablen hat sich die Berechnung *Spearman'scher Rangkorrelationskoeffizienten* als zweckmäßig erwiesen, die für Anwendungen in der Verkehrswertermittlung bereits durch ZADDACH ET AL. (2010) genutzt wurden. Rangkorrelationskoeffizienten bezeichnen ein parameterfreies Maß für die Bestimmung der Abhängigkeit und verzichten so auf explizite Annahmen über die Wahrscheinlichkeitsverteilung der eingeführten Größen. Zusätzlich sind keine Annahmen über Linearität in den Variablen notwendig, ebenso handelt es sich um eine gegenüber Ausreißern robuste Maßzahl. Der Spearman'sche Rangkorrelationskoeffizient ρ^s ergibt sich allgemein zu

$$\rho^s = \frac{\sum_{i=1}^n [R(x_i) - \bar{R}(x)] [R(y_i) - \bar{R}(y)]}{\sqrt{\sum_{i=1}^n [R(x_i) - \bar{R}(x)]^2 \sum_{i=1}^n [R(y_i) - \bar{R}(y)]^2}} = 1 - \frac{6 \sum_{i=1}^n \Delta_i^2}{n[n^2 - 1]}, \quad (4.2)$$

wobei $\Delta_i = R(x_i) - R(y_i)$ gilt. Wie der Name bereits impliziert, beruht die Berechnung der Korrelationen auf Ranginformationen. Zu diesem Zweck werden Rangzahlen $R(x_i)$ bzw. $R(y_i)$ vergeben, sodass in jeder der beiden Reihen x_i und y_i die kleinste Realisation den Rang 1 und die größte den Rang n erhält oder für beide Reihen umgekehrt. Mit $\bar{R}(\cdot)$ wird der Mittelwert der Ränge bezeichnet. Für den Fall, dass sogenannte Bindungen auftreten (d. h. mehrere Realisationen einer Zufallsvariablen sind gleich) wird als Rang für x_i und y_i das arithmetische Mittel der in Frage kommenden Ränge gesetzt. In diesem Fall ergibt sich der Spearman'sche Korrelationskoeffizient zu

$$\rho_t^s = \frac{n[n^2 - 1] - 6 \sum_{i=1}^n \Delta_i^2 - \frac{1}{2} [D_1 + D_2]}{\sqrt{[n[n^2 - 1] - D_1][n[n^2 - 1] - D_2]}} \quad (4.3)$$

$$\text{mit } D_j = \sum_{k=1}^{p_j} [d_{jk}^3 - d_{jk}] \quad \text{für } j = 1, 2, \quad (4.4)$$

wobei p_j die Anzahl verschiedener Werte der Reihen $x_1, \dots, x_n$ und $y_1, \dots, y_n$ und d_{jk} die Häufigkeit des k -ten Wertes in der ersten Reihe $x_1, \dots, x_n$ bzw. in der zweiten Reihe $y_1, \dots, y_n$ bezeichnet (HARTUNG & ELPELT, 2007, S. 191 f.).

Neben den Korrelationen von stetigen Einflussgrößen können in einem Datensatz auch dichotome Merkmale (Dummy-Variablen) auftreten, die i. d. R. mit den Ausprägungen 0 und 1 codiert sind. Soll der Zusammenhang zwischen einer stetigen Einflussgröße und einem dichotomen Merkmal bestimmt werden, kann die *punktbiseriale Korrelation* berechnet werden. Diese ergibt sich zu

$$\rho^{pb} = \frac{\bar{x}_0 - \bar{x}_1}{\sigma_x} \sqrt{\frac{n_0 n_1}{n(n-1)}}, \quad (4.5)$$

in der n_0 und n_1 die Anzahl der Objekte in der Kategorisierung nach dem dichotomen Merkmal bezeichnen und $n = n_0 + n_1$ gilt. Mit $\bar{x}_0$ sowie $\bar{x}_1$ werden die entsprechenden durchschnittlichen Ausprägungen der stetigen Größe x angegeben, σ_x stellt die Standardabweichung aller Werte von x ohne Berücksichtigung der Zugehörigkeit zu einer Gruppe dar (BORTZ, 2005, S. 171).

Während für den Fall stetiger und ordinalskaliert Einflussgrößen die Rangkorrelationskoeffizienten nach Spearman genutzt werden können, lässt sich der Zusammenhang zweier dichotome Merkmale durch den sogenannten ϕ -Koeffizienten bestimmen. Ohne Herleitung lässt sich dieser durch die Formel

$$\phi = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}} \quad (4.6)$$

angeben, in der die Buchstaben a , b , c und d die Häufigkeiten einer 2×2 -Tabelle der beiden Merkmalsalternativen ergibt. Abschließend lassen sich für die Untersuchung des Zusammenhangs eines dichotomen und eines ordinal skalierten Merkmals biseriale Rangkorrelationen berechnen:

$$\rho^b = \frac{2}{n} (\bar{x}_1 - \bar{x}_2). \quad (4.7)$$

Hier bezeichnet $\bar{x}_1$ den durchschnittlichen Rangplatz der zum Merkmal x_1 gehörenden Kategorie, x_2 den durchschnittlichen Rangplatz der zum Merkmal x_2 gehörenden Kategorie und n wiederum den Umfang der Stichprobe. Für weitergehende Ausführungen und Rechenbeispiele zu den einzelnen Korrelationskoeffizienten wird an dieser Stelle auf die weiterführende Literatur verwiesen (vgl. u. a. BORTZ (2005), HARTUNG & ELPELT (2007)).

Zusammenfassend werden in Tabelle 4.3 die verschiedenen Arten der Korrelationsberechnungen im Verhältnis zur spezifischen Datenart aufgezeigt. Sind die Korrelationen zwischen der Ziel- und den Einflussgrößen sowie zwischen den Einflussgrößen bestimmt, ist anhand der genannten Korrelationsschranken zu bestimmen, ob eine Einflussgröße aus dem Ansatz eliminiert werden sollte. Ergeben sich keine Auffälligkeiten, liegt als Ergebnis in Schritt 2.4 nach Abbildung 4.3 eine repräsentative Stichprobe des gewählten sachlichen und räumlichen Teilmarktes vor, die mittels der Bayesschen Modelle ausgewertet werden kann. Die erforderlichen einzelnen Schritte werden im folgenden Abschnitt beschrieben.

Tabelle 4.3.: Übersicht über die Berechnungen von Korrelationen in der Bayesschen Auswertestrategie

Merkmal x_1	Merkmal x_2		
	stetig	dichotom	ordinal
stetig	Rangkorrelation	punktbiseriale Korrelation	Rangkorrelation
dichotom	punktbiseriale Korrelation	ϕ -Koeffizient	biseriale Rangkorrelation
ordinal	Rangkorrelation	biseriale Rangkorrelation	Rangkorrelation

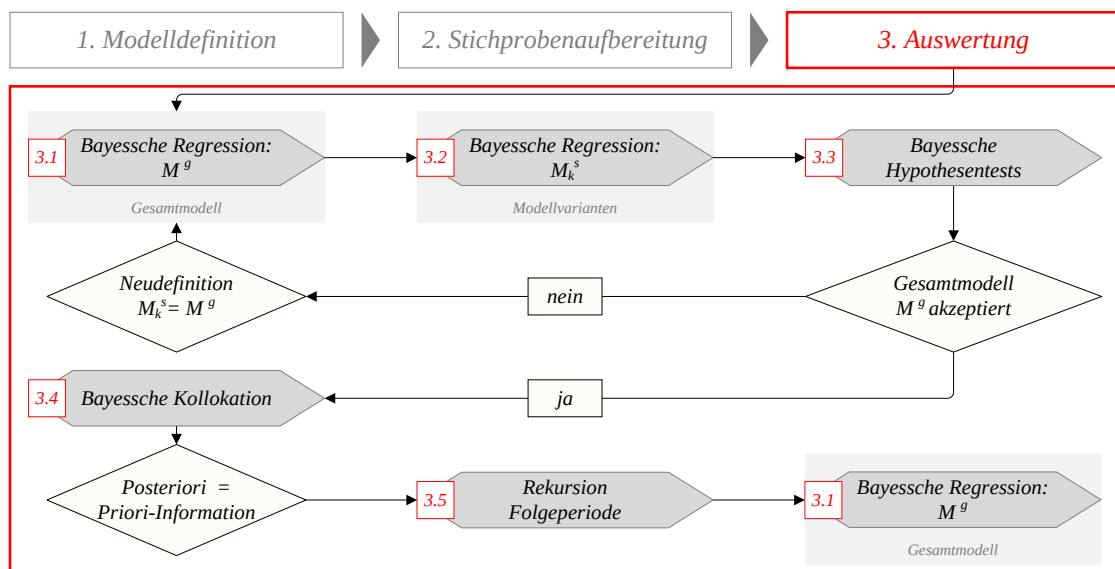


Abbildung 4.4.: Auswertestrategie für das indirekte Vergleichswertverfahren basierend auf der Bayes-Theorie

4.3. Bayessche Auswertung einer Stichprobe

Ist die Stichprobenaufbereitung abgeschlossen, kann aufbauend der dritte Schritt der Auswertung erfolgen. Der Auswerteprozess wird in Abbildung 4.4 aufgezeigt. Schematisch dargestellt werden dabei die Einzelschritte, die (teilweise iterativ) durchlaufen werden, um einen konkreten Datensatz auszuwerten. Wie in Schritt 3.1 aufgeführt, wird zu diesem Zweck zunächst ein sogenanntes Gesamtmodell geschätzt. Das Gesamtmodell enthält die in der Mo-

dellddefinition festgelegte Zielgröße sowie sämtliche potenziell wertrelevanten Einflussgrößen; es basiert ausschließlich auf den Ergebnissen einer informativen Bayesschen Regression, da dieses den Ausgangspunkt für die folgende Modellwahl darstellt. Da von vornherein nicht feststeht, welche Einflussgröße tatsächlich einen Erklärungsbeitrag zum Zielmodell enthält, werden in Schritt 3.2 Modellvarianten geschätzt. Die Modellvarianten setzen sich aus Kombinationen von Einflussgrößen des Gesamtmodells oder zusätzlichen Einflussgrößen zusammen und werden nach demselben Prinzip geschätzt wie das Ausgangsmodell. Sämtliche Modellschätzungen enthalten dabei eine Ausreißersuche basierend auf der Bayes-Theorie. Sind das Gesamtmodell sowie die Modellvarianten bestimmt, erfolgen in Auswerteschritt 3.3 Bayessche Hypothesentests, die die jeweilige Modellgüte angeben mit dem Ziel, das plausibelste der eingeführten Modelle (Gesamtmodell oder eine der Modellvarianten) zu bestimmen. Mit diesem Schritt ist die Bayessche Auswertung für die vorliegende Stichprobe abgeschlossen und ermöglicht die Bestimmung der Lösung der Bayesschen Kollokation in Schritt 3.4 sowie die Ermittlung der relevanten Unsicherheitsmaße. In Schritt 3.5 schließt sich die Rekursion an, indem die Posteriori-Lösung als Priori-Information für die folgende Auswerteepoche gesetzt wird. Für den Fall, dass das Gesamtmodell in Schritt 3.3 als Ergebnis des Hypothesentests abgelehnt wird, erfolgt eine Neudefinition durch Substitution des Gesamtmodells durch die wahrscheinlichere Modellvariante sowie ein weiterer iterativer Berechnungsdurchlauf. Der konkrete Ablauf der Schätzungen und die aufeinander aufbauenden Schritte werden in Abschnitt 4.3.1 ausführlich behandelt.

4.3.1. Ablauf der Bayesschen Modellschätzungen

Unabhängig von der Problematik der Modellwahl sind sowohl Gesamtmodell als auch die Modellvarianten nach der gleichen Vorgehensweise zu schätzen. Gleiches trifft auf die Berechnungen bzgl. der informativen und nichtinformativen Bayesschen Kollokation zu, die als erweiterte funktionale Modellierungen der Regressionslösungen aufzufassen sind. Aus diesem Grund werden in den folgenden Unterabschnitten zunächst die allgemeingültigen Prinzipien der Ausreißersuche in den Bayesschen Regressionsmodellen sowie die Ableitung von Autokovarianzfunktionen für die Bayesschen Kollokationsmodelle behandelt.

4.3.1.1. Ausreißersuche in der Bayesschen Regression

Die Ermittlung eines optimalen Bayesschen Regressionsmodells erfordert die Elimination von Ausreißern in den Daten. Wie erläutert, dienen die Residuen der Bayesschen Regression als Grundlage für die Bayessche Kollokation, sodass für zutreffende Ergebnisse eine möglichst ausreißerbereinigte Datengrundlage anzustreben ist.

Für die Ausreißersuche in Bayesschen Modellen existiert eine Vielzahl von Ansätzen, die üblicherweise auf der Definition eines Ausreißers nach FREEMAN (1985, S. 350) beruhen: Demnach handelt es sich bei einer Beobachtung um einen Ausreißer, sofern die betreffende Beobachtung nicht durch die gleichen Rahmenbedingungen wie die Mehrheit der übrigen Beobachtungen entstanden ist. Diesem Grundsatz folgen gängige methodische Ansätze, wie sie z. B. von HAWKINS (1980), BECKMAN & COOK (1983), BARNETT & LEWIS (1984) sowie PETTIT & SMITH (1985) diskutiert werden. Für die Auswertestrategie wird ein Ansatz gewählt, der auf den Forschungen von ZELLNER & MOULTON (1985) beruht und durch CHALONER & ROLLIN (1988) für lineare Modelle mit normalverteilten Residuen ϵ_i entwickelt wurde. Grundannahme ist ein Modell der Form $\epsilon = \mathbf{y} - \mathbf{X}\beta$ mit der Verteilungsannahme $\epsilon \sim N(0, \sigma^2 \mathbf{I})$ analog zu (3.31). Basierend auf dieser Annahme ist das Prinzip der Ausreißersuche, dass ein beliebiges ϵ_i als Ausreißer identifiziert wird, sobald die Beziehung $|\epsilon_i| > k\sigma$ erfüllt ist und in der k einen festzulegenden Parameter darstellt. Um die Posteriori-Wahrscheinlichkeit berechnen zu können, dass $|\epsilon_i|$ größer als der Term $k\sigma$ ist, muss zunächst die Posteriori-Verteilung $p(\epsilon|\mathbf{y})$ bestimmt werden. Diese beruht auf den für lineare Modelle getroffenen Verteilungsannahmen über den Präzisionsparameter τ , den unbekannten Parametern β sowie den Beobachtungen $\mathbf{y}$. Wie durch die Beziehung $\epsilon = \mathbf{y} - \mathbf{X}\beta$ ersichtlich, ergeben sich die Residuen als lineare Funktion der unbekannten Parameter. Wird im informativen Fall mit

$$\mathbf{H} = \mathbf{X} \left(\mathbf{X}'\mathbf{X} + \mathbf{Q}_{\beta\beta}^{-1} \right)^{-1} \quad (4.8)$$

die sogenannte Prädiktionsmatrix eingeführt, lässt sich der Vektor der Residuen ϵ bedingt durch τ ohne Herleitung als multivariate Normalverteilung mit dem Schätzwert $\bar{\epsilon} = \mathbf{y} - \mathbf{X}\bar{\beta}$ und der VKM $\tau^{-1}\mathbf{H}$ angeben. Für jedes Ele-

ment ϵ_i ergibt sich dann eine t-Verteilung mit dem Schätzwert $\bar{\epsilon}_i$, der Präzision $\bar{p}/\bar{b} h_{ii}$ (mit $h_{ij}, i, j \in \{1, \dots, n\}$ als Elementen der Prädiktionsmatrix) und den Freiheitsgraden $2\bar{p}$ (CHALONER & ROLLIN, 1988, S. 652), (DEGROOT, 1970, S. 42). Für den nichtinformativen Fall ergibt sich lediglich eine Änderung in der Formulierung der Prädiktionsmatrix, in der entsprechend der Term $\mathbf{Q}_{\beta\beta}^{-1}$ entfällt. Weiterhin ist zu beachten, dass sich die Berechnung der Parameter $\bar{b}$ und $\bar{p}$ ändert:

$$\bar{b} = \frac{1}{2} (n - u), \quad \bar{p} = \frac{1}{2} (n - u) \bar{\sigma}^2. \quad (4.9)$$

Im Folgenden muss die Posteriori-Wahrscheinlichkeit bestimmt werden, mit der die i -te Beobachtung als Ausreißer identifiziert wird. Diese lässt sich mit der bedingten Wahrscheinlichkeit $p_i = p(|\epsilon_i| > k\sigma | \mathbf{y})$ angeben. Bezeichnet weiterhin $\Phi(z)$ die Standard-Normalverteilung mit den Hilfsgrößen

$$z_1 = (k - \bar{\epsilon}_i \sqrt{\tau}) / (\sqrt{h_{ii}}), \quad z_2 = (-k - \bar{\epsilon}_i \sqrt{\tau}) / (\sqrt{h_{ii}}), \quad (4.10)$$

in denen h_{ii} als Hebelwerte bezeichnet werden, lässt sich die Posteriori-Wahrscheinlichkeit ermitteln zu

$$p_i = p(|\epsilon_i| > k\sigma | \mathbf{y}) = \int [1 - \Phi(z_1) + \Phi(z_2)] p(\tau | \mathbf{y}) d\tau. \quad (4.11)$$

Die berechneten Wahrscheinlichkeiten p_i können im Anschluss mit der Priori-Wahrscheinlichkeit $2\Phi(-k)$ verglichen werden: Residuen mit einer hohen Posteriori-Wahrscheinlichkeit p_i weisen i. d. R. ein hohes $|\bar{\epsilon}_i|$, ein hohes h_{ii} oder beides auf. Im Fall eines hohen (geschätzten) $|\bar{\epsilon}_i|$ ist auch die Wahrscheinlichkeit für eine hohe tatsächliche Realisierung $|\epsilon_i|$ gegeben. Ein hoher Wert für h_{ii} hingegen impliziert eine hohe Unsicherheit über ϵ_i . In Ergänzung zu der Berechnung der Posteriori-Wahrscheinlichkeiten lassen sich die HPDIs der Residuen angeben mit

$$P\left(\bar{\epsilon}_i - t_{1-\alpha; n-p}(\bar{\sigma}\sqrt{h_{ii}}) < \epsilon_i < \bar{\epsilon}_i + t_{1-\alpha; n-p}(\bar{\sigma}\sqrt{h_{ii}})\right) = 1 - \alpha, \quad (4.12)$$

wobei diese Grenzen für die Beschreibung der einzelnen Randverteilungen, bspw. durch Visualisierung in Residuenplots, herangezogen werden können (CHALONER & ROLLIN, 1988, S. 653).

Für die konkrete Bestimmung der Posteriori-Wahrscheinlichkeiten und der Identifikation der Ausreißer muss weiterhin der Parameter k bestimmt werden, der die Ausreißergrenzen festlegt. Dieser wird i. d. R. derart gewählt, dass eine hohe Priori-Wahrscheinlichkeit dafür vorliegt, keinen Ausreißer in den Daten vorzufinden. Wird bspw. ein Wert von 95 % angenommen, ergibt sich die Beziehung

$$k = \Phi^{-1}\left(0,5 + 0,5\left(0,95^{1/n}\right)\right), \quad (4.13)$$

mittels derer sich dann in Abhängigkeit der Anzahl der Beobachtungen n ein Wert für den Parameter k ermitteln lässt. Beispielhafte Ergebnisse für Priori-Wahrscheinlichkeiten von 95 % sind in Tabelle 4.4 aufgeführt. Jede Beobachtung mit einer Posteriori-Wahrscheinlichkeit nach (4.11), die über der Priori-Wahrscheinlichkeit von $2\Phi(-k)$ liegt, kann als Ausreißer gewertet werden (CHALONER & ROLLIN, 1988, S. 653).

Tabelle 4.4.: Parameter k in der Bayesschen Ausreißersuche für die Priori-Wahrscheinlichkeit 95 % nach CHALONER & ROLLIN (1988)

1- α	Anzahl Beobachtungen n					
	20	50	100	150	200	250
0,950	3,0	3,3	3,4	3,5	3,7	3,7

Durch die Ausreißer-Suche können die Priori-Stichprobe des Zeitpunktes $(t - 1)$ sowie die Likelihood-Stichprobe zum Zeitpunkt t jeweils auf Ausreißer überprüft werden. Hierbei handelt es sich im Auswerteprozess um einen iterativen Ablauf: Liegt ein Ausreißer vor, wird dieser aus dem Datensatz eliminiert. Für den Fall, dass mehrere Ausreißer identifiziert werden, wird die Beobachtung aus dem Datensatz entfernt, deren Residuum ϵ_i die höchste Posteriori-Wahrscheinlichkeit aufweist. Im Anschluss ist die Bayessche Regression erneut unter Ausschluss der

betreffenden Beobachtung zu berechnen. Dieser Vorgang setzt sich fort, bis kein Ausreißer mehr im Datensatz identifiziert werden kann.

Sind auf diese Weise sämtliche Ausreißer entfernt, können die endgültigen Schätzwerte der Bayesschen Regression – informativ wie nichtinformativ – nach den Ausführungen in Abschnitt 3.2 berechnet werden. Die aus diesen Modellen resultierenden Residuen sind die wesentliche Grundlage für die Berechnung der Bayesschen Kollokationsansätze. Die Ableitung von Autokovarianzfunktionen auf Basis der ermittelten Residuen werden im folgenden Abschnitt erläutert.

4.3.1.2. Ableitung von Autokovarianzfunktionen

In der klassischen Zeitreihenanalyse finden als Maßzahl für die Beschreibung des Zusammenhangs zweier Beobachtungen – hier der gegenseitigen Abhängigkeit zweier Vergleichsobjekte i und k – Autokovarianz- und Autokorrelationsfunktionen Anwendung. Grundlegend ist die Definition eines stochastischen Prozesses: Während in der klassischen Zeitreihenanalyse die zufällige Komponente lediglich eine untergeordnete Rolle spielt, stellt das Konzept der stochastischen Prozesse Zeitreihen als dynamische Vorgänge mit Zufallscharakter in den Vordergrund (HOWIND, 2005, S. 12). Da im Rahmen einer Kollokation insbesondere die nach Trendabspaltung in den Residuen verbleibenden stochastischen Abhängigkeiten als Restinformationen von Interesse sind, kann der Residuenvektor ϵ als Realisierung eines stochastischen Prozesses aufgefasst werden, die eine Zufallsstichprobe aus der Grundgesamtheit des Teilmarktes darstellt. Für die Anwendbarkeit des Modells der Kollokation stellen sich dann aus der Sichtweise stochastischer Prozesse Anforderungen, die für die zutreffende Formulierung der Autokovarianzfunktion erfüllt sein müssen. Im Einzelnen sind die Eigenschaften der (schwachen) Stationarität, der Isotropie sowie der Ergodizität zu nennen. Die folgenden Ausführungen setzen für Residuen die Verteilungsannahme $\epsilon \sim (\mathbf{0}, \sigma^2 \mathbf{P}^{-1})$ voraus.

1. *Schwache Stationarität*: Ein stochastischer Prozess ist dann schwach stationär, wenn

$$E(\epsilon_i) = \mu \quad \forall i \in \{1, \dots, n\}, \quad E(\epsilon_i - E(\epsilon_i))^2 = \text{var}(\epsilon_i) = \sigma^2 \quad \forall i \in \{1, \dots, n\} \quad (4.14)$$

gilt, wobei μ einen konstanten Erwartungswert bezeichnet. Charakteristisch für schwach stationäre Prozesse sind somit konstante erste und zweite Verteilungsmomente der Residuen, was gleichbedeutend mit translatorischer Invarianz ist. In Bezug zur Verkehrswertermittlung müssen sich für beliebige Stichproben aus der gewählten Grundgesamtheit des Teilmarktes gleiche Erwartungswerte und Varianzen ergeben; die Autokovarianzfunktion ist dann unabhängig vom Ort konstant (KOCH & SCHMIDT, 1994, S. 166).

2. *Isotropie*: Der Begriff der Isotropie ergänzt die translatorische Invarianz um die räumliche Invarianz der Verteilungsfunktion der Residuen, was sich als zweidimensionales Stationaritätsproblem formulieren lässt. Praktisch bedeutet dies, dass die Autokovarianzfunktion des gewählten Teilmarktes richtungsunabhängig gültig sein muss (ZIEGENBEIN & HAWERK, 1978, S. 256).
3. *Ergodizität*: Für die Analyse stochastischer Prozesse stehen zumeist nur wenige und teils beschränkte Realisierungen zur Verfügung. Daher ist eine wesentliche Voraussetzung für die Auswertungen, dass sich alle statistischen Eigenschaften des Prozesses selbst bereits aus einer einzelnen Realisierung des Prozesses bzw. aus der aktuell gewählten Stichprobe ergeben (HOWIND, 2005, S. 12).

Eine Übersicht über die Möglichkeiten zur Überprüfung der Annahmen für die klassischen Anwendungsgebiete der Zeitreihenanalyse geben u. a. KOCH & SCHMIDT (1994), HOWIND (2005) sowie TEUSCH (2006). Im Vergleich stellen die Daten der Verkehrswertermittlung ungleich höhere Anforderungen an die Überprüfung. ZIEGENBEIN & HAWERK (1978, S. 256) verweisen darauf, dass sich die empirische Prüfung der Annahmen für Anwendungen in der Verkehrswertermittlung als sehr schwierig darstellt. Die Annahmen werden hier als erfüllt betrachtet und ausschließlich durch Plausibilitätserklärungen belegt. Auch UHDE (1982, S. 45 f.) beschränkt sich größtenteils auf Plausibilität und setzt dafür eine sorgfältige Auswahl der Stichprobenelemente voraus. Für die Stationarität des ersten Verteilungsmoments kann demnach die Forderung

$$E(\epsilon_i) = E(v_i) = E(s_i) = 0 \quad \forall i \in \{1, \dots, n\} \quad (4.15)$$

formuliert werden, die als erfüllt angesehen wird, sofern der deterministische Anteil des Trends in den Kaufpreisen richtig erfasst ist. Das zweite zentrale Verteilungsmoment lässt sich nach UHDE (1982, S. 46) zerlegen in

$$E(\epsilon_i \epsilon_j) = E(v_i v_j) + E(s_i s_j) \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, n\}. \quad (4.16)$$

Für den Varianzanteil der Signalkomponente ($E(s_i s_j)$) ergeben sich die Elemente der VKM theoretisch ausschließlich durch eine Funktion der relativen Nachbarschaft, was bedeutet, dass zumindest in Teilbereichen gleiche Autokovarianzfunktionen gelten müssen. Eine Überprüfung der Annahme ließe sich daher durch einen Vergleich der empirischen Kovarianzen zufällig zusammengestellter Stichproben erreichen. Für den Varianzanteil des Rauschens ($E(v_i v_j)$) wird a priori stochastische Unabhängigkeit vorausgesetzt. Nach UHDE (1982, S. 46) lässt sich ein Nachweis der Stationarität hier nur dann erbringen, sofern sämtliche Beobachtungen gefiltert würden. Während die Eigenschaft der Isotropie in der Verkehrswertermittlung mangels praktischer Überprüfbarkeit nach ZIEGENBEIN & HAWERK (1978, S. 256) und UHDE (1982, S. 46) stets als erfüllt vorausgesetzt wird, lässt sich die Ergodizität dann unterstellen, wenn davon ausgegangen wird, dass für einen spezifischen Kauffall unter der Voraussetzung des gewöhnlichen Geschäftsverkehrs andere Verhandlungspartner auch zu ähnlichen Kaufpreisen gelangt wären. Da die Auswertungen grundsätzlich auf ausreißerbereinigten Daten basieren, ist von gewöhnlichem Geschäftsverkehr auszugehen, und der Schluss von einer einzelnen Stichprobe auf die Grundgesamtheit ist zulässig (UHDE, 1982, S. 46). Zusammenfassend lässt sich festhalten, dass die empirische Überprüfung der Eigenschaften eines stochastischen Prozesses für die Verkehrswertermittlung nur schwer realisierbar ist und als hinreichend erfüllt vorausgesetzt wird.

Im Vergleich zur klassischen Formulierung von Autokovarianzfunktionen erfordert deren Ableitung in der Verkehrswertermittlung eine modifizierte Vorgehensweise, da im Bewertungsfall keine klassischen Messwerte in Form von Zeitreihen vorliegen: Dabei wird der Residuenvektor ϵ als eine Realisierung eines stochastischen Prozesses aufgefasst, der eine Zufallsstichprobe aus der Grundgesamtheit des gewählten Teilmarktes darstellt. Da es sich nicht um Zeitreihen handelt und die Residuen nach Trendabspaltung kein Ordnungskriterium im Ablauf aufweisen, sind für die Ableitung einer Autokovarianzfunktion Gruppen auf Grundlage der wertmäßigen Ähnlichkeit der Vergleichsobjekte zu bilden. In einem Vorbereitungsschritt müssen die Residuen nach Trendabspaltung daher geeignet klassifiziert werden, um dann mittels der Autokovarianzfunktion eine lineare Abhängigkeit zwischen den Klassen bestimmen zu können. Durch diese Vorgehensweise kann (2.21) umformuliert werden zu

$$C(d) = \frac{1}{n-d} \sum_{i=1}^{n-d} \epsilon_i \epsilon_{i+d}, \quad (4.17)$$

wobei d die Klasse angibt, deren Klassenbreite Δ_d nach sachverständigem Ermessen gewählt werden muss, n die Anzahl der Klassenwerte und ϵ_i die Residuen aus der vorherigen Trendbestimmung bezeichnet. Analog lässt sich die Berechnung der Autokorrelationsfunktion nach (2.22) durch Substitution von $\tau = k$ herleiten.

Die Beurteilung der wertmäßigen Ähnlichkeit von Vergleichsobjekten erlaubt die Lösung der Gleichung nach (4.17) und damit die Besetzung der Kofaktormatrix des Signals $\mathbf{Q}_{ss}$ nach (2.20), die auf den Nebendiagonalen die Korrelationen zwischen zwei Wertermittlungsobjekten i und k enthält. Die Korrelation zweier Objekte kann dahingehend interpretiert werden, dass sich ähnelnde Kauffälle auch ein vergleichbares Verhalten aufweisen, während bei Kauffällen, die bezüglich der Ähnlichkeit weiter voneinander entfernt liegen, entsprechend von einem unabhängigen Verhalten ausgegangen werden kann. Die geeignete Definition der Nachbarschaft bzw. des Abstandes der Kauffälle muss anhand der Problemstellung erfolgen. Denkbar sind rein geografische Nachbarschaften, aber auch die Definition von problemorientierten Maßzahlen. Da sich die Ähnlichkeit zweier Objekte aus einer komplexen Zahl von Einflussgrößen mit teilweise unterschiedlichen Einheiten zusammensetzt, hat sich die Definition einer geeigneten Maßzahl etabliert: PELZER (1976) hat das Problem über die Einführung eines sogenannten *Wertabstands* gelöst. Nachfolgende Anwendungen in der Verkehrswertermittlung beruhen ausnahmslos auf dieser Vorgehensweise. Der Wertabstand wird dabei als Übereinstimmungsgrad vergleichbarer Immobilien bezogen auf ihre wesentlichen wertbeeinflussenden Merkmale interpretiert und ergibt sich als m -dimensionale Größe. Die Anzahl m der Dimensionen bezeichnet die Anzahl der Merkmale, die für eine vorliegende Stichprobe als wertrelevant angesehen werden. Ein Objekt i wird einem Objekt k umso ähnlicher sein, je besser die einzelnen wertbeeinflussenden Merkmale

übereinstimmen. Numerisch kann dies durch den Differenzvektor der entsprechenden Zeilen der Designmatrix $\mathbf{X}$ ausgedrückt werden:

$$\Delta \mathbf{x}'_{ik} = \mathbf{x}'_i - \mathbf{x}'_k \quad (4.18)$$

mit $\mathbf{x}_i = \begin{vmatrix} x_{i1} & x_{i2} & \dots & x_{im} \end{vmatrix}$, $\mathbf{x}_k = \begin{vmatrix} x_{k1} & x_{k2} & \dots & x_{km} \end{vmatrix}$.

Als Kennzahl ergibt sich der Wertabstand WA durch die Bildung des Skalarproduktes und dessen Wurzel:

$$\text{WA}_{ik} = \sqrt{\Delta \mathbf{x}'_{ik} \Delta \mathbf{x}_{ik}}. \quad (4.19)$$

Der so definierte Wertabstand kann bereits als Ähnlichkeitsmaß zur Beurteilung herangezogen werden, jedoch ergibt sich das Problem, dass die numerischen Ausprägungen der einzelnen wertbeeinflussenden Merkmale erheblich voneinander abweichen können. In der Folge würde der Wertabstand durch die Merkmale mit hohen Zahlenwerten dominiert, kleine Zahlenwerte würden keine nennenswerten Auswirkungen aufweisen. Der Wertabstand ist wahlweise durch einen der folgenden Ansätze zu modifizieren:

a) *Gewichteter Wertabstand*

Um die Unterschiede in den numerischen Ausprägungen zu beseitigen, werden zunächst alle Realisierungen der Einflussgrößen durch die Beziehung

$$\tilde{x}_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \quad \text{mit } i \in \{1, \dots, n\}, \quad j \in \{1, \dots, m\}$$

auf den Wertebereich zwischen 0 und 1 transformiert. Zusätzlich wird eine Gewichtsmatrix $\mathbf{D}$ der Dimension $m \times m$ eingeführt, die auf der Hauptdiagonalen mit individuellen Gewichten für jede Einflussgröße besetzt ist und damit die Bedeutung der wertbeeinflussenden Merkmale bezeichnet. Die Besetzung der Gewichtsmatrix kann dabei auf verschiedene Arten erfolgen: Gängige Vorgehensweisen sind die Besetzung nach gutachterlichem Sachverstand oder die Nutzung der partiellen Korrelationskoeffizienten ρ_{yx_j} zwischen der Zielgröße im Ansatz und der j -ten Einflussgröße (HÖPCKE, 1978, S. 3), (ZIEGENBEIN & HAWERK, 1978, S. 258), (ZADDACH & ALKHATIB, 2012a, S. 149). Der resultierende, gewichtete Wertabstand lässt sich dann formulieren als

$$\text{WA}_G = \sqrt{\Delta \tilde{\mathbf{x}}'_{ik} \mathbf{D} \Delta \tilde{\mathbf{x}}_{ik}}. \quad (4.20)$$

b) *Mahalanobis-Distanz*

Die Einführung einer Mahalanobis-Distanz ist im Vergleich zum gewichteten Wertabstand als strengere Lösung zu betrachten: Anstelle der Normierung werden durch die Nutzung der VKM der Einflussgrößen $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$ die Korrelationen zwischen den Einflussgrößen berücksichtigt:

$$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = \frac{1}{n-1} \mathbf{X}_z \mathbf{X}_z'. \quad (4.21)$$

Hier enthält $\mathbf{X}_z$ die spaltenweise zentrierten Mittelwerte der Einflussgrößen. Die Mahalanobis-Distanz berechnet sich unter Berücksichtigung einer Gewichtsmatrix $\mathbf{G} = \mathbf{G}^{1/2}$ dann nach HÖPCKE (1978, S. 3) zu

$$\text{WA}_M = \sqrt{\Delta \tilde{\mathbf{x}}'_{ik} \mathbf{G} \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}}^{-1} \mathbf{G} \Delta \tilde{\mathbf{x}}_{ik}}. \quad (4.22)$$

Da in beiden genannten Ansätzen die Wertabstände noch mit der Anzahl m der Merkmale wachsen und zudem vom Niveau des Gewichtsansatzes abhängen, schlägt HÖPCKE (1978, S. 3) die Reduktion der Gewichte auf die Summe 1 vor, sodass sich als modifizierte Gewichtsmatrizen der Ansätze nach (4.20) und (4.22)

$$\mathbf{D}^* = \frac{1}{sp(\mathbf{D})} \mathbf{D}, \quad sp(\mathbf{D}^*) = 1 \quad \text{bzw.} \quad \mathbf{G}^* = (\mathbf{D}^*)^{1/2} = \frac{1}{\sqrt{sp(\mathbf{D})}} \mathbf{G} \quad (4.23)$$

ergeben, die in den genannten Gleichungen entsprechend mit $\mathbf{D} = \mathbf{D}^*$ und $\mathbf{G} = \mathbf{G}^*$ zu substituieren sind. Je kleiner der berechnete Wertabstand WA, desto ähnlicher sind sich zwei Objekte in den wertbeeinflussenden Merkmalen. Die Entscheidung, welcher der beiden Ansätze zu wählen ist, kann nicht pauschal erfolgen und ist jeweils sachverständig zu treffen. Mögliche Unterschiede in den Resultaten sind dann auf den Ansatz zurückzuführen. Für die weiteren theoretischen Ausführungen wird aus Gründen der Lesbarkeit nicht zwischen den Ansätzen differenziert und als Bezeichnung der allgemeine Wertabstand $\delta_{ik} = \text{WA}_G = \text{WA}_M$ zwischen zwei Objekten i und k definiert. In den Untersuchungen des praktischen Teils in Kapitel 5 wird hingegen explizit auf die verwendete Berechnungsgrundlage verwiesen.

PELZER (1978) hat gezeigt, dass sich die gesuchte Korrelation ρ_{ik} in $\mathbf{Q}_{ss}$ nach (2.20) als Funktion des definierten Wertabstandes mit $\rho_{ik} = f(\delta_{ik})$ darstellen lässt. Offensichtlich gilt, dass $f(0) = 1$ (identische Wertmerkmale bewirken hohe Korrelation der Objekte) und $f(\infty) = 0$ (keine Übereinstimmung in den Wertmerkmalen führen zu keiner Korrelation zwischen den Objekten). Für die Bestimmung der Autokovarianzfunktion werden für jedes Objekt in der Stichprobe die Wertabstände δ_{ik} zu allen übrigen $n - 1$ Stichprobenelementen berechnet, woraus sich eine symmetrische Matrix der Dimension $[n \times n]$ ergibt. Für die nach (4.17) erforderliche Klassenbildung müssen die Wertabstände in einem weiteren Schritt zusammengefasst werden, was die Definition einer geeigneten Klassenbreite voraussetzt (ZADDACH & ALKHATIB, 2012a, S. 150). Die endgültige Festlegung der Klassenbreite Δ_d und damit die Anzahl d der Klassen muss empirisch auf Grundlage der Stichprobe und sachgerechter Einschätzung erfolgen. Gleichung (4.17) erlaubt dann unter den gegebenen Voraussetzungen die Bestimmung einer empirischen Autokovarianzfunktion, basierend auf den Vergleichsobjekten in der Stichprobe, sowie durch Nutzung der Autokorrelationsfunktion die Berechnung von deren Vergleichbarkeit. Anhand der Verläufe der empirischen Funktionen können analytische Autokovarianz- und Autokorrelationsfunktion geschätzt werden, die die Besetzung der Signalmatrix nach (2.20) und die Lösung des Ausgleichsalgorithmus der informativen sowie nichtinformativen Bayesschen Kollokation nach Abbildung 4.4 ermöglichen. Ein praktisches Beispiel der Berechnungen wird in Kapitel 5 gegeben und diskutiert.

Die bisherigen Ausführungen zum Auswerteprozess nach Abbildung 4.4 (S. 92) ermöglichen die Schätzung sämtlicher Bayesscher Modelle. Die Modellwahl der Auswertestrategie wird im folgenden Abschnitt behandelt.

4.3.2. Modellwahl mittels Hypothesentests

Für die Bayesschen linearen Modelle ist von besonderem Interesse, inwiefern die unabhängigen Größen die Zielgröße beeinflussen: Zu diesem Zweck ist ein geeignetes Modellwahlkriterium zu definieren, welches eine Beurteilung der Modellgüte und somit eine Entscheidung über relevante Einflussgrößen ermöglicht. Diese Fragestellung ist als (*Bayessche*) *Modellwahl* Thema zahlreicher Forschungsansätze und Veröffentlichungen. In den folgenden Abschnitten wird die Theorie der *Bayesschen Hypothesentests* über *Bayes-Faktoren* aufbereitet, wie sie in der Auswertestrategie angewendet wird und in den Grundzügen eng mit der Entwicklung Bayesscher Hypothesentests durch JEFFREYS (1935) sowie JEFFREYS (1961) verbunden ist. Die Nutzung von konjugierten linearen Modellen ermöglicht eine analytisch zugängliche Lösung bei der Bestimmung der Bayes-Faktoren. Liegen diese Voraussetzungen nicht vor, werden i. d. R. Verfahren genutzt, die eine numerische Approximation von Bayes-Faktoren erlauben. Als Beispiel für Gütemaße wird an dieser Stelle die Gruppe der Informationskriterien genannt, insbesondere das Bayesian Information Criterion (BIC) sowie das Deviance Information Criterion (DIC). Für einen Überblick über Modellwahlkriterien und deren Diskussion wird auf Grund der Vielzahl auf die weiterführende Literatur verwiesen (vgl. u. a. ANDO (2010), CHRISTENSEN (2011), LEE (2012), HELD & BOVÉ (2014)). In den folgenden Unterabschnitten werden die Grundlagen Bayesscher Hypothesentests und die Definition des Bayes-Faktors als Modellwahlkriterium thematisiert. Aufbauend erfolgt die praktische Übertragung für die Anwendung in der Auswertestrategie.

4.3.2.1. Posteriori-Wahrscheinlichkeit von Hypothesen

Parametrische statistische Inferenz beruht auf der Annahme eines Modells, welches vorliegende Daten oder Beobachtungen und deren stochastische Eigenschaften treffend beschreibt. Für einen konkret vorliegenden Datensatz von Beobachtungen stellt sich damit die Frage, welches Modell als treffend beurteilt wird und eine bestmögliche Erklärung liefert.

Allgemein lässt sich aus Bayesscher Sichtweise zunächst annehmen, dass beliebige Beobachtungen $\mathbf{y}$ unter einer der Hypothesen $H_1, \dots, H_r$ entstanden sind. Für die Beobachtungen können dann Wahrscheinlichkeiten der Form $p(\mathbf{y}|\boldsymbol{\theta}_k, H_k)$ formuliert werden, wobei $\boldsymbol{\theta}_k$ allgemein den Vektor der unbekannten Parameter mit einer Priori-Wahrscheinlichkeit $p(\boldsymbol{\theta}_k|H_k)$ unter der Hypothese H_k mit $k \in \{1, \dots, r\}$ repräsentiert. Unter Anwendung des Bayes-Theorems nach (3.17) ergibt sich als Posteriori-Wahrscheinlichkeit einer einzelnen Hypothesen dann

$$p(H_k|\mathbf{y}) = \frac{p(H_k) \int p(\mathbf{y}|\boldsymbol{\theta}_k, H_k) p(\boldsymbol{\theta}_k|H_k) d\boldsymbol{\theta}_k}{\sum_{i=1}^r p(H_i) \int p(\mathbf{y}|\boldsymbol{\theta}_i, H_i) p(\boldsymbol{\theta}_i|H_i) d\boldsymbol{\theta}_i} \quad (4.24)$$

mit $p(\mathbf{y}|\boldsymbol{\theta}_k, H_k)$ und $p(H_k)$ als Likelihood-Funktion und Priori-Wahrscheinlichkeit der Hypothese H_k . Die Priori-Wahrscheinlichkeiten der Hypothese $p(H_k)$ sowie der Parameter $p(\boldsymbol{\theta}_k|H_k)$ geben eine erste Information über die Modellunsicherheit, bevor die Daten $\mathbf{y}$ erhoben sind. Durch die Beziehung (4.24) werden die Priori-Information mit den Daten zusammengeführt, sodass mittels Bayes-Theorem die Posteriori-Wahrscheinlichkeit der Hypothese H_k , bedingt durch das zusätzliche Wissen durch $\mathbf{y}$, ausgedrückt werden kann. Grundprinzip des Bayesschen Hypothesentests ist es, das Modell mit der Hypothese zu wählen, die sich unter den gegebenen Posteriori-Wahrscheinlichkeiten am plausibelsten erweist, was gleichbedeutend mit der maximalen Posteriori-Dichte ist. In Bezug auf (4.24) bedeutet dies, dass diejenige Hypothese gewählt wird, die den Term

$$p(H_k) \int p(\mathbf{y}|\boldsymbol{\theta}_k, H_k) p(\boldsymbol{\theta}_k|H_k) d\boldsymbol{\theta}_k \quad (4.25)$$

maximiert. Dabei drückt das Integral

$$\int p(\mathbf{y}|\boldsymbol{\theta}_k, H_k) p(\boldsymbol{\theta}_k|H_k) d\boldsymbol{\theta}_k = p(\mathbf{y}|H_k) \quad (4.26)$$

die marginale Likelihood der Beobachtungen $\mathbf{y}$ aus und kann als Wahrscheinlichkeit der Beobachtungen bei gegebener Hypothese H_k interpretiert werden (HELD, 2008, S. 217), (ANDO, 2010, S. 101 f.). Wird (4.24) durch Wegfall des Nenners als Normierungskonstante vereinfacht, ergibt sich die Proportionalitätsbeziehung

$$p(H_k|\mathbf{y}) \propto p(H_k) p(\mathbf{y}|H_k), \quad (4.27)$$

aus der sich im Folgenden der sogenannte Bayes-Faktor entwickeln lässt, der für Hypothesentests und damit verbundene Modellvergleiche genutzt werden kann. Bayes-Faktoren können als klassische Herangehensweise zur Modellwahl in Bayesschen Ansätzen bezeichnet werden und beruhen auf einem paarweisen Vergleich von zwei beliebigen Modellen. Werden diese allgemein mit M^a und M^b bezeichnet, können die Hypothesen mit

$$H_0 : M^a, \quad H_A : M^b \quad (4.28)$$

bezeichnet werden, in der H_0 als Nullhypothese mit der Aussage interpretiert werden kann, dass das Modell M^a wahrscheinlicher ist, und H_A als Alternativhypothese mit der Aussage, dass das Modell M^b wahrscheinlicher ist. Es lassen sich damit die Annahmen treffen, dass die Beobachtungen entweder unter einer Hypothese H_0 mit der Wahrscheinlichkeit $p(\mathbf{y}|H_0)$ oder unter einer Hypothese H_A mit der Wahrscheinlichkeit $p(\mathbf{y}|H_A)$ entstanden sind. Die Beurteilung der Hypothesen erfolgt auf Grundlage von (4.27) und ist definiert als Verhältnis der Posteriori-Wahrscheinlichkeiten der einzelnen Modelle:

$$\underbrace{\frac{p(M^a|\mathbf{y})}{p(M^b|\mathbf{y})}}_{\text{Posteriori-Chance}} = \underbrace{\frac{p(M^a)}{p(M^b)}}_{\text{Priori-Chance}} \cdot \underbrace{\frac{p(\mathbf{y}|M^a)}{p(\mathbf{y}|M^b)}}_{\text{Bayes-Faktor}}. \quad (4.29)$$

Der letztgenannte Term in vorstehender Gleichung wird als Bayes-Faktor (BF) bezeichnet und stellt sich als Verhältnis der marginalen Likelihood der Beobachtungen $\mathbf{y}$ unter der Bedingung der formulierten Hypothesen dar. Der Bayes-Faktor gibt im Rahmen des Modellvergleichs an, wie wahrscheinlich das Modell M^a im Vergleich zu Modell

M^b ist und entscheidet damit als Hypothesentest zugunsten des Modells mit der größeren marginalen Likelihood (KASS & RAFTERY, 1995, S. 776). In praktischen Anwendungen ist es häufig üblich, dass die zu vergleichenden Modelle mit einer gleichen Priori-Chance in die Berechnung nach (4.29) eingehen, da entweder keine genaueren Kenntnisse vorliegen oder keines der Modelle von vornherein bevorzugt werden soll. Aus diesem Grund wird dann $p(M^a) = p(M^b) = 0.5$ gesetzt, sodass der Term als Konstante entfällt und sich die Posteriori-Chance direkt aus dem Bayes-Faktor ergibt.

Die Aussage des Bayes-Faktors als Hypothesentest kann lediglich als Wahrscheinlichkeit interpretiert werden und erlaubt keine strenge Aussage wie bspw. Hypothesentests in klassisch frequentistischen Ansätzen. Grundsätzlich wird in der Modellwahl eine Evidenz für die Gültigkeit des Modells M^a gegenüber dem Modell M^b gegeben. Die grundlegende Interpretation ergibt sich nach HELD & BOVÉ (2014, S. 232) durch die Beziehung

$$\left\{ \begin{array}{l} BF_{ab} > 1 \\ BF_{ab} < 1 \end{array} \right\} \text{ wenn } \mathbf{y} \text{ die Wahrscheinlichkeit von } M^a \left\{ \begin{array}{l} \text{erhöht} \\ \text{verringert} \end{array} \right\} \text{ hat,}$$

und in der die Abkürzung BF_{ab} den Bayes-Faktor aus dem Vergleich des Modells M^a mit dem Modell M^b bezeichnet. Die Interpretation von Bayes-Faktoren wurde von JEFFREYS (1961) eingeführt und als *Skala für die Evidenz* (*scale of evidence*) in Bezug auf eine Hypothese (hier die Wahrscheinlichkeit für das Modell M^a) bezeichnet. Tabelle 4.5 gibt eine Übersicht über die Skala.

Tabelle 4.5.: Skala für die Evidenz von Bayes-Faktoren nach JEFFREYS (1961)

Bayes-Faktor			Evidenz für M_a
BF_{ab}	< 1		negativ
1	$< BF_{ab} < 3$		kaum erwähnenswert
3	$< BF_{ab} < 10$		beachtlich
10	$< BF_{ab} < 30$		stark
30	$< BF_{ab} < 100$		sehr stark
100	$< BF_{ab}$		entscheidend

Bei der Interpretation ist zu beachten, dass es sich lediglich um Anhaltspunkte handelt: Mittels des Bayesschen Hypothesentests wird weder eine Hypothese angenommen noch verworfen, es werden lediglich Wahrscheinlichkeiten zugunsten von H_0 oder H_A angegeben. Zudem handelt es sich nicht um eine feststehende Skala, Modifikation haben sich im Laufe der Zeit durch weiterführende Forschungen, wie bspw. durch KASS & RAFTERY (1995), ergeben. In praktischen Anwendungen der Praxis hat sich die Einteilung nach JEFFREYS (1961) jedoch durchgesetzt, weshalb die Skala nach Tabelle 4.5 für die Modellwahl im Rahmen der Auswertestrategie nach Abbildung 4.4 zugrunde gelegt wird.

Die Definition und Interpretation des Bayes-Faktors ist bislang allgemein gehalten und wird im folgenden Abschnitt für die Modellwahl in der Auswertestrategie nach Abbildung 4.4 angewendet.

4.3.2.2. Bayes-Faktor in der Bayesschen Regression

Die in Abbildung 4.4 (S. 92) dargestellten Auswerteschritte 3.1 bis 3.3 beruhen auf dem Grundgedanken, dass ausgehend von einem beliebigen Gesamtmodell M^g Modellvarianten nach einem klassischen, sogenannten Branch-and-Bound Algorithmus gebildet werden, der seit den grundlegenden Forschungen von FURNIVAL & WILSON (1974) als Standardmethode für die Modellwahl basierend auf Regressionsanalysen bezeichnet werden kann. Branch-and-Bound bezeichnet dabei nicht nur eine einzelne Methode, vielmehr existieren zahlreiche Abwandlungen, die an dieser Stelle nicht vertieft behandelt werden (vgl. z. B. GATU & KONTOGHIOGHES (2006) oder HUO & NI (2007)). Für die Auswertestrategie nach Abbildung 4.4 wird eine Kombination aus Rückwärts- und Vorwärtsstrategie entwickelt, innerhalb derer die in dem vorherigen Unterabschnitt eingeführten Bayes-Faktoren Anwendung finden.

Der Algorithmus der *Rückwärtsstrategie* stützt sich in den Grundzügen auf die Ausführungen von BRUSCO & STAHL (2005, S. 187 ff.) und wird schematisch in Abbildung 4.5 dargestellt. Exemplarisch sind in dem Beispiel vier Einflussgrößen $x_1, \dots, x_4$ gegeben, die das Gesamtmodell M^g definieren. Ausgehend von M^g werden Submodelle M_i^s mit $i \in \{1, \dots, k\}$ gebildet, die sich jeweils durch die Elimination einer einzelnen Einflussgröße ergeben. Auf diese Weise werden insgesamt $k = 4$ weitere Modelle erhalten, die im Rahmen der Modellwahl über Bayes-Faktoren

dem Gesamtmodell gegenübergestellt werden. Durch die Selektion von Submodellen kann auf diese Weise geprüft werden, ob eine einzelne Einflussgröße aus dem Modellansatz entfernt werden kann. Im Umkehrschluss bedeutet dies, dass das entsprechende Merkmal mit der gegebenen Wahrscheinlichkeit keinen Erklärungsbeitrag zur Zielgröße leistet.

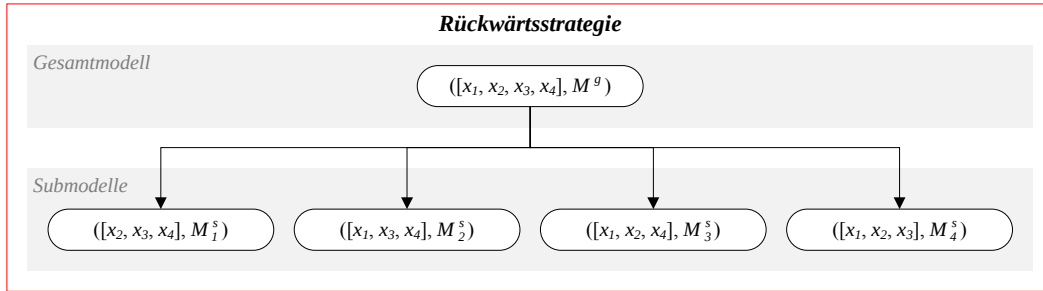


Abbildung 4.5: Grundprinzip der Rückwärtsstrategie für die Modellwahl im Branch-and-Bound Algorithmus

Nach dem gleichen Prinzip ergibt sich der Fall der *Vorwärtsstrategie*: Da sich im Laufe der rekursiven Anwendung der Auswertestrategie auch Einflussgrößen im Modellansatz ergeben können, die in den zurückliegenden Modellen bislang keine Berücksichtigung gefunden haben, müssen in einem zweiten Hypothesentest erweiterte Modelle gebildet werden; ein Beispiel wird in Abbildung 4.6 veranschaulicht. Ausgehend von einem Gesamtmodell M^g mit wiederum vier Einflussgrößen $x_1, \dots, x_4$ werden Hypermmodelle M_j^h mit $j \in \{1, \dots, u\}$ gebildet, wobei u die Anzahl der zusätzlichen, potenziell wertrelevanten Einflussgrößen bezeichnet. Im Beispiel sind dies die Einflussgrößen x_5 , x_6 und x_7 , die auf ihren Erklärungsbeitrag zum Modell hin untersucht werden.

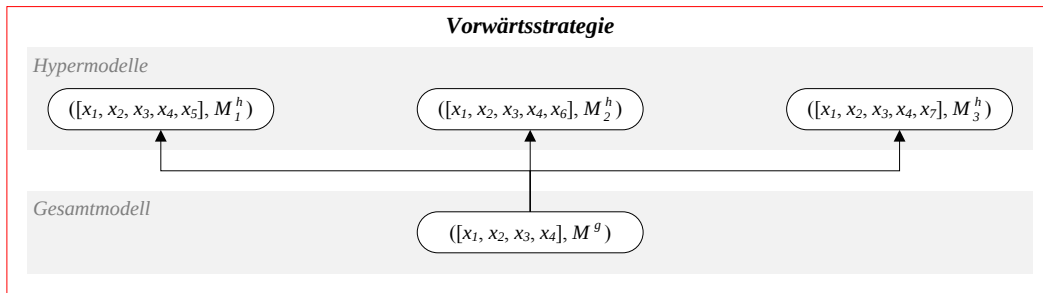


Abbildung 4.6: Grundprinzip der Vorwärtsstrategie für die Modellwahl im Branch-and-Bound Algorithmus

Grundsätzlich ist es bei der Anwendung von Bayes-Faktoren unerheblich, ob sich bei den Hypothesen um genestete oder um unterschiedliche Modelle handelt, welche durch abweichende Parameter und Wahrscheinlichkeitsdichten beschrieben werden. Ablauf der Bayesschen Hypothesentests bzw. die Berechnung des Bayes-Faktors nach (4.33) und Entscheidung über die Wahrscheinlichkeit eines der Modelle nach Tabelle 4.5 ergeben sich in der Rückwärtsstrategie und der Vorwärtsstrategie analog. Die folgenden Ausführungen basieren daher zunächst auf der Definition der Rückwärtsstrategie und dem Beispiel nach Abbildung 4.5.

Für jedes der Modelle $M^g, M_1^s, \dots, M_4^s$ wird zunächst eine eigene Schätzung basierend auf der Bayes-Strategie in Form einer informativen Bayesschen Regression durchgeführt. Bezieht sich die Auswertung auf einen Zeitpunkt t , ist zunächst für jedes Modell die Priori-Information zum Zeitpunkt $(t - 1)$ zu generieren. Die Generierung der Priori-Information vollzieht sich auf Grundlage einer nichtinformativen Lösung für die Bayessche Regression. Dabei ist zusätzlich eine iterative Ausreißersuche durchzuführen, wie sie in den vorherigen Ausführungen bereits beschrieben wurde. In einem weiteren Schritt wird die gewonnene Vorinformation für die Posteriori-Lösung zum Zeitpunkt t genutzt. Der Ablauf der Schätzungen ergibt sich dabei analog zu der Schätzung der Priori-Modelle mit dem Unterschied, dass nunmehr durch die Vorinformation eine informative Bayessche Regression zu schätzen ist.

Sind die Schätzungen erfolgt, können in Schritt 3.3 nach Abbildung 4.4 die Hypothesentests für die Modellwahl durchgeführt werden, in denen die jeweiligen Schätzergebnisse für $M^g, M_1^s, \dots, M_4^s$ gegenübergestellt werden. Wird die Modellwahl anhand des im vorherigen Unterabschnitt erläuterten Hypothesentests durchgeführt, ermöglicht

die Anwendung von Bayes-Faktoren eine Aussage darüber, welches der generierten Modelle $M^g, M_1^s, \dots, M_4^s$ sich für die vorliegenden Beobachtungen $\mathbf{y}$ als wahrscheinlichstes erweist. Im Mittelpunkt der Tests stehen damit die Parametervektoren $\bar{\beta}^g, \bar{\beta}_1^s, \dots, \bar{\beta}_4^s$ der einzelnen Modelle. Der Bayessche Hypothesentest für Schritt 3.3 der Auswertestrategie lässt sich im Hinblick auf den Branch-and-Bound Algorithmus formulieren, indem für jedes Submodell ein Bayes-Faktor mit Vergleich zum Gesamtmodell bestimmt wird. Dasjenige Modell, welches die nach Tabelle 4.5 (S. 100) größte Evidenz aufzeigt, wird als das wahrscheinlichste Modell gewählt. Handelt es sich dabei um das Gesamtmodell, ist die Auswertung des aktuellen Datensatzes abgeschlossen. Wird eines der Submodelle als das wahrscheinlichste Erklärungsmodell gewählt, wird ein neues Gesamtmodell gesetzt und eine iterative Berechnung über Hypothesentests durchgeführt, bis das wahrscheinlichste Modell ermittelt ist. Dabei ist zu beachten, dass sich die Interpretation der Evidenz von Modellen nach Tabelle 4.5 zwar in der praktischen Anwendung bewährt hat; jedoch kann sie keine abschließende Wertung darstellen. Ausgedrückt werden lediglich Wahrscheinlichkeiten, zudem sind die aufgeführten Begriffe mit linguistischen Unschärfen belegt. Auf Grund der gegebenen Heterogenität der Daten in der Verkehrswertermittlung und den daraus resultierenden Unsicherheiten ist eine strenge Anwendung der Tabelle in der Vergleichswertermittlung zu relativieren. Für die Modellwahl in der Auswertestrategie wird daher als Grenzwert für die Entscheidung zugunsten eines Modells M_k^s mindestens eine starke Evidenz ($BF_{M_k^s M^g} > 10$) gefordert, um den Gegebenheiten und den zugrunde liegenden Daten der Verkehrswertermittlung gerecht zu werden. Abbildung 4.7 gibt in Erweiterung zu 4.5 ein Beispiel für den Ablauf der iterativen Modellwahl. In jeder Iteration wird das Modell gewählt, welches den Bayes-Faktor maximiert und zusätzlich die Bedingung $BF_{M_k^s M^g} > 10$ erfüllt. Der Algorithmus wird abgebrochen, sobald das Modell auf eine wesentliche Einflussgröße reduziert wurde oder der Hypothesentest keines der Submodelle präferiert.

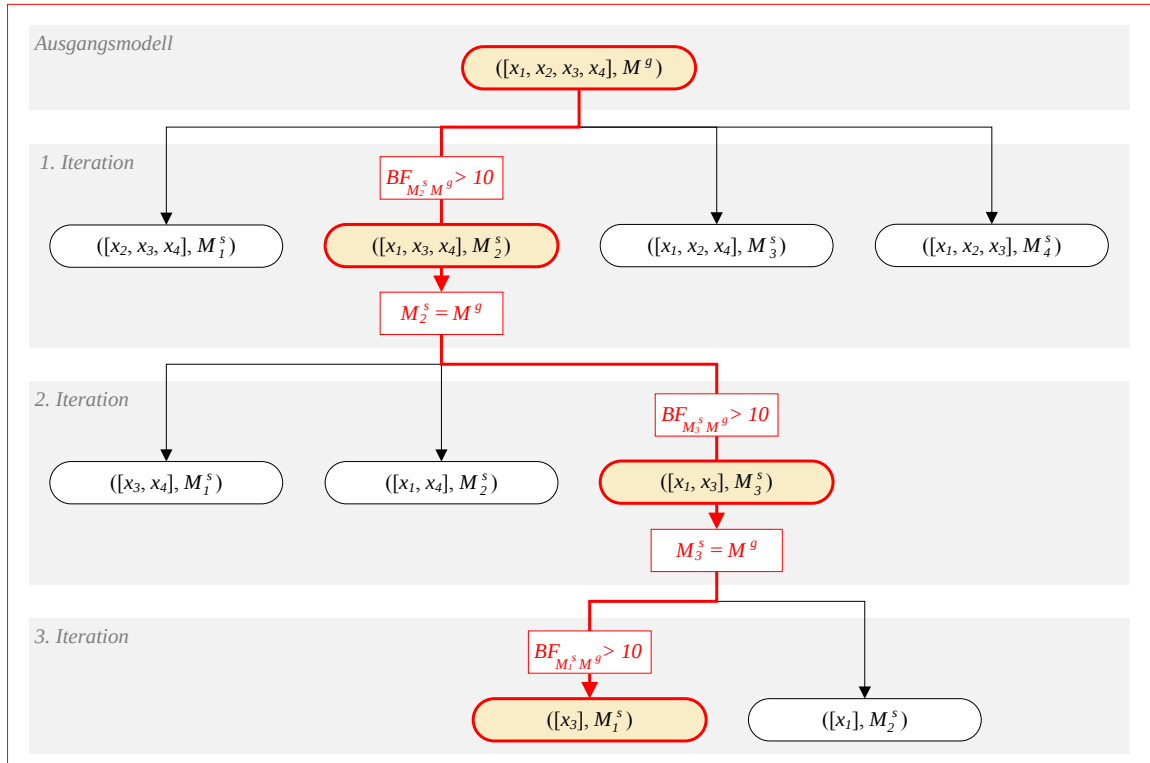


Abbildung 4.7: Beispiel des iterativen Branch-and-Bound Algorithmus für die Modellwahl

Bei Anwendungen im Zuge der Auswertestrategie ist zu beachten, dass der Bayessche Hypothesentest ausschließlich für die informative Bayessche Regression formuliert wird. ANDO (2010, S. 104) verweist auf die Tatsache, dass die marginale Likelihood für nichtinformative Ansätze nicht definiert ist und ausschließlich durch numerische Methoden gelöst werden kann. Als praktische Alternativen werden Pseudo-Lösungen vorgeschlagen, zu denen u. a. der Posteriori Bayes-Faktor nach AITKIN (1991), der fraktionale Bayes-Faktor nach O'HAGAN (1995) oder der intrinsische Bayes-Faktor nach BERGER & PERICCHI (1996) zählen. Für eine vertiefte Behandlung wird auf die genannte Literatur verwiesen.

Im vorliegenden Fall handelt es sich um eine Lösung basierend auf konjugierten Dichten, wie in Kapitel 3 ausführlich beschrieben. Aus diesem Grund lässt sich für die Bestimmung des Bayes-Faktors eine analytische Lösung herbeiführen (HELD & BOVÉ, 2014, S. 233). Werden die Hypothesen mit den korrespondierenden Modellen gleichgesetzt, ergeben sich die Hypothesen zu

$$M_k^s (= H_0) : \beta_k = 0 \quad M^g (= H_A) : \beta_k \neq 0 \quad \text{mit } k \in \{1, \dots, u\}, \quad (4.30)$$

wobei stets die Evidenz eines Submodells gegen das Gesamtmodell getestet wird, indem für das Submodell unterstellt wird, dass sich die k -te Einflussgröße zu Null ergibt. Auf diese Weise ist für die Bayessche Regression die Posteriori-Wahrscheinlichkeit der Hypothesen analog zu (4.27) und der Bayes-Faktor für den Hypothesentest mit

$$\frac{p(M_k^s | \mathbf{y})}{p(M^g | \mathbf{y})} = \frac{p(M_k^s)}{p(M^g)} \cdot \frac{\int p(\mathbf{y} | \beta_k^s, M_k^s) p(\beta_k^s | M_k^s) d\beta_k^s}{\int p(\mathbf{y} | \beta^g, M^g) p(\beta^g | M^g) d\beta^g}, \quad (4.31)$$

anzugeben, wobei β_k^s den Parametervektor des Submodells k und β^g den Parametervektor des Gesamtmodells bezeichnen. Wird zudem die Priori-Chance für beide zu testenden Modelle gleich gesetzt, ergibt sich der Bayes-Faktor als Verhältnis der marginalen Likelihood von M_k^s und M^g . Unter der Voraussetzung konjugierter linearer Modelle und den Verteilungsannahmen nach Abschnitt 3.2.1 (S. 58 ff.) lässt sich für ein beliebiges Modell M die marginale Likelihood in (4.31) analytisch durch

$$p(\mathbf{y} | M) = c \cdot \left(\frac{|\mathbf{Q}_{\bar{\beta}\bar{\beta}}|}{|\mathbf{Q}_{\beta\beta}|} \right)^{1/2} \cdot (\bar{p} \bar{b})^{-\bar{p}/2} \quad \text{mit} \quad c = \frac{\Gamma(\frac{\bar{p}}{2}) (\underline{p} \underline{b})^{\underline{p}/2}}{\Gamma(\frac{p}{2}) \pi^{n/2}} \quad (4.32)$$

angeben. Werden aus Gründen der Lesbarkeit für die Modelle M_k^s und M^g die Indizes a und b gewählt, ergibt sich der Bayes-Faktor damit zu

$$BF_{M_k^s M^g} = BF_{ab} = \frac{c_a \cdot \left(\frac{|\mathbf{Q}_{\bar{\beta}\bar{\beta},a}|}{|\mathbf{Q}_{\beta\beta,a}|} \right)^{1/2} \cdot (\bar{p}_a \bar{b}_a)^{-\bar{p}_a/2}}{c_b \cdot \left(\frac{|\mathbf{Q}_{\bar{\beta}\bar{\beta},b}|}{|\mathbf{Q}_{\beta\beta,b}|} \right)^{1/2} \cdot (\bar{p}_b \bar{b}_b)^{-\bar{p}_b/2}} \quad (4.33)$$

und kann für die Beurteilung der Modellwahrscheinlichkeit mit Tabelle 4.5 verglichen und nach dem Schema in Abbildung 4.7 angewendet werden. Analog zu den bisherigen Ausführungen erfolgen die Berechnungen der Bayes-Faktoren für den Vergleich des Gesamtmodells mit potenziellen Hypermodellen.

Durch die vorstehenden Ausführungen werden die Grundlagen für die Wahl des wahrscheinlichsten Modells gelegt. Ist im ggf. iterativen Auswerteprozess das ideale Modell gefunden, lassen sich die Ergebnisse der Bayesschen Auswertung für die Generierung von Priori-Information für zeitlich nachgeordnete Auswertepochen nutzen. Die auf diese Weise durchzuführende rekursive Bayessche Schätzung wird im folgenden Abschnitt behandelt.

4.3.3. Rekursive Bayessche Schätzung

Mittels der im vorstehenden Abschnitt erläuterten Hypothesentests ist die (ggf. iterativ durchzuführende) Auswertestrategie für die Ermittlung eines zutreffenden Gesamtmodells gegeben: In Schritt 3.4 der Abbildung 4.4 werden die vollständig spezifizierten Bayesschen Regressions- und Kollokationsmodelle erhalten. Auf Basis der Modelle können die funktionale Approximationsgüte sowie der Unsicherheitshaushalt der Modelle beurteilt werden.

Nach den in Kapitel 1 formulierten Forschungszielen ist ein weiterer wesentlicher Aspekt die Fortpflanzung der erhaltenen Unsicherheiten auf zeitlich nachfolgende Auswerteperioden. In Abbildung 4.4 (S. 92) wird dieser Schritt unter 3.5 als Rekursion der Folgeperioden bezeichnet und beinhaltet insbesondere die Nutzung der Posteriori-Lösung aus den Schritten 3.3 und 3.4 als Priori-Information für die weitere Auswertung. Auf diese Weise wird der wesentliche Vorteil der Bayes-Strategie deutlich: Die Ergebnisse vergangener Auswertungen können als Vorwissen in die Analyse aktueller Datensätze integriert werden. Für Anwendungen in der Vergleichswertermittlung bedeutet dies konkret, dass keine abgeschlossenen Analysen für betrachtete räumliche und sachliche Teilmärkte durchgeführt werden,

sondern einer dynamischen Entwicklung der Märkte Rechnung getragen wird. Die Wahl der Auswerteziträume spielt dabei keine Rolle, sie lassen sich individuell festlegen. Durch den Bayesschen Ansatz können unterschiedliche Auswertepochen integriert werden und stellen damit implizit eine Zeitreihe in der Entwicklung der Modelle zur Verfügung, in der sich die Dynamik der Märkte widerspiegelt.

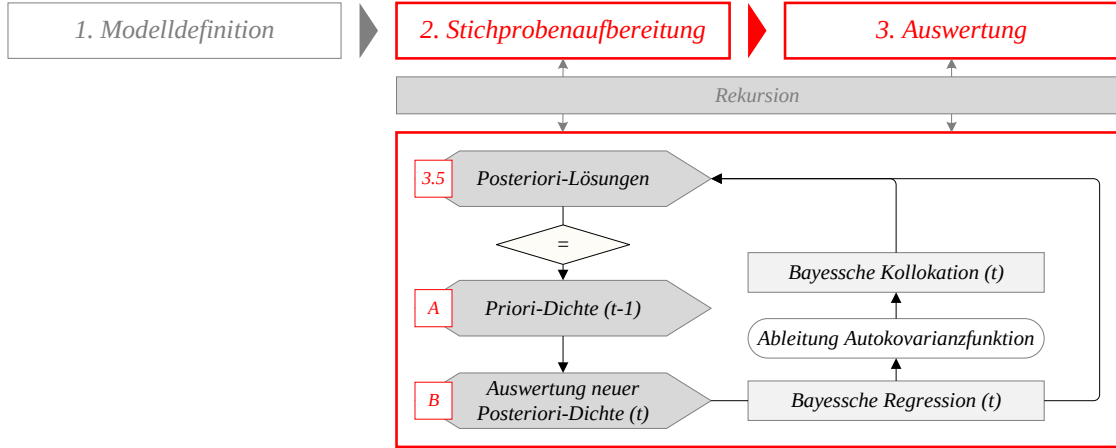


Abbildung 4.8: Schema der rekursiven Bayesschen Schätzung

Das Bayes-Theorem ermöglicht bei voneinander unabhängigen Informationen die rekursive Anwendung, wodurch sukzessive der Zustand des Wissens über ein beobachtetes Phänomen ergänzt wird. Wird ein beliebiger Beobachtungsvektor $\mathbf{y}$ partitioniert zu $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_h]'$, ergibt sich für das Bayes-Theorem anstelle von (3.18) die Beziehung

$$p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_h) \propto p(\boldsymbol{\theta}) p(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_h | \boldsymbol{\theta}). \quad (4.34)$$

Für den Fall voneinander unabhängiger Beobachtungsvektoren $\mathbf{y}_i$ und $\mathbf{y}_j$ mit $i \neq j$ und $i, j \in \{1, \dots, h\}$ gilt dann

$$p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_h) \propto p(\boldsymbol{\theta}) p(\mathbf{y}_1 | \boldsymbol{\theta}) p(\mathbf{y}_2 | \boldsymbol{\theta}) \dots p(\mathbf{y}_h | \boldsymbol{\theta}). \quad (4.35)$$

Für voneinander unabhängige Beobachtungen lässt sich das Bayes-Theorem daher rekursiv anwenden: Wird zunächst ausschließlich der Beobachtungsvektor $\mathbf{y}_1$ betrachtet, ergibt sich $p(\boldsymbol{\theta} | \mathbf{y}_1) \propto p(\boldsymbol{\theta}) p(\mathbf{y}_1 | \boldsymbol{\theta})$. Wird die erhaltene Posteriori-Dichte als Priori-Dichte für die Auswertung von $\mathbf{y}$ genutzt, folgt daraus $p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2) \propto p(\boldsymbol{\theta} | \mathbf{y}_1) p(\mathbf{y}_2 | \boldsymbol{\theta})$ und entsprechend bei weiterem Hinzufügen neuer Beobachtungen bis zu einem Zustand $\mathbf{y}_k$

$$p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k) \propto p(\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{k-1}) p(\mathbf{y}_k | \boldsymbol{\theta}) \quad (4.36)$$

für $k \in \{2, \dots, h\}$ (KOCH, 2000, S. 36 f.). Für die praktische Anwendung der Auswertestrategie im Vergleichswertverfahren bedeutet dies, dass durch zusätzliche Beobachtungen – hier in einem zeitlich nachfolgenden Auswertzeitraum – das Wissen über die unbekannten Parameter nach und nach vervollständigt wird: Die Posteriori-Lösungen, die nach Abbildung 4.4 mit den Schritten 3.3 und 3.4 erhalten werden, können als Priori-Information in die folgende Auswertung eingeführt werden. Konkret lässt sich das Vorgehen in der Auswertestrategie durch Abbildung 4.8 schematisch darstellen und als weitere Aufgliederung des Schrittes 3.5 in Abbildung 4.4 (S. 92) auffassen.

Als Ergebnis der Auswertungen in den Schritten 3.3 und 3.4 liegen die Posteriori-Lösungen für die Schätzung der unbekannten Parameter in den Modellen der Bayesschen Regression und der Bayesschen Kollokation zu einem aktuellen Auswertzeitpunkt t vor. Wie bereits erläutert und der Vergleich von (3.37) mit (3.52) zeigt, ergibt im konjugierten Modell die Dichte der Priori-Randverteilung der Parameter in Verbindung mit der Likelihood-Funktion (3.32) eine Posteriori-Dichte derselben Familie wie die Priori-Dichte. Im vorliegenden Fall bedeutet dies, dass sich sowohl Priori- als auch Posteriori-Dichte als multivariate t-Verteilungen ergeben. Die Posteriori-Randverteilung

der Parameter lässt sich so als Priori-Randverteilung nutzen, indem in Teilschritt A die Posteriori-Lösung zum Zeitpunkt $(t - 1)$ als neue Priori-Information für einen neuen Auswertezeitraum (t) gesetzt wird:

$$\underbrace{\beta_{(t-1)} | \mathbf{y}_{(t-1)} \sim t\left(\bar{\beta}_{(t-1)}, \Sigma_{\bar{\beta}_{(t-1)}}, \bar{\nu}_{(t-1)}\right)}_{\text{Alte Posteriori-Lösung}} = \underbrace{\beta_{(t)} | \mathbf{y}_{(t)} \sim t\left(\underline{\beta}_{(t)}, \Sigma_{\beta_{(t)}}, \underline{\nu}_{(t)}\right)}_{\text{Neue Priori-Information}} \quad (4.37)$$

Die Auswertung der Posteriori-Dichte in Teilschritt B erfolgt dann analog zur Beschreibung in Abschnitt 4.3.1 (S. 93) mit der iterativen Ausreißersuche sowie der Bestimmung der Autokovarianzfunktion. Als Ergebnis liegt wiederum eine Posteriori-Lösung vor, deren Ergebnisse durch die Beziehung (4.37) rekursiv fortgepflanzt werden können.

Zusammenfassend wird mit den Ausführungen in den Abschnitten 4.1, 4.2 sowie 4.3 eine Auswertestrategie gegeben, die basierend auf der Bayes-Strategie die Analyse von Kauffalldaten im Vergleichswertverfahren ermöglicht. Die theoretischen Grundlagen sollen im Folgenden auf ein praktisches Beispiel übertragen werden, in dem reale Kauffalldaten anhand des gegebenen Schemas ausgewertet werden. Zuvor werden abschließend für die theoretischen Ausführungen die methodischen Ansätze für die Validierung und Evaluierung der Modelle aufgezeigt.

4.4. Validierung und Wertung der Modelle und Ergebnisse

Für die Übertragung der Bayes-Theorie auf das Vergleichswertverfahren ist es nicht nur von Interesse, die reine Modellbildung auf Basis der Stichproben vorzunehmen, sondern darüber hinaus eine Aussage über die praktische Anwendbarkeit geben zu können. Unter diesem Gesichtspunkt sind zwei Aspekte zu differenzieren:

1. Validierung der Bayesschen Modelle durch die Prädiktion von Vergleichswerten,
2. Wertung der Bayesschen Modelle in Bezug auf Approximationsgüte und Unsicherheitshaushalt.

In praktischen Anwendungen der Verkehrswertermittlung werden die mathematisch-statistischen Modelle genutzt, um Vergleichswerte zu präzisieren, die der definierten Grundgesamtheit angehören, für die jedoch keine Kaufpreise, sondern lediglich die wesentlichen wertbeeinflussenden Merkmale vorliegen. Erst durch die Prädiktion wird das formulierte Vergleichswertverfahren für die praktische Anwendung sinnvoll nutzbar. Um die Effekte der Prädiktion beurteilen zu können, sollen die Modelle der Bayesschen Regression und Bayesschen Kollokation in einem Validierungsansatz genutzt werden: Das hier genutzte Prinzip der Kreuzvalidierung wird in Abschnitt 4.4.1 kurz beschrieben. Für eine abschließende Gegenüberstellung und Wertung der unterschiedlichen Modelle anhand von Vergleichskriterien wird die Methodik des Paarvergleichs angewendet. Die theoretischen Grundlagen sowie eine Übertragung auf die Bayesschen Modelle der Vergleichswertermittlung erfolgt in Abschnitt 4.4.2.

4.4.1. Methodik der Kreuzvalidierung

Das Prinzip der Kreuzvalidierung hat sich seit den Studien von STONE (1974) und GEISSER (1975) als ein Verfahren etabliert, um statistische Modelle bezüglich Schätz- und Prädiktionsergebnissen zu prüfen. Das Prinzip beruht darauf im Vorfeld der Modellbildung einen Teil der Stichprobendaten zu separieren und mit den übrigen Daten die Schätzung der Modellparameter vorzunehmen. Im Anschluss können die separierten Daten genutzt werden, um anhand der Modellparameter Beobachtungen zu schätzen. Durch die Differenz zwischen den auf diese Weise prädierten Werten sowie den Originaldaten können Aussage über die Prädiktionsgüte und letztlich eine adäquate Modellbildung getroffen werden (CRESSIE, 1993, S. 101). Ein Bezug zu praktischen Anwendungen in der Vergleichswertermittlung wird schnell deutlich: Das indirekte Vergleichswertverfahren wird i. d. R. dann angewendet, wenn ein konkretes Objekt zu bewerten ist. Die in den vorherigen Abschnitten definierten Modelle in den jeweiligen informativen und nichtinformativen Ansätzen werden herangezogen, um anhand der in den Gleichungen enthaltenen Einflussgrößen eine Schätzung des Vergleichswertes für das Objekt von Interesse vorzunehmen.

Dem Prinzip der Kreuzvalidierung folgend, können fiktive Wertermittlungsfälle simuliert werden, indem pro Stichprobe jeweils ein Datensatz aus der Gesamtstichprobe (d. h. des Likelihood-Datensatzes) pro Auswerteepoche entfernt wird. Die Schätzung der Modellparameter erfolgt anhand der verbleibenden $(n-1)$ Datensätze. Im Nachgang wird der Vergleichswert $\bar{y}_{p,k}$ mit $k \in \{1, \dots, n\}$ des zuvor entfernten Datensatzes prädiert und der zugehörigen

Kaufpreisbeobachtung $y_{p,k}$ gegenübergestellt. Auf diese Weise kann eine Aussage über die Prädiktionsgüte, insbesondere im Hinblick auf die Unsicherheit, getroffen werden. Der gesamte Prozess ist als iterativer Ablauf zu sehen. Sukzessive wird jeder Kauffall aus der Stichprobe entfernt und im Anschluss anhand der aktuellen Modellschätzungen prädiziert. Der Kauffall wird dann dem Gesamtdatensatz wieder hinzugefügt und der nächste Kauffall entfernt, sodass sich im Endeffekt durch die iterative Anwendung der Kreuzvalidierung ein Prädiktionswert für jeden Datensatz in der Stichprobe ergibt.

Für die eigentliche Modellschätzung der Parameter auf Basis der reduzierten Stichproben wird jeweils die Auswertestrategie nach den Ausführungen in den Abschnitten 4.2.2 bis 4.3 durchlaufen. Abweichungen von der Vorgehensweise der Auswertung über die gesamte Stichprobe ergeben sich damit nicht. Der iterative Ablauf wird zudem für jede Auswerteepoche wiederholt, wobei zu beachten ist, dass die Priori-Information stets aus der gesamten Stichprobe der Vorepoche übernommen wird, d. h., diese wird nicht reduziert.

Sowohl die Prädiktionsergebnisse als auch die Ergebnisse der Schätzung der Modelle basierend auf der gesamten Stichprobe können im Anschluss genutzt werden, um die unterschiedlichen Ansätze zu evaluieren. Zu diesem Zweck wird die Methodik des Paarvergleichs genutzt, die im folgenden Abschnitt beschrieben wird.

4.4.2. Methodik des Paarvergleichs

Um die unterschiedlichen Bayesschen Modelle gegenüberstellen zu können, ist eine geeignete Methodik zu wählen. Eine umfassende Übersicht über verschiedene Ansätze für die generelle Beurteilung von Varianten kann u. a. SCHÖNWANDT & MÜLLER-HERBERS (2007) entnommen werden. Für die Wertung der Bayesschen Modelle wird aus der Reihe der Methoden der Paarvergleich gewählt.

4.4.2.1. Theoretische Grundlagen

Die Methodik des Paarvergleichs lässt sich in den Kontext von Entscheidungsproblemen einordnen, deren Zielsetzung in der Gegenüberstellung unterschiedlicher Lösungsvarianten einer Fragestellung und einer objektiv begründbaren Empfehlung für die am besten geeignete Lösung besteht. Ein komplexer Entscheidungsprozess lässt sich nach SCHÖNWANDT & MÜLLER-HERBERS (2007, S. 9) durch zwei Kernaspekte charakterisieren, da jede Entscheidungsfindung unabhängig vom Grad der Komplexität eine gewisse Strukturierung aufweist. Zum einen zielt der Prozess auf die Auswahl einer Lösungsvariante aus einer Menge von Lösungsvarianten ab. Zum anderen lässt sich im Hinblick auf unterschiedliche Beurteilungsmethoden stets ein Grundmuster in den Teilschritten erkennen:

- i. Kriterienbestimmung zur Beschreibung der Lösungsvarianten,
- ii. Beschreibung und Darstellung der Lösungsvarianten (Ermittlung der Kriterienausprägungen),
- iii. vergleichende Betrachtung und Beurteilung der Lösungsvarianten sowie
- iv. Auswahl einer Lösungsvariante.

Im Kontext der systematischen Entscheidungsfindung existiert eine Vielzahl an Methoden, die sich an dem genannten Grundmuster orientieren und die an dieser Stelle nicht umfassend thematisiert werden können. Für eine Vertiefung der Grundlagen wird auf die weiterführende Literatur verwiesen (vgl. u. a. PATTON & SAWICKI (1993), SCHÖNWANDT & MÜLLER-HERBERS (2007), GRÜNIG & KÜHN (2013), OBERMAIER & SALIGER (2013), LAUX ET AL. (2014)). Für die Evaluierung der Bayesschen Modelle wird eine möglichst einfach zugängliche Methodik gewählt, anhand derer sich die definierten Forschungsziele beurteilen lassen. Die Methodik erhebt dabei nicht den Anspruch, eine komplexe Bewertung einer Vielzahl von Kriterien zu ermöglichen, sondern dient vielmehr in einem übersichtlichen Schema als Entscheidungshilfe und Argumentationsstütze für die Evaluierung. Aus der Vielzahl von Methoden wird für die vorliegende Arbeit der Paarvergleich gewählt.

Grundgedanke der Methodik ist, dass unterschiedliche Lösungsvarianten miteinander verglichen und anhand von einheitlichen Vergleichskriterien gewertet werden. Die Anzahl der zu vergleichenden Lösungsvarianten ist dabei unerheblich. Für eine strukturierte Darstellung wird eine sogenannte Paarvergleichsmatrix genutzt, die exemplarisch für eine Anzahl von vier Lösungsvarianten in Abbildung 4.9 dargestellt ist.

Varianten	Variante 1	Variante 2	Variante 3	Variante 4
Variante 1		1	0	0
Variante 2	1		1	0
Variante 3	2	1		2
Variante 4	2	2	0	
Summe	5	4	1	2
Rang	1.	2.	4.	3.

Legende:
0 - schlechter, 1 - neutral, 2 - besser

Abbildung 4.9: Beispiel einer Paarvergleichsmatrix

Die einzelnen Varianten werden zunächst senkrecht und waagrecht in die Zeilen und Spalten der Matrix eingetragen. Die unterschiedlichen Varianten können dann entweder in einem spaltenweisen oder einem zeilenweisen Vergleich gegenübergestellt werden. In der hier genutzten spaltenweisen Anwendung werden die Varianten der Kopfzeile nacheinander mit den Varianten der Kopfspalte verglichen. Anhand der gewählten Vergleichskriterien wird entschieden, welche der Varianten als besser eingestuft wird und mit einem entsprechenden Zahlenwert versehen. Die Summe der Zahlenwerte erlaubt dann die Festlegung einer Rangfolge in der Präferenz der Varianten. In praktischen Anwendungen hat sich eine Skalierung mit den Werten 0, 1 und 2 bewährt; eine feste Definition ist jedoch nicht vorgegeben und kann je nach Aufgabenstellung variieren. Für den Paarvergleich der Bayesschen Modelle wird die Skalierung der drei genannten Stufen gewählt. Erweist sich die Variante in der Kopfzeile als besser, wird in das entsprechende Feld der Matrix der Wert 2 eingetragen, erweist sie sich als schlechter, der Wert 0. Für den Fall, dass beide Varianten als gleichwertig beurteilt werden, ist der Wert 1 einzutragen. Aufgrund des Prinzips des Paarvergleichs ergeben sich für die untere und obere Dreiecksmatrix die jeweils entgegengesetzten Wertungen. Sind die Varianten auf diese Weise beurteilt, werden die spaltenweisen Summen gebildet und entsprechend der so erhaltenen Häufigkeiten mit Rangfolgen belegt. Die Variante mit der höchsten Häufigkeit kann dann als geeignetste Lösung ausgewählt werden; im Beispiel nach Abbildung 4.9 entsprechend Variante 1.

Mittels der Paarvergleichsmatrix kann über eine Rangfolge der Varianten entschieden werden. Als Nachteil kann sich ergeben, dass die Rangfolgen lediglich auf geringen numerischen Differenzen beruhen, wenn die Lösungsvarianten größtenteils ähnliche Ergebnisse zeigen. Ergebnis ist dann eine lediglich schwache Rangordnung, in der sich u. U. mehrere Varianten mit einer gleichen Rangstufe ergeben. Die Auswahl der besten Lösungsvariante ist in diesem Fall zu relativieren und entsprechend zu werten. Wesentlicher Vorteil der Methode ist hingegen, dass eine objektive und systematische Begründung für die Auswahl einer Lösungsvariante ermöglicht wird. Die Übertragung der theoretischen Grundlagen auf die Bayesschen Modelle wird im folgenden Unterabschnitt beschrieben.

4.4.2.2. Anwendung für die Bayesschen Modelle des Vergleichswertverfahrens

Für die Bayesschen Modelle ermöglicht die Nutzung von Paarvergleichsmatrizen eine Beurteilung der Güte der funktionalen Modellierung und des Unsicherheitshaushaltes. Ziel ist es, eine einfach zu handhabende Entscheidungshilfe zu formulieren, anhand derer die vier Modelle gegenübergestellt werden können. Dabei werden die einzelnen Modelle der informativen und nichtinformativen Bayesschen Regression sowie der informativen und nichtinformativen Bayesschen Kollokation als Varianten 1 bis 4 im Sinne der Abbildung 4.9 gegenübergestellt und anhand von Vergleichskriterien bewertet. Das Vorgehen orientiert sich an den Schritten i. bis iv. der Auflistung im vorherigen Unterabschnitt. Wird Schritt i. betrachtet, sind *Vergleichskriterien* zu definieren, wobei diese getrennt für die funktionale Modellierung und die Unsicherheitsmodellierung formuliert werden. Eine Übersicht über die in der Arbeit verwendeten Kriterien gibt Abbildung 4.10.

Für eine Beurteilung der Forschungsziele werden jeweils Paarvergleiche der vier genannten Modelle vorgenommen. Sowohl für die funktionale Modellierung als auch die Unsicherheitsmodellierung ergeben sich identische Vergleichsebenen: Im Einzelnen sind dies die Modellparameter (Trendkoeffizienten $\bar{\beta}$, Varianzfaktor $\bar{\sigma}^2$ und Signalkomponente $\bar{s}$ für die Bayesschen Kollokationsmodelle), die Posteriori-Schätzwerte $\bar{y}$ sowie die prädizierten Posteriori-Schätzwerte $\bar{y}_p$ und $\bar{s}_p$ für die Bayesschen Kollokationsmodelle.

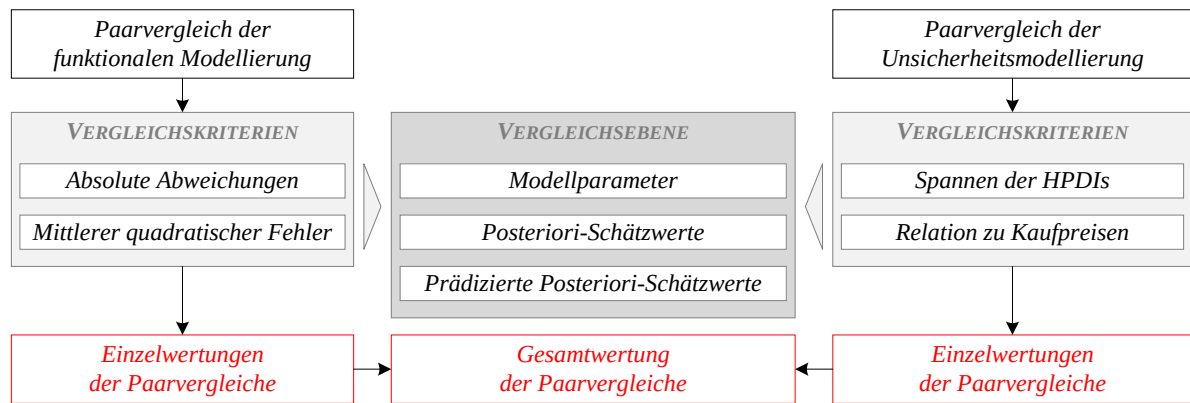


Abbildung 4.10: Vergleichskriterien für die Wertung der Bayesschen Modelle

Für die Beurteilung der funktionalen Modellierung werden die Vergleichskriterien der absoluten Abweichungen zwischen Posteriori-Schätzwerten und Beobachtungen sowie der mittlere quadratische Fehler über die Auswertepochen definiert. Durch die Ermittlung der numerischen Ergebnisse pro Modellvariante sowie eine vergleichende Betrachtung und Darstellung der Modelle untereinander werden die Schritte ii. und iii. der Aufzählung im vorherigen Unterabschnitt aufgegriffen. Der sich aus den Vergleichskriterien ergebende Gesamteindruck wird sodann in den Paarvergleichsmatrizen zusammengefasst. Zu beachten ist, dass die Modellparameter lediglich in Form der numerischen Ausprägungen verglichen werden können. Die Ableitung einer Paarvergleichsmatrix ist an dieser Stelle nicht möglich, da sich die Auswirkungen von unterschiedlichen Schätzungen ausschließlich in den Posteriori-Schätzungen der Zielgrößen und damit in deren Approximationsgüte erkennen lassen. Die Beurteilung der Annäherung von geschätzten Vergleichswerten zu den Kaufpreisen erfolgt getrennt für die Posteriori-Zielgrößen der vier Modelle sowie für die prädizierten Posteriori-Zielgrößen, die sich auf Grundlage der Kreuzvalidierung ergeben. Auf diese Weise lassen sich zwei getrennte Paarvergleichsmatrizen formulieren, die jeweils auf den genannten Vergleichskriterien basieren.

Eine Wertung der Resultate für die Unsicherheitsmodellierung in den vier Modellen erfolgt analog und anhand der Schritte i. bis iv. der Aufzählung des vorherigen Unterabschnittes. Als Vergleichskriterien dienen die absoluten Spannen der HPDIs der Modellparameter und Posteriori-Schätzgrößen sowie eine Beurteilung der Unsicherheitsspannen über ihre Relation zu den wahren Kaufpreisen. Entgegen der Ergebnisse für die funktionale Modellierung können an dieser Stelle für alle Aspekte der Vergleichsebene jeweils Paarvergleichsmatrizen formuliert werden.

Die Ergebnisse für die Variantenauswahl getrennt nach funktionaler Modellierung und Unsicherheitsmodellierung werden abschließend genutzt, um eine Gesamtwertung der Einzelergebnisse vornehmen zu können und eine Aussage über das Erreichen der Forschungsziele zu ermöglichen.

5. Praktische Anwendung der Auswertestrategie in der Vergleichswertermittlung

Die Verknüpfung der Vergleichswertermittlung mit dem Ansatz der Bayes-Strategie erfolgt im Rahmen der im vorherigen Kapitel geschilderten Auswertestrategie. Als innovative Herangehensweise wird ein besonderer Fokus auf die wahrscheinlichkeitstheoretische Modellierung gelegt. Für die Beurteilung, ob sich die Auswertestrategie und die Bayesschen Ansätze als geeignete Instrumente erweisen, um eine verbesserte funktionale Modellierung und Unsicherheitsmodellierung zu erreichen, sind sie anhand eines praktischen Datensatzes zu prüfen. In diesem Zusammenhang ist neben der Eignung von Interesse, wie sich die Ergebnisse der informativen Lösungsansätze zu denen der nichtinformativen verhalten.

Abbildung 5.1 gibt einen Überblick über die unterschiedlichen Modelle, die in der Arbeit gegenübergestellt werden. Die informative Bayessche Regression (fortan abgekürzt als *Modell R.1*) und die nichtinformativ Bayessche Regression (*Modell R.2*) sind als voneinander unabhängige Ansätze zu sehen. Das Modell *R.2* dient insbesondere Vergleichszwecken für eine Beurteilung von Vor- und Nachteilen bei der Integration von Priori-Wissen im Modell *R.1*. Die Ansätze der Bayesschen Kollokation sind als zusätzliche Varianten für das Vergleichswertverfahren zu betrachten. Die jeweiligen informativen und nichtinformativen Ansätze basieren auf den Residuen der Regressionsmodelle: Wie in der Abbildung dargestellt, ergibt sich dadurch eine direkte Verknüpfung des Modells *R.1* mit der informativen Bayesschen Kollokation (*Modell K.1*) sowie des Modells *R.2* mit der nichtinformativen Bayesschen Kollokation (*Modell K.2*). Im Mittelpunkt stehen in den vier Ansätzen jeweils die funktionale Modellierung sowie die Ermittlung des Unsicherheitshaushaltes.

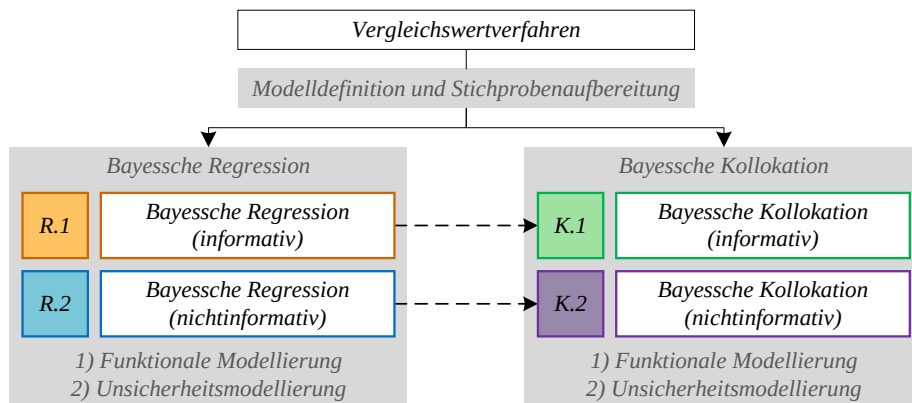


Abbildung 5.1: Modellübersicht in der Auswertestrategie

In den folgenden Abschnitten werden die Einzelschritte der Auswertestrategie anhand der Analyse einer Stichprobe von Kauffalldaten aufgezeigt. In Abschnitt 5.1 wird zunächst die Modelldefinition behandelt. Insbesondere wird auf die Definition der Grundgesamtheit sowie Definition und Aufbereitung der Stichprobendaten eingegangen. In Abschnitt 5.2 wird die Modellierung im Vergleichswertverfahren ausführlich anhand der initialen Epoche des Gesamtauswertezeitraumes gezeigt. Im Fokus stehen die Vergleiche der informativen und nichtinformativen Lösungen der Bayesschen Regression sowie die Ergebnisse der rekursiven Anwendung der Auswertestrategie. Abschnitt 5.3 ist der Modellierung im Ansatz der Bayesschen Kollokation gewidmet. Ausführlich behandelt werden die Ableitung von Autokovarianz- bzw. Autokorrelationsfunktionen, die Ergebnisse für die Schätzungen der Modellparameter und Zielgrößen sowie deren Unsicherheiten. Abschließend erfolgt in Abschnitt 5.4 die Validierung der geschätzten Modelle anhand des Ansatzes der Kreuzvalidierung. Zu diesem Zweck werden Vergleichswerte prädiziert, die sich aus den gegebenen Kauffalldaten simulieren lassen.

5.1. Modelldefinition und Stichprobenaufbereitung der Initialepoche

Wie in den theoretischen Grundlagen zur Auswertestrategie beschrieben, setzt deren Anwendung eine klare Modelldefinition voraus, um anhand derer eine geeignete Stichprobe von Kauffalldaten erheben zu können. Da die praktische Auswertung auf einer rekursiven Schätzung beruht, werden die Einzelschritte exemplarisch am Beispiel der initialen Auswertepoche gezeigt. Zwischenschritte für die Auswertungen der Folgeepochen können an dieser Stelle nicht im Detail dargestellt werden; an den entsprechenden Stellen wird jedoch auf die Ergebnisse verwiesen, die im Anhang zusammengestellt sind. In Abschnitt 5.1.1 wird zunächst die Definition der Grundgesamtheit sowie aufbauend in Abschnitt 5.1.2 die Definition der Stichprobendaten behandelt, bevor die Aufbereitung der erhobenen Stichprobendaten in Abschnitt 5.1.3 gezeigt wird.

5.1.1. Definition der Grundgesamtheit

Wie in Abschnitt 4.2.1 (S. 87 ff.) erläutert, ist es für eine zutreffende Abbildung von Teilmärkten notwendig, die Grundgesamtheit der am Grundstücksmarktmarkt gehandelten Objekte durch räumliche, sachliche und zeitliche Identifikationskriterien einzuschränken. Erst unter dieser Voraussetzung lassen sich eine adäquate Modellbildung sowie ein zutreffendes Abbild realer Zusammenhänge formulieren.



Abbildung 5.2: Räumlicher Teilmarkt für die praktischen Untersuchungen: Stadt Osnabrück und Stadtrandgemeinden (Kartengrundlage: Amtliche Geobasisdaten des LGLN, 2015)

Als übergeordnete räumliche Gliederung für die praktischen Untersuchungen wird das Bundesland Niedersachsen gewählt. Der räumliche Teilmarkt innerhalb Niedersachsens bezieht sich weiterhin auf ein Konzept, welches der für das Bundesland eingerichtete Obere Gutachterausschuss (OGA) in Kooperation mit der Investitions- und Förderbank Niedersachsen (NBank) erarbeitet hat. Im Zuge des Konzepts wurden Wohnungsmarktregionen gebildet, deren Abgrenzung sich aus allgemeinen Indikatoren – bspw. Bevölkerungsdichte- und -entwicklung – sowie aus der Einschätzung der Gutachterausschüsse des Landes Niedersachsens ergeben. Dabei ist zu beachten, dass es sich bei den Abgrenzungen um eine hohe Aggregationsstufe handelt: Abweichende und räumlich wesentlich differenziertere

Entwicklungen sind die Regel, sollen jedoch durch die gebildeten Wohnungsmarktregionen nicht widerspiegelt werden.

Um eine möglichst homogene und repräsentative Stichprobe zu erhalten, wird in den praktischen Untersuchungen eine vergleichsweise kleinräumige, städtische Region gewählt, die im Gegensatz zu großflächigen Wohnungsmarktregionen eine differenzierte Modellbildung erlaubt. Im vorliegenden Fall wird als räumlicher Teilmarkt die Stadt Osnabrück betrachtet. Die im Konzept der Wohnungsmarktregionen beschriebene Region Osnabrück umfasst neben dem reinen Stadtgebiet ebenfalls die Stadtrandgemeinden, die auf administrativer Ebene dem Landkreis zuzuordnen sind. Im Vordergrund sollen damit ähnliche Entwicklungstendenzen des Immobilienmarktes stehen, sodass der Einfluss und die Wirkung des eigentlichen Stadtgebietes sachgerecht berücksichtigt werden. Der *räumliche Teilmarkt der praktischen Untersuchungen* wird daher auf die Wohnungsmarktregion Osnabrück bezogen, die sich aus dem Stadtgebiet und den Stadtrandgemeinden zusammensetzt und die in Abbildung 5.2 dargestellt ist.

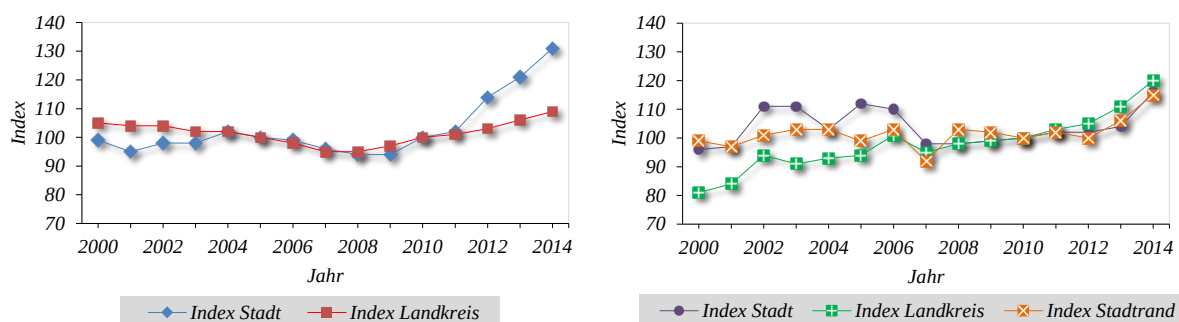


Abbildung 5.3: Indexreihen für freistehende Ein- und Zweifamilienhäuser (links) und Bodenpreisindexreihen für individuelles Wohnbauland in Osnabrück (rechts)

Neben der räumlichen Identifikation wird der *sachliche Teilmarkt der praktischen Untersuchungen* auf freistehende Ein- und Zweifamilienhäuser festgelegt, die in den Analysen von Gutachterausschüssen traditionell gemeinsam ausgewertet werden. Der Teilmarkt der Ein- und Zweifamilienhäuser hat sich in der Wertentwicklung in den vergangenen Jahren mit aufsteigender Tendenz gezeigt: Abbildung 5.3 (links) gibt einen Überblick über die konjunkturelle Entwicklung in Form einer aktuellen Indexreihe über den Zeitraum der vergangenen Jahre ab 2000. Die Indexreihe wurde durch den örtlichen Gutachterausschuss für Grundstückswerte (GAG) Osnabrück-Meppen abgeleitet und im aktuellen Grundstücksmarktbericht 2015 veröffentlicht. Da sich die Auswertungen des Gutachterausschusses an Stadt- und Landkreisgebiet orientieren, liegen für das Stadtrandgebiet keine separaten Auswertungen vor.

Als Tiefpunkt der Entwicklung sind die Jahre 2007 bis 2008 zu erkennen. Dieser Zeitraum deckt sich mit dem Beginn der Finanzkrise als Teil der Weltwirtschaftskrise, die ihren Ursprung im Sommer 2007 im Zusammenbruch des spekulativ aufgeblähten US-Immobilienmarktes fand (Subprime-Krise). Während sich die Krise in vielen Ländern weltweit massiv auf die Wirtschaft ausgewirkt hat, sind die Auswirkungen auf den deutschen Immobilienmarkt hingegen moderat ausgefallen. Wie anhand der Indexreihe zu erkennen ist, kann im hier betrachteten sachlichen Teilmarkt ab 2009 ein aufsteigender Trend sowohl für den Landkreis als auch die Stadt Osnabrück verzeichnet werden. Bestätigt wird die allgemeine Entwicklung durch die Bodenpreisindexreihe für individuelles Wohnbauland, die in Abbildung 5.3 (rechts) dargestellt ist und in der zusätzlich zu Stadt und Landkreis Osnabrück die Stadtrandgemeinden differenziert betrachtet werden. Als wesentlicher Indikator für preisliche Entwicklungen wird die Bodenpreisindexreihe an dieser Stelle lediglich nachrichtlich herangezogen, eine Verknüpfung zum hier betrachteten sachlichen Teilmarkt ergibt sich lediglich durch die Berücksichtigung des Bodenrichtwertes als Einflussgröße im Modellansatz. Dieser Aspekt wird im folgenden Abschnitt 5.1.2 aufgegriffen.

5.1.2. Definition der Stichprobendaten

Die Festlegung des räumlichen und sachlichen Teilmarktes ermöglicht im Weiteren die Stichprobenbildung. Eine wesentliche Voraussetzung in diesem Zusammenhang ist gemäß Schritt 1.2 nach Abbildung 4.2 (S. 87) die Wahl der Ziel- und Einflussgrößen, die in die Auswertung eingeführt werden sollen. Zielgröße sowie Art und Anzahl der Einflussgrößen, denen im Schritt der Modelldefinition zunächst ein wertrelevanter Zusammenhang unterstellt wird,

sind in besonderem Maße abhängig von der verfügbaren Datenbasis. Wie bereits erläutert, bietet die AKS der GAG eine umfangreiche Datenbank, die sämtliche Kauffalldaten innerhalb der jeweiligen Zuständigkeitsgrenzen enthält. Aus diesem Grund wird für die folgenden Auswertungen auf die AKS des GAG Osnabrück-Meppen als für den gewählten räumlichen Teilmarkt zuständige Ortsinstanz zurückgegriffen.

5.1.2.1. Vertrags- und Ordnungsmerkmale

Für die Erhebung der Stichprobenelemente werden zunächst allgemein die Vertrags- und Ordnungsmerkmale festgelegt, die einzelnen Kriterien können Tabelle 5.1 entnommen werden.

Für die *Vertragsmerkmale* erfolgt in einem ersten Schritt die Eingrenzung des betrachteten Auswertzeitraumes über das *Datum der Kaufverträge* sowie die Untergliederung in Auswertepochen. Wie den Ausführungen in Kapitel 4 entnommen werden kann, ermöglicht die Auswertestrategie eine rekursive Anwendung: Die Definition von mehreren Auswertepochen ist damit eine notwendige Voraussetzung der praktischen Untersuchungen. In der Theorie können die Auswertepochen jedes beliebige Intervall abdecken; denkbar sind Wochen-, Monats- oder Jahresabstände. Der konkrete Zeitraum muss im Einzelfall und im Hinblick auf die praktischen Erfordernisse sachverständig beurteilt werden. So kann sich bspw. für die Ableitung von Bodenrichtwerten in den vorgestellten Modellen entsprechend der gesetzlichen Vorgabe durch § 196 Abs. 1 BauGB eine Auswertepoche von zwei Jahren empfehlen, während sich für die Ableitung von Marktanpassungsfaktoren in Abhängigkeit der Datenverfügbarkeit ein Aktualisierungszeitraum von einem halben Jahr als günstig erweisen kann. Die Definition von Auswertepochen bewegt sich damit stets im Rahmen von fachlichen Anforderungen und Datenverfügbarkeit. Für den vorliegenden Fall der Auswertung von Ein- und Zweifamilienhäusern in der Wohnungsmarktregion Osnabrück liegen keine fachlich zwingend einzuhaltenden Anforderungen vor, wodurch vornehmlich die Datenverfügbarkeit in den Vordergrund rückt. Für eine ausreichende Datenbasis werden daher Auswertepochen von jeweils einem Jahr (Datum des Kaufvertrages (DAT) 01.01. bis 31.12.) gewählt, innerhalb derer die Modellbildung nach Kapitel 4 vorgenommen wird. Um sowohl die Gegenüberstellung und Beurteilung der Ergebnisse von Einzelauswertungen als auch der rekursiven Anwendung zu ermöglichen, wird neben der Definition von Jahresepochen ein Gesamtbetrachtungszeitraum von 10 Jahren festgelegt. Die praktischen Untersuchungen basieren auf den Daten der Jahre 2005 bis 2014, wobei das Jahr 2004 für die Generierung von Priori-Information für die Auswertung 2005 zusätzlich Berücksichtigung finden muss. Ergänzend für die Vertragsmerkmale nach Tabelle 5.1 wird der *Anlass des Eigentumsübergangs* festgelegt: Es wird gefordert, dass sich ausschließlich Objekte in der Stichprobe befinden, denen ein Eigentumswechsel durch Kauf zugrunde liegt. Objekte, deren Eigentumsübergang bspw. auf einem Tausch oder ähnlichem basiert, werden damit ausgeschlossen. Zudem werden ausschließlich Kauffälle der AKS selektiert, bei denen auf Grundlage des Kaufvertrages ein gewöhnlicher Geschäftsverkehr unterstellt wird.

Im Hinblick auf die *Ordnungsmerkmale* werden zunächst die *Kauffallkennzeichen* selektiert. Diese dienen ausschließlich der eindeutigen Zuordnung der Kauffallmerkmale und werden im eigentlichen Auswertalgorithmus lediglich zur Identifikation genutzt. Im Gegensatz zu dieser rein formalen Größe werden die *Objektkoordinaten* der einzelnen Kauffälle benötigt, um die Auswertung im Bayesschen Kollokationsansatz zu ermöglichen: In die Ermittlung der Autokovarianzfunktionen wird als zusätzliche Einflussgröße die geografische Distanz zwischen den Stichprobenobjekten genutzt, die sich aus den Objektkoordinaten berechnen lässt. Ergänzt werden die Ordnungsmerkmale abschließend durch die Vorgabe des Selektionskriteriums für die *Nutzungsart*. Die Stichprobenelemente werden ausschließlich aus Gebieten gewählt, denen im B-Plan die Nutzungsarten Wohnbaufläche (W), allgemeines Wohngebiet (WA) oder reines Wohngebiet (WR) zugeordnet sind. Diese Auswahl ist in Abhängigkeit zum gewählten sachlichen Teilmarkt zu sehen und reflektiert eine für Ein- und Zweifamilienhäuser typische Lage.

Tabelle 5.1.: Vertrags- und Ordnungsmerkmale für die Stichprobenerhebung

Vertragsmerkmale	Ordnungsmerkmale
a) Datumsbereich der Kaufverträge	a) Kauffallkennzeichen
b) Anlass des Eigentumsübergangs	b) Objektkoordinaten
	c) Nutzungsart

5.1.2.2. Ziel- und Einflussgrößen

Durch die Vertrags- und Ordnungsmerkmale ergibt sich bereits eine erste Einschränkung der Stichprobe für den gewählten Teilmarkt. Für die Definition der eigentlichen Stichprobendaten erfolgt zunächst die Bestimmung einer geeigneten Zielgröße. Da die Bayesschen Modellansätze in der vorliegenden Arbeit explizit auf die Vergleichswertermittlung übertragen werden, wird die Zielgröße durch den jeweiligen Kaufpreis (KPR) der Stichprobenobjekte als Beobachtung y repräsentiert. Wie bereits erläutert, wird diese Größe i. d. R. nicht als absolute Größe in Auswertansätze übernommen, sondern aus numerischen Gründen als normierter Wert eingeführt. Im Hinblick auf den hier betrachteten sachlichen Teilmarkt der Ein- und Zweifamilienhäuser lassen sich in den klassischen Auswertungen der GAG zwei Normierungsvarianten finden: Zum einen das Verhältnis *Kaufpreis/Sachwert*, zum anderen das Verhältnis *Kaufpreis/Bezugseinheit*. Erstgenannte Größe wird als Sachwertfaktor bezeichnet und im Sachwertverfahren nach der *Richtlinie zur Ermittlung des Sachwerts - Sachwertrichtlinie (SW-RL)* genutzt, um theoretisch berechnete Herstellungskosten eines Wertermittlungsobjektes an die tatsächliche Marktlage anzupassen. Der Sachwertfaktor stellt sich damit als unverzichtbare Größe dar, die regelmäßig im Zuge der Ableitung sonstiger für die Wertermittlung erforderlicher Daten nach § 193 Abs. 5 Nr. 2 BauGB durch Gutachterausschüsse abzuleiten und zu veröffentlichen ist. Die zweitgenannte Größe stellt die Definition eines Vergleichsfaktors nach der VW-RL dar: Nach Nr. 6 Abs. 1 handelt es sich bei Vergleichsfaktoren um durchschnittliche, auf eine Einheit bezogene Werte für Grundstücke mit bestimmten wertbeeinflussenden Grundstücksmerkmalen. Als geeignete Bezugseinheiten werden als Beispiele der jährliche Ertrag (Ertragsfaktor) sowie alternativ eine Flächen- oder Raumeinheit der baulichen Anlagen (Gebädefaktor) genannt.

Für die vorliegende Arbeit wird als *Zielgröße* der *Gebädefaktor als Vergleichsfaktor* (Kaufpreis pro Wohnfläche (KPW)) gewählt; die Diskussion der Übertragbarkeit auf abweichende Fragestellungen erfolgt in Kapitel 6. Die Bezugseinheit für den Kaufpreis soll durch die Wohnfläche der baulichen Anlage repräsentiert werden. Die Einheit der Bezugsgröße für die Beobachtungen ergibt sich damit zu Euro pro Quadratmeter (€/m^2).

Im engen sachlichen Zusammenhang mit der Definition der Zielgröße kann die Bestimmung der (vermuteten) wertrelevanten Einflussgrößen erfolgen. Neben der rein funktionalen und stochastischen Modellierung in den Bayesschen Ansätzen sollen insbesondere Modelle genutzt werden, die durch einige wenige signifikante Einflussgrößen ein zutreffendes Abbild des betrachteten Teilmarktes wiedergeben. Die Reduktion der Modellkomplexität auf ein notwendiges Maß ermöglicht zudem eine einfache praktische Handhabbarkeit, da im Idealfall umfangreiche Nach- bzw. Neuerfassungen von Einflussgrößen vermieden werden können. Aus diesem Grund wird in der folgenden praktischen Untersuchung ausschließlich auf verfügbare Daten der AKS zurückgegriffen. Eine Erhebung zusätzlicher Daten für Modellerweiterungen ist grundsätzlich denkbar, soll an dieser Stelle jedoch nicht weiter verfolgt werden. Um ein möglichst einfaches Trendmodell zu entwickeln, werden als Einflussgrößen für den Teilmarkt der Ein- und Zweifamilienhäuser die klassischen Größen gewählt, deren Werteeinfluss sich bereits regelmäßig in praktischen Anwendungen erwiesen hat:

- *Bodenrichtwert (BRW)*: Der Bodenrichtwert als durchschnittlicher Wert für eine Zone mit vergleichbaren wertbeeinflussenden Merkmalen gilt als wichtigster Indikator für die Lagewertigkeit eines Objekts. Die Ermittlung von Bodenrichtwerten erfolgt in Niedersachsen flächendeckend zum Stichtag 31.12. jedes Jahres.
- *Grundstücksfläche (GRF)*: Als preisbeeinflussendes Merkmal stellt sich i. d. R. die verfügbare Grundstücksfläche dar. Insbesondere in Bezug auf Ein- und Zweifamilienhäuser ist die Grundstücksgröße ein entscheidender Faktor für die Erklärung des Kaufpreises.
- *Wohnfläche (WOF)*: Ebenso wie der Grundstücksfläche ist der eigentlichen Wohnfläche eine Wertrelevanz zu unterstellen. Eine höhere verfügbare Wohnfläche wird sich i. d. R. auch in höheren Kaufpreisen niederschlagen. Zudem findet sich die Wohnfläche als Bezugseinheit in der Definition der Zielgröße wieder. An dieser Stelle sei hingewiesen auf den Methodenstreit über die Verwendung einer Einflussgröße, die sich zugleich in der Zielgröße wiederfindet. Diese Thematik ist nicht abschließend geklärt und variiert je nach Auffassung des Auswerters: Im Rahmen der vorliegenden Arbeit wird die Wohnfläche als zusätzliche Einflussgröße in den Modellansatz übernommen.

- *Ausstattungsstandard (AST)*: Der Ausstattungsstandard einer Immobilie wird in der Anwendung der AKS als kategoriale Variable erhoben. Die Definition der Kategorien richtet sich nach den Vorgaben der SW-RL und wird mit den Standardstufen 1 bis 5 bezeichnet. Zu beachtende Aspekte sind in erster Linie Qualität der verwendeten Materialien und der Bauausführung, energetische Eigenschaften sowie Standardmerkmale für spezifische sachliche Teilmärkte.
- *Unterkellerung (KEL)*: Als weitere Einflussgröße wird der Grad der Unterkellerung erfasst. Durch eine Unterkellerung ist i. d. R. mit erhöhten Baukosten eines Objekts zu rechnen, eine Wertrelevanz kann durch diesen Umstand gegeben sein. Eingeschlossen sind in der Stichprobe Objekte mit einer Voll- oder Teilunterkellerung sowie Objekte ohne Unterkellerung.
- *(Modifiziertes) Alter (ALT)*: Abschließend wird das Alter der Objekte zum jeweiligen Vertragsdatum in der Selektion berücksichtigt. Es ist zu unterstellen, dass ältere Objekte im Hinblick auf die wirtschaftliche Gesamtnutzungsdauer regelmäßig einen höheren Abschlag auf den Kaufpreis zu verzeichnen haben als neuere Objekte, die erst wenige Jahre alt sind. Abhängig sind Einflüsse auf den Kaufpreis jedoch auch vom Grad der regelmäßigen Instandhaltungsmaßnahmen, Modernisierung oder eventuell durchgreifenden Kernsanierungen. Um diesem Umstand gerecht zu werden, lassen sich der AKS Angaben zum modifizierten Baujahr entnehmen: Hier werden anhand von vorgenommenen Modernisierungsarbeiten fiktive Baujahre angenommen, da unterstellt wird, dass die Maßnahmen eine Verlängerung der wirtschaftlichen Gesamtnutzungsdauer bewirken. Im Zuge der praktischen Auswertungen wird das jeweilige Gebäudealter daher basierend auf einem fiktiven Baujahr berechnet. Zu diesem Zweck werden das tatsächliche Baujahr und das auf Einschätzungen des GAG beruhende, unterstellte modifizierte Baujahr in ein gewichtetes Verhältnis gesetzt. Das tatsächliche Baujahr (BAJ) ist dabei höher zu gewichten, da auch durch Modernisierungsarbeiten i. d. R. die eigentliche, tragende Bausubstanz erhalten bleibt und um punktuelle Maßnahmen ergänzt wird (außer Acht bleiben Kernsanierungen). Das Alter der Objekte wird daher nach der Gewichtung $ALT = (Tats\ddot{a}chliches\ Baujahr \cdot 0,75) + (modifiziertes\ Baujahr \cdot 0,25)$ in die Modellansätze eingeführt.

Auf Grundlage der bisherigen Ausführungen zu Vertrags- und Ordnungsmerkmalen sowie Ziel- und Einflussgrößen erfolgt die Erhebung der Stichprobendaten aus der AKS. Eine Einschränkung der Selektion ergibt sich bei Anwendungen der AKS durch die Auswahl der Selektionsgrenzen, die als Modellgrenzen interpretiert werden können und nach sachverständigem Ermessen gesetzt werden. Ziel der Selektionsgrenzen ist es, bereits im Vorfeld der Datenerhebung eine möglichst homogene Stichprobe anzustreben, innerhalb derer von einer gewissen Vergleichbarkeit der Objekte ausgegangen werden kann. Grobe Ausreißer werden durch diese Maßnahme erheblich reduziert, jedoch nicht vollständig eliminiert. Das Erkennen und Entfernen von sowohl groben als auch sonstigen Ausreißern erfolgt daher zusätzlich im Rahmen der Auswertestrategie. Da die einzelnen Auswertepochen der Jahrgänge 2004 bis 2014 den jeweils gleichen Selektionsgrenzen unterliegen, können die Stichproben für die einzelnen Auswertepochen als Gesamtstichprobe erhoben werden; die Einhaltung der allgemein gültigen Selektionsgrenzen der Einzelepochen ist gewährleistet. Die Auswahl der Stichprobengrenzen kann Tabelle 5.2 entnommen werden und entspricht den in der Praxis verwendeten Ansätzen der GAG.

Tabelle 5.2.: Selektionsgrenzen für die Gesamtstichprobe der Ein- und Zweifamilienhäuser (Stadt und Stadtrandgemeinden Osnabrück)

Merkmal	Abkürzung	Untere Grenze	Obere Grenze
Vertragsdatum	DAT	01.01.2004	31.12.2014
Kaufpreis	KPR	50.000 €	500.000 €
Bodenrichtwert	BRW	55 €/m ²	-
Grundstücksfläche	GRF	-	1.500 m ²
Wohnfläche	WOF	50 m ²	-
Ausstattungsstandard	AST	-	-
Unterkellerung	KEL	-	-
Baujahr	BAJ	1950	2011

Mit den vorstehend beschriebenen Schritten ist die Modelldefinition nach Abbildung 4.2 (S. 87) abgeschlossen und die definierte Gesamtstichprobe für die AuswerteePOCHEN 2004 bis 2014 selektiert. Gemäß Abbildung 4.3 (S. 88) folgt als nächster Schritt die Aufbereitung der Stichprobendaten für die Auswertung der EinzelePOCHEN. Die Vorgehensweise am numerischen Beispiel der initialen Epoche 2004/2005 wird im folgenden Abschnitt aufgezeigt.

5.1.3. Stichprobenaufbereitung der initialen Epoche

Für die Stichprobenaufbereitung werden jeweils die Einzelstichproben der Jahre 2004 und 2005 betrachtet. Die Kauffalldaten des Jahres 2004 dienen ausschließlich der Generierung von empirischer Priori-Information für die Daten des Jahres 2005. Dies bedeutet, dass im Rahmen der Auswertestrategie für diese Epoche ausschließlich Bayessche Modelle basierend auf einer nichtinformativen Lösung berechnet werden können. Tabelle 5.3 fasst die wesentlichen Kenngrößen der Epochen 2004 und 2005 zusammen. Die Stichprobenaufbereitung selbst vollzieht sich nach den in Abbildung 4.3 (S. 88) dargestellten Einzelschritten 2.1 bis 2.3.

Tabelle 5.3.: Deskriptive Statistik der Epochen 2004 und 2005

Epoche	Merkmal	DAT	KPW	BRW	GRF	WOF	AST	KEL	ALT
	Einheit	[-]	[€/m ²]	[€/m ²]	[m ²]	[m ²]	[-]	[%]	[Jahre]
2004	Anzahl (belegt)	94	94	94	94	94	87	94	94
	Minimum	02.01.2004	769	65	305	90	1	0	4
	Maximum	28.12.2004	2.867	340	1.499	265	5	100	54
	Mittelwert	-	1.458	148	774	161	3	77	30
	Median	-	1.394	140	726	150	3	100	30
2005	Anzahl (belegt)	105	105	105	105	105	100	105	105
	Minimum	05.01.2005	813	75	267	92	1	0	3
	Maximum	30.12.2005	2.667	360	1.414	280	5	100	55
	Mittelwert	-	1.481	164	780	162	3	75	32
	Median	-	1.436	145	761	155	3	100	33

5.1.3.1. Stichprobenbildung

Für die weitere Datenauswertung wird im ersten Schritt die Stichprobenbildung mit der Prüfung auf Vollständigkeit der Kauffalldaten vorgenommen. Wie der Vergleich der Anzahl der belegten Kauffälle in Tabelle 5.3 zeigt, sind die Zielgröße sowie die einzelnen Einflussgrößen nahezu gleichmäßig mit Werten in der AKS hinterlegt. Während sich unter den Selektionsgrenzen für die Epoche 2004 insgesamt 94 Kauffälle ergeben, liegt die Zahl für die Epoche 2005 bei 105 Datensätzen. Die Ausnahme in der Belegung stellt die Einstufung des Ausstattungsstandards dar, die in jeder AuswerteePOCHe unter der Anzahl der Kauffälle insgesamt liegt (2004: 87 Kauffalldaten, 2005: 100 Kauffalldaten). Der Ausstattungsstandard wird in der praktischen Führung der AKS manuell nachgeführt und durch sachverständiges Ermessen anhand der übrigen vorliegenden Objektdaten ergänzt. Somit liegt hier eine Größe vor, die nicht aus den Kaufvertragsdaten erfasst werden kann und ausschließlich bedarfsgerecht – bspw. für die Ableitung von Sachwertfaktoren – ergänzt wird. Wird durch den Auswerter ein Objekt als nicht geeignet für die Ableitung von Sachwertfaktoren eingestuft, erfolgt keine zusätzliche Beurteilung des Ausstattungsstandards: Als Folge ergeben sich in einzelnen Fällen die Datenlücken in der Einflussgröße des Ausstattungsstandards. Da es sich insgesamt um eine vergleichsweise geringe Anzahl von Datenlücken handelt, wird der Ausstattungsstandard in die weiteren Auswertungen eingeführt. Für die weiteren Auswertungen bedeutet dies, dass im Modellansatz diejenigen Kauffälle aus den Epochen zu eliminieren sind, bei denen keine Belegung des Ausstattungsstandards vorliegt. Sofern sich in den Modellansätzen die Größe AST als nicht relevant für das Gesamtmodell erweist, können die zuvor eliminierten Kauffalldaten den Auswertungen wieder hinzugefügt werden, da unter dessen Ausschluss eine vollständige Belegung der übrigen Einflussgrößen gewährleistet ist und der Gesamtdatensatz der Epoche zur Verfügung steht.

Insgesamt profitiert die Stichprobenbildung von der Wahl derjenigen Einflussgrößen, die sich in praktischen Untersuchungen regelmäßig als potentiell wertrelevant erwiesen haben. Da diese Größen standardmäßig in der AKS mit Werten erfasst werden, resultieren für die Auswertepochen Datensätze, die nahezu vollständig belegt sind und einen größtmöglichen Stichprobenumfang gewährleisten.

5.1.3.2. Stichprobenaufbereitung – Zielgröße

Gemäß Schritt 2.2 der Stichprobenaufbereitung nach Abbildung 4.3 (S. 88) erfolgt die Untersuchung der Zielgröße auf Gleichgewichtigkeit. Wie erläutert, ist als ein erster Indikator für die Beurteilung der Gleichgewichtigkeit das Verhältnis zwischen minimalem und maximalem Wert der Zielgröße heranzuziehen. Bei einer Gesamtspanne von 769 €/m^2 bis 2.867 €/m^2 liegt das Verhältnis für die Priori-Stichprobe der Epoche 2004 bei 3,73 und damit unter der Forderung $y_{\min} : y_{\max} < 5$. Für die Stichprobe 2005 ergibt sich ein Verhältnis von 3,28 bei einer Spanne von 813 €/m^2 bis 2.667 €/m^2 , womit auch in diesem Fall von Gleichgewichtigkeit ausgegangen werden kann. Als zusätzliche Kenngröße wird in beiden Fällen zudem die Schiefe nach (4.1) bestimmt: Für den Datensatz 2004 ergibt sich ein Wert von $v_3^{2004} = 1,08$, der sich außerhalb der pragmatisch gewählten Grenzen von $0,00 \pm 0,30$ bewegt und damit auf eine schiefsymmetrische Verteilung hindeutet. Visuell wird diese Tatsache ebenfalls durch das Histogramm der Zielgröße nach Abbildung 5.4 (links) bestätigt. Ein ähnliches Bild ergibt sich für die Zielgröße der Epoche 2005: Während das Verhältnis der minimalen und maximalen Werte auch hier Gleichgewichtigkeit suggeriert, ergibt sich die Schiefe zu $v_3^{2005} = 0,80$ und liegt damit ebenfalls außerhalb des gewählten Grenzbereiches. Die grafische Darstellung kann Abbildung 5.5 (links) entnommen werden. Die Zielgrößen der betrachteten Epochen entsprechen damit nicht der Forderung nach einer näherungsweisen symmetrischen Verteilung, die jedoch für die konjugierten Bayesschen Modelle unterstellt wird.

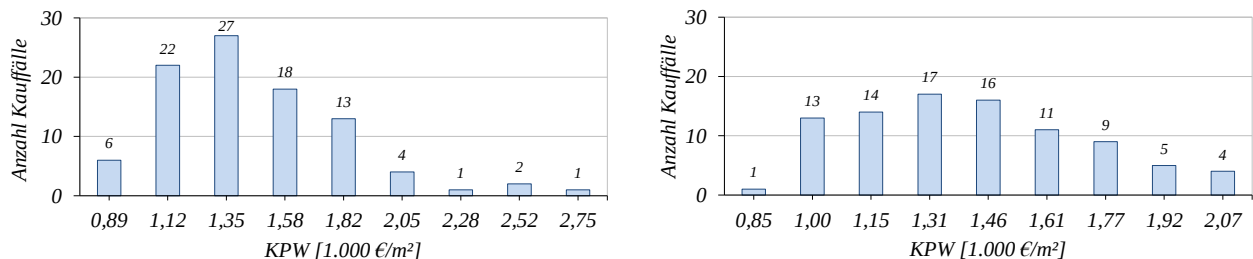


Abbildung 5.4: Histogramme der Zielgröße KPW der Epoche 2004 (links: Originaldaten, rechts: Aufbereitete Daten)

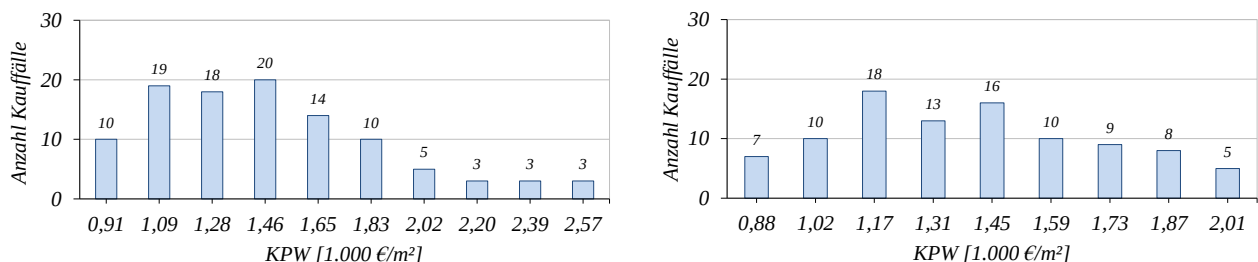


Abbildung 5.5: Histogramme der Zielgröße KPW der Epoche 2005 (links: Originaldaten, rechts: Aufbereitete Daten)

Die Histogramme der Zielgrößen für die Epochen 2004 und 2005 können dahingehend interpretiert werden, dass die Stichprobenabgrenzung optimiert werden muss. Während sich die Masse der Kaufpreise pro Wohnfläche auf einen Bereich bis etwa 2.000 €/m^2 konzentriert, sind im Vergleich nur relativ wenige Objekte mit einem höheren Quadratmeterpreis veräußert worden. Im statistischen Sinne und im Hinblick auf die überwiegende Preisspanne werden diese Kauffälle als Ausreißer interpretiert. Als weiterer Schritt in der Iteration der Datenaufbereitung erfolgt daher die Ausreißersuche anhand einer $2,5\sigma$ -Regel, d. h., es werden diejenigen Kauffalldatensätze ausgeschieden, die um mehr als das 2,5-fache der Standardabweichung vom Mittelwert abweichen. Für die Epoche 2004 werden auf diese Weise insgesamt 4 Ausreißer aufgedeckt, die aus dem Datensatz eliminiert werden. Die Gesamtzahl reduziert sich dementsprechend von 94 auf 90 Datensätze. Das Verhältnis von minimalem (769 €/m^2) zu maximalem

Wert (2.148 €/m^2) reduziert sich auf 2,79. Die Neuberechnung der Schiefe ergibt sich zu $v_3^{2004} = 0,28$, liegt unterhalb der gewählten Grenze und erfüllt damit die Anforderungen. Eine zusätzliche grafische Darstellung kann dem Histogramm nach Abbildung 5.4 (rechts) entnommen werden.

Für die Stichprobe 2005 wird in einem ersten Ansatz ebenfalls eine Ausreißersuche nach der $2,5\sigma$ -Regel vorgenommen. Insgesamt können so 6 Kauffälle als Ausreißer identifiziert werden, sodass sich der Gesamtdatensatz von 105 auf 99 Kauffälle reduziert. Als neuer Wert für das Verhältnis Minimum zu Maximum ergibt sich 2,75 bei einer Spanne von 813 €/m^2 bis 2.238 €/m^2 . Die Neuberechnung der Schiefe nach Ausreißersuche ergibt sich zu $v_3^{2005} = 0,37$, was nach wie vor außerhalb der geforderten Grenze von 0,30 liegt und einen weiteren Iterationsschritt in der Stichprobenaufbereitung erfordert. Nach einem erneuten Durchlauf werden weitere 3 Kauffälle als Ausreißer aus dem Gesamtdatensatz entfernt, sodass sich die Anzahl auf 96 reduziert. Die Spanne der Zielgröße reicht nunmehr von 813 €/m^2 bis 2.083 €/m^2 , was einem Verhältnis von 2,56 entspricht. Da sich die Schiefe auf Grundlage des neuen Datensatzes zu $v_3^{2005} = 0,21$ ermitteln lässt, kann auch für die Epoche 2005 von einer symmetrischen Verteilung ausgegangen werden. Bestätigt wird diese Annahme durch das Histogramm der Zielgröße nach Abbildung 5.5 (rechts). Tabelle 5.4 gibt einen abschließenden Überblick über die wesentlichen Kennzahlen und die Iterationsschritte für die Aufbereitung der Zielgrößen der Epochen 2004 und 2005.

Tabelle 5.4.: Iterationen in der Aufbereitung der Zielgröße der Initialepochen

Epoche	Iteration	Maßnahme Datensätze	Anzahl Minimum	Gleichgewichtigkeit KPW			Schiefe
				Maximum	Verhältnis	v_3	
2004	0	-	94	769	2.867	3,73	1,08
	1	Ausreißersuche ($2,5\sigma$)	90	769	2.148	2,79	0,28
2005	0	-	105	813	2.667	3,28	0,80
	1	Ausreißersuche ($2,5\sigma$)	99	813	2.238	2,75	0,37
	2	Ausreißersuche ($2,5\sigma$)	96	813	2.083	2,56	0,21

Bei der Datenaufbereitung ist zu beachten, dass es sich um eine statistisch nicht strenge Vorgehensweise handelt. Durch die beschriebene Aufbereitung werden die praktischen Belange der Verkehrswertermittlung in den Mittelpunkt gestellt sowie der immobilienpezifischen, heterogenen Datengrundlage Rechnung getragen. Nachdem die Zielgröße durch die beschriebenen Maßnahmen aufbereitet ist, wird im folgenden Abschnitt die Aufbereitung der Einflussgrößen thematisiert.

5.1.3.3. Stichprobenaufbereitung - Einflussgrößen

Nach der Untersuchung der Zielgröße müssen die gewählten Einflussgrößen in Abhängigkeit des jeweiligen Datentyps auf eine ausreichende Anzahl in der Belegung sowie auf grobe Ausreißer geprüft werden. In den Einflussgrößen für den hier gewählten Modellansatz lassen sich die Merkmale BRW, GRF, WOF, KEL sowie ALT als stetige Daten kategorisieren. Das Merkmal AST ist den kategorialen Daten zuzuordnen, dichotome Merkmale sind nicht im Ansatz vorhanden.

Wird zunächst das Merkmal AST betrachtet, ergibt sich eine Einteilung von 5 Klassen, die den jeweiligen Ausstattungsstandard repräsentieren: Während die Stufe 1 eine einfache Ausstattung darstellt, wird die Stufe 5 mit einer stark gehobenen Ausstattung verbunden. Wie in Abschnitt 4.2.2 (S. 88 ff.) erläutert, wird an die jeweilige Belegung der Klassen die Forderung einer Anzahl von $n/3d$ gestellt, was im Fall der Epoche 2004 einer Anzahl von $87/15 = 5,8 \approx 6$ entspricht, im Fall der Epoche 2005 $100/15 = 6,6 \approx 7$. In Tabelle 5.5 wird die Häufigkeit der Klassenbelegungen der einzelnen Epochen aufgezeigt. Wie anhand der Tabelle zu erkennen ist, wird die Forderung nach mindestens sechs Belegungen in der Epoche 2004 in zwei der fünf Fälle nicht erfüllt; in den Klassen 2 und 4 findet sich jeweils nur ein Kauffall. Während die Klassen 1 und 5 mit jeweils acht Kauffällen knapp über dem Grenzwert liegen, konzentriert sich die Masse der Kauffälle mit 69 Kauffällen im Bereich der Standardstufe 3.

In der Epoche 2005 zeigt sich ein ähnliches Bild: Die Randklassen 1 und 5 fallen mit sieben bzw. fünf Fällen stark gegenüber Klasse 3 mit 88 Kauffällen ab, zudem erfüllt Klasse 5 nicht die Mindestanforderung an deren Belegung. Die Standardstufen 2 und 4 sind im gesamten Datensatz nicht belegt. Aufgrund der ungleichmäßigen

Verteilung und der Konzentration auf Standardstufe 3 wird die Einflussgröße aus dem Modellansatz entfernt: Die Klassen 2 und 4 sind zum einen nicht ausreichend belegt, um Aussagen über signifikante Werteinflüsse zu ermöglichen, zum anderen ergibt sich selbst bei der Zusammenfassung der Randklassen 1 und 2 bzw. 4 und 5 ein erhebliches Ungleichgewicht zugunsten der Klasse 3. Eine Aussage über einen signifikanten Werteinfluss einzelner Standardstufen erscheint zumindest fraglich und kann unter diesen Umständen nicht allgemeingültig formuliert werden. Nach der Entfernung der Einflussgröße des Ausstattungsstandards verbleiben daher ausschließlich die Merkmale BRW, GRF, WOF, KEL und ALT als Einflussgrößen vom stetigen Datentyp im Modellansatz. Zu beachten ist, dass sich durch die Elimination des Merkmals AST der Datenumfang pro Epoche entsprechend der Tabelle 5.3 wieder erhöht.

Tabelle 5.5.: Häufigkeiten der Klassenbelegung der Einflussgröße AST

AST (Klasse)	1	2	3	4	5
Epoche 2004	8	1	69	1	8
Epoche 2005	7	0	88	0	5

Für die Prüfung der stetigen Einflussgrößen auf grobe Ausreißer wird nach den Ausführungen in Abschnitt 4.2.2 (S. 88 ff.) eine Unter- bzw. Obergrenze eingeführt, die sich anhand einer 4σ -Regel ergibt. Im Vergleich zur Aufbereitung der Zielgröße ($2,5\sigma$ -Regel) wird an dieser Stelle ein größeres Intervall gewählt: In der Auswertestrategie sollen bewusst wenige Kauffälle als Ausreißer eliminiert werden. Da in den Bayesschen Modellansätzen lediglich für die Zielgröße Verteilungsannahmen zu treffen sind, können sich die Auswirkungen schiefsymmetrischer Einflussgrößen im Modellansatz widerspiegeln, ohne dass die Modellannahmen verletzt werden. Mit dieser Betrachtungsweise ist es u. a. Ziel der Modellierung, die Einflussgrößen möglichst unverfälscht und ohne weitere Aufbereitungsschritte in den Modellansatz zu übernehmen. Ähnlich wie im Fall der Aufbereitung der Zielgröße werden die stetigen Variablen jeweils in iterativen Schritten aufbereitet, d. h., dass für jede Einflussgröße zunächst die Unter- und Obergrenzen ermittelt werden. Liegen einzelne Realisierungen außerhalb der Grenzen, wird der zugehörige Kauffall aus dem Datensatz entfernt und eine Prüfung unter neu berechneten Bereichsgrenzen vorgenommen.

In Tabelle 5.6 sind die Ergebnisse für die betrachteten Epochen 2004 und 2005 aufgelistet: Als Kennwerte der Einflussgrößen werden jeweils Minimum, Mittelwert und Maximum angegeben. Die letzten beiden Spalten der Tabelle zeigen die Unter- und Obergrenzen, die sich anhand der 4σ -Regel ergeben, wobei negative Untergrenzen aus sachlogischen Gründen auf den Wert 0 gesetzt werden. Wie zu erkennen ist, sind sowohl innerhalb der Epoche 2004 als auch der Epoche 2005 keine Iterationen notwendig, sämtliche Realisierungen liegen innerhalb der ermittelten Bereichsgrenzen – grobe Ausreißer in den stetigen Daten werden für den weiteren Prozess der Auswertung daher ausgeschlossen.

Abschließend ist zu prüfen, ob zwischen den Einflussgrößen Korrelationen vorliegen, die für die Vermeidung von Multikollinearität den Ausschluss einer oder mehrerer Merkmale erfordert. Da im Modellansatz nach der Eliminierung des Ausstattungsstandards ausschließlich stetige Daten vorhanden sind, werden Spearmansche Korrelationskoeffizienten nach (4.2) bzw. (4.3) berechnet. Die Ergebnisse lassen sich den Tabellen 5.7 und 5.8 entnehmen. Insgesamt zeigt sich, dass sich keine Korrelationen $\rho > |0,7|$ ergeben, in deren Fall über den Ausschluss einer Einflussgröße zu entscheiden wäre. Die höchste Korrelation ergibt sich in der Epoche 2005 mit einem Wert von $\rho = 0,42$ für die Einflussgrößen der Unterkellerung und des Alters, in der Epoche 2004 liegt der Wert bei $\rho = 0,29$. Zumindest für die Epoche 2005 ergibt sich damit, dass der Wert im Bereich von $|0,3| < \rho < |0,7|$ liegt mit der Konsequenz, dass die Koeffizienten nicht getrennt voneinander interpretiert werden sollten. Gleiches trifft auf die Kombination des Alters und der Grundstücksfläche ($\rho = 0,31$) sowie der Grundstücksfläche und der Wohnfläche ($\rho = 0,32$) in der Epoche 2004 und auf die Kombination der Einflussgrößen des Alters und der Grundstücksfläche ($\rho = 0,32$) in der Epoche 2005 zu. Im Vergleich der Epochen bestätigen sich die Tendenzen der Korrelationen in ihren Vorzeichen und variieren nur in ihrer numerischen Ausprägung. Eine Ausnahme bilden die Korrelationen, die sich auf den Keller als Einflussgröße beziehen; hier wechseln die Vorzeichen und Höhen der Korrelationen in

Tabelle 5.6.: Iterationen in der Aufbereitung der stetigen Einflussgrößen in den Epochen 2004 und 2005

Epoche	Iteration	Einflussgröße	Einheit	Anzahl Datensätze	Kennwerte Stichprobe			Grenzwerte (4σ)	
					Minimum	Mittel	Maximum	Untere Grenze	Obere Grenze
2004	0	BRW	€/m ²	90	65	142	340	0	343
	0	GRF	m ²	90	305	768	1.499	0	1.647
	0	ALT	Jahre	90	4	29	54	0	86
	0	WOF	m ²	90	90	161	265	0	320
	0	KEL	-	90	0	77	100	0	100
2005	0	BRW	€/m ²	96	75	154	340	0	359
	0	GRF	m ²	96	356	779	1.389	0	1.696
	0	ALT	Jahre	96	3	32	55	0	84
	0	WOF	m ²	96	92	163	280	0	322
	0	KEL	-	96	0	80	100	0	100

Einzelfällen. Da sich jedoch keine Korrelationen ergeben, die den gesetzten Grenzwert von $\rho > |0,7|$ überschreiten, werden sämtliche Einflussgrößen im Modellansatz belassen.

Mit der Untersuchung der Korrelationen ist der Schritt der Stichprobenaufbereitung abgeschlossen, es sind die Voraussetzungen für die Schätzung in den Bayesschen Modellen gegeben. Die funktionale Modellierung der Bayesschen Regressionsansätze wird im folgenden Abschnitt behandelt.

Tabelle 5.7.: Korrelationen zwischen den Einflussgrößen (Epoche 2004)

Einflussgröße	BRW	GRF	ALT	WOF	KEL
BRW	1,00	0,00	0,07	0,16	-0,03
GRF	0,00	1,00	0,31	0,32	0,12
ALT	0,07	0,31	1,00	-0,08	0,29
WOF	0,16	0,32	-0,08	1,00	-0,02
KEL	-0,03	0,12	0,29	-0,02	1,00

Tabelle 5.8.: Korrelationen zwischen den Einflussgrößen (Epoche 2005)

Einflussgröße	BRW	GRF	ALT	WOF	KEL
BRW	1,00	0,02	0,19	0,11	0,20
GRF	0,02	1,00	0,32	0,09	-0,04
ALT	0,19	0,32	1,00	-0,30	0,42
WOF	0,15	0,09	-0,30	1,00	-0,08
KEL	0,20	-0,04	0,42	-0,08	1,00

5.2. Modellierung im Ansatz der Bayesschen Regression

Die Auswertung der aufbereiteten Stichprobendaten stellt sich als iterativer Ablauf dar: Kern des Auswerteprozesses ist die Wahl eines Modellansatzes, der i. S. der Qualitätsansprüche an die mathematische Modellierung eine bestmögliche Beschreibung des räumlichen und sachlichen Teilmarktes erlaubt. Als Grundmodell dient die informative Bayessche Regression; die Wahl eines geeigneten Modellansatzes basiert auf den Bayesschen Hypothesentests. Ist in diesem Zuge das wahrscheinlichste Modell für die Beschreibung der vorliegenden Daten gefunden, können neben der endgültigen Schätzung im informativen Regressionsmodell die nichtinformativen Lösung sowie die informative und nichtinformative Bayessche Kollokation als erweiterte Varianten der funktionalen Modellierung geschätzt werden. Im Ergebnis ermöglicht die Vorgehensweise den direkten Vergleich der jeweiligen informativen und nichtinformativen Schätzergebnisse.

In Abschnitt 5.2.1 wird zunächst die Modellwahl für die vorliegende Stichprobe aufbereitet. Da insgesamt ein Zeitraum über 10 Epochen abgedeckt wird, erfolgt die initiale Modellwahl in der Epoche 2005. Die Schätzergebnisse für die informative und nichtinformative Bayessche Regression werden in Abschnitt 5.2.2 vorgestellt, die resultierenden Unsicherheiten werden aufbauend in Abschnitt 5.2.3 behandelt.

5.2.1. Modellwahl mittels Bayes-Faktor

Die Auswertestrategie nach Abbildung 4.4 (S. 92) erfordert in Schritt 3.1 für die Ermittlung eines zutreffenden Modells zunächst die Schätzung eines Gesamtmodells. In der Modellwahl wird dieses dann mittels Bayes-Faktoren den ebenfalls zu schätzenden Modellvarianten gegenübergestellt. Dabei steht zunächst nicht die Darstellung der resultierenden Schätzfunktion, der Schätzergebnisse sowie der Unsicherheiten im Vordergrund, sondern die für die Berechnung der Bayes-Faktoren erforderlichen Daten nach den Erläuterungen in Abschnitt 4.3.2 (S. 98 ff.) sowie insbesondere Gleichung (4.32). Im Einzelnen sind dies für die Priori-Information die Anzahl $\underline{n}$ der Datensätze, die Parameter $\underline{b}$ und $\underline{p}$ der Normal-Gammaverteilung sowie die Kofaktormatrix $\mathbf{Q}_{\beta\beta}$; für die Posteriori-Lösung entsprechend $\bar{n}$, $\bar{b}$ und $\bar{p}$ und $\mathbf{Q}_{\bar{\beta}\bar{\beta}}$. Die Berechnung der einzelnen Größen und die Modellwahl für die Epoche 2005 wird im folgenden Unterabschnitt behandelt, bevor im zweiten Unterabschnitt die Ergebnisse für die Modellwahl der Folgeepochen 2006 bis 2014 dargestellt werden.

5.2.1.1. Modellwahl der initialen Epoche

Der Modellansatz für die Schätzung des Gesamtmodells $M^{g,2005}$ basiert auf den Definitionen der Bayesschen Regression nach (3.30) und (3.31). Wie anhand der Definition des Bayes-Faktors nach (4.32) zu erkennen ist, sind für das Gesamtmodell zunächst zwei Lösungen zu bestimmen: Zum einen die nichtinformativ Lösung der Epoche 2004, die zur Generierung empirischen Vorwissens genutzt wird, zum anderen die informative Lösung der Epoche 2005 auf Grundlage der Priori-Information 2004. Ist das Gesamtmodell geschätzt, können im zweiten Schritt die zugehörigen Modellvarianten gebildet werden. Da sämtliche verfügbaren Einflussgrößen in den Modellansatz eingeführt wurden, werden ausschließlich Submodelle gebildet. Hypermodelle finden an dieser Stelle daher keine weitere Anwendung. Für die weiteren Berechnungen werden die Zielgröße und die Einflussgrößen mit den Einheiten nach Tabelle 5.9 eingeführt.

Tabelle 5.9.: Einheiten der Zielgröße und der Einflussgrößen in den praktischen Untersuchungen

Merkmal	Einheit
KPW	1.000 €/m ²
BRW	100 €/m ²
GRF	100 m ²
ALT	(ALT-1900/10)
WOF	100 m ²
KEL	10 ⁻² %

Wird zunächst die Lösung des Gesamtmodells betrachtet, ergeben sich die Schätzungen der nichtinformativen und informativen Lösung als iterative Prozesse, innerhalb derer die Schätzung der unbekannten Parameter vorgenommen und im gleichen Zuge eine Überprüfung auf verbliebene Ausreißer in den Daten vorgenommen wird. Für die Epoche 2004 vollzieht sich die Schätzung der unbekannten Parameter nach den Ausführungen in Abschnitt 3.2.2.2 (S. 63 f.) unter Annahme der Priori-Dichten (3.43) und (3.44). Die Epoche 2005 basiert auf den Schätzungen eines informativen Regressionsmodells nach den Ausführungen in Abschnitt 3.2.2.1 (S. 60 f.) mit der Priori-Information nach (3.35) als Ergebnis der Auswertung der Epoche 2004. Für die Aufdeckung von Ausreißern wird ein Wert nach Tabelle 4.4 gewählt: Da der Datensatz 2004 mit einer Anzahl vollständig belegter Kauffälle von $\underline{n} = 90$ sowie 2005 mit $\bar{n} = 96$ zur Verfügung steht, wird entsprechend der vertafelten Werte ein Faktor für die Ausreißergrenze von $k = 3, 4$ bei einer Priori-Wahrscheinlichkeit von 95 % genutzt.

Die Ergebnisse der nichtinformativen und informativen Schätzung werden in Tabelle 5.10 dargestellt: Wie zu erkennen ist, haben sich in dem Auswerteprozess keine weiteren Ausreißer in den jeweiligen Epochen ergeben, sodass die Schätzungen auf den vollständigen Datensätzen basieren. Ebenfalls zeigt sich eine ähnliche Tendenz in der numerischen Ausprägung und dem Vorzeichen der Trendkoeffizienten. Die auf Grundlage der Schätzungsergebnisse erhaltenen Kofaktormatrizen der einzelnen Epochen können Tabelle 5.11 entnommen werden. Mit den ermittelten Werten liegen die bezüglich des Gesamtmodells erforderlichen Werte für die Berechnung von Bayes-Faktoren vor und können entsprechend in Gleichung (4.32) eingesetzt werden.

Tabelle 5.10.: Ergebnisse für die Schätzung des Gesamtmodells M_g^{2005}

Einflussgröße	Koeffizient Trendmodell	Bayessche Regression (nichtinformativ) Epoche 2004	Bayessche Regression (informativ) Epoche 2005
Intercept	$\bar{\beta}_0$	1.649	1.518
BRW	$\bar{\beta}_1$	0.349	0.377
GRF	$\bar{\beta}_2$	0.017	0.039
ALT	$\bar{\beta}_3$	-0.103	-0.117
WOF	$\bar{\beta}_4$	-0.402	-0.417
KEL	$\bar{\beta}_5$	0.107	0.069
Anzahl Datensätze		90	96
Anzahl Ausreißer		0	0
Parameter $\bar{p}$		47	54
Parameter $\bar{b}$		1,856	2,177

Tabelle 5.11.: Übersicht über die Kofaktormatrizen des Gesamtmodells $M^{g,2005}$

$Q_{\beta\beta}^{2004} =$						$Q_{\beta\beta}^{2005} =$					
0,3493	-0,0484	-0,0106	-0,0063	-0,0840	-0,0446	0,1903	-0,0219	-0,0053	-0,0048	-0,0490	-0,0222
	0,0492	0,0018	-0,0044	-0,0173	-0,0069		0,0227	0,0000	-0,0023	-0,0066	0,0000
		0,0031	-0,0017	-0,0065	0,0000			0,0013	-0,0000	-0,0022	0,0000
			0,0075	0,0063	-0,0081				0,0042	0,0039	-0,0049
				0,0897	-0,0061					0,0408	-0,0016
					0,0885						0,0460

Für die Berechnung der äquivalenten Kenngrößen der Submodelle müssen zunächst nach dem Prinzip des Branch-and-Bound Algorithmus die Modellvarianten $M_1^{s,2005}$ bis $M_5^{s,2005}$ durch die Elimination jeweils einer Einflussgröße pro Submodell gebildet werden. Eine Übersicht über die Kombinationen im vorliegenden Fall gibt Tabelle 5.12.

Tabelle 5.12.: Übersicht über die Submodelle auf Basis des Gesamtmodells $M_g^{g,2005}$ (1. Iteration)

Modell	$M_1^{s,2005}$	$M_2^{s,2005}$	$M_3^{s,2005}$	$M_4^{s,2005}$	$M_5^{s,2005}$
Einflussgröße	$\bar{\beta}_0$	$\bar{\beta}_0$	$\bar{\beta}_0$	$\bar{\beta}_0$	$\bar{\beta}_0$
	Intercept	Intercept	Intercept	Intercept	Intercept
	$\bar{\beta}_1$	BRW	BRW	BRW	GRF
	$\bar{\beta}_2$	GRF	GRF	GRF	ALT
	$\bar{\beta}_3$	ALT	ALT	WOF	WOF
	$\bar{\beta}_4$	WOF	KEL	KEL	KEL

Analog zu den bisherigen Beschreibungen des Gesamtmodells werden die Schätzungen der Modellvarianten vorgenommen: Insbesondere sind dieselben Annahmen über die Verteilungsdichten sowohl der Priori- als auch der Posteriori-Lösungen gültig. Die Ergebnisse für die Schätzung der Submodelle lassen sich Tabelle 5.13 entnehmen; für eine Darstellung der zugehörigen Kofaktormatrizen wird auf Tabelle 5.14 verwiesen.

Mittels der durch die Schätzungen der Modelle erhaltenen Kenngrößen können durch Einsetzen in (4.33) die für die vorliegenden Daten und Modellansätze gültigen Bayes-Faktoren berechnet werden. Eine vergleichende Übersicht der Resultate kann Tabelle 5.15 entnommen werden. Insgesamt zeigt sich, dass in allen fünf Fällen das Gesamtmodell als das wahrscheinlichere eingestuft wird: Zwar weichen die Bayes-Faktoren a) bis e) numerisch voneinander ab, befinden sich jedoch deutlich unter dem Grenzwert von 10. Wie der Vergleich mit Tabelle 4.5 (S. 100) für die Interpretation der Bayes-Faktoren zeigt, ergibt sich damit eine negative Evidenz für die Submodelle $M_1^{s,2005}$ bis $M_5^{s,2005}$ – das Gesamtmodell wird im Vergleich zu den unterschiedlichen Kombinationen der Einflussgrößen präferiert und als wahrscheinlicher eingestuft. Im Umkehrschluss bedeutet dieses Testergebnis, dass die zugehörigen Submodelle als nicht wahrscheinliche Erklärungsansätze verworfen werden können.

Tabelle 5.13.: Ergebnisse für die Schätzung der Submodelle

Koeffizient	$M_1^{s,2005}$		$M_2^{s,2005}$		$M_3^{s,2005}$		$M_4^{s,2005}$		$M_5^{s,2005}$	
	2004	2005	2004	2005	2004	2005	2004	2005	2004	2005
$\hat{\beta}_0$	1,703	1,551	1,273	1,017	1,564	1,386	1,707	1,673	1,992	1,881
$\hat{\beta}_1$	0,340	0,377	0,271	0,310	0,289	0,313	0,339	0,355	0,004	0,027
$\hat{\beta}_2$	0,017	0,039	-0,012	0,016	-0,007	0,014	-0,093	-0,090	-0,072	-0,078
$\hat{\beta}_3$	-0,093	-0,109	-0,075	-0,077	-0,315	-0,311	-0,366	-0,349	-0,279	-0,309
$\hat{\beta}_4$	-0,394	-0,415	0,080	0,053	-0,003	-0,066	0,107	0,055	0,059	0,067
n	90	96	90	96	90	96	90	96	90	96
Ausreißer	0	0	0	0	0	0	0	0	0	0
$\bar{p}$	47	88,62	47	88,62	47	88,62	47	88,62	47	88,62
$\bar{b}$	1,921	3,772	2,755	5,816	2,559	5,292	1,902	4,295	3,092	6,791

Tabelle 5.14.: Übersicht über die Kofaktormatrizen der Submodelle $M_1^{s,2005}$ bis $M_5^{s,2005}$

$M_1^{s,2005}: Q_{\beta\beta}^{2004} =$					$M_1^{s,2005}: Q_{\beta\beta}^{2005} =$				
<u>0,3268</u>	-0,0449	-0,0106	-0,0103	-0,0871	<u>0,1796</u>	-0,0218	-0,0051	-0,0072	-0,0498
	<u>0,0486</u>	0,0018	-0,0038	-0,0168		<u>0,0227</u>	0,0000	-0,0023	-0,0066
		<u>0,0031</u>	-0,0017	-0,0065			<u>0,0013</u>	-0,0000	-0,0023
			<u>0,0068</u>	0,0058				<u>0,0037</u>	0,0037
				<u>0,0893</u>					<u>0,0407</u>
$M_2^{s,2005}: Q_{\beta\beta}^{2004} =$					$M_2^{s,2005}: Q_{\beta\beta}^{2005} =$				
<u>0,2705</u>	-0,0646	-0,0167	-0,0000	-0,0503	<u>0,1315</u>	-0,0298	-0,0081	-0,0000	-0,0241
	<u>0,0458</u>	0,0000	-0,0032	0,0057		<u>0,0217</u>	0,0000	-0,0017	-0,0000
		<u>0,0026</u>	-0,0013	-0,0000			<u>0,0012</u>	-0,0000	0,0000
			<u>0,0071</u>	-0,0076				<u>0,0039</u>	-0,0048
				<u>0,0881</u>					<u>0,0459</u>
$M_3^{s,2005}: Q_{\beta\beta}^{2004} =$					$M_3^{s,2005}: Q_{\beta\beta}^{2005} =$				
<u>0,3441</u>	-0,0520	-0,0121	-0,0788	-0,0513	<u>0,1849</u>	-0,0245	-0,0063	-0,0446	-0,0278
	<u>0,0466</u>	0,0000	-0,0136	0,0022		<u>0,0214</u>	0,0000	-0,0044	-0,0026
		<u>0,0027</u>	-0,0050	-0,0019			<u>0,0011</u>	-0,0015	-0,0000
			<u>0,0843</u>	0,0000				<u>0,0373</u>	0,0029
				<u>0,0798</u>					<u>0,0403</u>
$M_4^{s,2005}: Q_{\beta\beta}^{2004} =$					$M_4^{s,2005}: Q_{\beta\beta}^{2005} =$				
<u>0,3126</u>	-0,0422	-0,0123	-0,1064	-0,0446	<u>0,1694</u>	-0,0189	-0,0084	-0,0582	-0,0204
	<u>0,0481</u>	-0,0034	-0,0135	0,0069		<u>0,0223</u>	-0,0018	-0,0052	-0,0000
		<u>0,0065</u>	0,0027	-0,0081			<u>0,0036</u>	0,0023	-0,0046
			<u>0,0760</u>	-0,0061				<u>0,0368</u>	-0,0000
				<u>0,0885</u>					<u>0,0458</u>
$M_5^{s,2005}: Q_{\beta\beta}^{2004} =$					$M_5^{s,2005}: Q_{\beta\beta}^{2005} =$				
<u>0,3017</u>	-0,0088	-0,0106	-0,1010	-0,0379	<u>0,1693</u>	-0,0046	-0,0071	-0,0553	-0,0221
	<u>0,0030</u>	-0,0016	-0,0058	-0,0000		<u>0,0013</u>	-0,0000	-0,0021	-0,0000
		<u>0,0071</u>	0,0048	-0,0075			<u>0,0040</u>	0,0032	-0,0049
			<u>0,0836</u>	-0,0036				<u>0,0389</u>	-0,0016
				<u>0,0875</u>					<u>0,0460</u>

Tabelle 5.15.: Bayes-Faktoren für den Vergleich der Submodelle $M_k^{s,2005}$ mit dem Gesamtmodell $M^{g,2005}$

	Vergleichs- modelle	Bayes- faktor	Evidenz für Submodell $M_k^{s,2005}$
a)	$M_1^{s,2005} : M^{g,2005}$	$BF_{M_1^s M^g}$ 0,8141	negativ
b)	$M_2^{s,2005} : M^{g,2005}$	$BF_{M_2^s M^g}$ $1,9 \cdot 10^{-5}$	negativ
c)	$M_3^{s,2005} : M^{g,2005}$	$BF_{M_3^s M^g}$ $2,0 \cdot 10^{-4}$	negativ
d)	$M_4^{s,2005} : M^{g,2005}$	$BF_{M_4^s M^g}$ 0,0022	negativ
e)	$M_5^{s,2005} : M^{g,2005}$	$BF_{M_5^s M^g}$ $3,0 \cdot 10^{-7}$	negativ

Als Ergebnis der Auswertung der initialen Epoche 2005 steht ein aufbereiteter Datensatz zur Verfügung, für den in einem iterativen Prozess das wahrscheinlichste Modell ermittelt wurde. Im Einzelnen sind unter Angabe des Modellrahmens, der Modellbeziehungen und des Funktionstyps gemäß der Ausführungen in Abschnitt 2.2.1 (S. 31 ff.) die unbekannten Parameter in Form der Trendkoeffizienten sowie des Varianzfaktors mittels der Bayesschen Regression geschätzt. Auf Grundlage der bisherigen Ergebnisse werden im folgenden Unterabschnitt die Resultate für die Modellwahl innerhalb des rekursiven Ablaufs der Folgeepochen dargestellt.

5.2.1.2. Modellwahl der Folgeepochen

Für die Modellwahl der Folgeepochen ergibt sich ein analoges Vorgehen zu der gezeigten exemplarischen Vorgehensweise für die Epoche 2005. Im Einzelnen sind die Stichprobenbildung für die Epochen 2006 bis 2014, die Stichprobenaufbereitung der Ziel- und Einflussgrößen sowie die Prüfung von Korrelationen zwischen den Einflussgrößen vorzunehmen. Während sich die Kenndaten der aufbereiteten Stichproben in Tabelle 5.16 finden, sind die zugehörigen deskriptiven Statistiken sowie die Ergebnisse der Korrelationsberechnungen in den Anhängen A.1 und A.2 (S. 205 ff.) aufgeführt.

Wird zunächst die Stichprobenbildung betrachtet, sind die Kauffälle auf ihre vollständige Belegung zu prüfen. Explizit betrachtet werden dieselben Einflussgrößen, die als Ausgangspunkt der Stichprobenezusammenstellung der initialen Epoche dienen: Insbesondere schließt dies die Einflussgröße des Ausstattungsstandards mit ein, die aufgrund ihrer ungleichmäßigen Belegung entfernt wurde. Die ungünstige Verteilung erstreckt sich über sämtliche betrachteten Epochen und deutet auf eine generelle Schwachstelle für statistische Auswertungen in der Erfassung der AKS hin. Für die weiteren Auswertungen wird das Merkmal AST daher wie zuvor ausgeschlossen. Für die übrigen, vollständig belegten Kauffälle zeigt Tabelle 5.16 in Zeile 1 die Anzahl der verfügbaren Datensätze pro Epoche. Es zeigt sich, dass die Anzahl relativ homogen über die Epochen verteilt ist. Das Maximum findet sich mit 134 Kauffällen im Jahr 2012, das Minimum mit 75 in 2014. Auswirkungen durch die unterschiedliche Belegung pro Epoche werden zunächst nicht erwartet.

Nachdem die reine Stichprobenbildung abgeschlossen ist, wird in einem weiteren Schritt die Zielgröße aufbereitet. Werden die Auswerteschritte analog nach dem Beispiel in Abschnitt 5.1.3.2 (S. 116 ff.) angewendet, ergeben sich die in Zeile 2, 3 und 4 der Tabelle 5.16 dargestellten Ergebnisse. Wie zu erkennen ist, werden durch den (teilweise iterativen) Auswerteprozess die Zielgrößen der Einzelepochen so aufbereitet, dass sie den gestellten Anforderungen des Verhältnisses $y_{min} : y_{max}$, der 2,5 σ -Regel sowie der Schiefe der Verteilung gerecht werden. Nachdem die Zielgröße in den einzelnen Epochen aufbereitet ist, werden die Einflussgrößen auf grobe Ausreißer untersucht. Anhand der gewählten 4 σ -Regel lassen sich die in Zeile 5 der Tabelle 5.16 aufgeführten Ausreißer ermitteln und aus dem Datensatz entfernen. Insgesamt reduziert sich der Ausgangsdatsatz durch die iterative Stichprobenaufbereitung der Ziel- und Einflussgrößen von den in Zeile 1 genannten Werten auf die in Zeile 6 dargestellten Summen. Die verbliebenen Datensätze können im Folgenden für die rekursive Modellwahl genutzt werden.

Die rekursive Modellwahl basiert auf der fortlaufenden Schätzung informativer Bayesscher Regressionen über die Epochen 2006 bis 2014. Unter Nutzung der Posteriori-Lösung der Vorgängerepoche als Priori-Information der

Tabelle 5.16.: Stichprobenaufbereitung der Epochen 2006 bis 2014

Zeile	Epoche	2006	2007	2008	2009	2010	2011	2012	2013	2014
1	Anzahl (belegt)	84	95	94	105	120	114	134	129	75
	<i>Aufbereitung der Zielgröße</i>									
2	Ausreißerbereinigung (Zielgröße)	2	5	3	8	10	6	9	10	11
3	Verhältnis (Minimum:Maximum)	4,38	3,44	3,73	3,35	2,49	3,83	3,00	2,91	2,89
4	Schiefe v_3 (Zielgröße)	0,22	0,23	0,29	0,20	0,04	0,23	0,26	0,27	0,30
	<i>Aufbereitung der Einflussgrößen</i>									
5	Ausreißerbereinigung (Einflussgrößen)	0	0	0	0	2	3	0	1	0
6	$\sum$ Datensätze	82	90	91	97	108	105	125	118	64

aktuellen Auswerteepoche werden die so als jeweilige Gesamtmodelle M^g betrachteten Lösungen nach dem in Abschnitt 5.2.1.1 (S. 120 ff.) vorgestellten Prinzip ihren Submodellen gegenübergestellt. Die Submodelle werden zu diesem Zweck eigens berechnet und beruhen auf der empirisch generierten Priori-Information einer nichtinformativen Bayesschen Regression der Vorgängerepoche – analog zu den Ausführungen über die Initialepoche 2005 mit der Vorinformation der Epoche 2004. Die Annahme eines stabilen Modells über den Gesamtzeitraum der AuswerteePOCHEN wird bestätigt durch die numerischen Ergebnisse: Die Testergebnisse können Tabelle 5.17 entnommen werden. Wie zu erkennen ist, ergeben sich innerhalb jeder Epoche negative Evidenzen für die gebildeten Submodelle.

Nachdem die Modellwahl abgeschlossen ist, können die endgültigen Schätzwerte für die Initialepoche sowie der Folgeepochen zusammengestellt und die zusätzlichen Modelle der nichtinformativen Bayesschen Regression und der informativen sowie nichtinformativen Bayesschen Kollokation berechnet werden. Im folgenden Abschnitt werden die Schätzergebnisse dargestellt.

Tabelle 5.17.: Bayes-Faktoren der Epochen 2006 bis 2014 unter rekursiver Auswertung (Submodelle)

Epoche	2005	2006	2007	2008	2009
$BF_{M_1^s M^g}$	0,8141	0,6176	0,6144	0,4477	0,9401
$BF_{M_2^s M^g}$	$1,9 \cdot 10^{-5}$	$6,9 \cdot 10^{-10}$	$4,2 \cdot 10^{-4}$	$5,3 \cdot 10^{-4}$	0,0011
$BF_{M_3^s M^g}$	$2,0 \cdot 10^{-4}$	$1,5 \cdot 10^{-5}$	$9,4 \cdot 10^{-4}$	0,0044	0,0056
$BF_{M_4^s M^g}$	0,0022	$2,4 \cdot 10^{-5}$	0,0592	0,0136	0,4236
$BF_{M_5^s M^g}$	$3,0 \cdot 10^{-7}$	$4,4 \cdot 10^{-9}$	$1,9 \cdot 10^{-4}$	0,0047	$1,8 \cdot 10^{-4}$
Epoche	2010	2011	2012	2013	2014
$BF_{M_1^s M^g}$	0,5604	0,5228	1,1381	1,0205	0,4864
$BF_{M_2^s M^g}$	0,0014	$1,2 \cdot 10^{-4}$	$1,9 \cdot 10^{-5}$	0,0021	0,0156
$BF_{M_3^s M^g}$	0,0010	$2,9 \cdot 10^{-6}$	$2,4 \cdot 10^{-7}$	$9,6 \cdot 10^{-7}$	0,6735
$BF_{M_4^s M^g}$	0,0285	0,0013	0,3227	0,0364	0,1595
$BF_{M_5^s M^g}$	$9,6 \cdot 10^{-4}$	$4,3 \cdot 10^{-7}$	$2,9 \cdot 10^{-4}$	$2,2 \cdot 10^{-6}$	$6,2 \cdot 10^{-4}$

5.2.2. Schätzergebnisse der Bayesschen Regression

Durch die Modellwahl mittels Bayesscher Hypothesentests über die Auswertepochen 2005 bis 2014 wird die abschließende Schätzung der Bayesschen Regressionsmodelle ermöglicht. Neben der Schätzung der informativen Modelle wird für die Untersuchung des Einflusses von integriertem Vorwissen für jede Epoche eine nichtinformativ Bayessche Regression berechnet. Für den direkten Vergleich bezieht sich dieses dabei auf denselben Modellrahmen und Funktionstyp sowie dieselben Modellbeziehungen der informativen Lösung; lediglich die Verteilungsannahmen werden entsprechend der Ausführungen zur nichtinformativen Bayesschen Regression getroffen. In den folgenden Unterabschnitten werden die Ergebnisse der Schätzung der Modellparameter sowie der Posteriori-Schätzung der ausgeglichenen Zielgröße dargestellt.

5.2.2.1. Schätzung der Modellparameter

Für den gewählten räumlichen und sachlichen Teilmarkt der Ein- und Zweifamilienhäuser in der Wohnungsmarktregion Osnabrück lassen sich durch Anwendung der vorgestellten Auswertestrategie und unter Annahme des für die Epoche 2005 ermittelten Gesamtmodells die unbekannten Parameter für die Epochen 2005 bis 2014 als Posteriori-Schätzwerte angeben. Im Einzelnen sind dies die Trendkoeffizienten $\bar{\beta}$ sowie der Varianzfaktor $\bar{\sigma}^2$ und die daraus abgeleitete Standardabweichung $\bar{\sigma}$ für die jeweils informative und nichtinformativ Lösung. Die Ergebnisse der Schätzung werden in Tabelle 5.18 gegenübergestellt. In den Zeilen 1 bis 6 werden die Trendkoeffizienten der jeweiligen Modelle aufgeführt, getrennt nach den jeweiligen informativen und nichtinformativen Lösungen der Auswertepochen. Ergänzend werden in den Zeilen 7 bis 8 die Varianzinformationen angegeben.

Tabelle 5.18.: Modellparameter der Epochen 2005 bis 2014, basierend auf Bayesscher Regression

	Epoche	2005		2006		2007		2008		2009	
Zeile	Parameter	R.1	R.2	R.1	R.2	R.1	R.2	R.1	R.2	R.1	R.2
1	$\bar{\beta}_0$	1,518	1,365	1,489	1,379	1,477	1,380	1,467	1,385	1,455	1,353
2	$\bar{\beta}_1$	0,377	0,413	0,411	0,495	0,399	0,366	0,371	0,305	0,373	0,375
3	$\bar{\beta}_2$	0,039	0,057	0,044	0,048	0,043	0,044	0,049	0,078	0,049	0,038
4	$\bar{\beta}_3$	-0,117	-0,128	-0,113	-0,097	-0,117	-0,125	-0,122	-0,143	-0,123	-0,104
5	$\bar{\beta}_4$	-0,417	-0,414	-0,460	-0,524	-0,444	-0,369	-0,440	-0,440	-0,429	-0,345
6	$\bar{\beta}_5$	0,069	0,038	0,067	0,054	0,065	0,026	0,074	0,113	0,061	-0,034
7 *	$\bar{\sigma}^2$	0,043	0,040	0,050	0,067	0,052	0,055	0,056	0,069	0,056	0,057
8 *	$\bar{\sigma}$	0,207	0,199	0,224	0,260	0,229	0,235	0,237	0,262	0,237	0,239

	Epoche	2010		2011		2012		2013		2014	
Zeile	Parameter	R.1	R.2	R.1	R.2	R.1	R.2	R.1	R.2	R.1	R.2
1	$\bar{\beta}_0$	1,428	1,297	1,427	1,376	1,457	1,668	1,485	1,534	1,481	1,336
2	$\bar{\beta}_1$	0,369	0,290	0,377	0,447	0,372	0,335	0,373	0,476	0,377	0,540
3	$\bar{\beta}_2$	0,046	0,034	0,048	0,064	0,045	0,028	0,043	0,046	0,043	0,046
4	$\bar{\beta}_3$	-0,119	-0,087	-0,121	-0,137	-0,123	-0,142	-0,126	-0,165	-0,123	-0,077
5	$\bar{\beta}_4$	-0,403	-0,249	-0,403	-0,408	-0,399	-0,376	-0,392	-0,359	-0,388	-0,399
6	$\bar{\beta}_5$	0,046	-0,055	0,028	-0,081	0,037	0,082	0,025	0,013	0,019	-0,187
7 *	$\bar{\sigma}^2$	0,054	0,041	0,053	0,041	0,051	0,041	0,053	0,053	0,054	0,061
8 *	$\bar{\sigma}$	0,233	0,202	0,229	0,203	0,226	0,203	0,231	0,231	0,233	0,247

* $\bar{\sigma}^2$: Angaben in $[1.000 \text{ €/m}^2]^2$, $\bar{\sigma}$: Angaben in $[1.000 \text{ €/m}^2]$

Werden zunächst die Posteriori-Trendkoeffizienten betrachtet, zeigen sich im Vergleich $R.1$ zu $R.2$ der informativen und nichtinformativen Lösungen pro Auswertepoche numerische Unterschiede, die auf die Auswirkung der Integration von Priori-Wissen zurückzuführen sind. In beiden Modellansätzen sind jedoch ähnliche Entwicklungstendenzen festzuhalten: Sowohl in der Höhe der numerischen Ausprägungen der Trendkoeffizienten als auch im Vorzeichen ergibt sich über den gesamten Auswertzeitraum ein konstantes Bild. Dies deutet in der Entwicklung

der nichtinformativen Lösungen als von der Rekursion unabhängige Berechnung auf eine gleichgerichtete Entwicklung in dem gewählten räumlichen und sachlichen Teilmarkt hin. In beiden Modellansätzen ist zudem festzustellen, dass die Trendkoeffizienten in einem vergleichbaren Bereich liegen.

Einzige Ausnahme im Einfluss auf das Schätzergebnis ist das Merkmal der Unterkellerung (Zeile 6, $\bar{\beta}_5$), bei dem sich in der nichtinformativen Lösung *R.2* für die Epochen 2005 bis 2008 sowie 2012 und 2013 ein positives Vorzeichen ergibt, in den Epochen 2009 bis 2011 sowie 2014 hingegen ein negativer Einfluss zu verzeichnen ist. Durch die rekursive Fortpflanzung der Priori-Information ergibt sich im Fall der informativen Lösung *R.1* hingegen kein Vorzeichenwechsel, die Einflussgröße weist einen konstant positiven Einfluss auf. Aus sachlogischen Gründen ist für eine Unterkellerung ein positiver Werteinfluss zu erwarten, sodass die Plausibilität der nichtinformativen Lösung an dieser Stelle in Frage zu stellen ist. Da sich die Bayesschen Hypothesentests, wie vorstehend beschrieben, ausschließlich auf die informative Lösung stützen, ist diese als plausible Schätzung einzustufen. Es ist zu vermuten, dass im Rahmen der hier nicht thematisierten Hypothesentests basierend auf nichtinformativen Priori-Dichten ein entsprechendes Submodell ohne die Einflussgröße der Unterkellerung zu wählen wäre. Für die übrigen Einflussgrößen können die Einflüsse, die durch die Vorzeichen repräsentiert werden, als konstant über die Auswertepochen beurteilt werden.

Bezüglich der numerischen Ausprägungen ist festzuhalten, dass diese im Fall der nichtinformativen Modelle stärker variieren als in den informativen Modellen. An dieser Stelle wird der Einfluss der rekursiven Anwendung deutlich: Durch die Fortpflanzung von Posteriori-Ergebnissen als Priori-Information der folgenden Auswertepochen wird im informativen Fall eine Konstanz im Sinne der zeitlichen Entwicklung numerischer Ausprägungen erreicht. Für die nichtinformativ Lösung ist eine stärkere Schwankung der Trendkoeffizienten über die einzelnen Auswertepochen zu verzeichnen. Da hier die Posteriori-Schätzung ausschließlich auf dem aktuellen Datensatz der Auswertepoche beruht, reagiert die Schätzung der Trendkoeffizienten erwartungsgemäß stärker auf veränderte Datensätze als bei der informativen Lösung, bei der die Ergebnisse der vorherigen Auswertepochen einfließen und so Konstanz durch Fortpflanzung erzeugen. Die Auswirkungen der Variationen der Trendkoeffizienten auf die Posteriori-Schätzung der Zielgröße werden im folgenden Unterabschnitt untersucht.

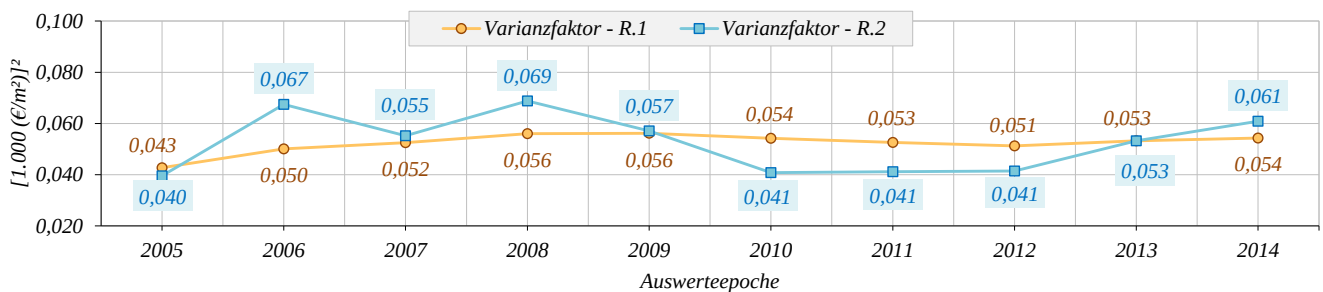


Abbildung 5.6: Posteriori-Schätzung des Varianzfaktors σ^2 der Epochen 2005 bis 2014, Bayessche Regression

Werden die Varianzinformationen der Auswertepochen nach Tabelle 5.18 (Zeilen 7 und 8) betrachtet, zeigt sich eine heterogene Entwicklung. Wie in Abbildung 5.6 dargestellt wird, ergibt sich für die Modelle der nichtinformativen Lösung *R.2* keine konstante Entwicklung der Varianzfaktoren über die Auswertepochen, für die Modelle *R.1* hingegen ist eine gleichmäßigere Schätzung zu erkennen. Wie auch im Fall der geschätzten Trendkoeffizienten kann demnach festgehalten werden, dass die Schätzungen im informativen Fall aufgrund der rekursiven Fortpflanzung in den Epochen vergleichbar sind. Im Mittel beträgt die Varianz $\bar{\sigma}_{R.1}^2 = 0,052$ bei einem minimalen Wert von $\bar{\sigma}_{R.1, \min}^2 = 0,043$ und einem maximalen Wert von $\bar{\sigma}_{R.1, \max}^2 = 0,056$ (jeweils in der Einheit $[1.000 \text{ €/m}^2]^2$). Die Varianz der nichtinformativen Lösung schwankt erwartungsgemäß stärker: Diese liegt zwar im Mittel mit $\bar{\sigma}_{R.2}^2 = 0,053$ ebenfalls im Bereich der informativen Lösung, der minimale Wert liegt jedoch bei $\bar{\sigma}_{R.2, \min}^2 = 0,040$, der maximale bei $\bar{\sigma}_{R.2, \max}^2 = 0,069$ (jeweils $[1.000 \text{ €/m}^2]^2$). Wiederum wird an dieser Stelle die Sensibilität gegenüber aktuellen Datensätzen ohne den Einfluss von Priori-Informationen deutlich. Wie im Vergleich der Epochen 2005 sowie 2010 bis 2012 zu erkennen ist, liegen hier die Schätzungen des Varianzfaktors auf Basis der nichtinformativen Lösung unterhalb der der informativen Lösung. Begründet werden kann dies über die Berechnungsformeln der Varianz nach (3.54) und (3.69), in denen insbesondere der Einfluss der Priori-Information deutlich wird. Die weitere Diskussi-

on erfolgt im Zusammenhang mit der Darstellung der Modellunsicherheiten sowie der Wertung der Ergebnisse in Kapitel 6. Im folgenden Unterabschnitt werden die Ergebnisse der Posteriori-Schätzung der Zielgröße dargestellt.

5.2.2.2. Schätzung der Zielgröße

Auf Basis der Modellschätzungen nach Tabelle 5.18 lassen sich ausgeglichene Schätzwerte für die Zielgröße ermitteln, die insbesondere eine Beurteilung der funktionalen Approximationsgüte sowie deren Unterschiede durch die Verwendung eines informativen oder nichtinformativen Ansatzes erlauben. Werden die Trendfunktionen für die Schätzung der ausgeglichenen Zielgröße durch die Beziehungen

$$\bar{\mathbf{y}}_{R.1} = \mathbf{X}\bar{\boldsymbol{\beta}}_{R.1}, \quad \bar{\mathbf{y}}_{R.2} = \mathbf{X}\bar{\boldsymbol{\beta}}_{R.2} \quad (5.1)$$

genutzt, lassen sich die Posteriori-Schätzwerte für die Zielgrößen ermitteln. Zusätzlich zur Schätzung der Zielgröße können die Residuen ermittelt werden durch

$$\bar{\mathbf{e}}_{R.1} = \mathbf{y} - \bar{\mathbf{y}}_{R.1}, \quad \bar{\mathbf{e}}_{R.2} = \mathbf{y} - \bar{\mathbf{y}}_{R.2}. \quad (5.2)$$

Die Ergebnisse der Posteriori-Schätzungen auf Grundlage der informativen und nichtinformativen Lösungen lassen sich den Abbildungen 5.7 bis 5.9 entnehmen. Für die Darstellung werden exemplarisch drei Epochen gewählt: Die initiale Epoche *2005*, die mittlere Epoche *2009* sowie die abschließende Epoche *2014*. Aufgrund des Umfangs wird für die restlichen Abbildungen der übrigen ausgewerteten Epochen auf Anhang B.1 (S. 209 ff.) verwiesen. Aus Gründen der Übersichtlichkeit sind die Kaufpreise nach aufsteigender Größe sortiert, entsprechend zugeordnet werden die jeweiligen Posteriori-Schätzwerte der Zielgröße.

Die oberen beiden Grafiken zeigen die Ergebnisse der informativen Lösung (links) und der nichtinformativen Lösung (rechts). Unterhalb der Ergebnisgrafiken sind die Differenzen der Spannen der HPDIs zwischen den Ergebnissen der Modelle *R.2* zu *R.1* angegeben, die im nachfolgenden Abschnitt beschrieben werden. Wird zunächst die rein funktionale Modellierung in den Mittelpunkt gerückt, lässt sich anhand einer ersten Beurteilung der Grafiken erkennen, dass die Ergebnisse der informativen Modelle *R.1* zu denen der nichtinformativen Modelle *R.2* vergleichbar sind; es ergeben sich keine auffälligen Unterschiede in der Streuung der Schätzwerte. Dieser Eindruck wird zudem durch die Auswertungen der übrigen Epochen bestätigt. In sämtlichen Auswertungen wird jedoch eine Systematik der Schätzergebnisse deutlich: Während im tiefpreisigen Segment tendenziell die Schätzwerte über den Kaufpreisen liegen, ist im hochpreisigen Segment eine Unterschätzung zu beobachten. Im mittleren Bereich streuen die Schätzwerte hingegen unregelmäßig um das Wertniveau der betrachteten Objekte. Entsprechend gestaltet sich die Auswirkung auf die prozentualen Abweichungen zwischen Kaufpreisen und Vergleichswerten, die für die Epochen *2005*, *2009* und *2014* den Abbildungen 5.10 bis 5.12 entnommen werden können. Die Darstellung der übrigen Epochen findet sich in Anhang B.2 (S. 213 f.).

Insbesondere im unteren Drittel der Kaufpreisspannen ergeben sich hohe Differenzen. Die größte Abweichung der informativen Lösung *R.1* liegt in der Epoche *2005* bei circa (ca.) 56,5 %, *2009* bei 71,3 % und *2014* bei 47,8 %. Sämtliche Werte ergeben sich dabei durch Schätzungen im niedrigpreisigen Segment. Minimale Werte für die informative Lösung sind hingegen im mittleren Drittel der Kaufpreisspanne festzustellen und liegen jeweils bei unter einem Prozent (*2005*: 0,3 %, *2009*: 0,1 %, *2014*: 0,4 %). Es zeigt sich so eine hohe Übereinstimmung zwischen Kaufpreisen und Vergleichswerten. Ähnliche Ergebnisse können für die nichtinformativ Lösung *R.2* festgehalten werden. Als maximale Abweichungen ergeben sich für die Epochen *2005* 52,4 %, *2009* 68,9 % sowie *2014* 52,2 %, als minimale Abweichungen *2005* 0 %, *2009* 0 % sowie *2014* 0,1 %.

Wie anhand der Medianwerte zu erkennen ist, weisen die Modelle insgesamt eine sehr geringe durchschnittliche Abweichung auf. Für die hier betrachteten repräsentativen Epochen liegen die Medianwerte im Vergleich informativer zu nichtinformativer Lösung bei 9,9 % zu 9,4 % (*2005*), 13,5 % zu 14,6 % (*2009*) und 13,2 % zu 11,6 % (*2014*), was generell für eine gute Näherung in der Vergleichswertermittlung spricht. Wie zu erkennen ist, gleichen sich die Ergebnisse innerhalb der Auswertungen einer einzelnen Epochen. Unterschiede sind lediglich von Epoche zu Epoche festzustellen. Werden die Auswertungen der übrigen Epochen betrachtet (vgl. Anhang B.2, S. 213 f.), ergibt sich im Schnitt in sämtlichen Auswertezeiträumen ein ähnliches Bild für die Ergebnisdarstellung. Wie die

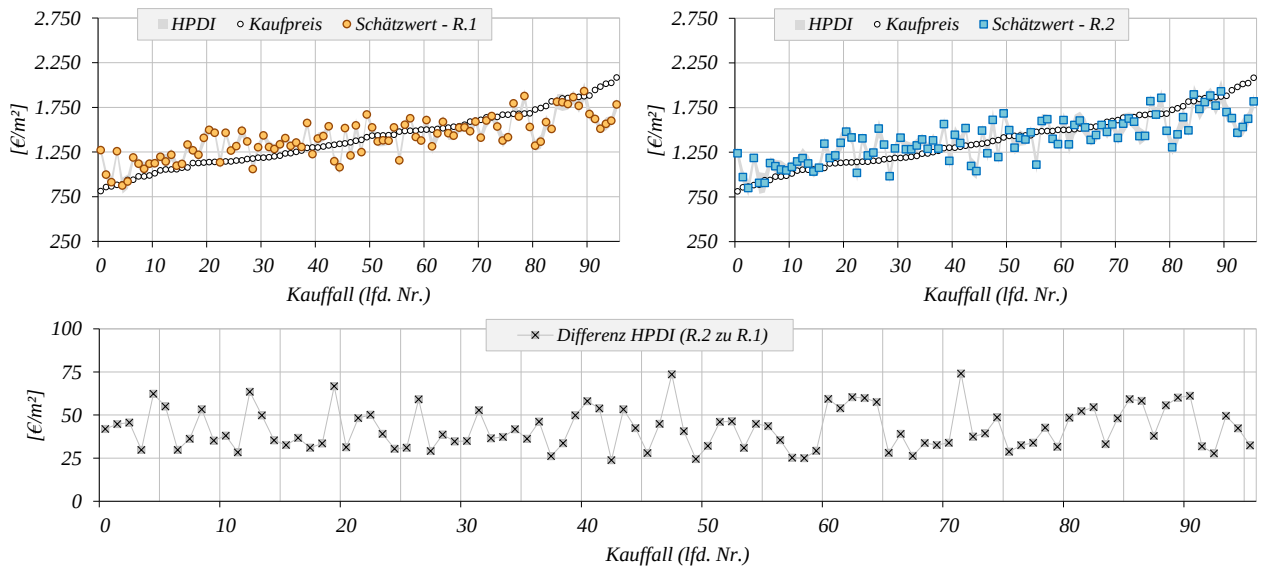


Abbildung 5.7: Posteriori-Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

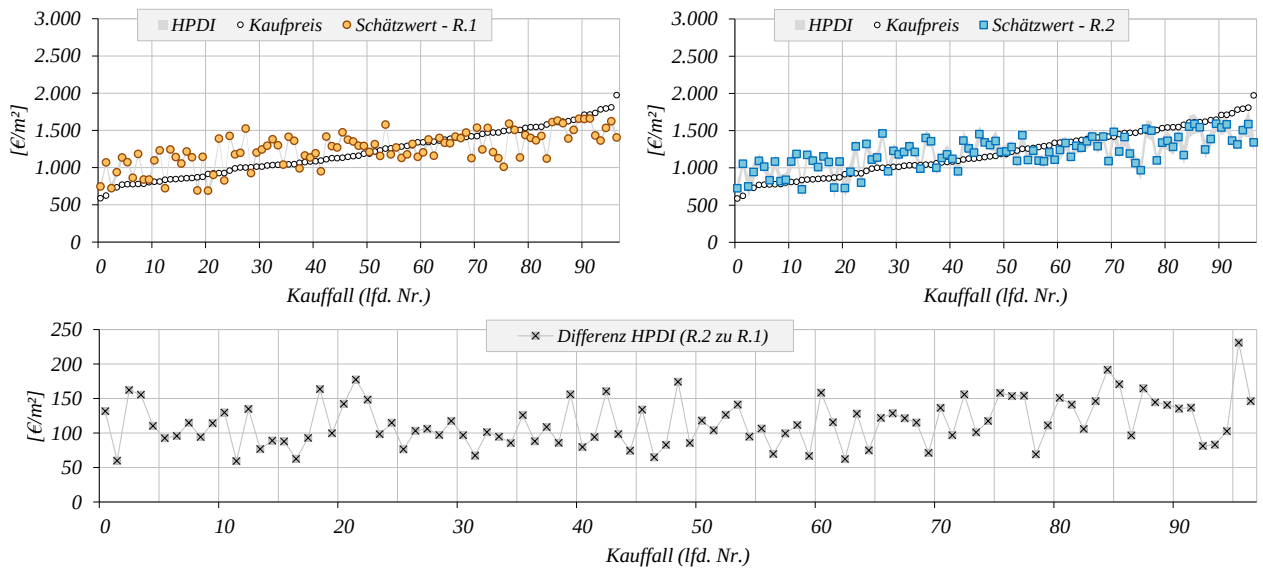


Abbildung 5.8: Posteriori-Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

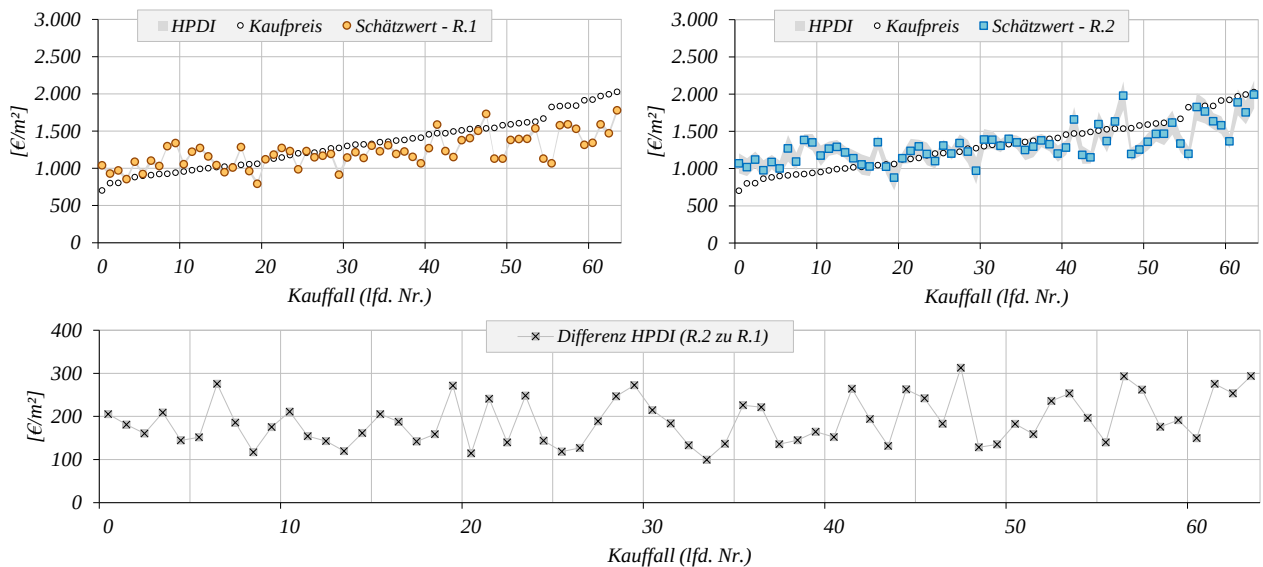


Abbildung 5.9: Posteriori-Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

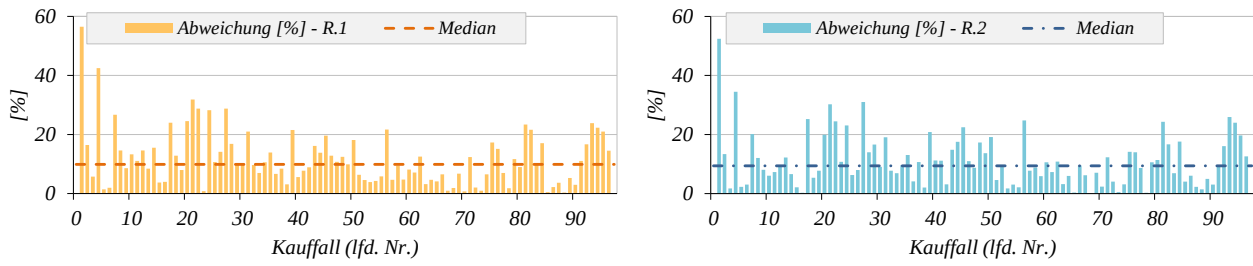


Abbildung 5.10: Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2005

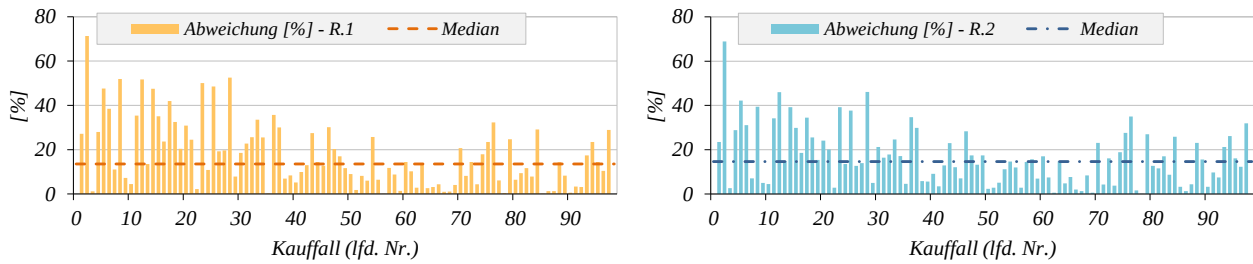


Abbildung 5.11: Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2009

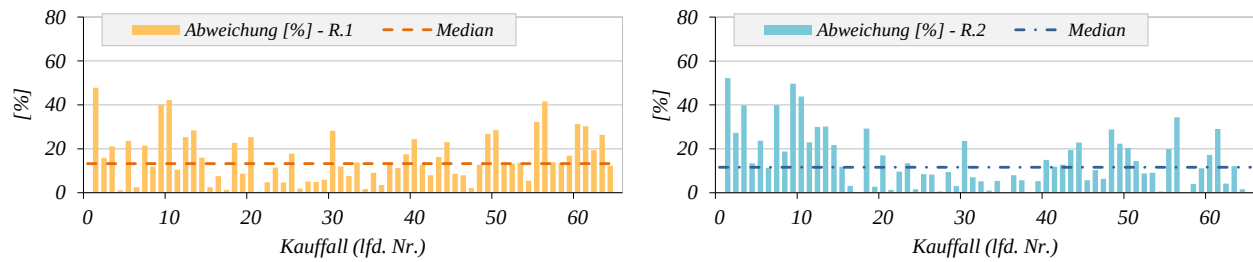


Abbildung 5.12: Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2014

Übersicht in Abbildung 5.13 verdeutlicht, zeigen sich sowohl im Fall der informativen wie auch der nichtinformativen Lösung konstante Entwicklungen der prozentualen Abweichungen (gemessen am Medianwert). Während die Epochen 2006 bis 2008 sowie 2013 bis 2014 tendenziell höhere Abweichungen aufweisen, sind die jeweiligen Tiefpunkte in den Epochen 2005, 2010 und 2011 zu erkennen. Dabei liegen die Medianwerte der nichtinformativen Lösung bei 6 von 10 Auswerteziträumen unterhalb der entsprechenden Werte der informativen Lösung. Gleichwohl ist zu berücksichtigen, dass die prozentualen Abweichungen zwischen den Modellansätzen teilweise lediglich in den Nachkommastellen variieren.

Die Diskussion der funktionalen Modellierung und der erreichten Approximationsgüte in der Bayesschen Regression erfolgt in Kapitel 6. In Ergänzung zu den Ergebnissen der funktionalen Modellierung der Bayesschen Regression werden im folgenden Abschnitt die Ergebnisse für die Ermittlung der Unsicherheiten behandelt.

5.2.3. Unsicherheitsmodellierung mittels Bayesscher Regression

Im folgenden Unterabschnitt werden zunächst die Unsicherheiten der Modellparameter in Form der Trendkoeffizienten $\bar{\beta}$ sowie des Varianzfaktors $\bar{\sigma}^2$ behandelt, bevor im zweiten Unterabschnitt ein Überblick über die HPDIs der Posteriori-Schätzwerte der Zielgröße gegeben wird.

5.2.3.1. Unsicherheit der Modellparameter

Für die Schätzung der Trendkoeffizienten wurden im vorstehenden Abschnitt bereits die numerischen Ausprägungen sowie deren Variationen über den gesamten Zeitraum der Auswertepochen gezeigt. Für die informative Bayessche Regression lassen sich ergänzend die HPDIs angeben. Wird eine Irrtumswahrscheinlichkeit von $\alpha = 5\%$ angesetzt,

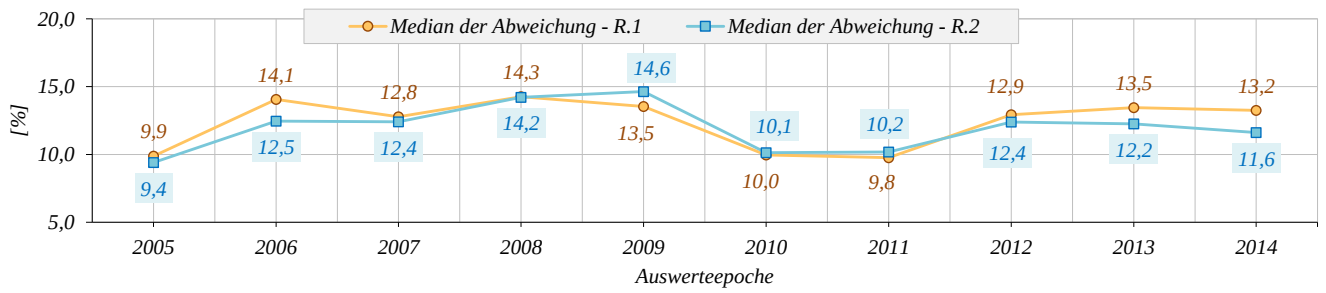


Abbildung 5.13: Medianwerte der Abweichungen zwischen Kaufpreisen und Posteriori-Schätzwerten $\bar{y}$ der Epochen 2005 bis 2014, Bayessche Regression

kann (3.78) für die Berechnung gelöst werden; eine grafische Übersicht über die Resultate kann den Abbildungen 5.14 bis 5.19 entnommen werden.

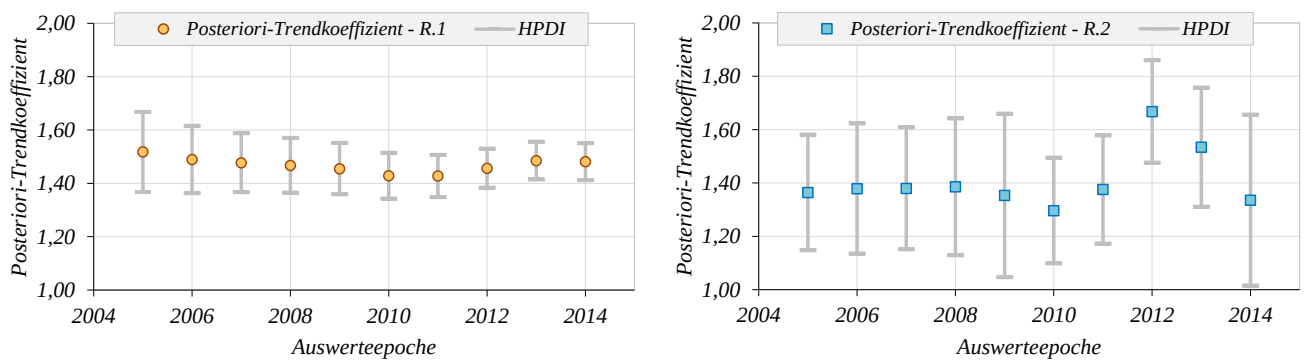


Abbildung 5.14: HPDI des Posteriori-Trendkoeffizienten $\hat{\beta}_0$ der Epochen 2005 bis 2014, Bayessche Regression

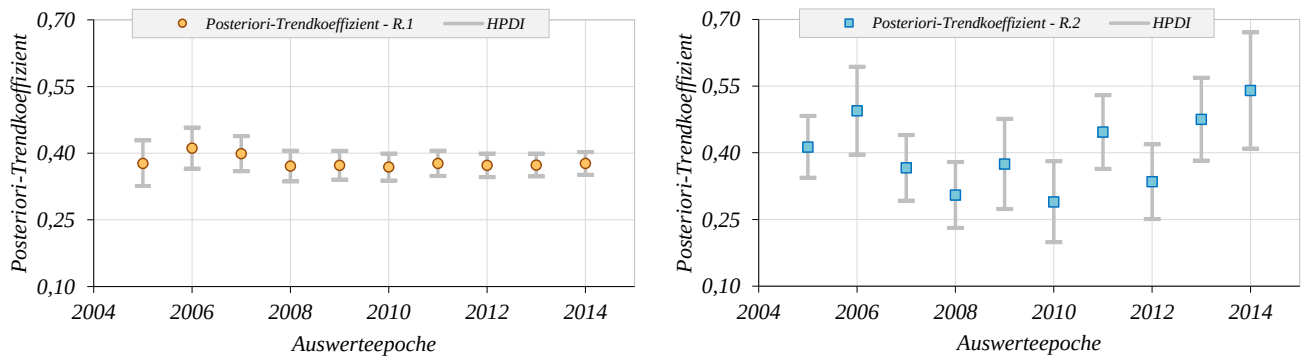


Abbildung 5.15: HPDI des Posteriori-Trendkoeffizienten $\hat{\beta}_1$ der Epochen 2005 bis 2014, Bayessche Regression



Abbildung 5.16: HPDI des Posteriori-Trendkoeffizienten $\hat{\beta}_2$ der Epochen 2005 bis 2014, Bayessche Regression

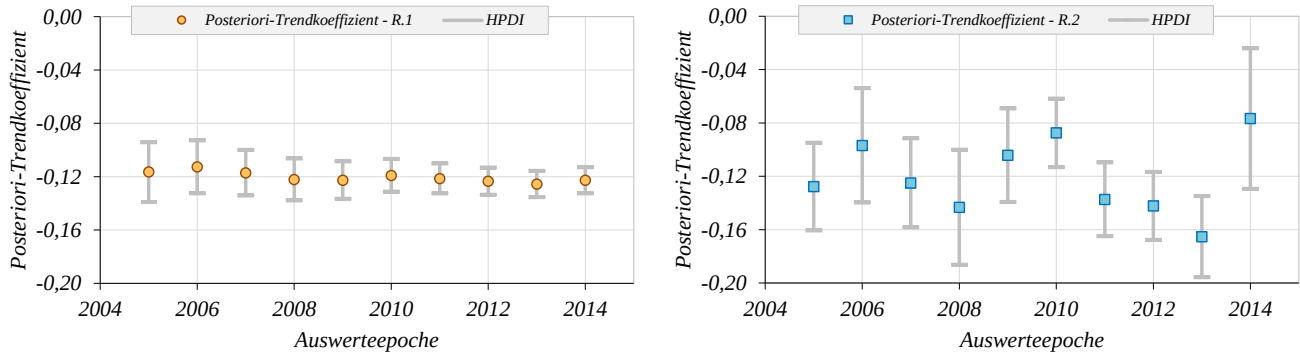


Abbildung 5.17: HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_3$ der Epochen 2005 bis 2014, Bayessche Regression

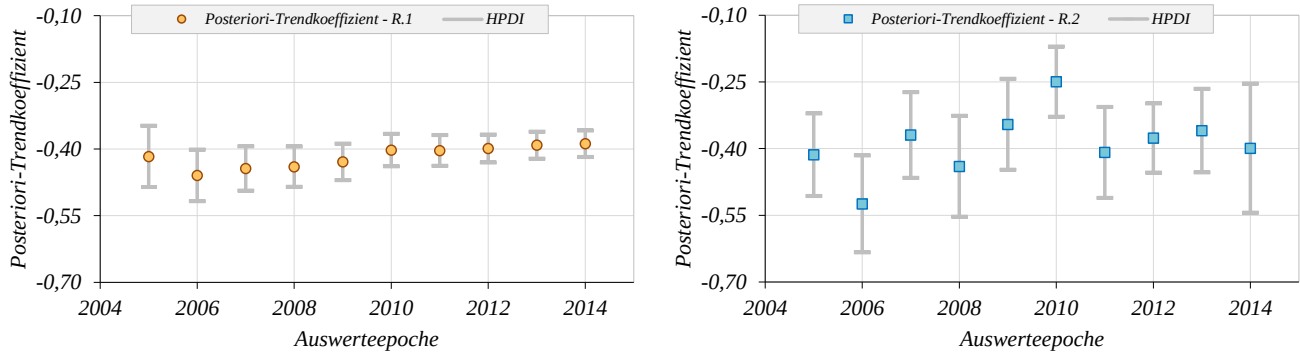


Abbildung 5.18: HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_4$ der Epochen 2005 bis 2014, Bayessche Regression

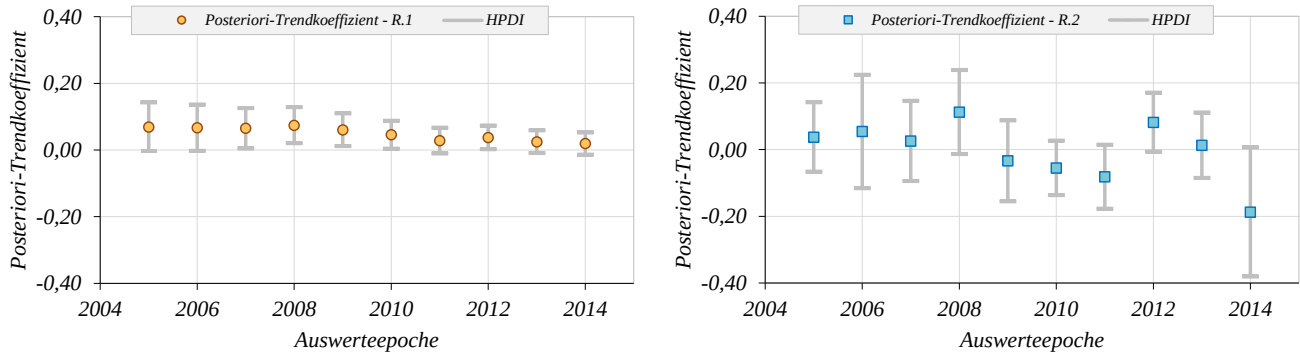


Abbildung 5.19: HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_5$ der Epochen 2005 bis 2014, Bayessche Regression

Dargestellt werden in den Abbildungen zunächst die Posteriori-Schätzwerte der Trendkoeffizienten $\bar{\beta}_0$ bis $\bar{\beta}_5$ aufgegliedert nach informativen und nichtinformativen Lösungen sowie deren Verlauf über die AuswerteePOCHen 2005 bis 2014. Veranschaulicht werden die Variationen, die bereits im Zuge der funktionalen Modellierung thematisiert wurden. Als zusätzliche Information werden in Bezug auf die Konfidenzregionen die Spannen der HPDIs gezeigt, die sich durch den Minimalwert $HPDI_{min}^{R.1}$ bzw. $HPDI_{min}^{R.2}$ und den Maximalwert $HPDI_{max}^{R.1}$ bzw. $HPDI_{max}^{R.2}$ ergeben. Deutlich zu erkennen sind die Variationen der Trendkoeffizienten über die einzelnen AuswerteePOCHen. Für die nichtinformativen Lösungen schwanken die numerischen Ausprägungen insbesondere bezüglich der Koeffizienten $\bar{\beta}_1$ bis $\bar{\beta}_5$. Ebenso wird erneut auf die Konstanz in den Schätzungen der informativen Modellansätze hingewiesen.

Werden explizit die Konfidenzregionen betrachtet, zeigt sich der Einfluss der rekursiven Nutzung von Priori-Informationen: Die HPDIs erweisen sich nicht nur als deutlich kleiner als die vergleichbaren Spannen der nichtinformativen Lösungen, vielmehr ist auch zu erkennen, dass sich durch die Fortpflanzung der Unsicherheiten von Epoche zu Epoche kontinuierlich verringerte Spannen ergeben. Die Wahrscheinlichkeit, dass durch die Posteriori-Schätzung der wahre Wert der Koeffizienten möglichst gut approximiert wird, erhöht sich dadurch, dass der Wahrscheinlichkeitsbereich auf einen minimalen Bereich eingegrenzt werden kann. Im Gegensatz hierzu kann diese Aussage für

die nichtinformativ Lösung nicht getroffen werden. Zum einen weisen die Konfidenzregionen im Vergleich einen ungleich höheren Wahrscheinlichkeitsbereich auf, zum anderen ergibt sich keine Kontinuität in der Entwicklung über den Gesamtzeitraum der Auswertepochen. Ist die informative Lösung durch eine sukzessive Verringerung der HPDIs durch Informationsfortpflanzung gekennzeichnet, variieren die Intervalle der nichtinformativen Lösung von Epoche zu Epoche und ausschließlich in Abhängigkeit des jeweils ausgewerteten Datensatzes.

Deutlich sei an dieser Stelle auf die Problematik der Entwicklung der Einflussgröße $\tilde{\beta}_5$ (Abbildung 5.19) hingewiesen. Wie bereits im Kontext der funktionalen Modellierung festgestellt, ergibt sich hier in den Epochen 2009 bis 2011 sowie 2014 ein Vorzeichenwechsel, der nicht plausibel erklärt werden kann. Durch die zusätzliche Information über die Konfidenzregionen wird deutlich, dass auch für die übrigen Epochen die Wahrscheinlichkeit für einen negativen Koeffizienten $\tilde{\beta}_5$ nicht ausgeschlossen werden kann. Im Gegensatz dazu liegen die Posteriori-Schätzungen desselben Koeffizient für die informative Lösung zwar, wie bereits erläutert, ausschließlich im positiven Bereich, jedoch ist hier im Verlauf bereits eine Tendenz zum negativen Vorzeichen zu erkennen. Die Entwicklung ist über die nichtinformativ Lösung zu begründen, für die ohne die Integration von Priori-Daten ein Abwärtstrend aus den Auswertungen zu erkennen ist. Diese schlagen sich verzögert auch in den informativen Schätzungen nieder und stellen damit kein Problem der numerischen Schätzung dar, sondern sind vielmehr über veränderte Gegebenheiten am Grundstücksmarkt zu begründen. Lediglich starke Variationen in den Ergebnissen werden durch die Nutzung von Priori-Wissen zugunsten einer gleichmäßigen Entwicklung vermieden. Ein genereller, langfristiger Trend bildet sich jedoch stets auch in rekursiven Schätzungen ab. Für die HPDIs ist zudem anzumerken, dass auch im Fall der informativen Lösung in den Epochen 2005, 2006 sowie 2011 bis 2014 bereits die Wahrscheinlichkeit für einen negativen Koeffizienten gegeben ist, da die unteren Grenzen der Konfidenzregionen hier bereits im negativen Bereich liegen.

Insgesamt lässt sich anhand der ausgewerteten Daten festhalten, dass die Integration von Priori-Information neben einer kontinuierlichen Entwicklung der numerischen Ausprägungen ebenfalls zu deutlich verringerten Konfidenzregionen für die Posteriori-Schätzwerte der Trendkoeffizienten führen und so zu einer signifikanten Verringerung der Modellunsicherheiten beitragen. Bestätigt wird diese Aussage bei Betrachtung des Unsicherheitshaushaltes des Modellparameters der Varianz $\tilde{\sigma}^2$, die Abbildung 5.20 entnommen werden kann. Während die Schätzwerte der Varianz im Vergleich der minimalen Werte zwischen informativer und nichtinformativer Lösung schwanken (vgl. Abbildung 5.6, S. 126), zeigt sich bei Betrachtung der Varianz des Varianzfaktors, dass die Nutzung von Priori-Wissen zu einer kontinuierlichen Verringerung führt. Während die Unsicherheit in der Bestimmung auf Grundlage der nichtinformativen Lösung stets über der zugehörigen Schätzung der informativen Lösung liegt, variiert sie zudem auch stärker zwischen den einzelnen Auswertepochen. Wie auch im Fall der Trendkoeffizienten wird hier die Bayessche Vorgehensweise über die Integration und Fortpflanzung von Priori-Informationen durch eine signifikante Reduzierung der Unsicherheiten deutlich.

Durch die Verringerung der Unsicherheiten in den Modellparametern ist ebenfalls eine Auswirkung auf die Posteriori-Schätzung der Zielgröße zu erwarten. Die wesentlichen Ergebnisse werden im folgenden Unterabschnitt gezeigt.

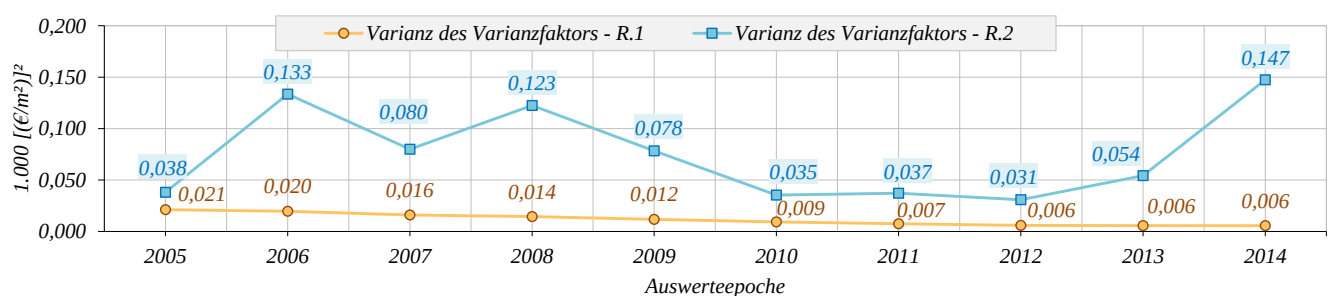


Abbildung 5.20: Varianz $V_{\tilde{\sigma}^2}$ des Varianzfaktors $\tilde{\sigma}^2$ der Epochen 2005 bis 2014, Bayessche Regression

5.2.3.2. Unsicherheit der Zielgröße

Anhand von Gleichung (3.79) zur Berechnung der HPDIs der geschätzten Zielgröße $\bar{y}$ im Modell *R.1* lassen sich in Verbindung mit der Ermittlung der Kofaktormatrix nach (3.49) direkt die Einflüsse der Schätzung der Modellparameter ablesen: In Form der Kofaktormatrix $Q_{\bar{\beta}\bar{\beta}}$ finden die Unsicherheiten der Trendkoeffizienten unmittelbar Eingang in die Schätzung der Konfidenzregionen. Für die Bayessche Regression des Modells *R.2* entfällt die Vorinformation entsprechend. Anhand der bisherigen Ergebnisse des vorherigen Unterabschnitts ist daher zu erwarten, dass die HPDIs der informativen Lösungen sich deutlich unterhalb der Spannen der nichtinformativen Lösung bewegen.

Werden die Konfidenzregionen nach den genannten Gleichungen ermittelt, ergeben sich die in den Abbildungen 5.7 bis 5.8 (S. 128) grafisch dargestellten Ergebnisse der Epochen 2005, 2009 und 2014, einen Gesamtüberblick über sämtliche Auswertepochen gibt wiederum Anhang B.1 (S. 209 ff.). Neben der Darstellung der funktionalen Schätzung sind in grauer Farbe die HPDIs der einzelnen Schätzwerte hinterlegt. Streng genommen dürften diese nicht miteinander verbunden dargestellt werden; aus Gründen der Übersichtlichkeit wird an dieser Stelle unter dem Hinweis auf die eigentliche Definition von Konfidenzregionen jedoch diese Darstellungsform gewählt.

Wie zu erkennen ist, sind sowohl die informative wie auch die nichtinformativ Lösung in den drei dargestellten Epochen durch sehr schmale HPDIs gekennzeichnet. Ein vergleichbares Bild ergibt sich ebenfalls für die übrigen Auswertepochen, deren Darstellungen entsprechend im Anhang eingesehen werden können. Für eine detailliertere Ansicht vermittelt die jeweils zweite Zeile der Abbildungen einen direkten Vergleich der Spannen der HPDIs: Gezeigt werden die Differenzen, die sich aus den absoluten Spannen der HPDIs des Modells *R.2* zu denen des Modells *R.1* ergeben. Durch den Einfluss der Vorinformation und die Fortpflanzung der Unsicherheit sind die HPDIs des Modells *R.1* durchgehend schmaler als die des Modells *R.2* und erlauben damit den Schluss, dass durch die Rekursion die angestrebte Verringerung in den Unsicherheitsmaßen der Vergleichswerte erreicht wird. Insbesondere im direkten Vergleich der Epochen 2005 und 2014 zeigt sich, dass sich die Unsicherheit in der Schätzung der Zielgröße erheblich vermindert hat. Eine Bestätigung dieser Ergebnisse findet sich in Abbildung 5.21, in der die Medianwerte der Größen

$$\begin{aligned} HPDI_{\Delta}^{R.1} &= HPDI_{max}^{R.1} - HPDI_{min}^{R.1} \quad \text{und} \\ HPDI_{\Delta}^{R.2} &= HPDI_{max}^{R.2} - HPDI_{min}^{R.2} \end{aligned}$$

pro Epoche in ihrer zeitlichen Entwicklung über den Gesamtzeitraum 2005 bis 2014 betrachtet werden. Demnach verringert sich die Konfidenzregion der informativen Lösung stetig von 112 €/m² in der Epoche 2005 auf 48 €/m² in der Epoche 2014. Ein Anstieg ist lediglich 2010 zu erkennen, die Differenz zum Vorjahr beträgt jedoch einen zu vernachlässigenden Betrag von 1 €/m² gegenüber 2009. Im Gegensatz zu diesem Ergebnis variieren die Konfidenzregionen der nichtinformativen Lösung stärker, eine Systematik ist nicht zu erkennen, was durch eine starke Abhängigkeit von den jeweiligen Datensätzen der Epochen zu erklären ist. Als Minimalwert ergibt sich eine Spanne von 132 €/m² in 2012, maximal wird ein Wert von 233 €/m² in 2014 geschätzt.

Eine weitergehende Diskussion der Ergebnisse der Unsicherheitsmodellierung durch den Bayesschen Regressionsansatz erfolgt in Kapitel 6. Für einen direkten Vergleich der Modellansätze werden zunächst im folgenden Abschnitt die Ergebnisse für die Modellierungen mittels Bayesscher Kollokation zusammengestellt.

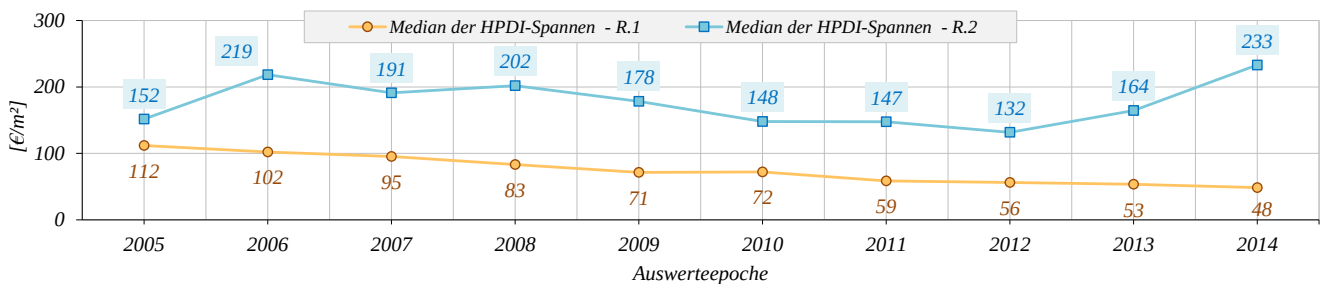


Abbildung 5.21: Medianwerte der HPDI-Spannen von $\bar{y}$ der Epochen 2005 bis 2014, Bayessche Regression

5.3. Modellierung im Ansatz der Bayesschen Kollokation

Als Erweiterung der funktionalen und stochastischen Modellierung der Regressionsansätze ist für die Schätzergebnisse der informativen und nichtinformativen Bayesschen Kollokation *K.1* und *K.2* eine Verbesserung in der Approximationsgüte zu erwarten. Wie auch im Fall der Bayesschen Regression werden für Vergleichszwecke sowohl eine informative als auch eine nichtinformativ Lösung nach der entwickelten Auswertestrategie berechnet. Die Residuen der jeweiligen zugehörigen Regressionsmodelle bilden dabei die Grundlage für die Ableitung von Autokorrelationsfunktionen; wie im Zuge der Erläuterungen zur Auswertestrategie dargelegt, beruhen die Filterungsansätze auf der Ausschöpfung von Restinformationen in den Residuen. Entsprechend der beschriebenen Grundprinzipien eines Kollokationsansatzes sind zunächst die Autokorrelationsfunktionen zu ermitteln, bevor die eigentliche Schätzung der Modellparameter durchgeführt werden kann.

5.3.1. Ableitung von Autokorrelationsfunktionen

Für die Ableitung der Autokorrelationsfunktionen nach (4.17) in Verbindung mit (2.22) ist in einem ersten Schritt über freie Parameter zu verfügen. Dies sind nach ZIEGENBEIN & HAWERK (1978, S. 258) im Einzelnen

- die Form der Autokorrelationsfunktion,
- das Verhältnis von Rausch- zu Signalanteilen sowie
- die Merkmale für die Berechnung von Wertabständen inklusive Gewichtungsansätzen.

Für die Ableitung von Korrelationen zur Besetzung der Signalmatrix Q_{ss} sind aus den empirischen Verläufen der Autovarianz- bzw. Autokorrelationsfunktionen pro Epoche analytische Funktionen zu schätzen. Für die *Form der Autokorrelationsfunktionen* werden in den bisherigen praktischen Anwendungen der einfachen Kollokation als Vergleichswertverfahren Funktionen vom Gaußschen Typ genutzt, nach der sich die Korrelationen ρ_{ik} zwischen zwei Objekten i und k durch die Funktionsform $\rho_{ik} = f(\delta_{ik}) = e^{-c\delta_{ik}^2}$ beschreiben lassen. Hierbei bezeichnet c den Parameter der Funktion, der mittels Ausgleichung aus den empirischen Daten zu schätzen ist, und δ_{ik} den Wertabstand zwischen den betreffenden Objekten. Der theoretische Verlauf einer solchen Funktionsform kann Abbildung 5.22 (rechts) entnommen werden.

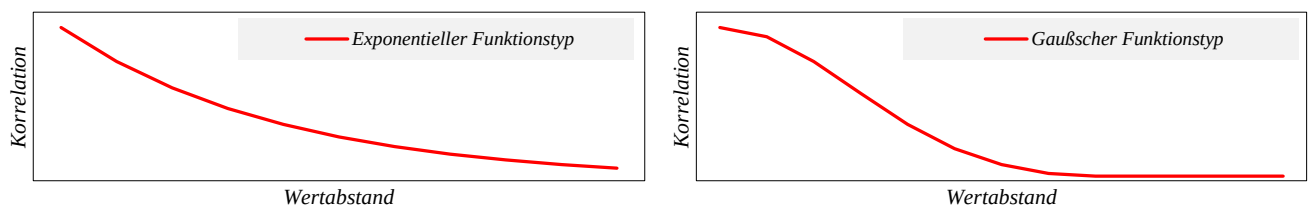


Abbildung 5.22: Theoretische Verläufe von Funktionen des exponentiellen und des Gaußschen Typs

Der Funktion des Gaußschen Typs liegt in Anwendungen der Verkehrswertermittlung der Gedanke zugrunde, dass die Korrelationen zwischen zwei Objekten bei wachsendem Wertabstand vergleichsweise schwach abfallen, sodass auch mit wachsender Verschiedenartigkeit der Objekte ein höherer Einfluss in Form hoher Korrelationen bleibt. Während dieser Sachverhalt aus fachlicher Sicht und insbesondere im Hinblick auf die Heterogenität der Datengrundlage zunächst gerechtfertigt erscheint, haben Studien gezeigt, dass dieser Verlauf nicht den realen Gegebenheiten entspricht. Wie u. a. durch ZIEGENBEIN & HAWERK (1978), ZADDACH & ALKHATIB (2012a), ZADDACH & ALKHATIB (2012b), ZADDACH & ALKHATIB (2013a) sowie ZADDACH & ALKHATIB (2014a) gezeigt wird, gleicht der nach der Beziehung (4.17) ermittelte Verlauf von Autokovarianz- bzw. Autokorrelationsfunktionen basierend auf realen Kauffalldaten einer Funktion exponentiellen Typs. Dieser ist in Abbildung 5.22 (links) abgebildet und lässt sich dahingehend interpretieren, dass mit steigendem Wertabstand ein schnellerer Abfall der Korrelationen zwischen den Objekten einhergeht. Der Heterogenität im Datenmaterial wird auf diese Weise eher Rechnung getragen, indem größtenteils die unmittelbar vergleichbaren Objekte hohe Korrelationen aufweisen. Unter Berücksichtigung der bisherigen Ergebnisse, basierend auf Daten unterschiedlicher räumlicher und sachlicher Teilmärkte, sowie der

Tatsache, dass auch in denjenigen Anwendungen, in denen für die analytische Autokorrelationsfunktion eine Funktion Gaußschen Typs verwendet wurde, ein dem exponentiellen Typ entsprechender empirischer Verlauf festzustellen war, wird für die folgenden Untersuchungen eine analytische Funktion der Form

$$\rho_{ik} = f(\delta_{ik}) = e^{-c\delta_{ik}} \quad (5.3)$$

gewählt, in der der Parameter c durch eine einfache Ausgleichung nach der MkQ geschätzt werden kann.

Wird als zweiter freier Parameter der Bestimmung von Autokorrelationsfunktionen das *Verhältnis zwischen Signal- und Rauschanteil* betrachtet, ergeben sich in bisherigen Anwendungen zwei Untersuchungsansätze: Durch ZIEGENBEIN & HAWERK (1978) werden als eine grobe Annäherung für das Verhältnis der Varianzen empirische Untersuchungen durchgeführt. Durch Variation der Anteile ergibt sich, dass eine optimale Schätzung bei einem Verhältnis Rauschen:Signal von 4:1 bis 9:1 erhalten wird. Wird die Einheitsmatrix im Ausgleichungsalgorithmus der Kollokation entsprechend mit einem Faktor 4 bis 9 multipliziert, ergibt sich nach ZIEGENBEIN & HAWERK (1978) eine optimale Schätzung für die Zielgröße (in Form minimaler Residuen). UHDE (1982) nähert sich der Problematik durch das Verhältnis der Varianzen des zugrunde liegenden Regressionsmodells σ_R^2 und des Trendmodells der Kollokation σ_K^2 , welches eine Gewichtung Rausch- zu Signalanteil durch $\sigma_R^2 / (\sigma_K^2 - \sigma_R^2)$ erlaubt. Eine dritte Alternative stellt in klassisch ingenieurgeodätischen Anwendungen die Varianzkomponentenschätzung dar, die durch KOCH (2007) ebenfalls für die Bayessche Modellbildung formuliert wird. Durch die Schätzung der Varianzkomponenten wird es ermöglicht, für Teilmengen gegebener Datensätze eigene Varianzfaktoren zu schätzen. Aufgrund des Umfangs wird an dieser Stelle zugunsten der Grundlagenentwicklungen einer Auswertestrategie und der reinen Modellbildung im Bayesschen Kontext auf die zusätzliche Einführung Bayesscher Varianzkomponenten explizit verzichtet. Die Verfeinerung der Auswertung sowie die Untersuchung der Auswirkungen auf die Schätzungsergebnisse sind künftigen Forschungsarbeiten zu überlassen und werden als Ausblick für Modelloptimierungen in Kapitel 6 erneut aufgegriffen. Im vorliegenden Fall wird für das grundlegende Modell der Bayesschen Kollokation zunächst ein Gewichtungsansatz von 1:1 für das Signal-Rausch-Verhältnis angehalten. Gleichwohl bleibt der Nachweis über positive Auswirkungen von Varianzkomponenten auf das Schätzungsergebnis eine offene Fragestellung.

Abschließend ist über die Merkmale zu entscheiden, die Eingang in die *Berechnung von Wertabständen* finden. Da sich die Trendmodelle der Bayesschen Regression und Bayesschen Kollokation, wie bereits erläutert, auf einige wenige, gut zu erfassende Determinanten beschränken sollen, können diese ebenfalls als wesentliche Basis für den Wertabstand δ_{ik} herangezogen werden. Wie sich durch empirische Untersuchungen in den vorstehend benannten Publikationen gezeigt hat, ergibt sich eine Verbesserung der Schätzergebnisse durch die Einführung einer zusätzlichen Einflussgröße von Geodistanzen. Auf diese Weise werden die rein geografischen Distanzen berücksichtigt, die sich aus Koordinatendifferenzen zwischen den einzelnen Objekten berechnen lassen. Aus fachlicher Sicht wird durch diese funktionale Erweiterung den Einflüssen der unmittelbaren Nachbarschaft von Objekten Rechnung getragen, denen in der Praxis der Vergleichswertermittlung traditionell eine hohe Bedeutung zugemessen wird. Dieser Zugang bleibt zudem dem Ansatz der (Bayesschen) Regression verwehrt, was einen wesentlichen Vorteil in der Nutzung eines Kollokationsansatzes darstellt. Wie in Abschnitt 4.3.1.2 (S. 95 ff.) erläutert, können die auf diese Weise zusammengestellten Einflussgrößen im Rahmen der Berechnung des Wertabstandes zusätzlich gewichtet werden, um die tatsächlichen Einflüsse auf die Zielgröße abbilden zu können. Im vorliegenden Fall werden die partiellen Korrelationskoeffizienten zwischen Ziel- und Einflussgrößen genutzt; in Bezug auf die Geodistanzen – für die die Abhängigkeiten nicht empirisch ermittelt werden können – wird ein Wert von 1,0 für das entsprechende Gewicht angenommen. Im Verhältnis der Einflussgrößen wird den Geodistanzen so die größte Bedeutung zugeordnet. Dies kann neben der Betonung der unmittelbaren Nachbarschaft auch aus statistischer Sicht damit begründet werden, dass die Einflüsse der übrigen Merkmale bereits durch das Trendmodell größtenteils erfasst sind.

Die Berechnung der empirischen Autokovarianz- und Autokorrelationsfunktion vollzieht sich entsprechend der Festlegung der freien Parameter anhand der Gleichung (4.17) unter einer angenommenen Klassenbreite von $\Delta_d = 0,1$ für die Einteilung der Wertabstände. Nicht vertieft werden an dieser Stelle Ergebnisse zusätzlicher Untersuchungen über die Nichtberücksichtigung von Gewichten in der Berechnung des Wertabstandes und die alternative Einführung der Mahalanobis-Distanz als Gewichtsansatz, da sich in praktischen Berechnungen keine nennenswerten Änderungen durch die Ansätze ergeben haben.

Werden wie im Zuge der Darstellung der Berechnungen mittels Bayesscher Regression lediglich die ausgewählten Auswertepochen 2005, 2009 und 2014 dargestellt, ergeben sich für die Autokorrelationsfunktionen der Signalkomponenten die empirischen und analytischen Verläufe, die in Abbildung 5.23 bis 5.25 skizziert sind. Die Parameter der analytischen Funktion, die durch die MkQ basierend auf (5.3) ermittelt wurden, sind ergänzend Tabelle 5.19 zu entnehmen. Dargestellt sind jeweils die informative und nichtinformative Lösung einer einzelnen Epoche. Die Ergebnisse der übrigen Epochen können in Anhang C.1 (S. 219 f.) eingesehen werden.

Tabelle 5.19.: Geschätzte Parameter $\hat{c}$ der analytischen Autokorrelationsfunktionen der Epochen 2005 bis 2014

Auswertepoche	Parameter $\hat{c}$		Auswertepoche	Parameter $\hat{c}$	
	K.1	K.2		K.1	K.2
2005	-15,24	-14,15	2010	-11,72	-12,55
2006	-14,37	-14,18	2011	-14,18	-13,71
2007	-6,67	-12,57	2012	-11,93	-11,52
2008	-13,89	-14,04	2013	-4,87	-12,31
2009	-10,76	-11,80	2014	-6,16	-14,00

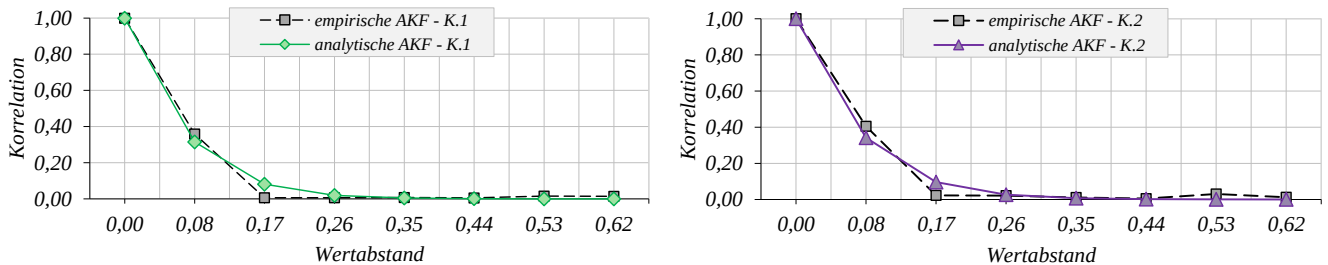


Abbildung 5.23: Empirische und analytische Autokorrelationsfunktionen der Epoche 2005

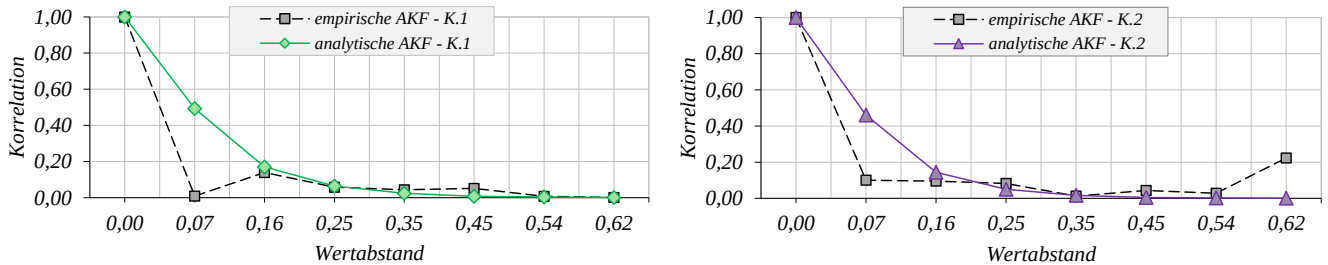


Abbildung 5.24: Empirische und analytische Autokorrelationsfunktionen der Epoche 2009

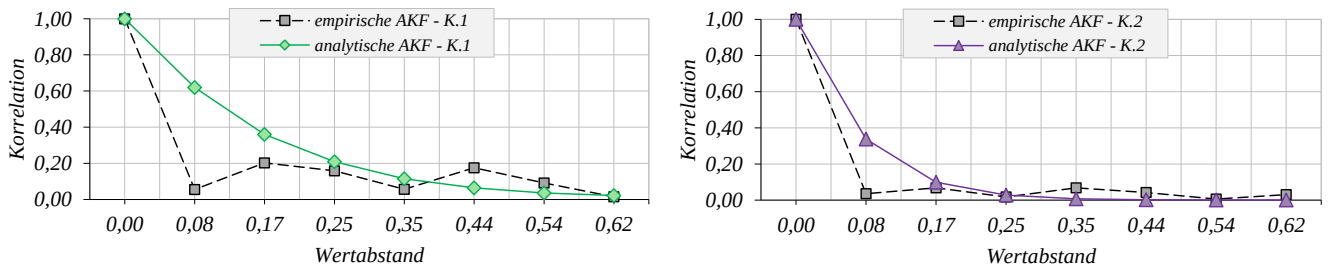


Abbildung 5.25: Empirische und analytische Autokorrelationsfunktionen der Epoche 2014

Deutlich zu erkennen ist in sämtlichen Epochen, dass durch die Schätzung der analytischen Funktionsverläufe die empirischen Gegebenheiten gut approximiert werden: Der Ansatz einer exponentiellen Funktion lässt sich auf diese Weise plausibilisieren. Gleichwohl wird anhand der Verläufe der empirischen Autokorrelationsfunktionen deutlich, dass diese ab der dritten Klasse kaum noch Korrelationen aufweisen. Diese Grenze entspricht einem ungefähren Wertabstand von 0,17 in nahezu sämtlichen Auswertungen, der je nach Epoche minimal variieren kann. Ausnahmen

bilden lediglich die Epochen *2006*, *2013* und *2014*. Für praktische Anwendungen bedeutet dies, dass ausschließlich diejenigen Vergleichsobjekte über einen starken Einfluss auf die Schätzung eines betrachteten Objekts verfügen, die lediglich geringe Differenzen in den wertbeeinflussenden Merkmalen aufweisen; Objekte mit einem Wertabstand $> 0,17$ zeigen demzufolge einen vernachlässigbaren Einfluss auf die Schätzergebnisse. Ausnahmen bilden hier die drei genannten Epochen, deren Funktionsverläufe sich erheblich flacher darstellen und aus höheren Korrelationen in den einzelnen Klassen resultieren.

Ebenfalls wird anhand der Grafiken deutlich, dass es sich bei den empirischen Funktionen nicht um konstant abfallende Funktionen handelt – so zeigen Klassen mit einem höheren Wertabstand mitunter höhere Korrelationen auf als Klassen mit einem geringeren Wertabstand. Dies ist aus sachlogischen Erwägungen ungewöhnlich, jedoch auf die heterogene Datengrundlage zurückzuführen. Ebenso ist es möglich, dass sich hier die Auswirkungen von Merkmalen zeigen, die nicht im Modellansatz enthalten sind. Durch die Schätzung und weitere Verwendung der analytischen Autokorrelationsfunktion werden diese nicht plausiblen Einzelergebnisse jedoch relativiert. Im Vergleich der informativen und nichtinformativen Lösungen zeigen sich pro Epoche nur geringe Unterschiede. Wie bereits anhand der Schätzergebnisse der Bayesschen Regression festzustellen ist, variieren die Ergebnisse der Schätzung nicht sehr stark in Abhängigkeit der Nutzung von Priori-Information. Da sich die Signale in den Kollokationsansätzen unmittelbar aus den Residuen der Trendfunktionen der Modelle *R.1* und *R.2* ergeben, liegt das Ergebnis in dem zu erwartenden Rahmen. Ausnahme bilden hier die Epochen *2007*, *2013* und *2014*, deren Verläufe im Vergleich informativer und nichtinformativer Lösung voneinander abweichen. Da in die Berechnung der Signalkomponenten ausschließlich Vorinformation über die Trendkoeffizienten einfließt, ist dieses Resultat insbesondere auf die unterschiedlichen Schätzergebnisse der Bayesschen Regressionen – bzw. der Residuen – zurückzuführen. Wie ein Vergleich mit den Ergebnissen nach Abbildung 5.13 (S. 130) zeigt, liegen die Medianwerte der prozentualen Abweichungen der informativen Lösung in den genannten Epochen oberhalb derer der nichtinformativen Lösung. Da die prozentualen Abweichungen als eine Kenngröße für die Streuung interpretiert werden, kann daraus geschlossen werden, dass ein höherer Restinformationsgehalt in den Daten der betreffenden Epochen vorhanden ist. Die Auswirkungen auf die Schätzergebnisse werden im weiteren Verlauf diskutiert.

Durch die Ermittlung der analytischen Autokorrelationsfunktionen wird die Besetzung der Signalmatrizen $\mathbf{Q}_{ss}$ sowie folgend die Lösung der Bayesschen Kollokationen ermöglicht. Die Ergebnisse der Modellschätzung werden im folgenden Abschnitt dargestellt.

5.3.2. Schätzergebnisse der Bayesschen Kollokation

Analog zu den Ausführungen zu der Modellschätzung der Bayesschen Regression lassen sich die Ergebnisse für die Bayesschen Kollokationen *K.1* und *K.2* in die Schätzung der eigentlichen Modellparameter sowie der Posteriori-Zielgröße $\bar{\mathbf{y}}$ aufteilen. In dem folgenden Unterabschnitt werden zunächst die Ergebnisse für die Trendkoeffizienten aufgezeigt. Zusätzlich werden die Schätzungen für die Varianzfaktoren gegeben. Im zweiten Unterabschnitt werden die Posteriori-Schätzwerte der Zielgröße den Kaufpreisen der Stichprobendaten gegenübergestellt.

5.3.2.1. Schätzung der Modellparameter

Durch die Besetzung der Signalmatrix mittels der definierten Autokorrelationsfunktionen wird die Schätzung des Ausgleichsalgorithmus nach Abschnitt 3.3.2 (S. 67 ff.) ermöglicht. Als Ergebnis werden wie im Fall der Bayesschen Regression die Parameter der Trendkoeffizienten $\bar{\beta}$, der Varianzfaktor $\bar{\sigma}^2$ sowie dessen Varianz $V_{\bar{\sigma}^2}$ erhalten. Dem Prinzip des Kollokationsansatzes folgend, gelten als zusätzliche Parameter die Signalkomponenten $\bar{\mathbf{s}}$, die die stochastische Restinformation aus den Residuen repräsentieren. Die Ergebnisse der Modellschätzung für die Bayesschen Kollokationsmodelle der Epochen *2005* bis *2014* werden in Tabelle 5.20 gezeigt. Hierbei handelt es sich in den Zeilen 1 bis 6 um die Koeffizienten des Trendmodells, in den Zeilen 7 bis 9 werden ergänzend die Varianzinformationen aufgeführt. Da die Signalkomponenten zur Verbesserung der Posteriori-Schätzungen der Zielgröße dienen, werden diese im nachfolgenden Unterabschnitt dargestellt.

Ähnliche Entwicklungstendenzen in der numerischen Ausprägung sind auch im Rahmen der Bayesschen Kollokation festzustellen. Wie ein direkter Vergleich der Posteriori-Koeffizienten pro Auswerteepoche zeigt, sind wie im Fall der Bayesschen Regression numerische Unterschiede zu erkennen. Diese sind ausschließlich auf die Integration des

Tabelle 5.20.: Modellparameter der Epochen 2005 bis 2014, basierend auf Bayesscher Kollokation

	Epoche	2005		2006		2007		2008		2009	
Zeile	Parameter	K.1	K.2	K.1	K.2	K.1	K.2	K.1	K.2	K.1	K.2
1	$\bar{\beta}_0$	1,529	1,382	1,499	1,394	1,489	1,383	1,477	1,379	1,473	1,460
2	$\bar{\beta}_1$	0,375	0,416	0,407	0,477	0,394	0,357	0,365	0,298	0,363	0,330
3	$\bar{\beta}_2$	0,038	0,058	0,042	0,047	0,042	0,044	0,048	0,079	0,048	0,044
4	$\bar{\beta}_3$	-0,118	-0,131	-0,114	-0,102	-0,114	-0,123	-0,120	-0,141	-0,122	-0,120
5	$\bar{\beta}_4$	-0,413	-0,420	-0,451	-0,508	-0,432	-0,363	-0,432	-0,437	-0,423	-0,366
6	$\bar{\beta}_5$	0,073	0,042	0,072	0,068	0,061	0,019	0,073	0,111	0,059	-0,023
7 *	$\bar{\sigma}^2$	0,022	0,021	0,025	0,034	0,027	0,029	0,029	0,037	0,029	0,029
8 *	$\bar{\sigma}$	0,149	0,145	0,160	0,184	0,164	0,169	0,170	0,192	0,170	0,171

	Epoche	2010		2011		2012		2013		2014	
Zeile	Parameter	K.1	K.2	K.1	K.2	K.1	K.2	K.1	K.2	K.1	K.2
1	$\bar{\beta}_0$	1,441	1,308	1,439	1,382	1,465	1,688	1,470	1,528	1,463	1,315
2	$\bar{\beta}_1$	0,360	0,293	0,370	0,446	0,367	0,325	0,370	0,469	0,376	0,548
3	$\bar{\beta}_2$	0,046	0,035	0,048	0,065	0,045	0,028	0,045	0,047	0,044	0,044
4	$\bar{\beta}_3$	-0,119	-0,097	-0,122	-0,138	-0,125	-0,147	-0,126	-0,161	-0,124	-0,074
5	$\bar{\beta}_4$	-0,395	-0,251	-0,397	-0,417	-0,392	-0,369	-0,388	-0,354	-0,387	-0,391
6	$\bar{\beta}_5$	0,047	-0,041	0,031	-0,074	0,039	0,079	0,031	0,014	0,024	-0,200
7 *	$\bar{\sigma}^2$	0,028	0,021	0,027	0,021	0,026	0,020	0,027	0,028	0,027	0,032
8 *	$\bar{\sigma}$	0,167	0,146	0,164	0,146	0,161	0,141	0,164	0,168	0,166	0,179

* $\bar{\sigma}^2$: Angaben in $[1.000 \text{ €/m}^2]^2$, $\bar{\sigma}$: Angaben in $1.000 \text{ [€/m}^2]$

Vorwissens bzw. die rekursive Anwendung des Bayes-Theorems zurückzuführen. Für die Entwicklungstendenzen der Ausprägungen über die Auswertepochen ist zu erkennen, dass die Vorzeichen größtenteils gleichgerichtet sind und sich somit Einflüsse auf die Schätzung der Zielgröße in jeder Epoche gleich auswirken. Ausnahme bildet bei der nichtinformativen Lösung wiederum das Merkmal der Unterkellerung (Zeile 6, $\bar{\beta}_5$), welches in den Epochen 2009 bis 2011 sowie 2014 das Vorzeichen von positiv zu negativ wechselt. Dieses Ergebnis entspricht den Erkenntnissen aus den Schätzungen der Bayesschen Regression und ist aufgrund der lediglich stochastischen Verfeinerung der funktionalen Modelle zu erwarten.

Wie auch im Fall der Bayesschen Regressionen ist für die numerischen Ausprägungen der Trendkoeffizienten ein Unterschied in den Schwankungen von Epoche zu Epoche festzustellen: Wird die zeitliche Entwicklung der einzelnen Trendkoeffizienten über den gesamten Zeitraum der Auswertepochen betrachtet, ergeben sich für die nichtinformativen Lösungen deutliche Schwankungen, während für die Schätzungen der informativen Lösungen eine gewisse Konstanz erreicht wird, was sich in deutlich geringeren Schwankungen zeigt.

Ein Quervergleich mit den Schätzergebnissen der Bayesschen Regressionsmodelle nach Tabelle 5.18 lässt erkennen, dass sich innerhalb der Auswertepochen – und in Abhängigkeit der informativen und nichtinformativen Lösungen – Abweichungen in den numerischen Ausprägungen zeigen, da die Kollokationsmodelle in der Schätzung der Trendkoeffizienten ein erweitertes stochastisches Modell berücksichtigen. Nichtsdestoweniger weichen die Trendmodelle lediglich geringfügig voneinander ab, sodass eine Verbesserung der funktionalen Modellierung hauptsächlich durch die Berücksichtigung der Signalkomponente zu erwarten ist.

In der Varianzinformation sind deutliche Unterschiede zu den Ergebnissen der Bayesschen Regression zu erkennen. Wie Abbildung 5.26 zeigt, ergibt sich über den Zeitraum der Auswertepochen zwar ein ähnlicher Verlauf in der Entwicklung der Varianzfaktoren der informativen und der nichtinformativen Lösungen $K.1$ und $K.2$ zu denen der Lösungen $R.1$ und $R.2$, jedoch liegen die Werte deutlich unterhalb derjenigen der Regressionsmodelle. Die Varianz hat sich durch die Modellierung in den Bayesschen Kollokationsansätzen signifikant verringert: Wie zuvor zeigt sich für die informative Lösung $K.1$ zwar ein vergleichsweise konstanter Verlauf über die Auswertepochen, der minimale Wert liegt jedoch nunmehr bei $\bar{\sigma}_{K.1, \min}^2 = 0,022$, der maximale bei $\bar{\sigma}_{K.1, \max}^2 = 0,029$

(jeweils in $[1.000 \text{ €/m}^2]^2$). Der Minimalwert der nichtinformativen Lösung $K.2$ liegt bei $\bar{\sigma}_{K.2, \min}^2 = 0,020$, der Maximalwert bei $\bar{\sigma}_{K.2, \max}^2 = 0,037$ (jeweils in $[1.000 \text{ €/m}^2]^2$). Insgesamt lässt sich festhalten, dass durch die Bayesschen Kollokationsmodelle eine Verringerung der Streuung in der funktionalen Modellierung erreicht wird. Im Vergleich der Bayesschen Kollokationsmodelle $K.1$ zu $K.2$ zeigt sich äquivalent zu den vorherigen Ergebnissen der Regressionsmodelle, dass der Varianzfaktor der informativen Lösung in den Epochen 2005 sowie 2010 bis 2012 unterhalb der Streuung der nichtinformativen Lösung liegt. In diesem Kontext werden im folgenden Unterabschnitt die Posteriori-Schätzungen der Zielgröße $\bar{\mathbf{y}}$ der einzelnen Auswertepochen aufgezeigt.

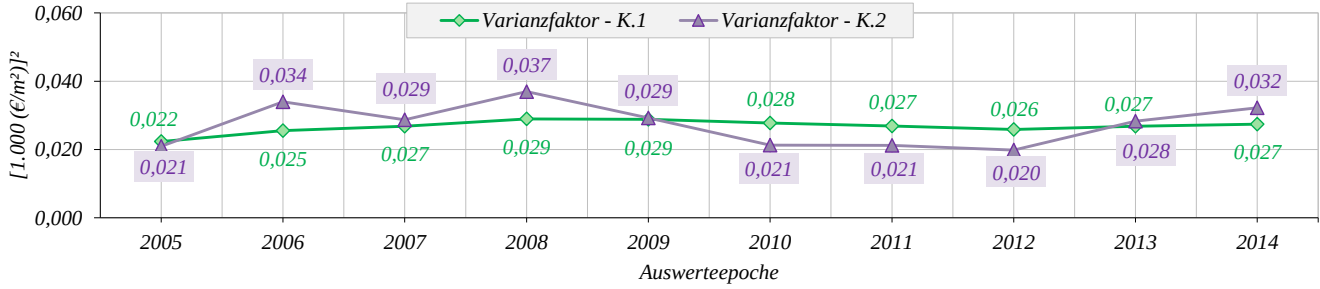


Abbildung 5.26: Posteriori-Schätzung des Varianzfaktors $\bar{\sigma}^2$ der Epochen 2005 bis 2014, Bayessche Kollokation

5.3.2.2. Schätzung der Zielgröße

Auf Grundlage der Ergebnisse nach Tabelle 5.20 lassen sich für die Bayesschen Kollokationen $K.1$ und $K.2$ Posteriori-Schätzwerte für die Zielgröße ermitteln. Wie auch im Fall der Bayesschen Regressionsmodelle können durch einen Vergleich der Ergebnisse Rückschlüsse auf die Auswirkung der Integration von Priori-Information gezogen werden. Die Posteriori-Schätzwerte der Zielgrößen können durch die Beziehungen

$$\bar{\mathbf{y}}_{K.1} = \mathbf{X}\bar{\boldsymbol{\beta}}_{K.1} + \bar{\mathbf{s}}_{K.1}, \quad \bar{\mathbf{y}}_{K.2} = \mathbf{X}\bar{\boldsymbol{\beta}}_{K.2} + \bar{\mathbf{s}}_{K.2} \quad (5.4)$$

bestimmt werden, wobei sich die Vektoren der Signalkomponenten nach (3.108) bzw. (3.129) ergeben. Die nach der Schätzung verbleibenden Rauschanteile, denen keine Restinformation zugeordnet werden können, ergeben sich entsprechend zu

$$\bar{\mathbf{v}}_{K.1} = \mathbf{y} - \bar{\mathbf{y}}_{K.1}, \quad \bar{\mathbf{v}}_{K.2} = \mathbf{y} - \bar{\mathbf{y}}_{K.2}, \quad (5.5)$$

und lassen sich wiederum in prozentuale Abweichungen zwischen Kaufpreis und Posteriori-Schätzwert umrechnen. Für die bereits im vorherigen Abschnitt dargestellten exemplarischen Epochen 2005, 2009 und 2014 können die Ergebnisse für die Posteriori-Schätzwerte der Signalkomponenten $\bar{\mathbf{s}}$ und die Zielgröße $\bar{\mathbf{y}}$ den Abbildungen 5.27 bis 5.32 entnommen werden, die Grafiken für die übrigen Epochen sind in Anhang C.2 (S. 221 ff.) zusammengestellt.

In den Abbildungen 5.27 bis 5.29 mit den Schätzungen der Signalkomponenten sind die systematischen Effekte zu erkennen, die bereits in den Ergebnissen für Schätzung der Zielgröße in den Modellen $R.1$ und $R.2$ beobachtet werden konnten: Da die Kaufpreise wie in den übrigen Abbildungen aufsteigend sortiert sind, zeigen sich größere, negative Signalkomponenten im niedrigpreisigen Segment, im höherpreisigen Segment sind überwiegend positive Signalkomponenten festzustellen. Das Ergebnis lässt sich sowohl für das Modell $K.1$ und das Modell $K.2$ über die gesamten Epochen des Auswertzeitraumes bestätigt. Da die Signalkomponenten der systematischen Über- und Unterschätzung der Trendmodelle mit entgegengesetzten Vorzeichen folgen, lässt sich bereits eine Verbesserung in der funktionalen Approximationsgüte der Kollokationsansätze erkennen. Bestätigt wird dies bei Betrachtung der Ergebnisse für die Posteriori-Schätzung der Zielgröße $\bar{\mathbf{y}}$ in den Abbildungen 5.30 bis 5.32.

Wie anhand der Abbildungen zu erkennen ist, zeigen sich keine auffälligen Unterschiede der Ergebnisse im Vergleich der jeweiligen informativen und nichtinformativen Modelle $K.1$ zu $K.2$. Bestätigt wird somit auch das Ergebnis für die Schätzung der Varianzfaktoren, bei denen teilweise die Ergebnisse der informativen Lösungen in ihrer Höhe über denen der nichtinformativen liegen und sich daher keine deutlichen Unterschiede in den Streuungen

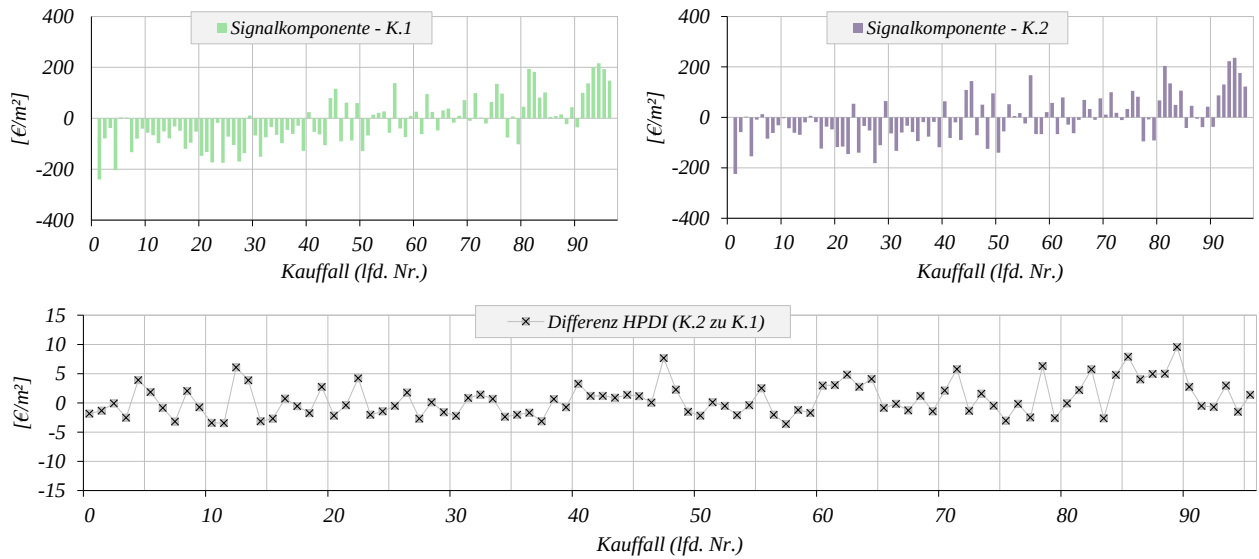


Abbildung 5.27: Posteriori-Schätzwerte $\bar{s}$, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

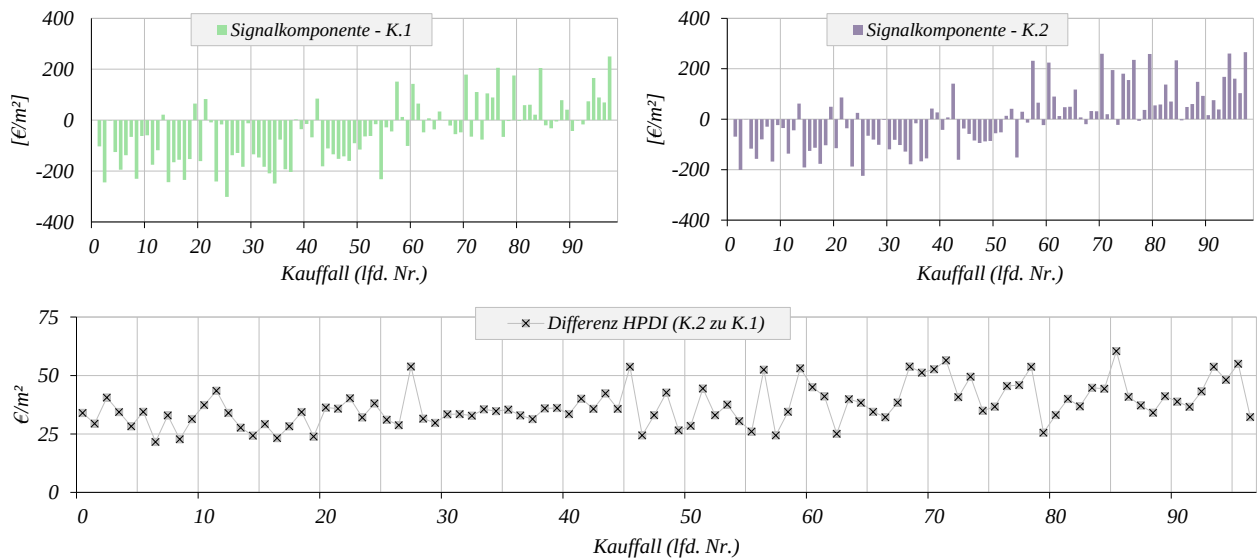


Abbildung 5.28: Posteriori-Schätzwerte $\bar{s}$, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

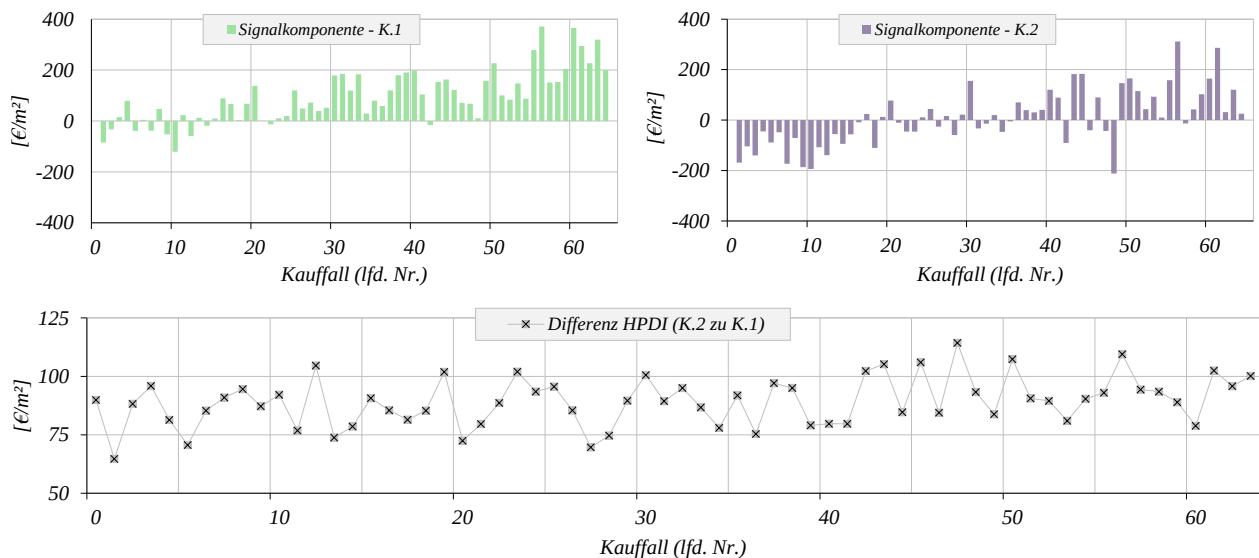


Abbildung 5.29: Posteriori-Schätzwerte $\bar{s}$, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

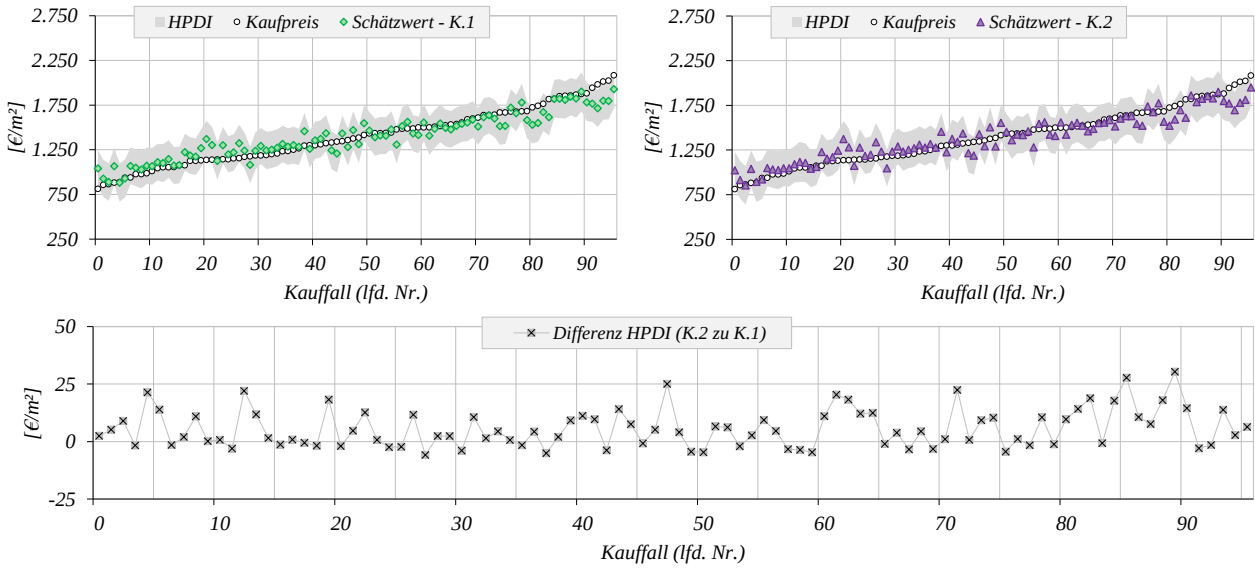


Abbildung 5.30: Posteriori-Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

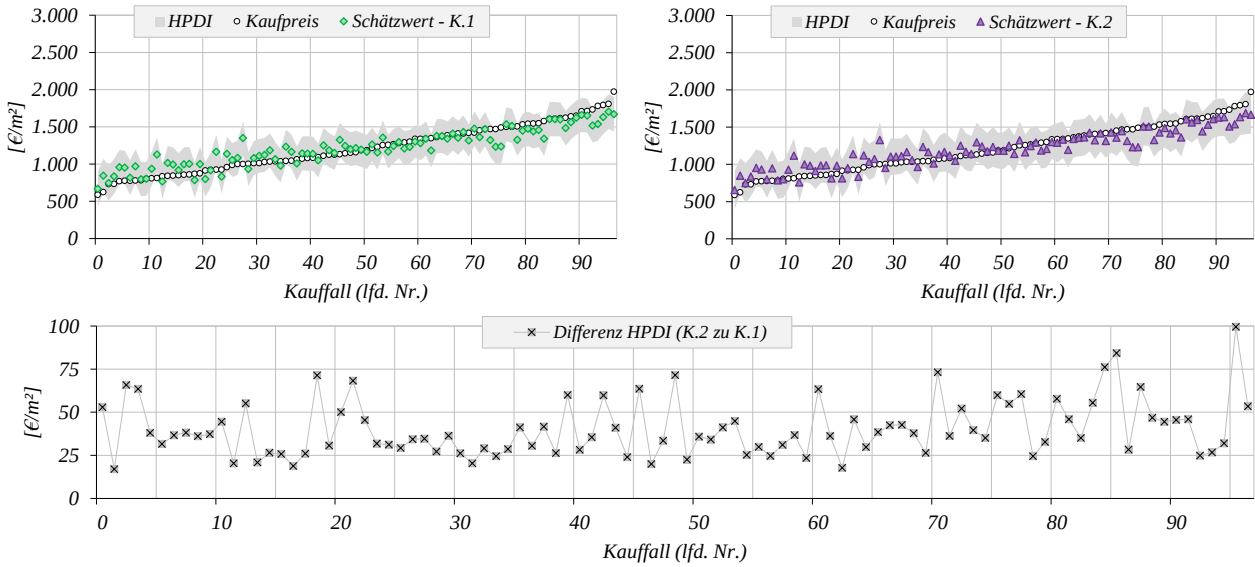


Abbildung 5.31: Posteriori-Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

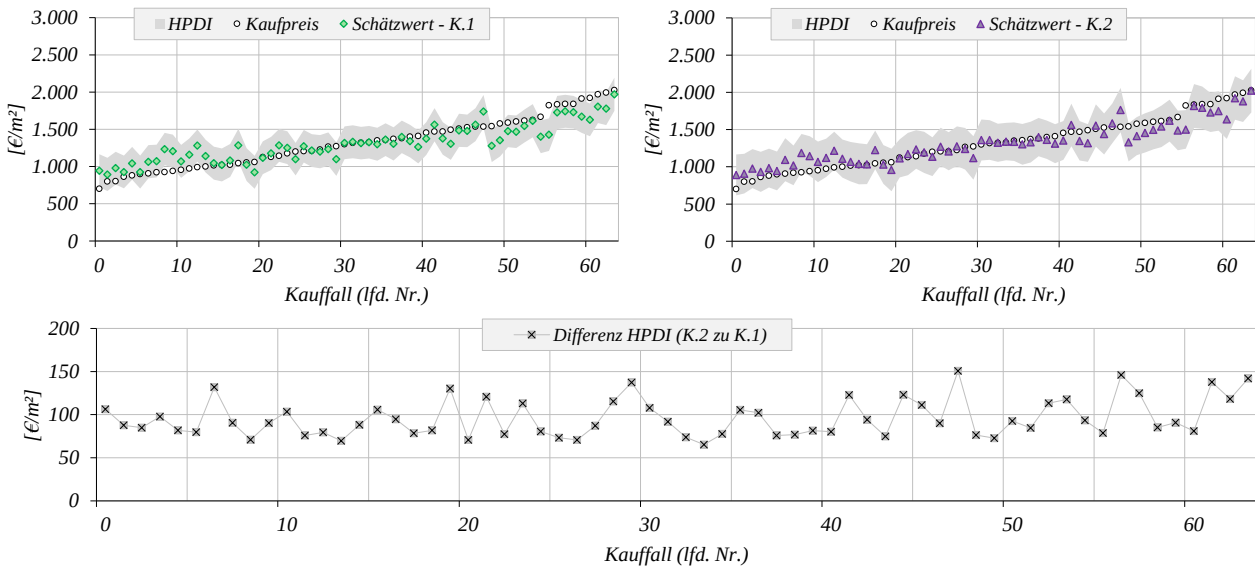


Abbildung 5.32: Posteriori-Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

beobachten lassen. Generell zeigt sich wie auch im Fall der Bayesschen Regression, dass sich im niedrigpreisigen Segment eine Überschätzung, im hochpreisigen Segment hingegen eine Unterschätzung der Vergleichswerte ergibt. Wiederum streuen die Schätzwerte im mittleren Bereich hingegen zufällig um die realen Kaufpreise und weisen hier entsprechend minimale Abweichungen bzw. die beste Approximationsgüte auf. Die Ergebnisse bestätigen sich über sämtliche Auswertepochen und sind daher unabhängig von der Nutzung des Modells *K.1* oder des Modells *K.2* zu sehen. Gegenüber den Ergebnissen der Bayesschen Regressionsansätze bewirkt die Berücksichtigung der Signalkomponenten eine Verbesserung in der Annäherung zwischen Kaufpreisen und Posteriori-Schätzwerten $\bar{y}$.

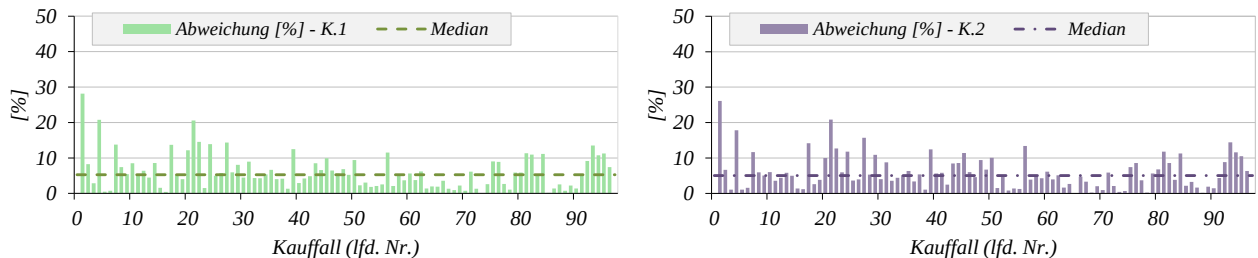


Abbildung 5.33: Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2005

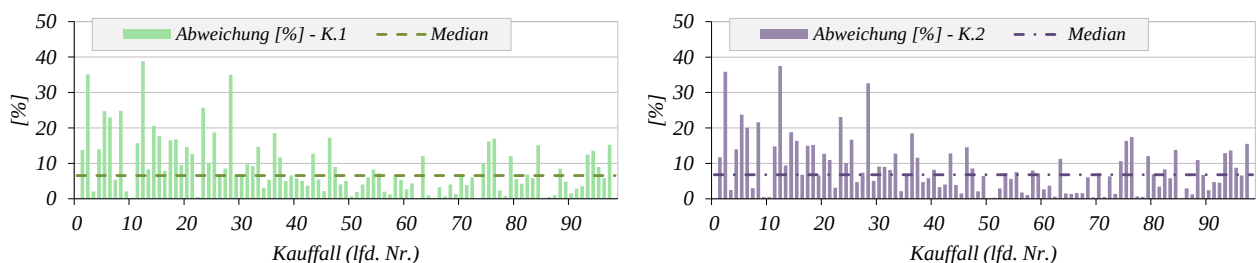


Abbildung 5.34: Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2009

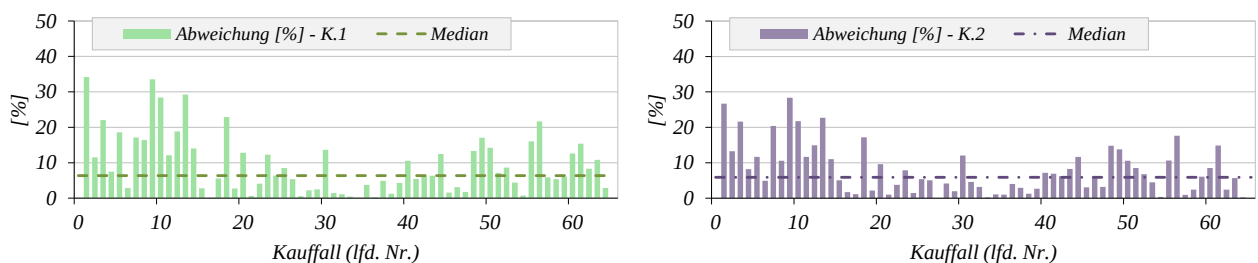


Abbildung 5.35: Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2014

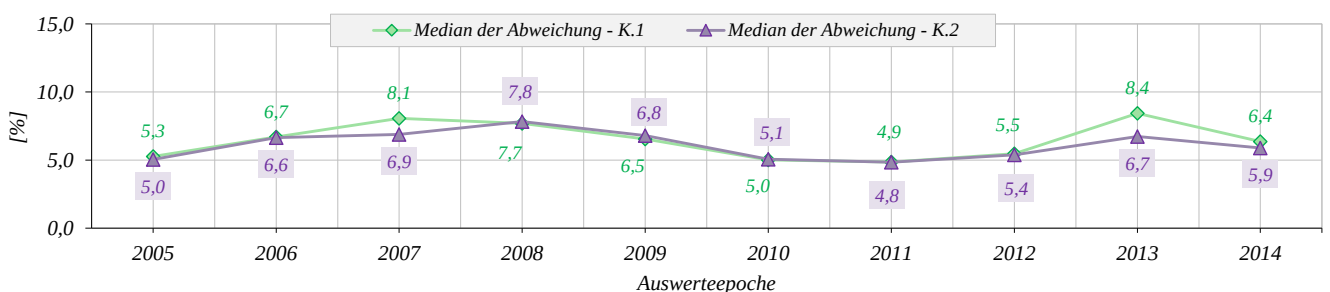


Abbildung 5.36: Medianwerte der prozentualen Abweichungen zwischen Kaufpreisen und Posteriori-Schätzwerten $\bar{y}$ der Epochen 2005 bis 2014, Bayessche Kollokation

Anhand der prozentualen Abweichungen lassen sich ergänzend die Extremwerte ablesen. Für die exemplarischen Epochen 2005, 2009 und 2014 sind diese in den Abbildungen 5.33 bis 5.35 zusammengestellt, für die übrigen Epochen in Anhang C.3 (S. 232 ff.). Als größte Abweichungen ergeben sich für die informativen Lösungen des Modells *K.1* in der Epoche 2005 28,2 %, 2009 38,8 % sowie 2014 34,2 %, was insgesamt eine deutliche Verringerung

im Vergleich zu den Ergebnissen der Bayesschen Regression nach den Modellen *R.1* und *R.2* bedeutet. Für die nichtinformativen Lösungen des Modells *K.2* können die Maximalwerte für 2005 mit 26,1 %, 2009 mit 37,5 % und 2014 mit 28,3 % angegeben werden, was sich ebenfalls als signifikante Verbesserung gegenüber den Ergebnissen der Regressionsmodelle darstellt. Wie auch zuvor sind die genannten maximalen prozentualen Abweichungen im niedrigpreisigen Segment zu finden. Minimale Differenzen ergeben sich entsprechend im mittleren Kaufpreisbereich und liegen für die informativen Lösungen bei nahezu 0 % (2005: 0,3 %, 2009: 0,1 %, 2014: 0,4 %), ebenso wie für die nichtinformativen Lösungen (2005: 0 %, 2009: 0 %, 2014: 0,1 %). Werden als allgemeinere Aussage die Medianwerte herangezogen, differieren diese zwischen informativer und nichtinformativer Lösung *K.1* zu *K.2* lediglich um wenige Prozentanteile: Für die betrachteten Epochen liegen diese bei 6,4 % zu 6,2 % (2005), 8,0 % zu 8,5 % (2009) und 9,6 % zu 8,8 % (2014). Es zeigt sich daher – wie auch eine Gegenüberstellung der Medianwerte der übrigen Auswertepochen nach Abbildung 5.36 bestätigt – eine signifikante Verringerung der Abweichungen zwischen Kaufpreis und Posteriori-Schätzwert $\bar{y}$ von Bayesscher Regression zu Bayesscher Kollokation durch eine Reduktion der Abweichungen um teilweise bis zu 50 %. Eine pauschale Aussage über die Bevorzugung eines informativen oder nichtinformativen Modells kann jedoch auch an dieser Stelle nicht getroffen werden. Wie Abbildung 5.36 zeigt, ergeben sich für die Auswertepochen unterschiedliche Ergebnisse, in denen sich teilweise die Berücksichtigung von Priori-Information positiv in Form geringerer Abweichungen auswirken, teilweise die Nichtberücksichtigung von Priori-Information. Rückschlüsse auf die funktionale Modellierung im Bayesschen Ansatz werden in Kapitel 6 gezogen.

Bislang werden ausschließlich die Ergebnisse der rein funktionalen Modellierung im Ansatz der informativen und nichtinformativen Bayesschen Kollokation dargestellt. Ein weiterer Aspekt ist die Betrachtung der Unsicherheitsmodellierung, die im folgenden Abschnitt behandelt wird.

5.3.3. Unsicherheitsmodellierung mittels Bayesscher Kollokation

In Erweiterung der Unsicherheitsbetrachtung der Bayesschen Regressionsansätze werden im Folgenden die Ergebnisse für die Bayesschen Modelle der Kollokation vorgestellt. Wie bereits gezeigt, hat sich durch die Integration von Priori-Informationen im Modell *R.1* eine deutliche Reduzierung der Unsicherheiten sowohl in den Modellparametern als auch den Posteriori-Schätzungen der Zielgröße erreichen lassen. Für die Bayessche Kollokation ist ein ähnliches Ergebnis zu erwarten; als zusätzliche Komponente ist der Signalanteil mit einem Einfluss auf die Gesamtunsicherheit zu berücksichtigen. Im nachfolgenden Unterabschnitt werden zunächst die Unsicherheiten der Modellparameter aufgezeigt, zu denen nach Definition in Abschnitt 3.3.1 (S. 65 f.) ebenfalls die Signalkomponente zu zählen ist. Aufgrund des unmittelbaren Einflusses auf die Posteriori-Schätzung der Zielgröße werden die Unsicherheiten ungeachtet dessen im zweiten Unterabschnitt im Zusammenhang mit der Erläuterung der Unsicherheiten der Zielgröße beschrieben.

5.3.3.1. Unsicherheit der Modellparameter

Die Schätzung der Konfidenzregionen der Trendkoeffizienten in den Modellen der Bayesschen Kollokation erfolgt anhand von (3.138) unter Beachtung der jeweils zugehörigen Varianzfaktoren und Kofaktormatrizen (informativ oder nichtinformativ). Als Irrtumswahrscheinlichkeit wird ein Wert von $\alpha = 5\%$ angenommen. Werden die HPDIs der Trendkoeffizienten in den Modellen *K.1* und *K.2* über die Auswertepochen 2005 bis 2014 berechnet, ergeben sich die in den Abbildungen 5.37 bis 5.42 dargestellten Entwicklungen. Analog zu den vergleichbaren Abbildungen der Bayesschen Regression finden sich die Schätzungen der informativen Lösung in der linken, die der nichtinformativen in der rechten Grafik.

Entsprechend der Erkenntnisse aus der Schätzung der Trendkoeffizienten zeigen sich in den Grafiken die Variationen in den numerischen Ausprägungen: Der durch die Priori-Information geprägte kontinuierliche Verlauf in der Entwicklung der informativen Modelle *K.1* steht den stärkeren Variationen in der Entwicklung der Schätzwerte der nichtinformativen Modelle *K.2* gegenüber. Für die HPDIs bestätigt sich im Vergleich zu den Resultaten der Regressionsansätze, dass die Integration von Vorinformationen zu einer erheblichen Reduzierung führt. Die Größenordnungen der HPDIs lassen sich in eine vergleichbare Größenordnung zu den jeweiligen Modellen der Bayesschen Regression einordnen. Lediglich geringe Differenzen sind zwischen den jeweiligen informativen und nichtinformativen

tiven Lösungen zu erkennen, sodass die Wahrscheinlichkeiten für die einzelnen Trendkoeffizienten innerhalb der gekennzeichneten Wertebereiche als übereinstimmend beurteilt werden können.

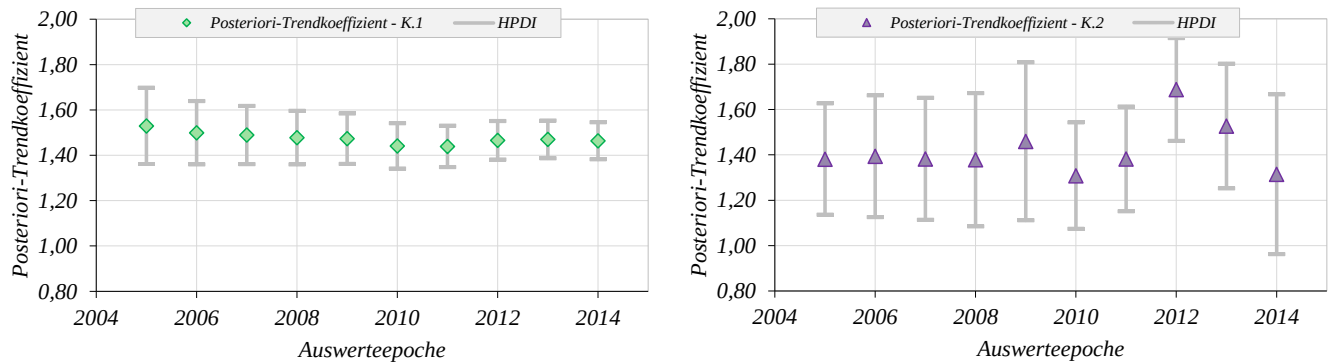


Abbildung 5.37: HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_0$ der Epochen 2005 bis 2014, Bayessche Kollokation

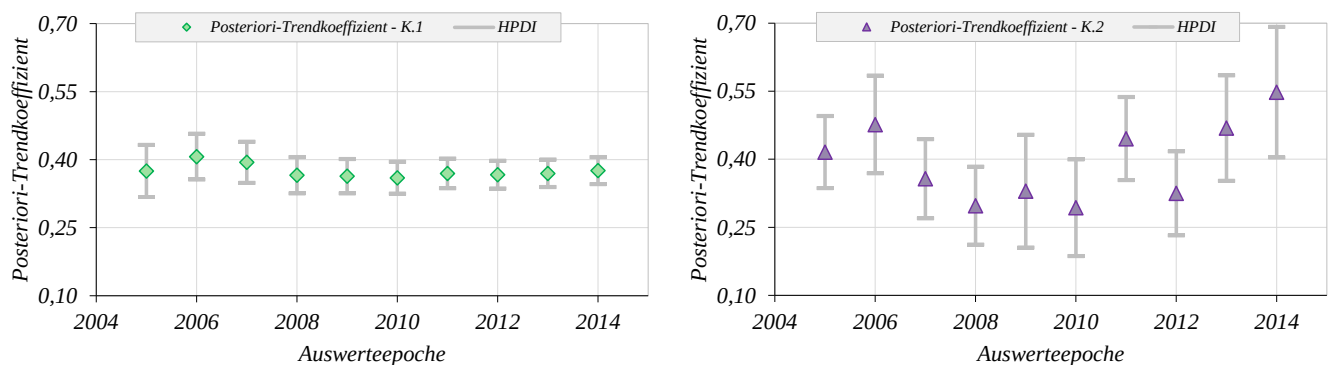


Abbildung 5.38: HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_1$ der Epochen 2005 bis 2014, Bayessche Kollokation

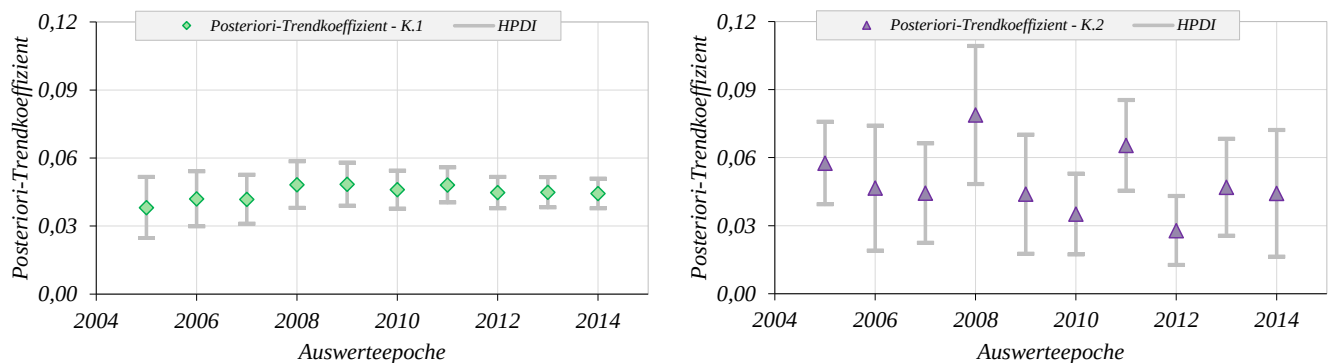


Abbildung 5.39: HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_2$ der Epochen 2005 bis 2014, Bayessche Kollokation

Es zeigen sich somit gleiche Tendenzen in den unterschiedlichen Modellen *R.1* und *K.1* sowie *R.2* und *K.2*. Wie auch in den Regressionsansätzen ist die Problematik des Merkmales der Unterkellerung erkennbar, die sich in dem mehrmaligen Vorzeichenwechsel des Trendkoeffizienten zeigt. Die Verfeinerung der funktionalen Modellierung in einem Kollkationsansatz führt an dieser Stelle zu keiner Verbesserung bzw. Plausibilisierung, vielmehr wird auch hier der Trend bestätigt, dass das Merkmal einen Wahrscheinlichkeitsbereich umfasst, der auch einen negativen Einfluss auf die Zielgröße zulässt.

In Ergänzung zu der Beobachtung eines signifikanten Einflusses informativer Modellansätze auf die Reduzierung der Unsicherheiten der Trendkoeffizienten wird durch die Schätzung der Varianz des Varianzfaktors ebenfalls eine Verringerung der Unsicherheiten deutlich. Wie in Abbildung 5.43 gezeigt, lassen sich die Ergebnisse in ihrem Verlauf über die AuswerteePOCHEN mit denen der Bayesschen Regressionsmodelle vergleichen. Die Unsicherheit der Varianzen zeigen einen kontinuierlich abfallenden Verlauf, was auf die stetige rekursive Anwendung der Auswertestrategie

zurückzuführen ist. Liegt die Unsicherheit des Varianzfaktors in der Epoche 2005 bei $0,006 [1.000 \text{ €/m}^2]^2$, fällt sie bis zur Epoche 2014 auf einen Wert von $0,001 [1.000 \text{ €/m}^2]^2$ ab. Im Gegensatz zu dieser Entwicklung ist in den Varianzen des Varianzfaktors der nichtinformativen Lösungen wiederum eine größere Variation zu erkennen, die aus dem Bezug auf ausschließlich aktuelle Datensätze resultiert. Als Minimalwert lässt sich in der Epoche 2012 ein Wert von $0,007 [1.000 \text{ €/m}^2]^2$ angeben, der maximale Betrag zeigt sich in der Epoche 2014 mit $0,041 [1.000 \text{ €/m}^2]^2$. Weiterhin ist zu erkennen, dass die Varianz der nichtinformativen Lösungen in sämtlichen Epochen über der der informativen Lösungen liegt, was für die Verbesserung in der Unsicherheitsmodellierung spricht.

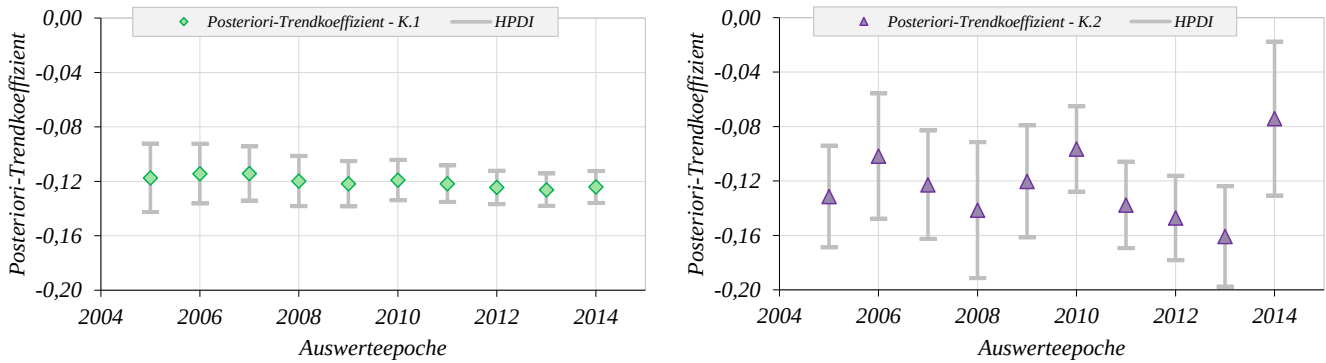


Abbildung 5.40: HPDI des Posteriori-Trendkoeffizienten β_3 der Epochen 2005 bis 2014, Bayessche Kollokation

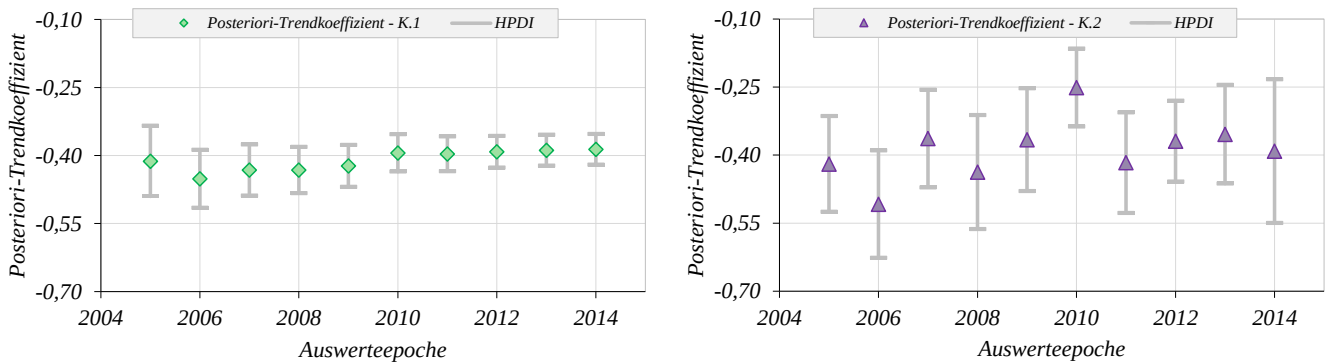


Abbildung 5.41: HPDI des Posteriori-Trendkoeffizienten β_4 der Epochen 2005 bis 2014, Bayessche Kollokation

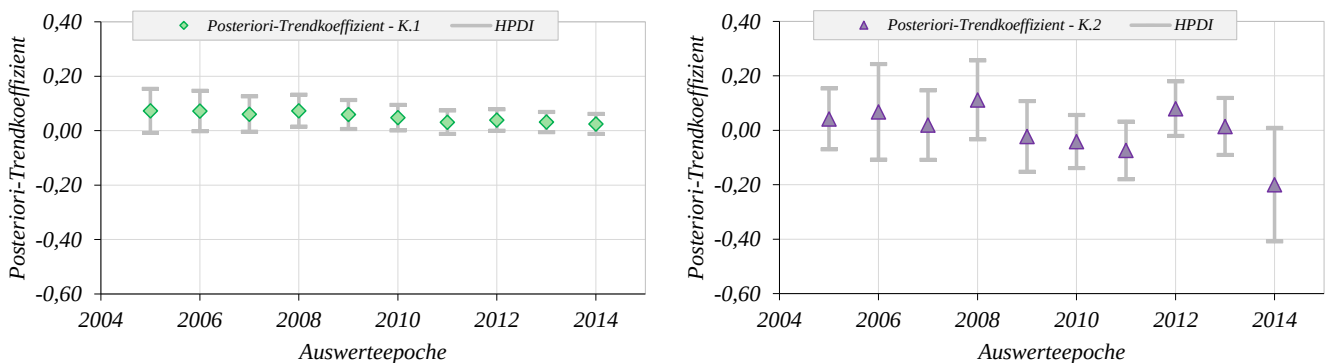


Abbildung 5.42: HPDI des Posteriori-Trendkoeffizienten β_5 der Epochen 2005 bis 2014, Bayessche Kollokation

Als weiterer Modellparameter der Bayesschen Kollokationsmodelle ist die Signalkomponente zu schätzen. Die Ergebnisse für deren Unsicherheiten sind im direkten Zusammenhang mit der Unsicherheitsmodellierung der Posteriori-Schätzung der Zielgröße zu sehen. Die Resultate werden im folgenden Unterabschnitt thematisiert.

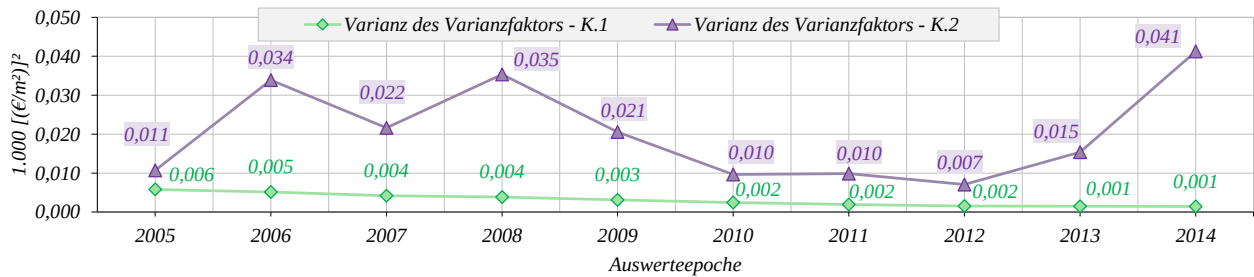


Abbildung 5.43: Varianz $V_{\bar{\sigma}^2}$ des Varianzfaktors $\bar{\sigma}^2$ der Epochen 2005 bis 2014, Bayessche Kollokation

5.3.3.2. Unsicherheit der Zielgröße

Wird zunächst der Unsicherheitshaushalt des Signalanteils betrachtet, lassen sich gemäß (3.139) die zugehörigen HPDIs berechnen. In den bisherigen Ausführungen zur Anwendung der Theorie des Kollokationsansatzes für Zwecke der Vergleichswertermittlung wurde bereits darauf hingewiesen, dass die stochastische Modellierung von Restinformationen auf Basis von Autokorrelationen in den Residuen erfolgt. Die Einführung von Priori-Information für das Signal soll damit explizit ausgeschlossen werden; dieser Annahme liegt der Gedanke zugrunde, dass sich die Restinformation durch die Schätzung von Autokorrelationsfunktionen basierend auf Wertabständen erfassen lässt. Wertabstände sind in der Definition für die Verkehrswertermittlung ausschließlich an die spezifischen Objekte der aktuellen Stichprobe gebunden, wodurch die Einführung von Priori-Information unmöglich wird. Die Integration von Vorwissen wäre lediglich dann sinnvoll, wenn im Voraus bekannt ist, welche Objekte zu welchen Kaufpreisen gehandelt werden – in der Realität ist diese Anforderung nicht zu erfüllen. Entsprechend fließt in die Schätzung der Konfidenzregionen für die Signalkomponente ausschließlich Vorwissen über das Trendmodell ein. Wie (3.112) entnommen werden kann, wird im Fall der informativen Lösung ausschließlich die Kofaktormatrix der Priori-Trendkoeffizienten für die Berechnung der Kofaktormatrix bzw. VKM des Signals genutzt, d. h., dass lediglich das Vorwissen über die Trendkoeffizienten Berücksichtigung in der Schätzung findet. Dementsprechend gering fallen die Differenzen in den HPDIs zwischen dem Modell *K.1* und dem Modell *K.2* aus. Für die Epochen 2005, 2009 und 2014 lassen sich diese den Abbildungen 5.27 bis 5.29 entnehmen, für die übrigen Epochen wird auf die Darstellungen des Anhangs C.2 (S. 221 ff.) verwiesen.

Anhand der Differenzplots ist zu beobachten, dass die HPDIs der Signalkomponenten in den informativen Lösungen im Vergleich zu denen der nichtinformativen Lösungen größtenteils einen geringeren Bereich umfassen und sich Vorinformation daher in Form verringerter Unsicherheiten auswirkt. Eine Ausnahme stellt die Epoche 2005 dar, hier liegen die Differenzen lediglich in einem Bereich von bis zu maximal $+10 \text{ €/m}^2$, wobei überwiegend negative Differenzen von bis zu -5 €/m^2 zu erkennen sind. Ähnliche Ergebnisse sind für die Epochen 2010, 2011 und 2012 festzustellen, bei denen die Differenzen *K.2* zu *K.1* teilweise bis zu -25 €/m^2 betragen. Wie ein Vergleich mit Abbildung 5.26 zeigt, deckt sich dies mit den Epochen, in denen die geschätzten Varianzfaktoren der nichtinformativen Lösungen unterhalb derer der informativen Lösungen liegen. Da der Varianzfaktor die maßgebliche Größe für die Berechnung der HPDIs ist, kann das Ergebnis aus mathematisch-statistischer Sicht begründet werden. Zudem ist daraus schließen, dass der Einfluss des Varianzfaktors größer ist als die verringerte Unsicherheit der Trendkoeffizienten, die als einzige Vorinformation in die Bestimmung der Unsicherheit der Signalkomponenten im Modell *K.1* einfließt (vgl. (3.112)). Die Entwicklung der Medianwerte der HPDIs pro Epoche und Modell über den gesamten Auswertzeitraum kann Abbildung 5.44 entnommen werden und fasst die Einzelergebnisse der Epochen 2005 bis 2014 zusammen.

Die Unsicherheit des Signals bildet sich ebenfalls in der Unsicherheit der Posteriori-Schätzung der Zielgröße ab. Diese wird nach (3.143) unter Beachtung der jeweils gültigen Kofaktormatrizen und Varianzfaktoren ermittelt. Die Ergebnisse für die gewählten Epochen 2005, 2009 und 2014 sind den Abbildungen 5.30 bis 5.32 (S. 141) zu entnehmen, die Grafiken der übrigen Epochen in Anhang C.2 (S. 221 ff.).

Wie bereits erläutert, werden durch die grau hinterlegten Bereiche in den oberen Zeilen der Abbildungen die HPDIs gekennzeichnet; auf die nicht allgemeingültige Darstellungsweise wird erneut hingewiesen. Im Vergleich zu den Ergebnissen der Bayesschen Regression ergeben sich für die Bayesschen Kollokationsansätze wesentlich größere

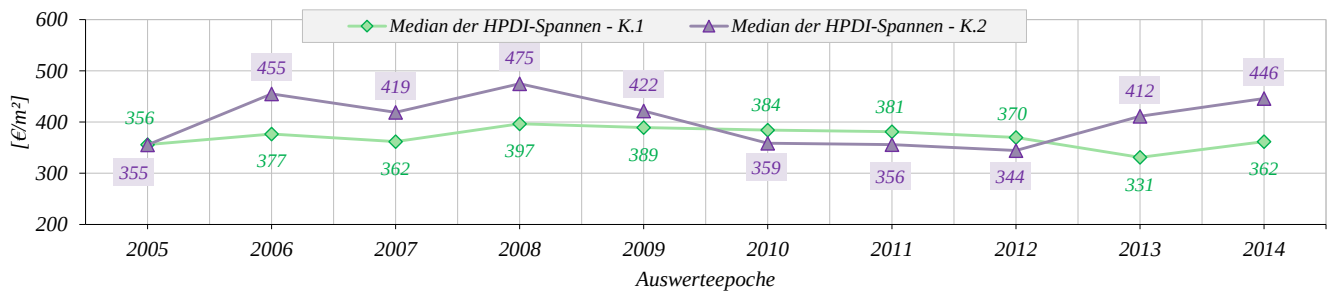


Abbildung 5.44: Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{s}$ der Epochen 2005 bis 2014

Spannen. Konzentrieren sich die Spannen der informativen und nichtinformativen Bayesschen Regression eng um die Posteriori-Schätzwerte der Zielgrößen, umfassen die HPDIs der Modelle *K.1* und *K.2* einen Bereich von mehreren Hundert Euro. Der Median der Spannen liegt für die Epoche 2005 bei 413 €/m² für die informative und 416 €/m² für die nichtinformative Bayessche Kollokation. Im Vergleich ergeben sich für die Epoche 2009 Werte von 458 €/m² zu 493 €/m² und 2014 440 €/m² zu 530 €/m². Entgegen den vorherigen Ergebnissen weisen die Spannen der HPDIs der Kollokationsansätze im direkten Vergleich geringere Differenzen als die Modelle *R.1* und *R.2* auf. Teilweise liegt die Spanne der nichtinformativen Lösung unter derjenigen der informativen, wie das Beispiel der Epoche 2005 sowie 2010 bis 2012 in Anhang C.2 (S. 221 ff.) zeigt und damit die Erkenntnisse aus der Modellierung der Signalkomponente bestätigt. Eine zusätzliche Übersicht über die Entwicklung der Medianwerte der HPDIs pro Modell und Auswerteepoche gibt Abbildung 5.45, welche die einzelnen Auswertungen zusammenfasst.

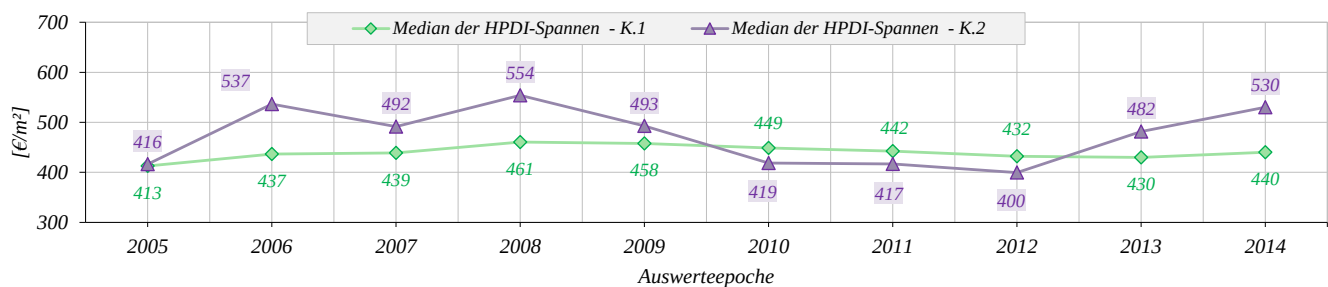


Abbildung 5.45: Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{y}$ der Epochen 2005 bis 2014, Bayessche Kollokation

Zu erkennen ist die bereits bei den Ansätzen der Bayesschen Regression festgestellte stetige Entwicklung in den Spannen der Konfidenzregionen für die informativen Lösungen: Es zeigt sich ein annähernd gleichbleibender Trend mit einem Minimalwert von 413 €/m² in 2005. Entgegen den vergleichbaren Ergebnissen der Bayesschen Regression liegen die Ergebnisse für die informative Lösung jedoch nicht durchgängig unter denjenigen der nichtinformativen Lösung, wie es durch die Integration von Priori-Information zu erwarten wäre: In den Epochen 2005 sowie 2010 bis 2012 ergeben sich – wie bereits bei der Betrachtung des Unsicherheitshaushaltes der Signalkomponente festgestellt – kleinere Wahrscheinlichkeitsbereiche für die Schätzwerte. Die Integration von Vorwissen führt demnach nicht in jedem Fall zu einer verbesserten Schätzung – mögliche Ursachen werden im Zuge der Ergebnisdiskussionen in Kapitel 6 aufgegriffen. Dennoch zeigt sich auch anhand der Ergebnisse der Bayesschen Kollokation, dass die HPDIs der nichtinformativen Lösung *K.2* stärker variieren. Eine Tendenz in der zeitlichen Entwicklung ist nicht zu erkennen. Es ergibt sich für 2008 ein Maximalwert mit 554 €/m². Im direkten Vergleich ergibt sich durch die Nutzung von Priori-Information somit eine dämpfende Wirkung auf die Wertebereiche der HPDIs für den Fall hoher Streuung in den Daten. Die Minimalwerte liegen für den nichtinformativen Fall in der Epoche 2005 (416 €/m²), 2010 (419 €/m²), 2011 (417 €/m²) und 2012 (400 €/m²).

Für die Wahrscheinlichkeitsbereiche ist anhand der Abbildungen 5.30 bis 5.32 (S. 141) festzustellen, dass die Kaufpreise in den einzelnen Epochen überwiegend innerhalb der HPDIs liegen. Hier ergibt sich wiederum ein Gegensatz zu den Ergebnissen der Bayesschen Regression, durch die zwar Schätzwerte mit einer sehr geringen Unsicherheitsspanne ermittelt werden können, jedoch den jeweiligen Kaufpreis als Zielgröße nicht in den Konfi-

denzregionen beinhalten. Die erweiterte Unsicherheitsspanne – welche insbesondere bedingt ist durch den Einfluss der Signalkomponente – trägt somit zu Schätzergebnissen mit einem plausibleren Bezug zu den realen Beobachtungen bei. Auswirkungen und weitere Diskussionen zu dieser Thematik finden sich in der Ergebnisdiskussion in Kapitel 6.

Zusammenfassend lässt sich festhalten, dass die informative Bayessche Regression (*R.1*) im Vergleich zur nicht-informativen Variante (*R.2*) eine deutliche Verringerung der Unsicherheiten sowohl in den Modellparametern als auch den Posteriori-Schätzwerten der Zielgröße mit sich bringt. Im Gegensatz hierzu liegen die Ergebnisse der informativen und nichtinformativen Bayesschen Kollokation (*K.1* und *K.2*) dichter beieinander; eine pauschale Aussage über die Verbesserung des Unsicherheitshaushaltes kann zunächst nicht getroffen werden und ist stark abhängig von der Streuung in den Daten. Überwiegend zeigen die Schätzwerte unter Einbeziehung von Priori-Information jedoch einen verbesserten Unsicherheitshaushalt – sowohl im Hinblick auf die Modellparameter (Trendkoeffizienten und Varianzfaktor) als auch die Posteriori-Schätzwerte der Zielgröße. Der Einfluss des Signals, welches selbst kaum durch die Integration von Priori-Wissen beeinflusst wird, kann als wesentliche Komponente für die Differenzen der Unsicherheitshaushalte im Quervergleich Bayesscher Regression und Bayesscher Kollokation interpretiert werden. Wie zu erwarten, zeigen sich größere Unsicherheiten als Resultat der Berücksichtigung eines erweiterten stochastischen Modells.

Wurden bislang ausschließlich die funktionale Modellierung und die damit verbundenen Unsicherheiten behandelt, stellt sich darüber hinaus die Frage, wie sich die Ergebnisse auf die Prädiktion von neuen Beobachtungen und insbesondere deren Unsicherheiten auswirken. Im Zuge der Modellvalidierung wird diese Thematik im folgenden Abschnitt behandelt, um die praktischen Untersuchungen abzuschließen.

5.4. Validierung der Modelle

Durch die Untersuchungen in den vorherigen Abschnitten wird anhand einer Stichprobe von Kaufpreisdaten die Übertragbarkeit der in Kapitel 4 entwickelten Auswertestrategie auf praktische Anwendungen demonstriert. Zusammenfassend wird gezeigt, wie sich im Kontext der Bayes-Strategie die Integration von Priori-Wissen in die Modellansätze der Regression und Kollokation auf die funktionale Modellierung und die resultierenden Unsicherheiten auswirkt. Für die hier thematisierte Übertragung auf das Vergleichswertverfahren ist es jedoch nicht nur von Interesse, die reine Modellbildung auf Basis der Stichproben abzubilden, sondern darüber hinaus Schätzwerte zu präzisieren. In praktischen Anwendungen der Verkehrswertermittlung werden die mathematisch-statistischen Modelle genutzt, um Vergleichswerte zu präzisieren, die der definierten Grundgesamtheit angehören, für die jedoch keine Kaufpreise, sondern lediglich die wesentlichen wertbeeinflussenden Merkmale vorliegen. Erst durch die Prädiktion wird das formulierte Vergleichswertverfahren für die praktische Anwendung sinnvoll nutzbar. Um die Effekte der Prädiktion beurteilen zu können, sollen die Modelle der Bayesschen Regression und Bayesschen Kollokation in einem Kreuzvalidierungsansatz genutzt werden, wie er in Kapitel 4.4.1 (S. 105 f.) beschrieben wurde. In den Abschnitten 5.4.1 und 5.4.2 werden die Prädiktionsergebnisse vorgestellt.

5.4.1. Prädiktionsergebnisse der Bayesschen Regression

Die Grundlagen der prädiktiven Schätzungen in Bayesschen Modellen wurden in Abschnitt 3.2.2 (S. 60 ff.) gegeben. Die Posteriori-Schätzung von Vergleichswerten $\bar{\mathbf{y}}_p$ erfolgt demnach anhand einer prädiktiven Dichte, deren Erwartungswert $E(\mathbf{y}_p | \sigma^2, \mathbf{y})$ durch (3.61) spezifiziert ist. Die zugehörige VKM $\Sigma_{\bar{\mathbf{y}}_p \bar{\mathbf{y}}_p}$ ergibt sich ergänzend durch (3.62). Die gegebenen Gleichungen sind sowohl für die informative als auch die nichtinformativ Lösung der Bayesschen Regression gültig, wobei zu beachten ist, dass die korrespondierenden Trendkoeffizienten $\bar{\beta}$ und Kofaktormatrizen $\mathbf{Q}_{\bar{\beta}\bar{\beta}}$ zu verwenden sind, die sich im Zuge der Kreuzvalidierung für jede Iteration sowie getrennt nach informativem und nichtinformativem Ansatz ergeben.

Wird das Prinzip der Kreuzvalidierung auf die einzelnen Auswertepochen angewendet, ergeben sich die für die exemplarischen Epochen 2005, 2009 und 2014 in den Abbildungen 5.46 bis 5.48 dargestellten Ergebnisse. In den oberen Zeilen finden sich die prädierten Schätzwerte der informativen und nichtinformativen Lösungen *R.1* und *R.2* im direkten Vergleich. Unterhalb der Abbildungen sind die Differenzplots der HPDIs dargestellt, die sich aus

dem Verhältnis $R.2$ zu $R.1$ ergeben. Anhang B.3 (S. 215 ff.) enthält darüber hinaus die Abbildungen der übrigen Auswertepochen.

Bei der funktionalen Modellierung zeigt sich wie erwartet, dass durch die Prädiktion keine größeren Abweichungen in den Schätzwerten auftreten, was auf die numerischen annähernd gleichen Trendmodelle in der Kreuzvalidierung zurückzuführen ist. In den drei dargestellten Epochen ergeben sich daher Differenzen zwischen prädizierten Posteriori-Schätzwerten $\bar{y}_p$ und Kaufpreisen, die denen in Bezug auf die Posteriori-Schätzungen der Zielgröße $\bar{y}$ gleichen (vgl. Abbildungen 5.7 bis 5.9, S. 128). Werden die Medianwerte der prozentualen Abweichungen über die Epochen 2005 bis 2014 betrachtet, zeigen sich im direkten Vergleich der Abbildungen 5.49 (für die Prädiktion) und 5.13 (S. 130, für die Posteriori-Schätzungen der Zielgröße $\bar{y}$) sowohl äquivalente Verläufe der informativen und nicht-informativen Lösungen der Modelle $R.1$ und $R.2$ als auch numerisch nahezu gleiche prozentuale Abweichungen. Da lediglich jeweils ein Kauffall für die Prädiktion aus der Stichprobe entfernt wird, kann die Prädiktionsgüte für die funktionale Modellierung von $\bar{y}_p$ mit der Schätzung von $\bar{y}$ gleichgesetzt werden. Aus diesem Grund kann auf eine Einzeldarstellung der prozentualen Abweichungen pro Epoche verzichtet werden. Eine weiterführende Diskussion bezüglich der Bedeutung für die Verkehrswertermittlung erfolgt in Kapitel 6.

Ungleich höhere Unterschiede zum Ausgangsmodell basierend auf den Gesamtstichproben ergeben sich bei Betrachtung der Konfidenzregionen. Es lässt sich in den Abbildungen 5.46 bis 5.48 erkennen, dass die HPDIs in den einzelnen Epochen einen wesentlichen größeren Wahrscheinlichkeitsbereich abdecken, als es in den Ergebnissen nach den Abbildungen 5.7 bis 5.9 (S. 128) der Fall ist. Neben den vergrößerten Unsicherheitsbereichen sowohl in den Lösungen nach Modell $R.1$ als auch Modell $R.2$ ist darüber hinaus festzustellen, dass die Differenzplots größere Unterschiede in den HPDIs zwischen informativer und nichtinformativer Lösung aufweisen. Diese Beobachtung kann über den gesamten Auswertzeitraum bestätigt werden und zeigt, dass die Prädiktion der informativen Bayesschen Regression geringere Unsicherheiten aufweist als die der nichtinformativen.

Werden in Bezug auf die absoluten Spannen die Beispielpochen 2005, 2009 und 2014 verglichen, ist zu sehen, dass in Abhängigkeit der Heterogenität in den Ausgangsdaten die HPDIs von Auswertepoche zu Auswertepoche variieren; während die Bereiche der Epochen 2005 und 2009 vergleichbar sind und in Bezug auf die Medianwerte für die informative Lösung im Bereich von ca. 800 €/m² und für die nichtinformativ Lösung im Bereich von ca. 900 €/m² liegen, weicht der Wertebereich der Epoche 2014 in der nichtinformativen Lösung mit ca. 1.160 €/m² stark ab. Für die informative Lösung ist hingegen zu beobachten, dass diese mit wiederum etwa 800 €/m² den Wahrscheinlichkeitsbereichen der übrigen Epochen gleicht; hier ergeben sich entsprechend die größten Unterschiede im Differenzplot der HPDIs mit einem Maximalwert von ca. 475 €/m². Werden die Entwicklungen über den gesamten Auswertzeitraum betrachtet, ergeben sich die in Abbildung 5.50 dargestellten Verläufe. Wie schon anhand der exemplarischen Epochen zu erkennen ist, kann in den informativen Modellen $R.1$ durch die Nutzung der Priori-Information eine konstante Entwicklung in den Medianwerten der absoluten HPDIs erreicht werden. Dieses ist für die nichtinformativen Lösungen nicht der Fall, die Konfidenzregionen variieren in ihrem Verlauf: Im Minimum lässt sich in der Epoche 2012 ein Wert von 845 €/m² ermitteln, das Maximum liegt in der Epoche 2014 mit vorgenanntem Betrag. Eine Systematik ist nicht zu erkennen und lässt sich wiederum durch die Sensibilität auf Streuungen im aktuell ausgewerteten Datensatz erklären.

Die abschließende Wertung der Prädiktionsergebnisse sowie die Verknüpfung zur Relevanz in der Verkehrswertermittlung erfolgt in Kapitel 6. Im folgenden Abschnitt werden zunächst die Prädiktionsergebnisse der Bayesschen Kollokation behandelt.

5.4.2. Prädiktionsergebnisse der Bayesschen Kollokation

Die prädiktive Schätzung der Posteriori-Zielgröße vollzieht sich für die Bayessche Kollokation nach den Grundlagen des Abschnitts 3.3.2 (S. 67). Anhand der prädiktiven Dichte für die Beobachtungen $\bar{y}_p$ lässt sich der Erwartungswert $E(\bar{y}_p | \sigma^2, \mathbf{y})$ durch (3.118) formulieren, die zugehörige VKM $\Sigma_{\bar{y}_p \bar{y}_p}$ mittels (3.120). Die gegebenen Gleichungen sind nach der Verfügbarkeit von Priori-Information zu unterscheiden, sodass sich durch entsprechende Anwendung Lösungen für den informativen und den nichtinformativen Ansatz erhalten lassen. Wie im Fall der Bayesschen Regression wird für die Prädiktion das Prinzip der Kreuzvalidierung angewendet, sodass sich zu jedem Kauffall der Stichproben ein prädizierter Schätzwert ermitteln lässt. Signifikanter Unterschied zu den bisherigen Ergebnissen

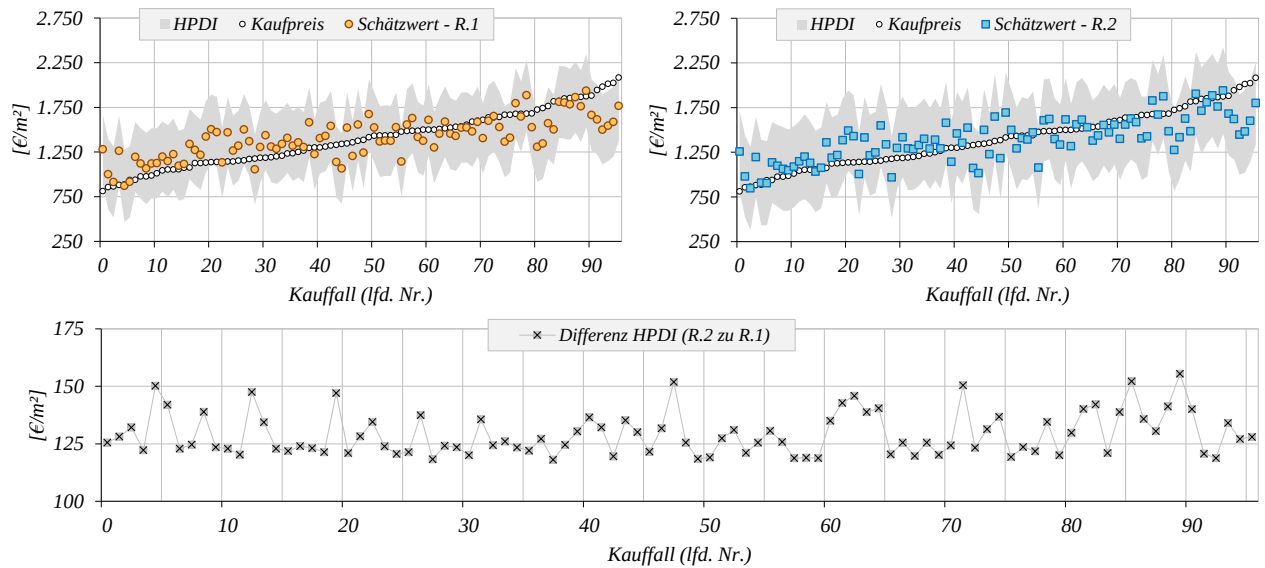


Abbildung 5.46: Posteriori-Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

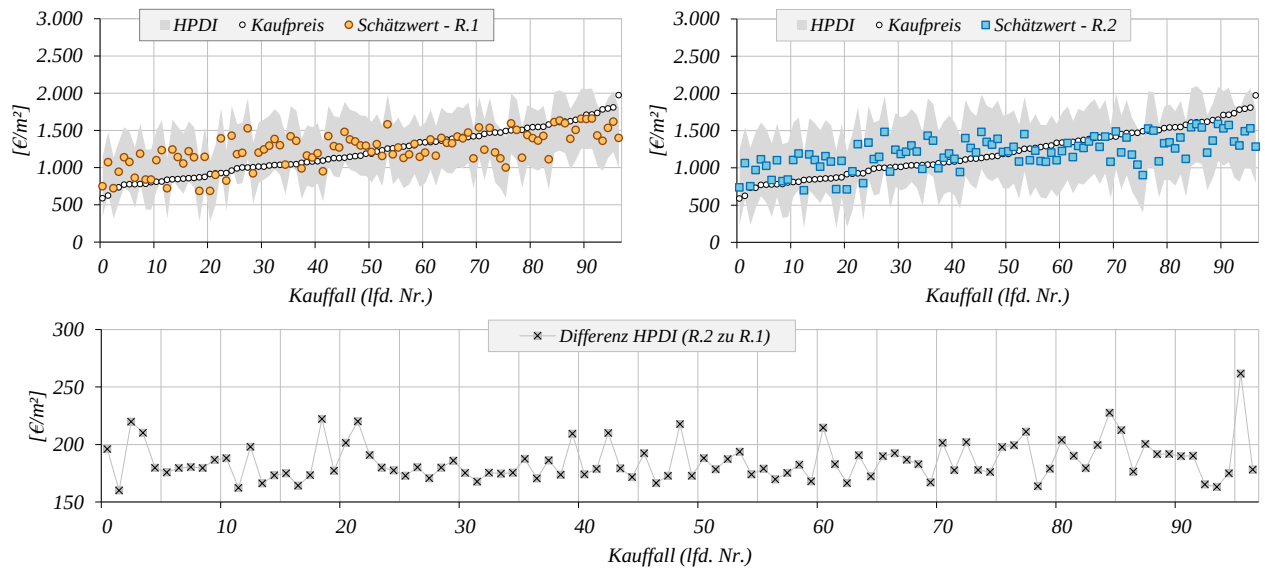


Abbildung 5.47: Posteriori-Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

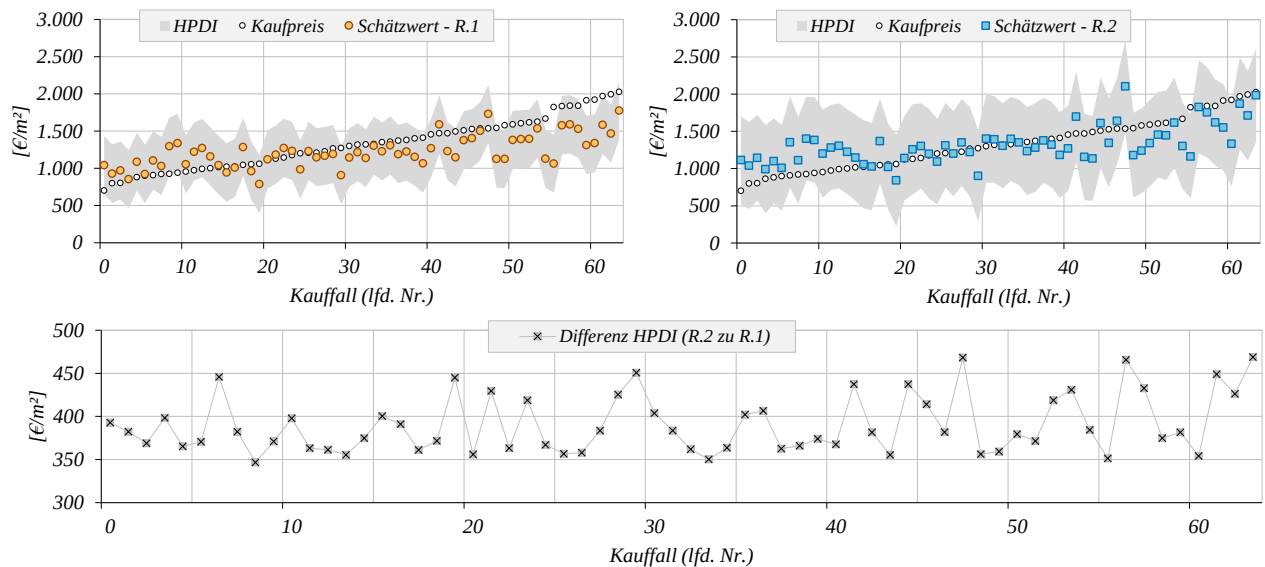


Abbildung 5.48: Posteriori-Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

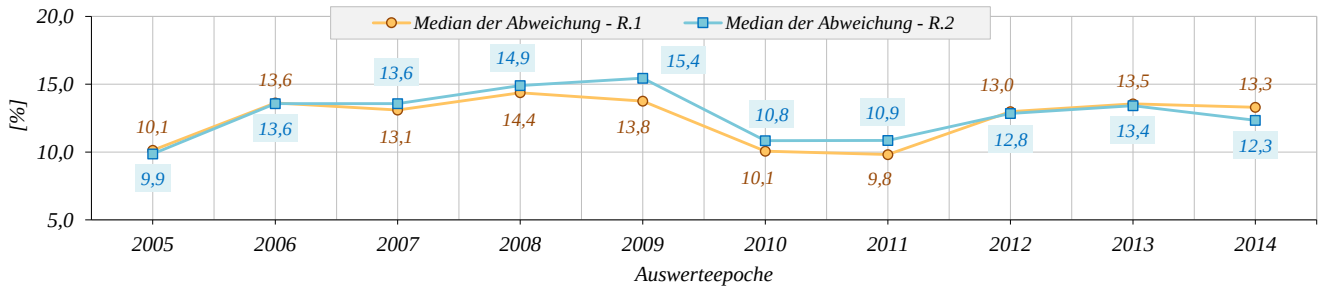


Abbildung 5.49: Prozentuale Abweichungen zwischen Kaufpreisen und prädizierten Posteriori-Schätzwerten $\bar{y}_p$ in den Epochen 2005 bis 2014, Bayessche Regression

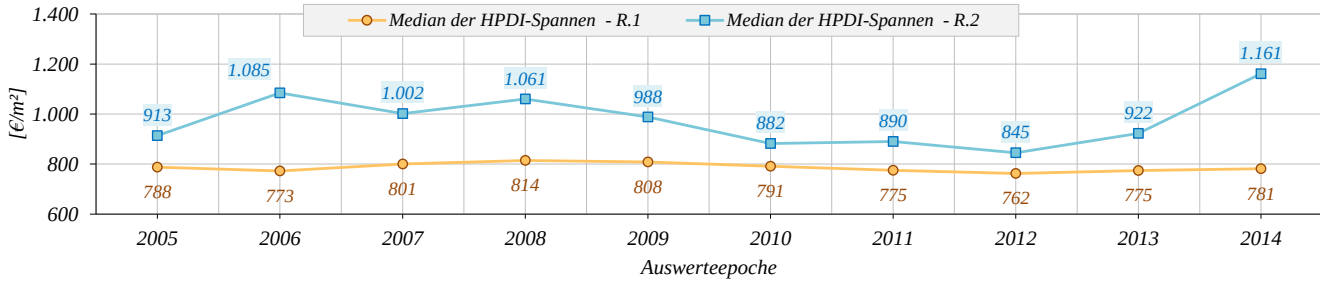


Abbildung 5.50: Median der HPDI-Spannen der Posteriori-Schätzwerte $\bar{y}_p$ in den Epochen 2005 bis 2014, Bayessche Regression

ist die Verwendung einer Signalmatrix für das prädizierte Signal, die durch die Kofaktormatrix $Q_{s_p s_p}$ Eingang in die Berechnungen findet. Unter diesen Voraussetzungen können prädizierte Posteriori-Schätzwerte bestimmt werden, deren Ergebnisse sich für die exemplarischen Epochen 2005, 2009 und 2014 in den Abbildungen 5.51 bis 5.53 (Signalkomponente $\bar{s}_p$) sowie den Abbildungen 5.54 bis 5.56 (Zielgröße $\bar{y}_p$) finden. Für die Darstellung der Schätzungen der übrigen Epochen wird auf Anhang C.4 (S. 234 ff.) verwiesen.

Werden zunächst die Schätzungen der prädizierten Signalkomponente $\bar{s}_p$ betrachtet, ist im Gegensatz zu den Schätzungen der Bayesschen Kollokationen basierend auf den Gesamtstichproben zu erkennen, dass keine Systematiken vorliegen: Im Vergleich zu den vorherigen Schätzungen lassen die Signalkomponenten sich bezüglich ihres Vorzeichens nicht nach niedrig- oder hochpreisigen Segmenten kategorisieren. Zudem ist auffällig, dass sich bei einer direkten Gegenüberstellung der Auswertungen der Einzelepochen die absolute Höhen der einzelnen prädizierten Signalkomponenten $\bar{s}_p$ numerisch geringer ergeben, als es bei der Schätzung von $\bar{s}$ der Fall war (vgl. Abbildungen 5.27 bis 5.29, S. 140). Die Unterschiede können für sämtliche Auswerteepochen 2005 bis 2014 bestätigt werden. Weiterhin zeigen sich größtenteils keine sich ähnelnden Schätzungen basierend auf dem informativen Modell $K.1$ und dem nichtinformativen Modellen $K.2$. Dieser Aspekt ist bislang auffällig gewesen im Zuge der Schätzungen basierend auf den Gesamtstichproben. Aus den Darstellungen können zunächst zwei Schlüsse gezogen werden:

1. Es ergeben sich für die Prädiktion der Signalkomponente $\bar{s}_p$ größere Abweichungen zwischen den Lösungen des Modells $K.1$ und des Modells $K.2$. Es kann daher gefolgert werden, dass die Prädiktion der Signalkomponente sensibel auf die Verfügbarkeit von Vorinformation reagiert.
2. Aufgrund der im Vergleich zu den Posteriori-Schätzungen von $\bar{s}$ geringeren absoluten Höhen der einzelnen prädizierten Signalkomponenten $\bar{s}_p$ in den Auswerteepochen 2005 bis 2014 ist eine funktionale Approximationsgüte für $\bar{y}_p$ zu erwarten, die unter der für $\bar{y}$ liegt.

Wird zunächst der zweite Aspekt betrachtet, bestätigt sich die Annahme bei Betrachtung der Abbildungen 5.54 bis 5.56. Entgegen der Erwartungen – die sich aus den guten Approximationsergebnissen der Modelle $K.1$ und $K.2$ für die Posteriori-Schätzwerte $\bar{y}$ ergeben haben – zeigt die Prädiktion eine Approximationsgüte, die denen der Prädiktionen mittels der Bayesschen Regressionsmodelle $R.1$ und $R.2$ entspricht. Die Ausschöpfung zusätzlicher stochastischer Informationen für eine verbesserte funktionale Modellierung ist nicht unmittelbar zu erkennen. Die Medianwerte der prozentualen Abweichungen zwischen den prädizierten Werten und den Kaufpreisen betragen in

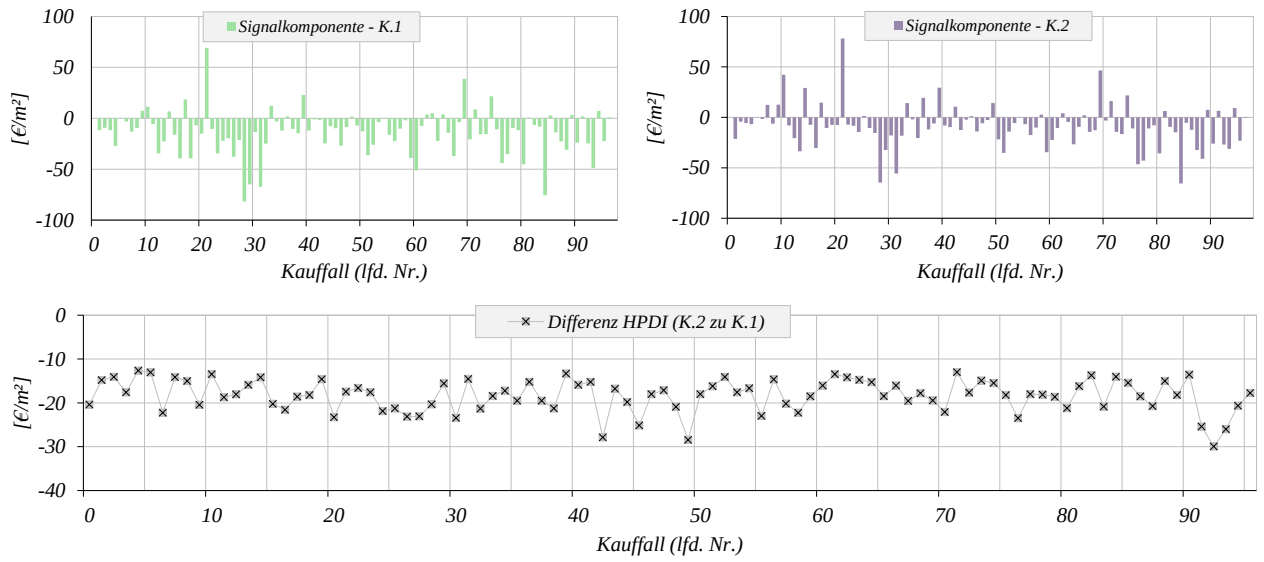


Abbildung 5.51: Posteriori-Schätzwerte $\bar{s}_p$, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

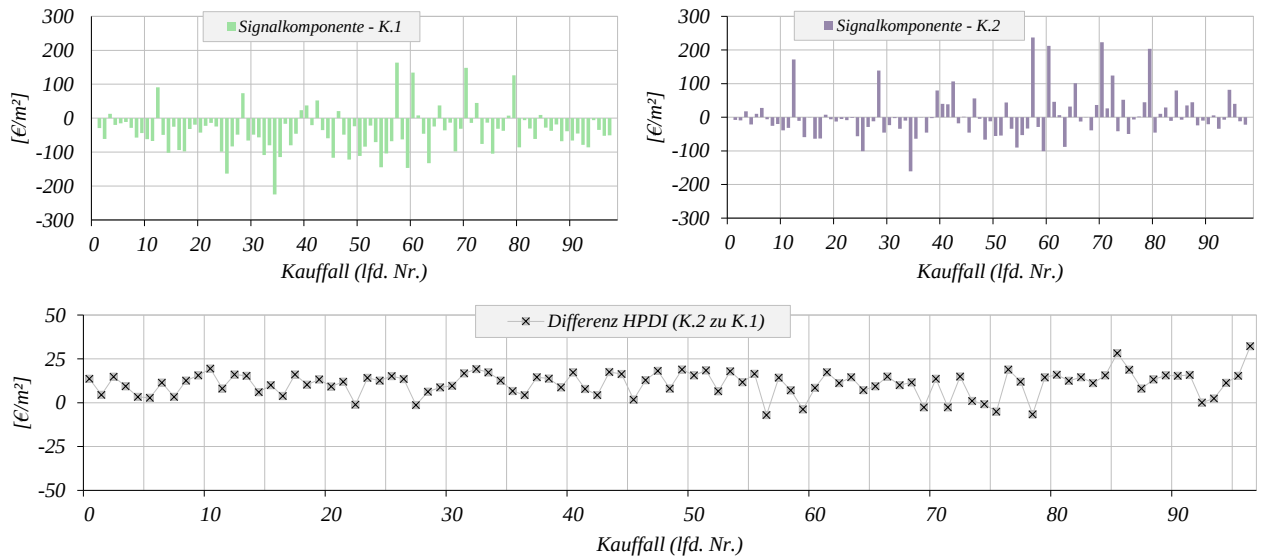


Abbildung 5.52: Posteriori-Schätzwerte $\bar{s}_p$, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

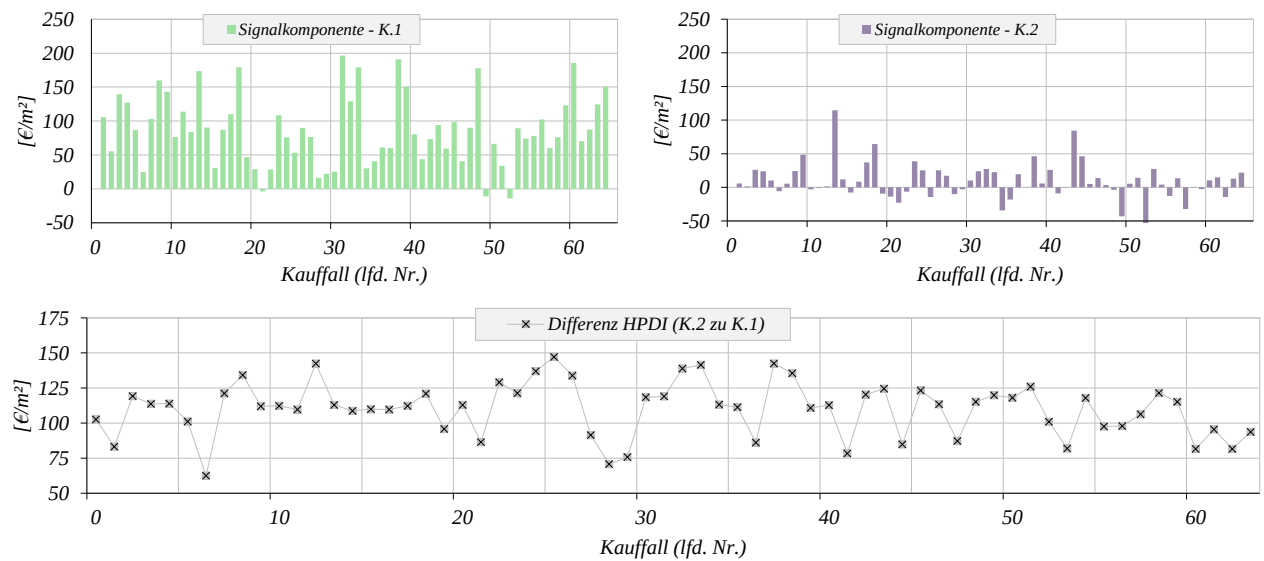


Abbildung 5.53: Posteriori-Schätzwerte $\bar{s}_p$, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

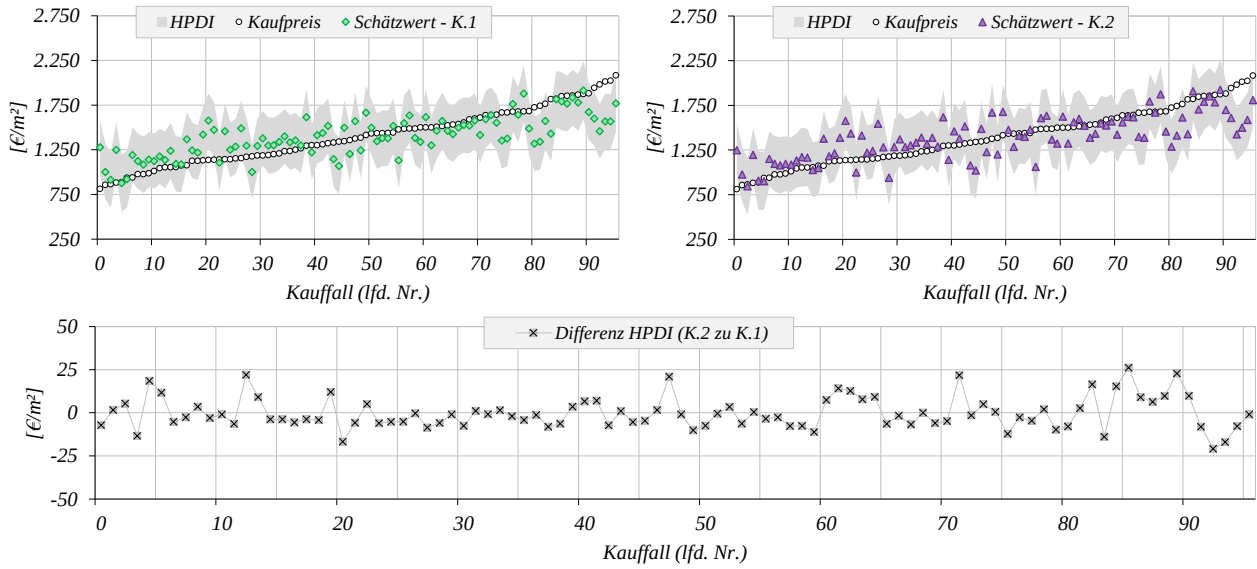


Abbildung 5.54: Posteriori-Schätzwerte $\bar{y}_p$ der Bayesschen Kollokation, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

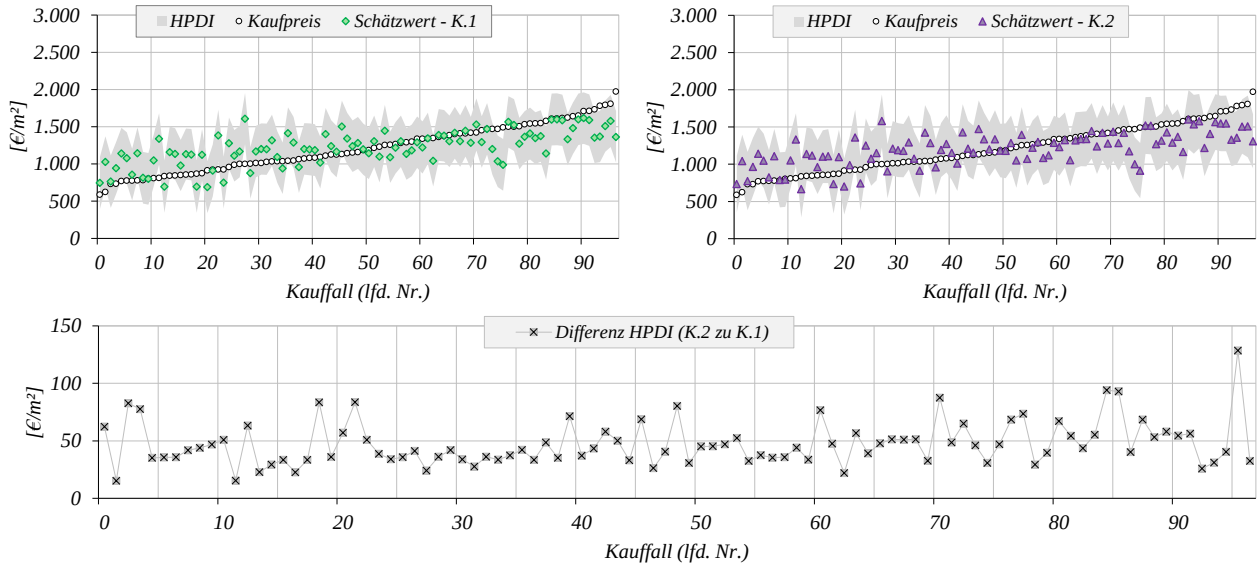


Abbildung 5.55: Posteriori-Schätzwerte $\bar{y}_p$ der Bayesschen Kollokation, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

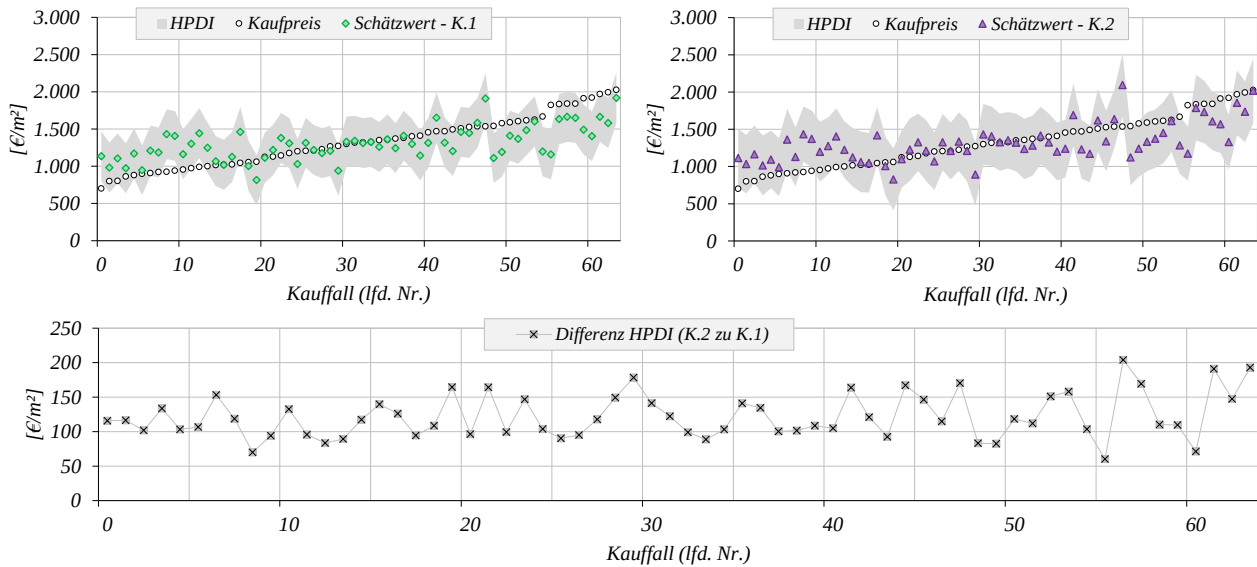


Abbildung 5.56: Posteriori-Schätzwerte $\bar{y}_p$ der Bayesschen Kollokation, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

der Epoche 2005 10,6 % für die informative und 10,1 % für die nichtinformative Lösung. In den Epochen 2009 und 2014 liegen die Differenzen bei 12,2 % zu 14,2 % und 11,3 % zu 12,8 %. Ähnliche Resultate zeigen sich bei der Betrachtung der prozentualen Entwicklung im gesamten Auswertezeitraum 2005 bis 2014, die mit Abbildung 5.57 gegeben ist. Werden die Werte mit den prozentualen Abweichungen basierend auf den Auswertungen der Gesamtstichproben verglichen (vgl. Abbildung 5.36, S. 142), wird eine deutliche Differenz in der funktionalen Anpassung deutlich: Im Quervergleich weichen die Prädiktionsergebnisse $\bar{y}_p$ der informativen Lösungen K.1 über die Auswertepochen 2005 bis 2014 um +3,6 % von den Schätzergebnissen $\bar{y}$ der entsprechenden informativen Lösungen basierend auf der Auswertung der Gesamtstichproben ab. Für die nichtinformativen Lösungen ist im Schnitt eine Differenz von +4,2 % festzustellen.

Die Streuung der prädierten Schätzwerte ist ebenfalls anhand der Abbildungen 5.54 bis 5.56 zu erkennen; die Ergebnisse sind als repräsentativ für den Gesamtzeitraum der Auswertepochen zu sehen, wie den Darstellungen für die übrigen Epochen entnommen werden kann (vgl. Anhang C.4, S. 234). Die prädierten Schätzwerte ähneln in ihrer absoluten Höhe den Prädiktionsergebnissen der Bayesschen Regressionen, informativ wie nichtinformativ. Bestätigt wird dies durch den Vergleich der Medianwerte der prozentualen Abweichungen (Abbildung 5.49, S. 151 zu 5.57). Größere Differenzen sind nicht festzustellen und lassen den Schluss zu, dass sich Unterschiede in der Approximationsgüte in einem für die Verkehrswertermittlung nicht relevanten Bereich bewegen. In Verbindung mit dem oben angeführten ersten Aufzählungspunkt bezüglich der Erkenntnisse zur Schätzung der prädierten Signalkomponente $\bar{s}_p$ lässt sich weiterhin festhalten, dass die absolute Höhe der Signalkomponenten zwar zwischen informativer und nichtinformativer Lösung stark schwanken, die Unterschiede jedoch durch die Trendmodellierung größtenteils ausgeglichen werden. So ergibt sich im Endeffekt eine annähernd gleiche Approximationsgüte.

Für die funktionale Modellierung lässt sich zusammenfassen, dass zwar durch die Kollokationsansätze eine verbesserte funktionale Approximation erreicht werden kann, sich die Vorteile der Berücksichtigung einer zusätzlichen Signalkomponente hingegen nicht unmittelbar auf Prädiktionsaufgaben übertragen lassen. Die Diskussionen über die Ursachen sowie über Auswirkungen auf die Verkehrswertermittlung werden in Kapitel 6 geführt. An dieser Stelle kann festgehalten werden, dass sich im Vergleich zu den Ergebnissen der Prädiktion in Bayesschen Regressionsmodellen eine größtenteils vernachlässigbare Differenz in der funktionalen Approximation zeigt.

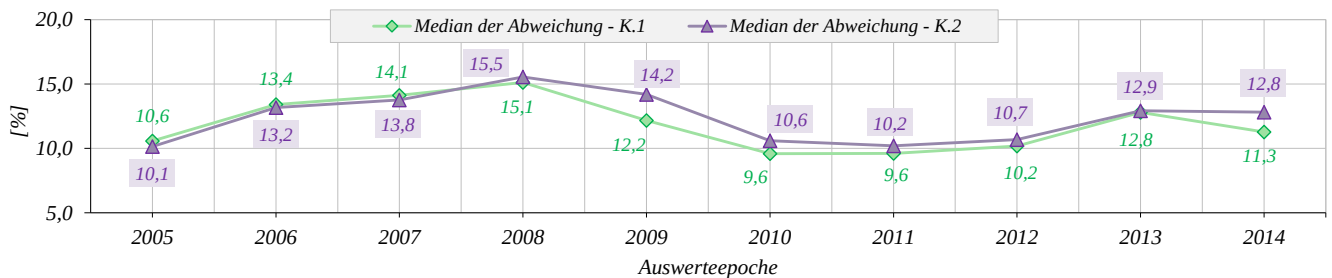


Abbildung 5.57: Prozentuale Abweichungen zwischen Kaufpreisen und prädierten Posteriori-Schätzwerten $\bar{y}_p$ in den Epochen 2005 bis 2014, Bayessche Kollokation

Für eine Betrachtung der Unsicherheiten sind in einem ersten Schritt die Differenzplots für die HPDIs der Signalkomponente $\bar{s}_p$ zu betrachten (vgl. Abbildungen 5.51 bis 5.53, S. 152). Hier bestätigt sich das Bild, welches sich bereits bei der Schätzung von $\bar{s}$ gezeigt hat: Während in der Epoche 2005 eine größere Unsicherheit in der informativen Lösung zu erkennen ist, wird in den Epochen 2009 und 2014 durch die Integration von Priori-Information eine Verbesserung der HPDIs erreicht. Die Größenordnungen der Differenzen sind weiterhin vergleichbar mit den jeweiligen Ergebnissen, die unter Berücksichtigung der Auswertungen der Gesamtstichproben erreicht werden konnten (vgl. Abbildungen 5.27 bis 5.29, S. 140). Aufgrund der äquivalenten Vorinformation über die Trendkoeffizienten, die als einzige Priori-Information einfließen, ist dieses Ergebnis zu erwarten. Das gleiche Bild ergibt sich für die übrigen Epochen, wie Anhang C.4 (S. 234) entnommen werden kann. Auch die Erkenntnis, dass die Unsicherheitsmaße der nichtinformativen Lösungen neben 2005 auch in den Epochen 2010 bis 2014 als geringer gegenüber denen der informativen Lösungen einzustufen sind, wird bestätigt. In Abbildung 5.58 werden als ergänzende Information die Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{s}_p$ der Epochen 2005 bis 2014 dargestellt. Die Werte

liegen über den vergleichbaren Werten der Schätzung von $\bar{s}$ (vgl. Abbildung 5.44, S.147), zeigen jedoch den gleichen Verlauf innerhalb der einzelnen Modelle *K.1* und *K.2* sowie in ihrem direkten Vergleich. Eine erhöhte Unsicherheit ist zu erwarten und bedingt durch die Verwendung prädiktiver Dichten.

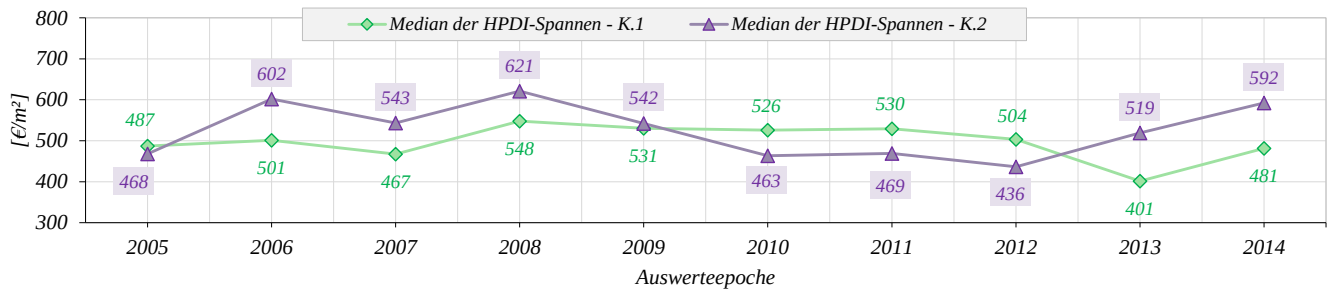


Abbildung 5.58: Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{s}_p$ in den Epochen 2005 bis 2014

In Erweiterung der Darstellung der HPDIs in Form von deren Differenzen lassen sich die Konfidenzregionen der Prädiktionswerte y_p beurteilen. Wie die Abbildungen 5.54 bis 5.56 für die Epochen 2005, 2009 und 2014 zeigen, ergeben sich im Vergleich zu den Ergebnissen der Prädiktion in den Bayesschen Regressionsmodellen erheblich schmalere HPDIs. Dies bedeutet, dass die Wahrscheinlichkeitsbereiche für die Schätzwerte und damit die resultierenden Vergleichswerte auf einen wesentlich geringeren Wertebereich eingegrenzt werden können. In diesem Zusammenhang zeigen sich im informativen Ansatz durch die Integration von Priori-Information größtenteils geringere Unsicherheitsspannen, wie den Differenzplots entnommen werden kann. Wie auch zuvor und den Abbildungen der übrigen Epochen in Anhang C.4 (S. 234) entnommen werden kann, bilden die Epochen 2005 sowie 2010 bis 2012 die Ausnahme: Die Unsicherheitsmaße liegen gemäß der Differenzplots für die nichtinformativen Ansätze unterhalb derer für die informativen Ansätze. Ergänzend lassen sich Abbildung 5.59 die Entwicklungen der Medianwerte der HPDI über den gesamten Auswertez Zeitraum entnehmen. Zu erkennen sind die Epochen, in denen die nichtinformativen Ansätze zu geringeren HPDIs führen.

Wird ein direkter Vergleich der exemplarischen Epochen 2005, 2009 und 2014 geführt, liegt der Medianwert der Konfidenzregionen der Bayesschen Kollokation im Modell *K.1* in der Epoche 2005 mit 617 €/m² insgesamt 168 €/m² unterhalb des Wertes der Bayesschen Regression im Modell *R.1*. Im nichtinformativen Fall *K.2* zu *R.2* beträgt die Differenz 296 €/m² zugunsten der Bayesschen Kollokation. Für die Epochen 2009 und 2014 ergeben sich ebenfalls Verbesserungen: Diese betragen 2009 116 €/m² für die informative bzw. 253 €/m² für die nichtinformativ Variante, in 2014 110 €/m² bzw. 374 €/m². Für die Unsicherheitsmodellierung in der Prädiktion ist somit eine Verbesserung durch den Bayesschen Kollokationsansatz zu erkennen, unabhängig von der Verwendung des Modells *K.1* oder *K.2*. Zusätzlich ist die Differenz zwischen den HPDIs der Schätzungen im Prädiktionsansatz und den jeweiligen Schätzungen basierend auf den Gesamtstichproben wesentlich geringer als im Fall der Bayesschen Regressionen, was für die erweiterte stochastische Modellierung im Kollokationsansatz spricht. Weiterhin ist anzumerken, dass die Kaufpreise nahezu vollständig innerhalb der HPDI liegen und so durch die Prädiktion eine realistische Angabe für Unsicherheitsbereich der Vergleichswerte erreicht wird, innerhalb derer die Kaufpreise mit hoher Wahrscheinlichkeit liegen. Im Vergleich zur Bayesschen Regression konnten die Unsicherheitsbereiche jedoch signifikant verringert werden.

Entgegen der Ergebnisse der Bayesschen Regression lassen sich in der Entwicklung der Medianwerte der informativen und der nichtinformativen Lösung über den Gesamtzeitraum keine größeren Unterschiede erkennen (vgl. Abbildung 5.59). Wie auch zuvor variieren die Ergebnisse der nichtinformativen Lösungen stärker, während durch die Integration von Priori-Wissen eine kontinuierliche Entwicklung in vergleichbarer Größenordnung erreicht wird. Die Unterschiede in den Spannen der Konfidenzregionen fallen jedoch geringer aus als in der Bayesschen Regression. Ein Erklärungsansatz ist, dass die Ergebnisse der informativen Lösung ausschließlich durch die Vorinformation über die Trendkoeffizienten beeinflusst werden. Für die zweite wesentliche Komponente des stochastischen Modells der Kollokationsansätze – der Signalkomponente – entsprechen sich die Modellierungen im informativen und nichtinformativen Ansatz. Insofern ist zu erwarten, dass sich die Ergebnisse annähern. Weiterführende Diskussionen zur Unsicherheitsmodellierung können Kapitel 6 entnommen werden.

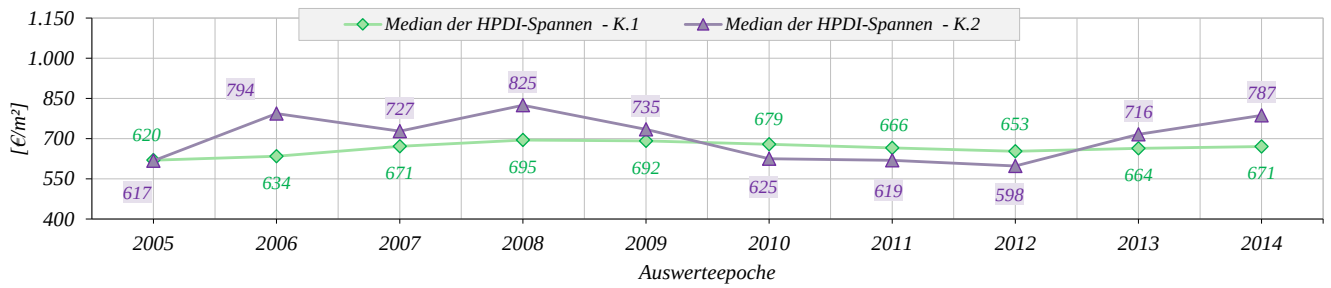


Abbildung 5.59: Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{y}_p$ in den Epochen 2005 bis 2014, Bayessche Kollokation

Mit der rekursiven Anwendung der Auswertestrategie sowie der Berechnung der Ergebnisse für die Kreuzvalidierung in den vier Modellen der informativen und nichtinformativen Bayesschen Regression und Bayesschen Kollokation werden die praktischen Untersuchungen abgeschlossen. Zusammenfassend lässt sich festhalten, dass sich die spezifischen Anforderungen der Verkehrswertermittlung nach Kapitel 2 mit der Theorie der Bayes-Strategie nach Kapitel 3 zu einer umfassenden Auswertestrategie nach Kapitel 4 verknüpfen lassen, die in der praktischen Anwendung für die Modellierung im Vergleichswertverfahren genutzt werden kann. Wie die Untersuchungen in den vorherigen Abschnitten zeigen, lassen sich die Bayesschen Modellansätze sinnvoll in die Vergleichswertermittlung integrieren. Die Diskussion der wesentlichen Ergebnisse sowie der praktische Bezug zur Verkehrswertermittlung und weiteren Anwendungen werden im folgenden Kapitel ergänzt.

6. Mehrwert Bayesscher Ansätze im Vergleich

Durch die Verknüpfung der Rahmenbedingungen der Verkehrswertermittlung mit der Bayes-Theorie in einer Auswertestrategie wurden in Form der Bayesschen Regression und der Bayesschen Kollokation Modellansätze formuliert, welche sich für die Schätzung und Prädiktion von Vergleichswerten nutzen lassen. Um eine abschließende Beurteilung der eingangs formulierten Forschungsziele zu ermöglichen, werden nachfolgend die Ergebnisse diskutiert, die durch Anwendung der Auswertestrategie auf eine exemplarische Stichprobe erhalten wurden. Abbildung 6.1 zeigt schematisch das Konzept der Ergebnisdiskussion.

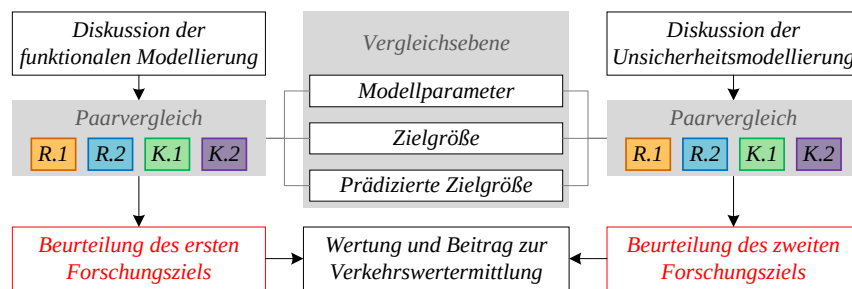


Abbildung 6.1: Konzept der Ergebnisdiskussion

In Bezug auf die Forschungsfragen steht zum einen die funktionale Modellierung, zum anderen die Unsicherheitsbetrachtung im Fokus. Um für die beiden Kernaspekte eine Wertung der Ergebnisse vornehmen zu können, wird die Methodik des Paarvergleiches genutzt, deren Theorie in Abschnitt 4.4.2 (S. 106 ff.) erläutert wurde und die als Grundlage für die Diskussion der Ergebnisse genutzt wird.

In Abschnitt 6.1 erfolgt die Diskussion und Wertung der funktionalen Modellierung. Differenziert wird nach den Einzelaspekten der Vergleichsebene, auf der die Paarvergleichsmethodik gemäß Abbildung 6.1 beruht. Im Einzelnen sind dies die Modellparameter, die Posteriori-Schätzwerte der Zielgröße sowie die prädizierten Posteriori-Schätzwerte der Zielgröße. Basierend auf den Erkenntnissen wird die Erreichung des ersten Forschungsziels gewertet. Abschnitt 6.2 ist der Diskussion der Unsicherheitsmodellierung gewidmet. Ebenso wie im Fall der funktionalen Modellierung werden die Ergebnisse für die Modellparameter und Zielgrößen getrennt betrachtet, bevor zusätzlich die Prädiktionsergebnisse betrachtet werden. Auf Grundlage der Ergebnisdiskussion kann im Anschluss die Erreichung des zweiten Forschungsziels beurteilt werden. Wie in Abbildung 6.1 gezeigt, werden abschließend die bisherigen Erkenntnisse durch die Wertung und die Einordnung der Ergebnisse in den Kontext der Verkehrswertermittlung verknüpft und in Abschnitt 6.3 dargelegt. Im Fokus steht der Beitrag Bayesscher Schätzverfahren für weiterführende Problemstellungen und ein Ausblick auf weitere Anwendungsmöglichkeiten.

6.1. Diskussion der funktionalen Modellierung

Im Rahmen der praktischen Untersuchung von Stichprobendaten des sachlichen Teilmarkts der Ein- und Zweifamilienhäuser des räumlichen Teilmarkts der Wohnungsmarktregion Osnabrück wurde gezeigt, dass sich die Formulierung Bayesscher Regressions- und Kollokationsansätze unter den vorgenommenen Modellspezifikationen grundsätzlich für die Vergleichswertermittlung eignet. Die erreichte Approximationsgüte der Schätzungsergebnisse wird im Folgenden beurteilt. Im Mittelpunkt stehen zum einen die Unterschiede zwischen den informativen Ansätzen mit der Integration von Priori-Wissen und den nichtinformativen Ansätzen, in denen auf Vorinformationen verzichtet wird. Zudem wird beurteilt, inwiefern die Nutzung eines Kollokationsansatzes zu einer verbesserten Anpassung zwischen Schätzwerten und Kaufpreisen gegenüber den Regressionsergebnissen führt. Die Beurteilungen stützen sich insbesondere auf die in Abschnitt 4.4.2 (S. 106 ff.) definierten Vergleichskriterien sowie die Ergebnisse der Paarvergleiche.

In Abschnitt 6.1.1 werden zunächst die Schätzungen der Modellparameter diskutiert. Die Wertungen der Ergebnisse für die Posteriori-Schätzung der Zielgröße sowie der Prädiktion auf Grundlage der Kreuzvalidierung werden im Anschluss in Abschnitt 6.1.2 behandelt. Die Diskussion der Ergebnisse soll abschließend eine Beurteilung hinsichtlich des ersten Forschungszieles ermöglichen, welches formuliert wurde zu:

Verbesserung der funktionalen Modellierung:

Die Modelle des Vergleichswertverfahrens sollen eine verbesserte Anpassung der geschätzten Vergleichswerte an die Realität ermöglichen.

6.1.1. Schätzung der Modellparameter

Basierend auf den theoretischen Komponenten ökonomischer Modellbildung kann die Güte einer funktionalen Modellierung im ersten Schritt auf die Modellparameter zurückgeführt werden (vgl. Abbildung 2.3, S. 32). Die Modellparameter sind im Fall der informativen und der nichtinformativen Bayesschen Regression mit den Trendkoeffizienten und dem Varianzfaktor gegeben, im Fall der informativen und nichtinformativen Bayesschen Kollokation zusätzlich durch die Signalkomponente.

Für die numerischen Ausprägungen der *Trendkoeffizienten* wurde anhand der praktischen Untersuchungen festgestellt, dass sich durch die Integration von Priori-Information sowohl in der Bayesschen Regression als auch der Bayesschen Kollokation konstante Verläufe im betrachteten Auswertzeitraum 2005 bis 2014 ergeben. Für die nichtinformativen Ansätze konnte eine stärkere Variation in den numerischen Ausprägungen der Trendkoeffizienten beobachtet werden (vgl. Abschnitte 5.2.2.1, S. 125 ff. und 5.3.2.1, S. 137 ff.).

Als wesentlicher Vorteil der informativen Ansätze können damit verzerrte Schätzungen durch kurzfristige Entwicklungen am Immobilienmarkt ausgeschlossen werden, die sich bspw. in lediglich einer einzelnen Epoche widerspiegeln. Die informative Modellierung im Rekursionsansatz bewirkt, dass sich das Vorwissen der Vorepochen in Form einer kontinuierlichen Entwicklung auf die nachfolgenden Schätzungen auswirkt. So können neben realen und zeitlich begrenzten Einflüssen auch Schwankungen gedämpft werden, die sich u. U. nur aus numerischen Gründen ergeben. Dies wäre der Fall, sofern in der Stichprobe Kauffalldaten enthalten sind, die durch einzelne Einflussgrößen eine zu den übrigen Epochen abweichende Schätzung bewirken. Haben sich die Verhältnisse am Grundstücksmarkt in der Gesamtbetrachtung nicht geändert, wären die Einflüsse der betreffenden Kauffalldaten und deren Auswirkungen rein numerisch und nicht nach sachlogischen Gesichtspunkten begründet. Als Beispiel ist die Einflussgröße der Unterkellerung zu nennen, welche sich in den nichtinformativen Lösungen der Modelle *R.2* und *K.2* als variierend im Vorzeichen gezeigt hat (vgl. Tabellen 5.18, S. 125 und 5.20, S. 138). Aus Wertermittlungssicht lässt sich keine sachliche Begründung für einen mehrmaligen Vorzeichenwechsel in dem Auswertzeitraum der betrachteten 10 Jahre finden. Es lässt sich daher schließen, dass dieses Ergebnis der jeweiligen Datenlage der Stichproben und damit dem Zufall zugeordnet werden muss. Durch die Rekursion in der Bayes-Strategie werden diese Auswirkungen in den informativen Modellen *R.1* und *K.1* abgefangen.

Neben numerisch begründbarer Schwankungen werden auch solche Variationen gedämpft, die sich durch kurz- und mittelfristige Entwicklungen am Immobilienmarkt ergeben. Dies erweist sich als Vorteil, wenn die auslösenden Gegebenheiten als lediglich zeitlich befristet beurteilt und von einer Entwicklung zum Ausgangszustand ausgegangen werden kann. Es handelt sich dann um einen kurzfristigen Strukturbruch in der Modellformulierung, dessen Auswirkungen jedoch absehbar sind. Als Beispiel lässt sich anführen, dass in den Untersuchungen teilweise Auffälligkeiten in den Auswertepochen in Form stärkerer Variationen der nichtinformativen Ergebnisse zu erkennen sind. Wie in Kapitel 5 ausführlich dargelegt, sind diese neben der initialen Epoche 2005 insbesondere in den Epochen 2010 bis 2012 zu erkennen. Stärkere Schwankungen waren in diesem Kontext in sämtlichen Ergebnissen – sowohl in der funktionalen Modellierung wie auch in der Unsicherheitsbetrachtung – zu erkennen. Eine Verbindung zu verzögerten Reaktionen auf die Immobilien- und Finanzkrise in den Epochen nach 2008 kann lediglich vermutet werden und wäre in Detailuntersuchungen aufzuarbeiten. Deutlich wird in der Modellierung jedoch, dass die Epochen 2013 und 2014 wieder die Ergebnisse der übrigen Epochen bestätigen. Eine abschließende Aussage lässt sich ausschließlich durch Langzeituntersuchungen treffen. Für die informativen Modelle ist ein Strukturbruch weder in den Trendkoeffizienten der Modelle *R.1* und *K.1* zu erkennen, noch wird er durch die Bayesschen Hypothesentests nachgewiesen. Im Fall der nichtinformativen Modelle *R.2* und *K.2* kann ein Nachweis über den nachhaltigen Ein-

fluss ebenfalls nicht nachgewiesen werden, da die Trendkoeffizienten über den gesamten Auswertzeitraum stärker variieren. Deutlich wird, dass kurzfristige Schwankungen durch die informativen Modelle gedämpft werden und sich nicht in voller Höhe in den Ergebnissen niederschlagen.

Die dämpfende Wirkung der informativen Bayesschen Modelle kann auch als Nachteil aufgefasst werden: Es ist zu vermuten, dass die Sensitivität der Hypothesentests gegenüber solchen Strukturbrüchen gering ist, die durch reale Entwicklungen bedingt sind. Tendenziell ist daher zu überlegen, inwiefern die Bayesschen Hypothesentests künftig auf nichtinformativen Ansätze wie das Modell *R.2* übertragen werden können, um die Signifikanz von einzelnen Trendkoeffizienten zu prüfen. Wie auch die Ergebnisse der Modellwahl nach Tabelle 5.17 (S. 124) zeigen, stellt sich das Merkmal der Unterkellerung als die Größe dar, welche durch Elimination in den Submodellen zu den größten Werten für die Bayes-Faktoren führt. Obwohl die Evidenz in keiner Epoche für die Elimination der Größe spricht, erweisen sich diejenigen Submodelle, in denen betreffende Merkmal entfernt wurde, als wahrscheinlicher gegenüber den Submodellen, in denen es Berücksichtigung gefunden hat. Diese Ergebnisse basieren jedoch ausschließlich auf dem informativen Modell, bei dem durch die Integration des Priori-Wissens eine Stabilität des betreffenden Koeffizienten erreicht wird. Die Entwicklung von Hypothesentests für Bayessche Modelle basierend auf nichtinformativen Priori-Dichten bleibt künftigen Forschungsaufgaben überlassen. Zudem werden in der vorliegenden Arbeit die nichtinformativen Modelle ausschließlich zu Vergleichszwecken berechnet. Dies geschieht unter der Prämisse, dass der Verzicht auf Vorinformationen zu Schätzergebnissen führt, die den bisherigen frequentistischen Ansätzen entsprechen.

Zusammenfassend sind im direkten Vergleich der Trendkoeffizienten der vier betrachteten Modellansätze numerische Differenzen in der funktionalen Modellierung zu erkennen, die ausschließlich durch die Vorinformation in den informativen Modellen bedingt sind. Dieses Ergebnis ist damit zu erwarten. Die Auswirkungen im Hinblick auf die erste Forschungsfrage lassen sich ausschließlich in Verbindung mit der Posteriori-Schätzung der Varianzen und der Zielgröße sowie dem Vergleich mit den realen Kaufpreisen beurteilen.

In den Kontext der vorherigen Ausführungen sind die Posteriori-Schätzungen für den *Varianzfaktor* als weiterer Modellparameter einzuordnen. Wie sich anhand der Gleichungen (3.54) und (3.69) für die informative und nichtinformative Bayessche Regression sowie den Gleichungen (3.110) und (3.131) für die informative und nichtinformative Bayessche Kollokation zeigt, sind die Schätzergebnisse insbesondere abhängig von den Anteilen der Residuen bzw. den Rauschanteilen der Kollokation als Differenz zwischen Kaufpreisen und Schätzwerten. Da sich die jeweiligen Schätzergebnisse der Modelle *R.1* und *K.1* sowie *R.2* und *K.2* für die Posteriori-Zielgröße annähernd gleichen, sind Unterschiede in den Varianzfaktoren insbesondere auf die Integration des Priori-Wissens zurückzuführen. Wie in den Abbildungen 5.6 (S. 126) und 5.26 (S. 139) gezeigt, ist zwar eine Kontinuität in der Entwicklung der Varianzfaktoren der informativen Lösungen *R.1* und *K.1* über die Auswertepochen 2005 bis 2014 zu erkennen, jedoch liegen die nichtinformativen Schätzungen der Modelle *R.2* und *K.2* teilweise unterhalb dieser Werte. Die Priori-Informationen bewirken in der rein funktionalen Modellierung damit einen Vorteil hinsichtlich einer über die Auswertepochen stabilen Schätzung der Parameter, zudem liegen die informativen Ergebnisse in der Mehrzahl der ausgewerteten Epochen unterhalb der Ergebnisse der nichtinformativen Ansätze. Die Ergebnisse für die nichtinformativen Lösungen variieren stärker und zeigen sich deutlicher beeinflusst durch aktuell ausgewertete Datensätze.

Zusammenfassend lässt sich in Bezug auf den Modellparameter des Varianzfaktors in den unterschiedlichen Modellansätzen der Schluss ziehen, dass bedingt durch die Heterogenität in den Stichprobendaten und damit den komplexen Verhältnissen im Zusammenspiel der Einflüsse auf den Grundstücksmarkt nicht zwingend jede Epoche eines Auswertzeitraumes miteinander vergleichbar ist. Informative Lösungen sorgen für eine kontinuierliche Entwicklung in der Streuung der Daten, nichtinformativen Lösungen können individuell auf veränderte Datensätze und unterschiedliche Streuungen in den Epochen reagieren. Vorteile für die informative Lösung ergeben sich dann, wenn Stichproben aufgrund stark heterogenen Datenmaterials hohe Streuungen aufweisen: Durch die Integration der vorherigen Auswertepochen können (eventuell rein numerisch bedingte) hohe Streuungen in einzelnen Stichproben relativiert werden.

Werden abschließend die *Signalkomponenten* der Modelle *K.1* und *K.2* betrachtet, zeigt sich durch die Gleichungen (3.108) für die informative Lösung und (3.129) für die nichtinformativen Lösung, dass sich der Einfluss von Priori-Information ausschließlich durch die geschätzten Trendkoeffizienten widerspiegelt. Der Hauptanteil in

der Schätzung der Signalkomponenten wird durch die Signalmatrix $\mathbf{Q}_{ss}$ sowie die Rauschmatrix $\mathbf{Q}_{vv}$ abgedeckt. Letztere wird in den Modellen *K.1* und *K.2* gleichsam als Einheitsmatrix eingeführt; ebenso entspricht sich in den Modellansätzen die Berechnung der Wertabstände und damit der numerische Aufbau der Signalmatrizen. Durch den Einfluss der Posteriori-Trendkoeffizienten sind daher auch lediglich geringe numerische Auswirkungen im Vergleich der Signalkomponente der informativen und nichtinformativen Lösung festzustellen. Eine Darstellung der geschätzten Signalkomponenten findet sich in Anhang C.2 (S. 221 ff.) und bestätigt die Aussage. Ein genereller Mehrwert der Integration von Priori-Information lässt sich auf diese Weise nicht ermitteln. Auswirkungen ergeben sich ausschließlich in der Unsicherheitsmodellierung, die in Abschnitt 6.2 diskutiert wird.

Die Auswirkungen der unterschiedlichen Ergebnisse der Schätzung der Modellparameter auf die Posteriori-Zielgröße und die prädizierte Posteriori-Zielgröße wird im folgenden Unterabschnitt diskutiert. Da sich Vor- und Nachteile der Schätzung der Modellparameter ausschließlich in Verbindung mit den Posteriori-Schätzwerten der Zielgröße und prädizierten Zielgröße sinnvoll beurteilen lassen, wird ein Paarvergleich nicht vorgenommen.

6.1.2. Schätzung und Prädiktion der Posteriori-Zielgröße

Anhand der Ergebnisse für die Schätzung der Posteriori-Zielgröße sowie für die prädizierte Posteriori-Zielgröße basierend auf der Kreuzvalidierung können die Modelle *R.1*, *R.2*, *K.1* und *K.2* im Paarvergleich gegenübergestellt und gewertet werden. Zunächst werden im folgenden Unterabschnitt die Schätzungen der Posteriori-Zielgröße diskutiert, bevor aufbauend im zweiten Unterabschnitt die Prädiktionsergebnisse verglichen werden.

6.1.2.1. Diskussion der Posteriori-Schätzwerte der Zielgröße

Anhand der praktischen Untersuchungen hat sich gezeigt, dass sowohl zwischen den Schätzwerten der informativen und nichtinformativen Bayesschen Regression (Modelle *R.1* und *R.2*) als auch denen der Bayesschen Kollokation (Modelle *K.1* und *K.2*) lediglich geringfügige Abweichungen zu erkennen sind; größtenteils entsprechen sich die Residuen bzw. Rauschanteile. Eine deutliche Verbesserung der nicht erklärbaren Restanteile ergibt sich hingegen durch die Erweiterung des Regressionsansatzes durch die Methode der Kollokation, d. h. bei direktem Vergleich des Modells *R.1* mit *K.1* sowie *R.2* mit *K.2*. Dieses Ergebnis war aufgrund der Ausnutzung stochastischer Restinformation in den Residuen zu erwarten. Für die exemplarischen Epochen 2005, 2009 und 2014 werden in den Abbildungen 6.2 bis 6.4 die absoluten Residuen der verschiedenen Ansätze *R.1* zu *K.1* und *R.2* zu *K.2* gegenübergestellt. In den oberen Grafiken finden sich dabei jeweils die Vergleiche der informativen, in den unteren die der nichtinformativen Lösungen.

Für die Epoche 2005 verbessert sich der Medianwert der absoluten Abweichungen im informativen Ansatz *R.1* von 122 €/m² auf 64 €/m² im Modell *K.1*, im nichtinformativen Ansatz *R.2* von 121 €/m² auf 62 €/m² im Modell *K.2*. In 2009 liegen die vergleichbaren Verbesserungen informativ bei 162 €/m² zu 80 €/m² und nichtinformativ bei 167 €/m² zu 85 €/m². Für die Epoche 2014 kann im informativen Fall eine Verringerung von 183 auf 96 €/m² und im nichtinformativen Fall von 150 €/m² auf 88 €/m² erreicht werden. Ähnliche Ergebnisse sind für die übrigen Auswertepochen festzustellen, deren Grafiken in Anhang C.5 (S. 245 ff.) zusammengestellt sind. Das Bild der absoluten Residuen bestätigt, dass im Hinblick auf die rein funktionale Modellierung ein genereller Mehrwert der Kollokationsansätze gegenüber der Regression sowohl informativ wie nichtinformativ nachgewiesen werden kann. Die vergleichbare Größenordnung der Residuen bzw. Rauschanteile im Quervergleich informative zu nichtinformativen Lösungen (Modelle *R.1* zu *R.2* sowie *K.1* zu *K.2*) verdeutlicht zusätzlich, dass die Integration von Priori-Wissen in den informativen Modellen im Vergleich zu den nichtinformativen zwar zu numerischen Unterschieden in den Schätzergebnissen der Posteriori-Zielgrößen führt, die Differenz allerdings marginal und vernachlässigbar ist. In der funktionalen Modellierung ist ein Mehrwert Bayesscher Ansätze damit nicht gegeben.

Anhand der absoluten Residuen sowie der Grafiken nach 5.7 bis 5.9 (S. 128) und 5.30 bis 5.32 (S. 141) ist weiterhin zu erkennen, dass sich in den Randbereichen der Kaufpreisspannen Systematiken in der Schätzung ergeben, unabhängig von einem Regressions- oder Kollokationsansatz. Dabei werden, wie bereits im Zusammenhang mit Kapitel 5 beschrieben, niedrige Kaufpreise zu hoch, hohe Kaufpreise zu niedrig geschätzt. Dieses Phänomen lässt sich aus statistischer Sicht damit erklären, dass Schätzungen basierend auf Regression und Kollokation sich stets auf den Schwerpunkt einer Stichprobe beziehen. Stärkere Abweichungen an den Rändern sind damit aufgrund ihrer

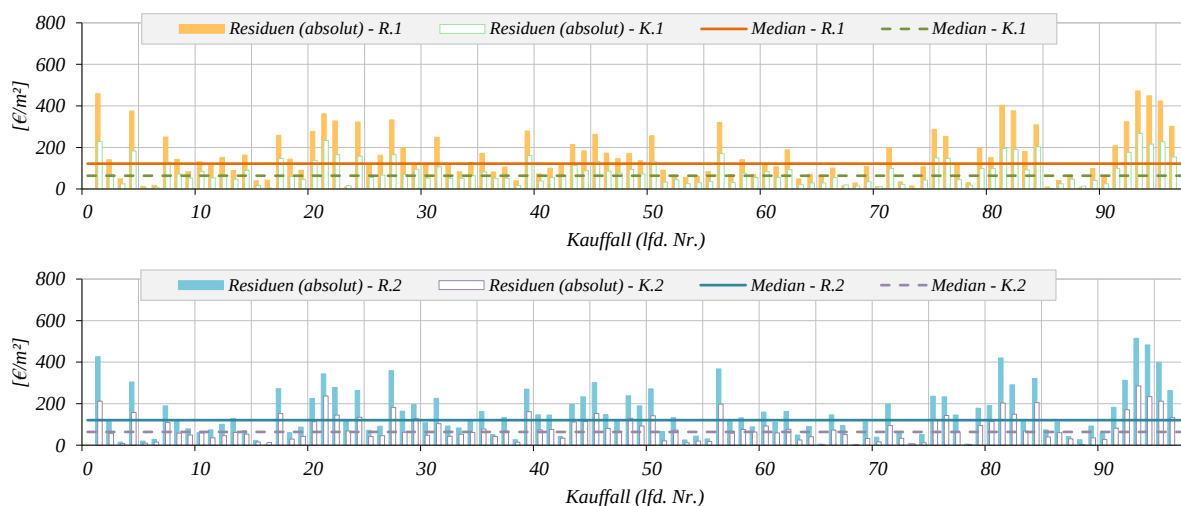


Abbildung 6.2: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2005

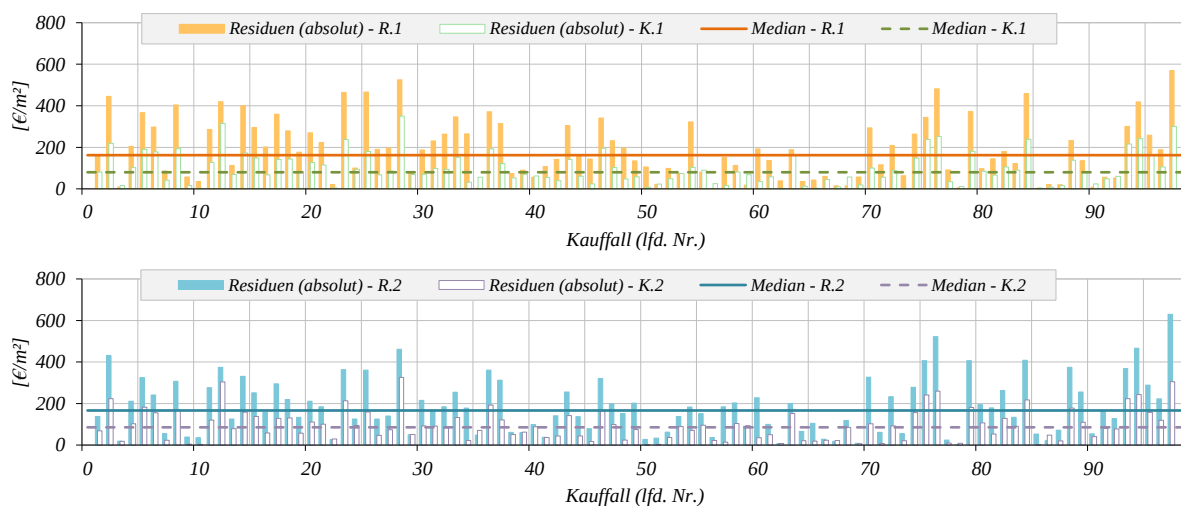


Abbildung 6.3: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2009

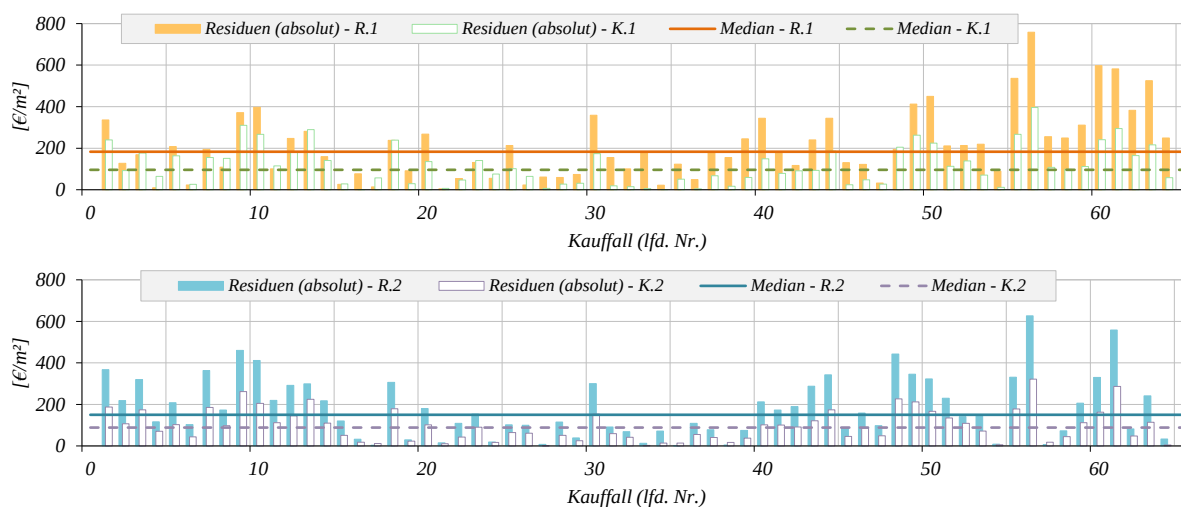


Abbildung 6.4: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2014

Extremelage und den hohen Differenzen in den hohen und niedrigen Kaufpreisen im Vergleich zum Mittelwert der Kaufpreise nicht zu vermeiden. Weiterhin werden im Zuge der Modellbildung und hinsichtlich der Auswertestrategie Annahmen über die Abhängigkeiten der Zielgröße von den Einflussgrößen getroffen. Tatsächliche Abhängigkeiten können jedoch von dieser Annahme abweichen, sodass Fehlspezifikationen im Residuum aufgefangen werden und sich zudem in der Unsicherheit über die geschätzten Parameter ausdrücken.

In Bezug auf die nicht erklärbaren Residuen in der mathematisch-statistischen Modellierung von Vergleichswerten hat bereits ZIEGENBEIN (1977, S. 27) auf den wesentlichen Kernpunkt hingewiesen, dass die Diskrepanz zwischen Trend und tatsächlicher Beobachtung in der Verkehrswertermittlung einen erheblichen Prozentsatz des Kaufpreises ausmachen kann. Diese Aussage kann durch die hier durchgeführten praktischen Untersuchungen bestätigt werden. Angesichts der in Kapitel 2 erläuterten Grundlagen zur Heterogenität von Daten des Immobilienmarktes in Kombination mit den Rahmenbedingungen ökonometrischer Modellbildung können in einem Modellansatz grundsätzlich nicht sämtliche wertrelevanten Einflussgrößen eines Objekts sowie deren Interaktionen mit dem Immobilienmarkt erfasst werden. Die Einflüsse, die aus unterschiedlichen Gründen keinen Eingang in das Modell finden, werden im Residuum bzw. Rauschanteil zusammengefasst. Werden die absoluten Residuen nach den vorstehenden Abbildungen unter diesem Gesichtspunkt beurteilt, spiegeln sich die Auswirkungen von Angebot und Nachfrage wider: Für Objekte, die im Modellansatz annähernd gleiche wertbeeinflussende Merkmale aufweisen, werden u. U. völlig unterschiedliche Kaufpreise gezahlt, ohne dass die Annahme des gewöhnlichen Geschäftsverkehrs verletzt wird. Ursachen liegen beispielsweise im Verhandlungsgeschick der Vertragsparteien. Ein Modell, welches sich auf den Schwerpunkt der Stichprobe und die damit verbundenen durchschnittlichen wertbeeinflussenden Merkmale bezieht, wird daher eine Differenz zwischen dem durchschnittlichen Kaufpreis und dem tatsächlich gezahlten nicht abfangen können. Weitere Gründe lassen sich in der nicht abschließenden Modellierung finden. So können für einen konkreten Kaufpreis Merkmale ausschlaggebend sein, die nicht im Modell abgebildet werden, aber insbesondere die niedrigen und hohen Kaufpreise am Rand der Stichprobe bedingen. Als Beispiel dient die Lage, die u. U. in den hier spezifizierten Modellansätzen nicht allumfassend durch den Bodenrichtwert widerspiegelt wird. Durch die funktionale Modellierung der Bayesschen Kollokation wird zumindest in den Randbereichen eine bessere Approximation erreicht, was dafür spricht, dass in den Residuen der Bayesschen Regression bislang im Modell nicht erfasste Resteinflüsse zusätzlicher wertbeeinflussender Merkmale vorhanden sind. Der Mehrwert in der Verbesserung durch die funktionale Modellierung in einem Kollokationsansatz wird an dieser Stelle deutlich.

In Ergänzung zu den Ergebnisdarstellungen von Residuen der Einzelepochen sowie der praktischen Untersuchungen in Kapitel 5 können für eine Darstellung der Approximationsgüte über den gesamten Auswertezeitraum die mittleren quadratischen Fehler (Root Mean Square Error (RMSE)) als Gütemaß herangezogen werden, die sich über die Beziehung

$$RMSE_{R.1} = \sqrt{(\mathbf{y} - \bar{\mathbf{y}}_{R.1})^2}, \quad RMSE_{R.2} = \sqrt{(\mathbf{y} - \bar{\mathbf{y}}_{R.2})^2} \quad (6.1)$$

angeben lassen. Werden die Werte pro Auswerteepoche berechnet, ergeben sich die in Abbildung 6.5 dargestellten Verläufe.

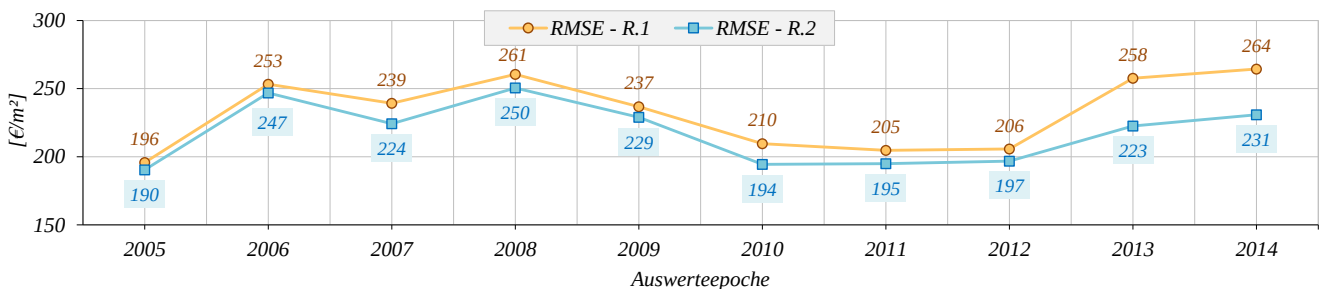


Abbildung 6.5: Mittlerer quadratischer Fehler der Posteriori-Schätzung $\bar{\mathbf{y}}$ der Epochen 2005 bis 2014, Bayessche Regression

Abbildung 6.5 zeigt, dass der RMSE des informativen Ansatzes *R.1* oberhalb der korrespondierenden Werte der nichtinformativen Lösung *R.2* liegt. Dies spricht grundsätzlich für eine höhere Streuung in den Schätzer-

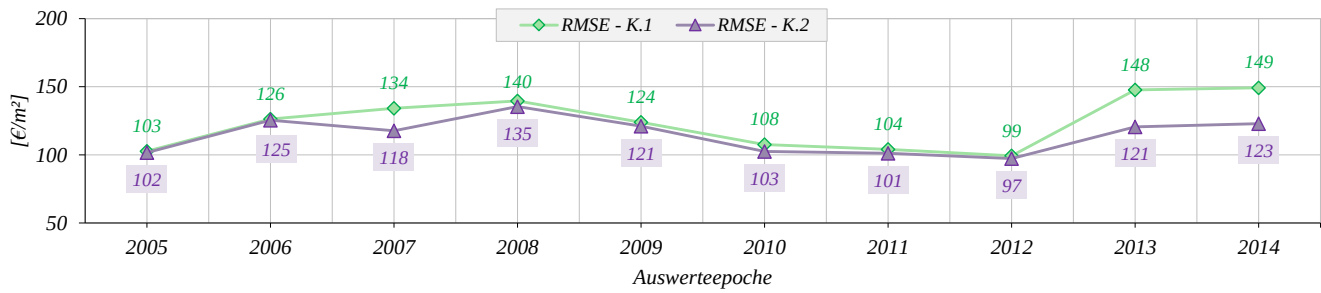


Abbildung 6.6: Mittlere quadratische Fehler der Posteriori-Schätzung $\bar{y}$ der Epochen 2005 bis 2014, Bayessche Kollokation

gebnissen; die Differenzen liegen jedoch im Bereich von wenigen Euro pro Quadratmeter, was im Rahmen der Vergleichswertermittlung und im Bezug auf die Heterogenität der Datengrundlage als vernachlässigbarer Betrag zu werten ist. In Anbetracht der aufgrund von Priori-Wissen modifizierten Schätzwerte der Trendkoeffizienten kann aus diesem Ergebnis der Schluss bestätigt werden, dass sich die funktionale Modellierung in der indirekten Vergleichswertermittlung durch einen informativen Ansatz nicht deutlich von einem nichtinformativen – und damit klassischen – Ansatz unterscheidet. Letztgenanntes Modell ist in der Lage, sich individueller an die Gegebenheiten der aktuellen Stichprobe anzupassen. Zurückzuführen ist dies auf die Zwänge, die im informativen Fall durch die Rekursion auf die Schätzwerte der Trendkoeffizienten ausgeübt werden; diese spiegeln sich insbesondere in den thematisierten konstanten Verläufen der einzelnen Trendkoeffizienten wider. Auch der Bezug zu den Verläufen der Posteriori-Schätzungen der Varianzfaktoren wird an dieser Stelle deutlich.

Ein äquivalentes Verhalten ergibt sich für die Schätzwerte, die auf den Modellen *K.1* und *K.2* der Bayesschen Kollokation beruhen. Wie bereits in den praktischen Untersuchungen beschrieben, ergeben sich auch hier die Abweichungen zwischen Kaufpreisen und Posteriori-Schätzungen der informativen und nichtinformativen Modellierung in einer vergleichbaren Größenordnung. Die RMSE lassen sich Abbildung 6.6 entnehmen und zeigen in Bezug auf die Ergebnisse der Bayesschen Regressionsansätze ähnliche Verläufe. Zusätzlich liegen die RMSE der informativen Lösung wiederum oberhalb derer der nichtinformativen Lösung. Im Gegensatz zu der vorherigen reinen Trendmodellierung hat sich die Größenordnung der RMSE signifikant verringert. Aus den Ergebnissen lässt sich schließen, dass wie zuvor keine Verbesserung in der funktionalen Modellierung durch die Integration von Priori-Wissen erreicht wird. Jedoch ergibt sich eine Steigerung der Approximationsgüte durch den Kollokationsansatz, die auf Grund der Ausschöpfung stochastischer Restinformationen in Form der Signalkomponente zu erwarten ist.

Werden die Erkenntnisse für eine Gegenüberstellung in einer Paarvergleichsmatrix zusammengefasst, ergibt sich die in Abbildung 6.7 dargestellte Belegung. Die Modelle *R.1* und *R.2* weisen in der Schätzung der Posteriori-Zielgröße Differenzen auf, deren Größenordnungen für Wertermittlungsanwendungen nicht relevant sind. Aus diesem Grund sind keine Präferenzen für das informative oder das nichtinformativ Bayessche Regressionsmodell auszumachen. Entsprechend werden die Vergleichsfelder mit dem Wert 1 belegt. Wie die Ergebnisse weiterhin gezeigt haben, weisen die Bayesschen Kollokationsansätze eine deutlich verbesserte Approximationsgüte gegenüber den Regressionsmodellen auf, so dass diese jeweils mit dem Wert 2 belegt werden können. Werden die Bayesschen Kollokationsmodelle hingegen untereinander verglichen, zeigt sich das gleiche Ergebnis wie bei den Regressionsansätzen: Eine Präferenz für eines der Modelle kann aufgrund der vernachlässigbaren Differenzen in den Posteriori-Schätzgrößen nicht bestimmt werden, so dass wiederum der Wert 1 zu setzen ist. In der Summe ergeben sich durch die Paarvergleiche die Rangstufe 1 für die Bayesschen Kollokationsmodelle *K.1* und *K.2* sowie die Rangstufe 2 für die Bayesschen Regressionsmodelle. Da sich für die Modelle auf den jeweiligen Rangstufen zudem die gleichen Häufigkeiten als Summe aus den Paarvergleichswerten ergeben, kann eine Differenzierung innerhalb der Ränge nicht festgestellt werden. Dies spricht für einen jeweils schwachen Rang, der durch die trotz Integration von Priori-Wissen unveränderte Approximationsgüte begründet ist.

Für eine abschließende Wertung des ersten Forschungszieles wird im folgenden Abschnitt die funktionale Modellierung in der Prädiktion diskutiert.

Modell	R.1	R.2	K.1	K.2
R.1		1	2	2
R.2	1		2	2
K.1	0	0		1
K.2	0	0	1	
Summe	1	1	5	5
Rang	2.	2.	1.	1.

Legende:
0 - schlechter, 1 - neutral, 2 - besser

Abbildung 6.7: Paarvergleichsmatrix für die Schätzung der Posteriori-Zielgröße $\bar{y}$

6.1.2.2. Diskussion der prädierten Posteriori-Schätzwerte der Zielgröße

In der praktischen Anwendung der entwickelten Auswertestrategie ist von vorrangigem Interesse, inwiefern sich die Modelle für die Schätzung der Vergleichswerte von Objekten eignen, die noch nicht am Markt gehandelt wurden. Im Fokus steht damit die Prädiktion, anhand derer die Eignung der Modelle und der Auswertestrategie für den praktischen Nutzen beurteilt werden kann. Im Sinne der VW-RL wurde durch das praktischen Beispiel nach Kapitel 5 die Schätzung von Vergleichsfaktoren gezeigt, die durch Multiplikation mit der Bezugsgröße (hier: Wohnfläche) direkt die Angabe eines realistischen und möglichst zutreffenden Vergleichswertes ermöglichen sollen. Anhand der Darstellungen in Abschnitt 5.4 (S. 148 ff.) wurden im Rahmen der Kreuzvalidierung die Prädiktionsergebnisse in den Modellen der Bayesschen Regression und Bayesschen Kollokation durch die Gegenüberstellung der Kaufpreise mit den prädierten Posteriori-Schätzwerten $\bar{y}_p$ gezeigt.

Für die Trendkoeffizienten sowie den Varianzfaktor sind weiterhin die Diskussionen des vorherigen Abschnitts gültig. Wie bereits erläutert, beruhen die Schätzungen auf dem Prinzip der Kreuzvalidierung. Aus diesem Grund sind auch im Rahmen der Kreuzvalidierung Schätzwerte für die Modellparameter zu erwarten, die den Ergebnissen unter Einbeziehung sämtlicher Kauffalldaten ähneln. Unterschiede können sich dann ergeben, wenn es sich bei dem eliminierten Kauffall der Kreuzvalidierung um ein Objekt handelt, welches in Bezug auf die Kombination aus Kaufpreis und wertbeeinflussenden Merkmalen dem Randbereich der Stichprobe zuzuordnen ist. In diesem Fall würde der Kauffall u. U. im Sinne eines Hebelwertes auf die Schätzung der Modellparameter wirken. Aufgrund der sorgfältigen Stichprobenaufbereitung im zweiten Schritt der Auswertestrategie sowie der formulierten Bayesschen Ausreißersuche werden diese jedoch für die weitere Ergebnisdiskussion ausgeschlossen.

Werden die resultierenden Modellparameter der einzelnen Modelle aus der Kreuzvalidierung genutzt, um den Vergleichswert des zuvor eliminierten Kauffalles zu schätzen, ergeben sich für die Beispielepochen 2005, 2009 und 2014 die in den Grafiken 5.46 bis 5.48 (S. 150) sowie 5.54 bis 5.56 (S. 153) dargestellten Ergebnisse. Für eine Übersicht über die übrigen Auswertepochen wird auf die Anhänge B.3 (S. 215 ff.) und C.4 (S. 234 ff.) verwiesen. Die Abbildungen 6.8 bis 6.10 ergänzen die bisherigen Darstellungen um die Gegenüberstellung der absoluten Residuen, die sich aus der Differenz der Kaufpreise zu den prädierten Posteriori-Schätzungen der Zielgröße ergeben. In den oberen Zeilen der Abbildungen finden sich die Ergebnisse der informativen Modellen *R.1* und *K.1*, in den unteren Zeilen die der nichtinformativen Modelle *R.2* und *K.2*. Die Darstellungen der übrigen Auswerteziträume lassen sich Anhang C.6 (S. 249ff.) entnehmen.

Im direkten Vergleich zu den Abbildungen 6.2 bis 6.4 ist zu erkennen, dass sich sowohl zwischen den Modellen *R.1* und *K.1* als auch zwischen den nichtinformativen Modellen *R.2* und *K.2* nur geringfügige Änderungen in den Residuen ergeben haben. Die Medianwerte der absoluten Residuen für die unterschiedlichen Modelle über die gesamten Auswertepochen können Tabelle 6.1 entnommen werden.

Durch die Medianwerte der gesamten Auswertepochen sowie die Abbildungen 6.8 bis 6.10 wird zum einen bestätigt, dass die funktionale Modellierung in den informativen Modellen *R.1* und *K.1* keine für Wertermittlungsanwendungen relevanten Verbesserungen gegenüber nichtinformativen Ansätzen bewirkt. Zum anderen ist zu beobachten, dass sich entgegen der Modellierungen ohne Kreuzvalidierungsansatz keine Verbesserung von Regressions- zu Kollokationsansatz ergibt; die Prädiktionen weisen die annähernd gleiche Approximationsgüte auf. Dieses Resultat

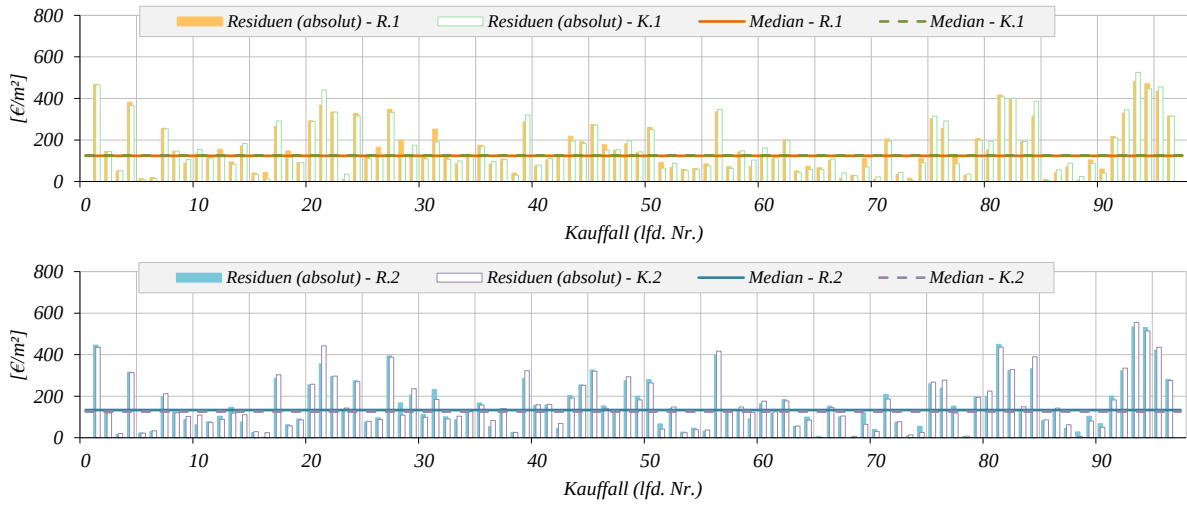


Abbildung 6.8: Absolute Residuen der prädizierten Posteriori-Schätzwerte $\bar{y}_p$ der Epoche 2005

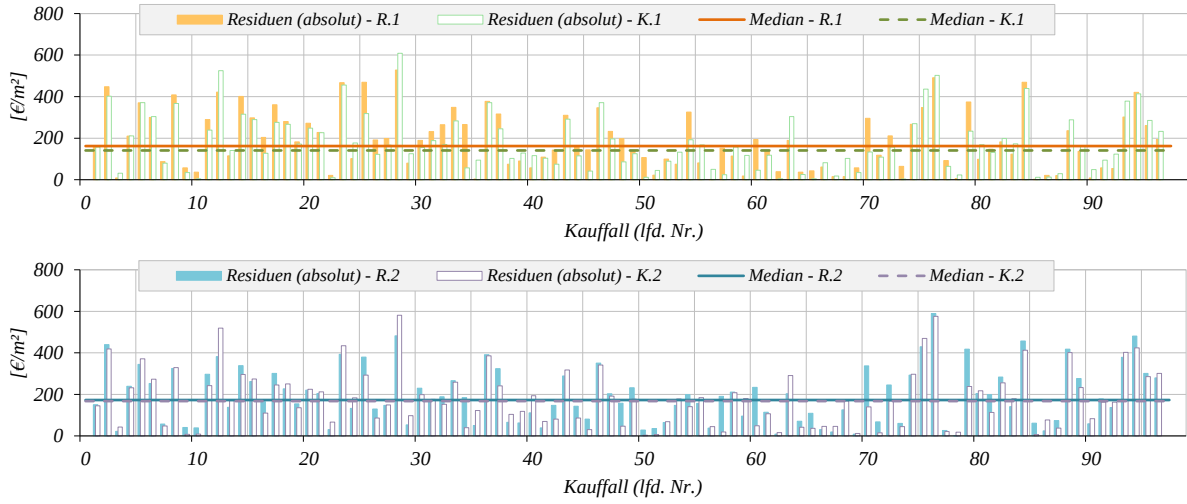


Abbildung 6.9: Absolute Residuen der prädizierten Posteriori-Schätzwerte $\bar{y}_p$ der Epoche 2009

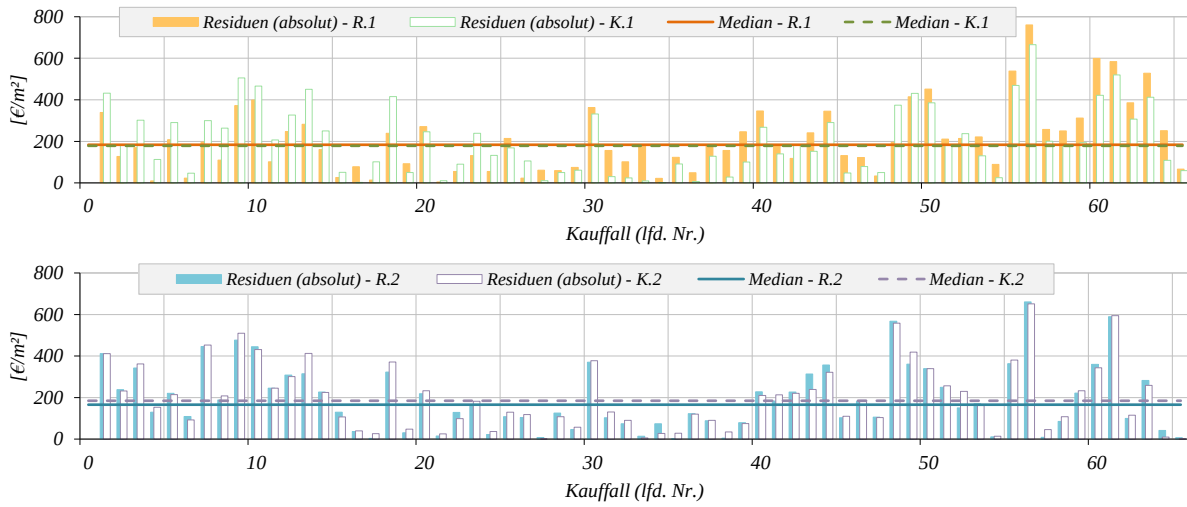


Abbildung 6.10: Absolute Residuen der prädizierten Posteriori-Schätzwerte $\bar{y}_p$ der Epoche 2014

Tabelle 6.1.: Medianwerte der absoluten Residuen in den Auswertepochen 2005 bis 2014

Modell	R.1	K.1	R.2	K.2
2005	124 €/m ²	126 €/m ²	133 €/m ²	125 €/m ²
2006	196 €/m ²	204 €/m ²	187 €/m ²	178 €/m ²
2007	167 €/m ²	163 €/m ²	177 €/m ²	168 €/m ²
2008	192 €/m ²	188 €/m ²	194 €/m ²	202 €/m ²
2009	164 €/m ²	155 €/m ²	178 €/m ²	168 €/m ²
2010	131 €/m ²	117 €/m ²	139 €/m ²	141 €/m ²
2011	120 €/m ²	119 €/m ²	138 €/m ²	139 €/m ²
2012	159 €/m ²	134 €/m ²	168 €/m ²	136 €/m ²
2013	188 €/m ²	175 €/m ²	172 €/m ²	168 €/m ²
2014	184 €/m ²	179 €/m ²	166 €/m ²	185 €/m ²

ergibt sich unabhängig von informativen oder nichtinformativen Modellen und kann als wesentliche Erkenntnis in Bezug auf den Aspekt einer erweiterten funktionalen Modellierung festgehalten werden.

Für die prädizierten Schätzwerte lassen sich analog zu der Vorgehensweise im vorherigen Unterabschnitt die RMSEs berechnen, um eine zusätzliche Kenngröße für die Approximationsgüte beurteilen zu können. Gleichzeitig dienen die RMSEs neben den absoluten Residuen als zweites Vergleichskriterium für den abschließenden Paarvergleich. Die Resultate für den gesamten Auswertzeitraum 2005 bis 2014 lassen sich den Abbildungen 6.11 und 6.12 entnehmen.

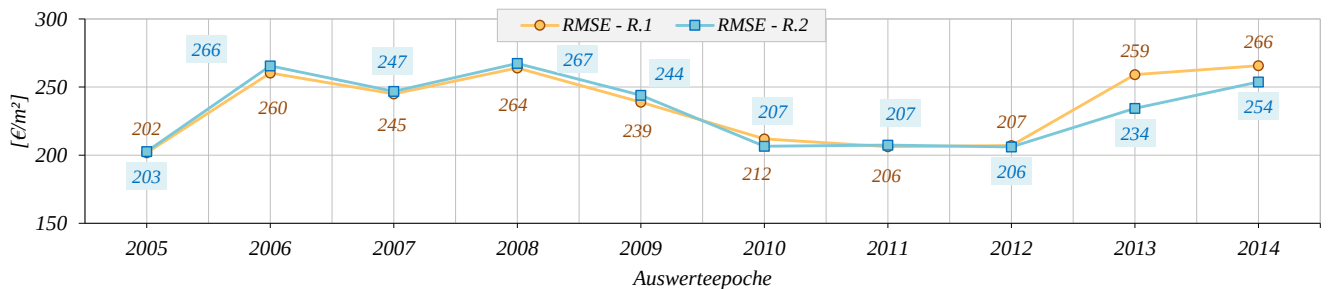
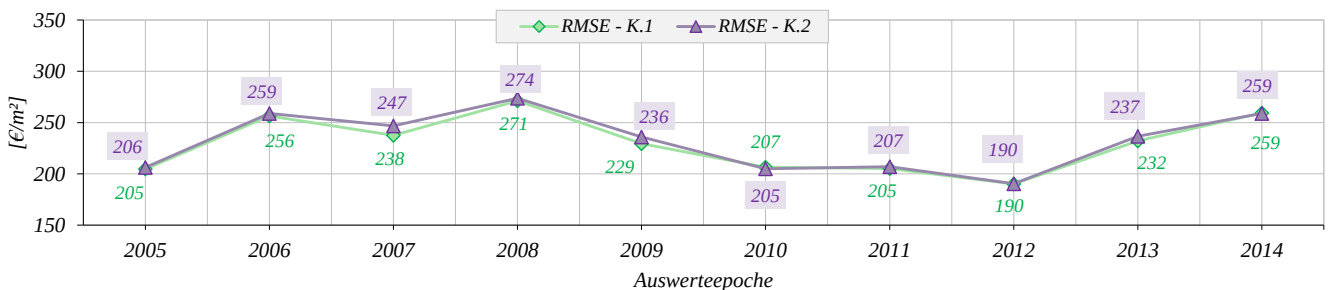
Abbildung 6.11: Mittlere quadratische Fehler der prädizierten Posteriori-Schätzung $\bar{y}_p$ der Epochen 2005 bis 2014, Bayessche RegressionAbbildung 6.12: Mittlere quadratische Fehler der prädizierten Posteriori-Schätzung $\bar{y}_p$ der Epochen 2005 bis 2014, Bayessche Kollokation

Abbildung 6.11 zeigt, dass die Prädiktionen $\bar{y}_p$ in den Bayesschen Regressionsansätzen *R.1* und *R.2* eine vergleichbare Approximationsgüte zu den Ergebnissen der Posteriori-Schätzung der Zielgröße $\bar{y}$ aufweisen. Differenzen pro Auswertepoche und Modellansatz liegen in einem Bereich von wenigen Euro pro Quadratmeter und damit wiederum in einem für Anwendungen der Verkehrswertermittlung vernachlässigbaren Bereich. Entgegen der Schätzungen von $\bar{y}$ liegen die RMSE der informativen Lösungen nicht mehr stetig oberhalb der Werte für die

nichtinformativ Lösung, mehrheitlich sind niedrigere Werte zu beobachten. Die Differenzen bewegen sich jedoch in einem geringen Bereich, womit die Verläufe über den Gesamtzeitraum der Auswertepochen als gleich anzusehen sind. Die Approximationsgüte der Prädiktion kann insgesamt als geeignet für die Anwendung im indirekten Vergleichsverfahren beurteilt werden und liegt im Bereich der Erwartungen auf Grundlage der Modelle ohne Prädiktionsansatz. In Abbildung 6.12 werden für den direkten Vergleich die RMSE der Bayesschen Kollokationsansätze gegenübergestellt. Wie bereits im Zusammenhang mit den Ergebnissen der Kreuzvalidierung nach Abschnitt 5.4.2 (S. 149 ff.) sowie den absoluten Residuen nach den Abbildungen 6.8 bis 6.10 beobachtet, lassen sich die Prädiktionsergebnisse nicht mit den Ergebnissen der funktionalen Modellierung ohne Ansatz der Kreuzvalidierung vergleichen. Auch im Vergleich zu Abbildung 6.11 zeigen sich kaum Unterschiede. Tendenziell liegen die RMSE von Bayesscher Regression und Bayesscher Kollokation in einer vergleichbaren Größenordnung, zeigen in der Bayesschen Kollokation mit Ausnahme der Epochen 2005 und 2008 jedoch eine Verbesserung in der funktionalen Modellierung gegenüber der reinen Trendmodellierung. Insofern kann durch die Verwendung der Kollokationsansätze *K.1* und *K.2* zwar eine leichte Verbesserung in der funktionalen Approximation erreicht werden, allerdings liegen diese unter den Erwartungen, die bezüglich der Steigerung zwischen der Bayesschen Regression und der Bayesschen Kollokation ohne Kreuzvalidierung gestellt werden konnten. In Bezug auf Abschnitt 5.3.2 (S. 137 ff.), in dessen Kontext eine Verbesserung in der funktionalen Anpassung erreicht werden konnte, stellt sich die Frage nach den Ursachen für die abweichenden Ergebnisse in der Prädiktion.

Da sich aufgrund des Prinzips der Kreuzvalidierung kaum numerische Differenzen in den Trendkoeffizienten der (informativen ebenso wie nichtinformativen) prädiktiven Bayesschen Kollokation ergeben, sind die Ursachen auf die Signalkomponente zurückzuführen. Wie aus den Gleichungen (3.116) und (3.135) interpretiert werden kann, sind als wichtigste Komponenten in der Bestimmung des prädizierten Signals $\bar{s}_p$ sowohl im informativen wie im nichtinformativen Fall deren Kofaktormatrix $\mathbf{Q}_{ss_p}$ auszumachen, welche die Korrelationen zwischen dem Prädiktionsobjekt und den Vergleichsobjekten enthalten. Die geringere Approximationsgüte in der funktionalen Modellierung deutet daher auf nicht ausreichende Korrelationen hin. Wird dieser Gedanke weiter verfolgt, lassen sich die ursächlichen Gründe in dem Wertabstand als maßgebende Größe für die Besetzung der Signalmatrizen finden.

Wie auch schon in den Ausführungen zur Schätzungen der Autokorrelationsfunktionen in Abschnitt 4.3.1.2 (S. 95 ff.) erläutert, weisen die empirischen Autokorrelationsfunktionen lediglich schwache Korrelationen in den Daten aus. Bereits ab einem Wertabstand von $> 0,17$ sind durchschnittlich kaum noch Zusammenhänge in den Daten nachweisbar. Für sämtliche ausgewertete Stichproben lässt sich feststellen, dass der Wertabstand lediglich schwach ausgeprägt ist. Die vergleichsweise gute funktionale Approximation in der reinen Modellschätzung ohne Prädiktion mit einer deutlichen Steigerung der Anpassungsgüte gegenüber den Regressionsansätzen ist daher größtenteils auf die Objekte in den ersten beiden Klassen einer jeden empirischen Autokorrelationsfunktion zurückzuführen. Ferner hat sich in praktischen Versuchsreihen gezeigt, dass die Signalkomponente $\bar{s}$ dominiert wird von den numerischen Eigenschaften der Signalmatrix $\mathbf{Q}_{ss}$ nach Gleichung (2.20). Auf der Hauptdiagonalen ist diese besetzt mit der Varianz bzw. der Eigenkorrelation des Objektes – für die Hauptdiagonalelemente ergibt sich daher das Element 1. Es hat sich gezeigt, dass aufgrund der geringen Wertabstände und damit der geringen Korrelation die Schätzung hauptsächlich durch dieses Element geprägt wird. Da die Signalmatrix des prädizierten Objekts ausschließlich Korrelationen zu den übrigen Vergleichsobjekten enthält, aufgrund eines fiktiv nicht vorhandenen Kaufpreises jedoch keine Aussage zur Eigenkorrelation enthält, entfällt an dieser Stelle das im Ausgleichsalgorithmus maßgebliche Element. Das Problem der verringerten Approximationsgüte ergibt sich als rein numerisches Problem, welches auf die Korrelationen innerhalb der Stichprobe und damit dem Wertabstand als Kenngröße für die Übereinstimmung von zwei Stichprobenobjekten zurückführt.

Als weiteres Problem im direkten Zusammenhang mit dem Wertabstand erweist sich die bereits angesprochene Problematik der Diskrepanz zwischen Kaufpreisen und wertbeeinflussenden Merkmalen. Objekte mit ähnlichen wertbeeinflussenden Merkmalen können erheblich differierende Kaufpreise aufweisen. Eine Systematik ist nicht immer erkennbar. Hinzu kommt, dass ein kausaler Zusammenhang zwischen den wertbeeinflussenden Merkmalen und dem Kaufpreis oftmals nur unterstellt wird. Die Realität muss sich in diesem Zusammenhang nicht zwingend widerspiegeln. Zudem sind nicht sämtliche relevanten Kauffallinformationen verfügbar. Beispielsweise werden Bau-

mängel und Bauschäden nicht in der AKS erfasst, so dass eine wesentliche Information über den Zustand eines Objektes fehlt. Während sich dieser signifikant auf einen Kaufpreis auswirkt, kann er nicht numerisch anhand der verfügbaren Daten nachvollzogen werden. Resultat sind Diskrepanzen zwischen den üblichen wertbeeinflussenden Merkmalen und einem für diese Größen am Markt üblichen Kaufpreis. Damit wird das wesentliche Prinzip der ökonometrischen Modelle bzw. das des Modellbegriffs aufgegriffen, welches in Abschnitt 2.2.1 (S. 31 ff.) ausführlich beschrieben ist. Wie an genannter Stelle erläutert, ist ein wesentliches Charakteristikum ökonometrischer Modelle, dass keine Gesetzmäßigkeiten in einem streng naturwissenschaftlichen Sinne gegeben sind. Verdeutlicht wird diese Aussage unter Berücksichtigung der Unterscheidung von Kaufpreis und Wert nach Abschnitt 2.1.2.1 (S. 26 ff.): Ein auf statistischen Grundlagen ermittelter Wert, der zudem die Masse der Stichprobendaten repräsentiert, muss nicht dem Kaufpreis entsprechen. Diese Aussage lässt sich sowohl anhand der Posteriori-Schätzwerte der Zielgrößen als auch den prädizierten Posteriori-Schätzwerten der Zielgrößen erkennen. Die Prädiktion innerhalb der Bayesschen Kollokationsmodelle ist ein zutreffendes Beispiel, da hier die Unterschiede zwischen Wertabstand und Kaufpreis deutlich zu erkennen sind. Eine umso größere Bedeutung kommt der Behandlung der Unsicherheiten zu, die die funktionale Modellierung entscheidend um die Aussage der Wahrscheinlichkeitsbereiche für die Schätzwerte ergänzen.

Werden die Ergebnisse für die Prädiktionen nach den Modellen *R.1*, *K.1*, *R.2* und *K.2* durch den systematischen Paarvergleich gegenübergestellt, ergibt sich die Paarvergleichsmatrix nach Abbildung 6.13. Da sich keines der Modelle anhand der Vergleichskriterien der absoluten Residuen und der RMSE als signifikante Verbesserung gegenüber den übrigen Modelle herausgestellt hat, ist für jeden Paarvergleich der Wert 1 einzusetzen. Dies ist gleichbedeutend mit der Aussage, dass die Modelle für die Prädiktion von Vergleichswerten in Bezug auf die Approximationsgüte als gleichwertig einzustufen sind. Es besteht somit keine Präferenz für eines der Modell. Als Konsequenz ergibt sich, dass die vier Modelle jeweils mit der Rangstufe 1 bewertet werden und sich somit keine eindeutige Rangfolge feststellen lässt.

Modell	R.1	R.2	K.1	K.2
R.1		1	1	1
R.2	1		1	1
K.1	1	1		1
K.2	1	1	1	
Summe	3	3	3	3
Rang	1.	1.	1.	1.

Legende:
0 - schlechter, 1 - neutral, 2 - besser

Abbildung 6.13: Paarvergleichsmatrix für die funktionale Modellierung der prädizierten Posteriori-Zielgröße $\bar{y}_p$

6.1.3. Fazit zum ersten Forschungsziel

Abschließend für die Diskussion der funktionalen Modellierung lassen sich die Ergebnisse im Hinblick auf das erste Forschungsziel wie folgt zusammenfassen:

Die Formulierung der klassischen multiplen linearen Regressionsanalyse und der einfachen Kollokation in Bayesschen Ansätzen lässt sich erfolgreich auf die Vergleichswertermittlung übertragen. Im Rahmen der Modellformulierung werden dabei die besonderen Belange der Heterogenität in der immobilien-spezifischen Datengrundlage durch die Entwicklung einer Auswertestrategie berücksichtigt. Gleichwohl kann das erste Forschungsziel einer verbesserten funktionalen Anpassung durch die hier erfolgten praktischen Untersuchungen nicht gelöst werden:

1. Mit Hilfe eines systematischen Paarvergleiches zeigt sich, dass die beabsichtigte Verbesserung in der funktionalen Modellierung nicht durch die Integration von Vorinformationen bedingt ist. Die Schätzungen in den Modellen der informativen und nichtinformativen Bayesschen Regression liefern annähernd gleiche Ergebnisse, deren Differenzen für die Verkehrswertermittlung als vernachlässigbar einzustufen sind. Gleiches trifft auf die Schätzungen der informativen und nichtinformativen Bayesschen Kollokation zu. **Durch die Nutzung**

informativer Bayesscher Modelle lässt sich keine Verbesserung in der funktionalen Modellierung erreichen. Verbesserungen sind ausschließlich durch die Erweiterung der Regressionsmodelle um einen Kollokationsansatz zu erreichen.

2. **Für die Prädiktion von Vergleichswerten ist ebenfalls keine signifikante Verbesserung in der Approximationsgüte zu erreichen.** Zudem zeigt sich, dass die zuvor festgestellten Verbesserungen durch die Einführung eines Kollokationsansatzes nicht auf die Prädiktionsergebnisse übertragen werden können. Diese Aussage trifft gleichermaßen auf die informative wie nichtinformative Bayessche Kollokation zu und liegt in der numerischen Bestimmung der prädizierten Signalkomponente begründet. Wie das Ergebnis eines Paarvergleichs zeigt, sind die vier untersuchten Modelle für Prädiktionszwecke als gleichwertig einzustufen.

Aus den Ergebnissen zur funktionalen Modellierung lässt sich zusammenfassen, dass eine nach den Ausführungen in Kapitel 4 entwickelte Auswertestrategie stets einen Kompromiss zwischen Pragmatismus in der Verkehrswertermittlung und mathematisch-statistischen Ansprüchen abbildet. Zu dem Ergebnis der funktionalen Modellierung ist zudem anzumerken, dass es sich bei den Schätzwerten um Vergleichswerte handelt, nicht um abschließende Verkehrswerte. Gemäß VW-RL ist jeder Vergleichswert für jedes Prädiktionsobjekt hinsichtlich weiterer Anpassungen an die realen Marktgegebenheiten zu prüfen, um letztlich zu einem zutreffenden Verkehrswert zu gelangen. Insofern erheben die Resultate der Bayesschen Modellansätze nicht den Anspruch, einen Verkehrswert abbilden zu können. Es lässt sich schließen, dass die Vorteile der Nutzung der Bayes-Strategie durch die Möglichkeit der Integration von Priori-Wissen unter den hier dargestellten Rahmenbedingungen nicht in der funktionalen Modellierung liegen.

6.2. Diskussion der Unsicherheitsmodellierung

Die funktionale Modellierung erlaubt die Ableitung von Vergleichsfaktoren aus der Beobachtung des realen Marktgeschehens. Die Plausibilität der Schätzwerte lässt sich ausschließlich durch die ergänzende Angabe von Unsicherheitsmaßen beurteilen, wie sie für die Vergleichswertmodelle *R.1*, *K.1*, *R.2* sowie *K.2* abgeleitet wurden. In Abschnitt 6.2.1 werden die Ergebnisse der Unsicherheitsmodellierung für die Modellparameter der informativen und nichtinformativen Bayesschen Regression und Bayesschen Kollokation gegenübergestellt. Diskutiert wird in diesem Kontext ebenfalls die Fortpflanzung von Unsicherheiten im Prozess der Rekursion. Abschnitt 6.2.2 rückt die Diskussion der Unsicherheitsspannen für die Posteriori-Schätzwerte sowie der prädizierten Posteriori-Schätzwerte der Zielgröße in den Mittelpunkt; insbesondere werden Auswirkungen der Rekursion behandelt. Aus den gewonnen Erkenntnissen folgt in Abschnitt 6.2.3 das Fazit zum zweiten Forschungsziel, welches formuliert wurde zu:

Verbesserung der Unsicherheitsmodellierung:

Die Modelle des Vergleichswertverfahrens sollen die Reduzierung auftretender Unsicherheiten in den Schätzungen der Modellparameter und der geschätzten Vergleichswerte sowie durch rekursive Anwendung die Fortpflanzung auf folgende Auswerteperioden ermöglichen.

6.2.1. Unsicherheit der Modellparameter

In Bezug auf die reinen Modellschätzungen spiegelt sich die Unsicherheit in den Eingangsdaten in der Unsicherheit der Modellparameter wider, welche im Fall der Bayesschen Regressionsmodelle durch die Trendkoeffizienten sowie die Varianz, im Fall der Bayesschen Kollokationsmodelle zusätzlich durch die Signalkomponenten angegeben werden. Anhand der Konfidenzregionen der unterschiedlichen Modellansätze kann daher abgelesen werden, wie sich die unsicheren Eingangsdaten auf die Modellbildung auswirken und wie plausibel die Schätzwerte der Parameter in ihrer numerischen Ausprägung letztlich sind - eine Sichtweise, die bislang im Rahmen der frequentistischen Modelle nicht für die Anwendungen in der Vergleichswertermittlung genutzt werden konnte und ausschließlich durch die Bayessche Modellierung ermöglicht wird. In den bisherigen (klassischen) Anwendungen der multiplen Regressionsanalyse und der einfachen Kollokation werden die Modellparameter als Konstanten aufgefasst, was den komplexen zufälligen Variabilitäten in Marktdaten und -analysen nicht umfassend gerecht wird. Anhand der exemplarischen Ergebnisse nach Abschnitt 5.2.3 (S. 129 ff.) für die Bayesschen Regression und Abschnitt 5.3.3 (S. 143 ff.) für die Bayessche Kollokation wird in diesem Zusammenhang eine signifikante Erweiterung der bisherigen Modellierung im

Vergleichswertverfahren deutlich. Im Sinne einer Steigerung der Markttransparenz können nunmehr als zusätzliche Information die Wahrscheinlichkeiten für die geschätzten Parameter in Form von Unsicherheitsspannen gegeben und veröffentlicht werden. Die Güte einer Modellierung lässt sich so direkt ablesen. Sachverständigen wird in praktischen Anwendungen auf diese Weise ermöglicht, direkt Schlüsse über den Einfluss unterschiedlicher Einflussgrößen im Modellansatz durch Würdigung des jeweiligen Grades der Plausibilität zu ziehen.

Bezüglich der Konfidenzregionen als *Unsicherheitsmaß der Trendkoeffizienten* hat sich durch die praktischen Untersuchungen gezeigt, dass mittels der rekursiven Anwendung der Auswertestrategie eine deutliche Verringerung in den informativen Modellen *R.1* und *K.1* der Bayesschen Regression und der Bayesschen Kollokation erreicht werden kann. Diese Ergebnisse konnten bereits in vorherigen Untersuchungen beobachtet werden, in denen ein abweichender Teilmarkt mit der hier dargestellten Methodik untersucht wurde (vgl. ZADDACH & ALKHATIB (2014b)). In den Abbildungen 5.14 bis 5.19 (S. 130 und 131) werden die für das Modell *R.1* sowie in den Abbildungen 5.37 bis 5.42 (S. 144 und 145) für das Modell *K.1* erhaltenen Schätzergebnisse dargestellt. Die Grafiken zeigen, dass die informativen Ansätze annähernd äquivalente Größenordnungen in den Wahrscheinlichkeitsbereichen aufweisen. Dieses Ergebnis ist auf die Fortpflanzung der Unsicherheiten zurückzuführen: Das Prinzip der Nutzung von Posteriori-Auswertungen als Priori-Information für nachfolgende Auswertepochen erlaubt eine Steigerung in der Ergebnisunsicherheit für die einzelnen Trendkoeffizienten. Implizit werden nicht nur die Informationen der direkten Vorgängerepoche fortgepflanzt, sondern durch die Rekursion auch die Unsicherheiten der weiter zurückliegenden Auswertepochen. Diese Tatsache findet sich nicht nur in der gezeigten Kontinuität der rein funktionalen Modellierung wieder, sondern auch in einer fortlaufenden Verringerung der HPDIs. Die Wahrscheinlichkeitsbereiche der informativen Modelle *R.1* und *K.1* konzentrieren sich auf numerisch sehr enge Spannen und sprechen so auch für eine sehr gute Approximationsgüte in der funktionalen Modellierung, da die Plausibilität für die einzelnen Trendkoeffizienten als hoch eingestuft werden kann.

Für die Ergebnisse der Unsicherheiten der Trendkoeffizienten in den nichtinformativen Modellen *R.2* und *K.2* lässt sich festhalten, dass auch diese zwischen Regressions- und Kollokationsansatz ähnliche Ergebnisse aufweisen. Im Quervergleich zu den informativen Modellen erweisen sich die HPDIs jedoch als wesentlich größer in den Spannen sowie variierend von Auswertepoche zu Auswertepoche. Zwar können sich die nichtinformativen Modelle durch Verzicht auf die Vorinformation und dadurch resultierende Zwänge individueller an die Daten anpassen, jedoch mit dem Ergebnis, dass die Modellparameter wesentlich unsicherer bestimmt werden. Insofern ist auch das Ergebnis der funktionalen Modellierungen zu relativieren: Insgesamt ist die Plausibilität der Trendkoeffizienten und damit letztlich auch die Abhängigkeit zur funktionalen Modellierung als wesentlich unsicherer einzustufen. Insbesondere durch die stark variierenden numerischen Ausprägungen wird eine zusätzliche Unsicherheit über die wahren Ausprägungen gegeben, was im Hinblick auf die sich kontinuierlich entwickelnden Schätzwerte und reduzierten Unsicherheitsmaße im informativen Ansatz vermieden wird.

Modell	R.1	R.2	K.1	K.2
R.1		0	1	0
R.2	2		2	1
K.1	1	0		0
K.2	2	1	2	
Summe	5	1	5	1
Rang	1.	2.	1.	2.

Legende:
0 - schlechter, 1 - neutral, 2 - besser

Abbildung 6.14: Paarvergleichsmatrix für die Unsicherheit der Trendkoeffizienten $\bar{\beta}$

Die Resultate für die Unsicherheit der Trendkoeffizienten lassen sich, wie in Abbildung 6.14 dargestellt, in einer Paarvergleichsmatrix zusammenfassen. Werden zunächst die Ergebnisse der Modelle *R.1* und *R.2* betrachtet, zeigt sich eine deutliche Präferenz für das Modell *R.1*. Aus diesem Grund wird das Modell *R.1* mit dem Wert 2 gegenüber dem Modell *R.2* gewertet. Das gleiche Bild ergibt sich im direkten Vergleich der Ergebnisse für *K.1* und *K.2*, da auch hier eine signifikante Verbesserung der Unsicherheiten durch die Integration und Fortpflanzung von Priori-

Information erreicht werden konnte. Für den Paarvergleich der informativen Modelle *R.1* und *K.1* wurde hingegen festgestellt, dass sich HPDIs mit vergleichbaren Spannen ergeben. Daher kann der Wert 1 für ein neutrales Ergebnis gesetzt werden. Ebenso ergibt sich keine Präferenz für das Modell *R.2* gegenüber dem Modell *K.2*, sodass ebenfalls der Wert 1 angenommen wird. Entsprechend der bisherigen Überlegungen ergibt sich abschließend, dass das Modell *R.1* dem Modell *K.2* vorzuziehen ist, ebenso wie das Modell *K.1* dem Modell *R.2*. Für die Paarvergleiche können deshalb die Werte 2 gesetzt werden. In der Summe ergeben sich die in Abbildung 6.14 dargestellten Rangfolgen. Wie erwartet, weisen die informativen Modelle Vorteile in der Unsicherheitsmodellierung gegenüber den nichtinformativen Modellen auf. Im Vergleich der Modelle *R.1* und *K.1* ist weiterhin keine weitere Differenzierung möglich. Gleiches trifft auf die Modelle *R.2* und *K.2* zu, die beide der Rangstufe 2 zuzuordnen sind.

Für die *Unsicherheit des Varianzfaktors* lassen sich durch die Integration von Priori-Information ebenfalls Vorteile erkennen. Die Varianz bzw. die daraus ableitbare Standardabweichung wurde in Abschnitt 2.2.2 (S. 34 ff.) als Maß für die Präzision definiert. Aus den in Abschnitt 6.1.1 (S. 158 ff.) genannten Gründen wird diese zwar in ihrer numerischen Ausprägung im Vergleich der informativen zur nichtinformativen Lösungen nicht zwingend verbessert, jedoch wird im Rahmen der informativen Modelle *R.1* und *K.1* der Modellparameter mit einer geringen Unsicherheit bestimmt. Bezogen auf die reine Modellformulierung ergeben sich damit auch hier Werte mit hoher Plausibilität unter den Rahmenbedingungen der Eingangsdaten und der Priori-Information. Die starken Variationen im nichtinformativen Fall deuten auf unsichere Schätzungen der Varianz hin. Dieser Umstand wird insbesondere bei der Bestimmung der HPDIs bedeutend, da die Varianz unmittelbar für deren Berechnung benötigt wird. Eine deutliche Verbesserung in der Unsicherheitsmodellierung lässt sich daher durch die Integration von Vorwissen erreichen. Weiterhin zeigt sich die Unsicherheit der Varianz in der informativen Bayesschen Kollokation gegenüber der in der informativen Bayesschen Regression deutlich verringert (vgl. Abbildung 5.43, S. 146 zu Abbildung 5.20, S. 132), was insbesondere auf die verbesserte funktionale Modellierung zurückzuführen ist. Werden die Ergebnisse im Paarvergleich gegenübergestellt, zeigt das Modell *K.1* in Bezug auf die Unsicherheitsmodellierung der Varianz die besten Ergebnisse; gegenüber den Modellen *R.1*, *R.2* und *K.2* ist damit der Wert 2 zu setzen. Das informative Modell *R.1* ist weiterhin gegenüber den Modellen *R.2* und *K.2* zu präferieren, da dieses durch die Integration von Priori-Information ebenfalls eine Verringerung der HPDIs bewirkt. Die Tatsache, dass durch einen Kollokationsansatz eine Verbesserung der funktionalen Modellierung zu erreichen ist und sich auf diese Weise die Varianz verringert, ergibt für den Paarvergleich der nichtinformativen Modelle *R.2* zu *K.2* einem Vorzug für die Bayessche Kollokation, ebenso wie zuvor im informativen Fall. Aus diesem Grund ist das Modell *K.2* mit einem Wert von 2 gegenüber *R.2* vorzuziehen. Werden die Wertungen auf diese Weise vorgenommen, ergibt sich die Rangfolge der Modelle gemäß Abbildung 6.15 mit den Stufen 1 bis 4. Die informative Bayessche Kollokation belegt dabei die erste Rangstufe, gefolgt von der informativen Bayesschen Regression. Auf der dritten Rangstufe findet sich die nichtinformativ Bayessche Kollokation, während die nichtinformativ Bayessche Regression die vergleichsweise größte Unsicherheit in der Schätzung des Varianzfaktors aufweist.

Modell	R.1	R.2	K.1	K.2
R.1		0	2	0
R.2	2		2	2
K.1	0	0		0
K.2	2	0	2	
Summe	4	0	6	2
Rang	2.	4.	1.	3.

Legende:
0 - schlechter, 1 - neutral, 2 - besser

Abbildung 6.15: Paarvergleichsmatrix für die Unsicherheit der Varianz σ^2

Abschließend für die Modellparameter lassen sich die *Unsicherheiten der Signalkomponente* in den Modellen *K.1* und *K.2* werten. Wie zu erwarten, zeigen sich auch hier die Vorteile der rekursiven Schätzung, wenn auch nicht so deutlich wie in den übrigen Modellparametern. Wie in den praktischen Untersuchungen beschrieben, ist

in den überwiegenden Fällen eine Verbesserung in den Unsicherheiten festzustellen. Zusammengefasst werden in Abbildung 5.44 (S. 147) die Medianwerte der HPDIs über die Auswertepochen 2005 bis 2014 gezeigt.

Insgesamt zeigen sich die HPDIs nicht so stark beeinflusst durch die Integration von Priori-Information wie bei den übrigen Modellparameter. Teilweise sind Epochen festzustellen, in denen die Konfidenzregionen der nichtinformativen Lösung unterhalb derer für die informative Lösung liegen (2005, 2010 bis 2012). Wie auch aus den zugehörigen Gleichungen zur Berechnung der entsprechenden Kofaktormatrizen deutlich wird, ergibt sich der Hauptanteil am Ergebnis durch die Kofaktormatrizen des Signals $\mathbf{Q}_{ss}$ und die Kofaktormatrix des Rauschanteils $\mathbf{Q}_{vv}$. Beide Matrizen sind im Vergleich der informativen und nichtinformativen Lösungen per Definition numerisch äquivalent. Die Auswirkungen der Kofaktormatrix $\mathbf{Q}_{\beta\beta}$ sind daher in ihrer Wirkung direkt aus den Abbildungen ablesbar: Da keine Vorinformation über das Signal selbst eingeführt wird, lässt sich schließen, dass die Priori-Information der Trendkoeffizienten zur Schätzung der Unsicherheiten der Signalkomponente insbesondere dann einen positiven Einfluss im Sinne schmalerer Konfidenzregionen aufweisen, wenn die nichtinformativ Lösung auf einem heterogeneren Datensatz beruht. Bestätigt wird dieser Schluss bei vergleichender Betrachtung der Abbildung 5.26 (S. 139), bei der für die Epochen 2005 sowie 2010 bis 2012 die vergleichsweise geringsten Varianzen in den Stichprobendaten festzustellen sind. Auch die Erkenntnisse bezüglich der Variationen der Trendkoeffizienten im Vergleich informativer zu nichtinformativer Lösung zeigen, dass die Integration von Priori-Information zu einer verbesserten Unsicherheit führt, wenn die Ausgangsdaten als heterogen einzustufen sind. Die Medianwerte der HPDIs zeigen weiterhin einen ähnlichen Verlauf im Vergleich zu denen der Posteriori-Schätzwerte der Zielgröße (vgl. Abbildung 5.45, S. 147). Daraus ist zu schließen, dass die Signalkomponente den Großteil der Ergebnisunsicherheit in den Werten $\bar{\mathbf{y}}$ trägt. Die Unsicherheit der Signalkomponenten tritt ausschließlich in den Modellen *K.1* und *K.2* der Bayesschen Kollokation auf; eine Gegenüberstellung mittels Paarvergleich wäre bei nur zwei Ansätzen nicht zielführend. Die Unsicherheit des Signals ist vielmehr als ein wesentlicher Aspekt der Unsicherheit in den Posteriori-Schätzwerten der Zielgröße zu sehen, weshalb dieser in den Paarvergleich bezüglich der Unsicherheit von $\bar{\mathbf{y}}$ bzw. $\bar{\mathbf{y}}_p$ integriert wird. Auf eine separate Wertung kann daher verzichtet werden.

Die Unsicherheitsmodellierung der vier Modelle ist als direktes Resultat der zufälligen Variabilität der Eingangsdaten aufzufassen. Im Sinne der einführenden Beschreibungen ist als weitere wesentliche Komponente die Unsicherheit der geschätzten Posteriori-Zielgrößen zu beurteilen. Eine Diskussion der Ergebnisse wird im folgenden Abschnitt gegeben.

6.2.2. Unsicherheit in der Schätzung und Prädiktion der Posteriori-Zielgröße

In Ergänzung zu den Ergebnissen der funktionalen Modellierung für Schätzung und Prädiktion der Posteriori-Zielgröße in den Ansätzen *R.1*, *K.1*, *R.2* und *K.2* lassen sich die resultierenden Unsicherheiten diskutieren und in einem Paarvergleich gegenüberstellen. Im folgenden Unterabschnitt werden zunächst die Unsicherheiten in der Posteriori-Zielgröße $\bar{\mathbf{y}}$ betrachtet. Im zweiten Unterabschnitt stehen die Unsicherheiten der Prädiktionsergebnisse $\bar{\mathbf{y}}_p$ im Fokus.

6.2.2.1. Diskussion der Unsicherheit in der Schätzung der Posteriori-Zielgröße

Die Unsicherheit der Posteriori-Schätzwerte der Zielgröße ergibt sich unmittelbar aus der Unsicherheit des Modellansatzes in Form der vorstehend beschriebenen Modellparameter. Insofern drücken sich neben der zufälligen Variabilität der Beobachtungen $\mathbf{y}$ hier insbesondere die Unsicherheiten der Modellierung aus. Konkret ergeben sich für die HPDIs der Bayesschen Regressionsansätze die in den Abbildungen 5.7 bis 5.9 (S. 128) sowie den Abbildungen des Anhangs B.1 (S. 209 ff.) dargestellten Ergebnisse, für die Bayesschen Kollokationsansätze die den Abbildungen 5.30 bis 5.32 (S. 141) und dem Anhang C.2 (S. 221 ff.) zu entnehmenden.

Im Kontext der Diskussionen zur funktionalen Modellierung wurde bereits festgestellt, dass in den Ergebnissen der teilweise hohe Unterschied zwischen Kaufpreis und geschätztem Vergleichswert zu erkennen ist. Für die Modelle *R.1* und *R.2* der informativen und nichtinformativen Bayesschen Regression hat sich dennoch ergeben, dass die Schätzwerte als sehr plausibel einzustufen sind. Wie aus den genannten Abbildungen hervorgeht, sind in beiden Modellen die Konfidenzregionen der Posteriori-Zielgröße $\bar{\mathbf{y}}$ durch sehr enge Wertebereiche gekennzeichnet, was zunächst für eine geringe Unsicherheit in der Schätzung der Vergleichswerte spricht. Bei einem direkten Vergleich

der HPDI-Differenzen lassen sich zudem pro Auswerteepoche geringere Unsicherheitsspannen in der Schätzung basierend auf informativen Bayesschen Regressionsmodellen feststellen. Weiterhin konnte beobachtet werden, dass sich die Medianwerte der HPDIs im informativen Ansatz über den gesamten Auswertezeitraum als konstant zeigen, was die Fortpflanzung der Unsicherheitsinformation im rekursiven Prozess bestätigt (vgl. Abbildung 5.50, S. 151). Stärkere Variationen sind wiederum in der nichtinformativen Lösung der Bayesschen Regression zu erkennen. Eine Tendenz in der Entwicklung ist hier nicht zu beobachten. Über den gesamten Auswertezeitraum ergeben sich die Spannen der Konfidenzregionen im Modell *R.1* als enger im Vergleich zum nichtinformativen Ansatz *R.2*. Die Ergebnisse zeigen neben der sich positiv auswirkenden Fortpflanzung von Unsicherheiten, dass die wahren Werte für die Schätzwerte mit hoher Wahrscheinlichkeit lediglich in sehr geringen Spannen liegen. Dies spricht im Prinzip für die hohe Güte der Modellierung: Unter den gegebenen Rahmenbedingungen und den vorgenommenen Modellspezifikationen lassen sich Schätzwerte mit hoher Plausibilität ableiten.

Wie in den Ausführungen zum Modellbegriff nach Abschnitt 2.2.1 (S. 31 ff.) deutlich wird, soll die Modellbildung ein möglichst übereinstimmendes Bild zwischen Realität und Approximation ergeben. Nach diesem ausschließlich theoretischen Anspruch müsste das Ergebnis für die Schätzung der Unsicherheiten als zu positiv für die Ermittlung von Vergleichswerten beurteilt werden. Die Schätzwerte weisen einen sehr geringen Wahrscheinlichkeitsbereich auf, die in den überwiegenden Fällen die realen Kaufpreisbeobachtungen nicht beinhalten. Insofern lässt sich der praktische Bezug der geschätzten Funktionen zum realen Marktgeschehen in Frage stellen. Wie jedoch bereits im Zuge der Diskussion der funktionalen Modellierung erläutert wird, ist der Modellbegriff im Hinblick auf ökonometrische Modelle zu relativieren: Kaufpreise und geschätzte Vergleichswerte müssen nicht zwingend übereinstimmen, da letztere ausschließlich Funktionswerte darstellen und wesentlich auf der angewendeten Auswertestrategie sowie auf sachverständigem Ermessen der auswertenden Person beruhen. In dieser Hinsicht werden die resultierenden Vergleichswerte stets mehr oder weniger große Differenzen zu den Kaufpreisen aufweisen; eine vollständige Automatisierung in der Ableitung von Verkehrswerten ist grundsätzlich nicht möglich.

Unter diesen Voraussetzungen können die Unsicherheiten in der Schätzung der Posteriori-Zielgröße sowohl im informativen Modell *R.1* als auch nichtinformativen Modell *R.2* in Bezug auf die reine Modellbildung betrachtet werden. Die wesentliche Aussage der Unsicherheiten ist demnach nicht, wie stark Vergleichswert und Kaufpreis voneinander abweichen, sondern dass die Modelle der informativen und nichtinformativen Bayesschen Regression unter den vorgenommenen Spezifikationen und bezogen auf die konkreten Eingangsdaten einer Stichprobe eine Modellierung von Vergleichswerten mit einer sehr geringen Unsicherheit erlauben. Bedingt werden diese durch die Modellbildung – oder konkreter durch die Ermittlung der Modellparameter – deren Unsicherheiten sich direkt in der Unsicherheit der geschätzten Zielgrößen abbildet. Soll hingegen der Bezug zum Kaufpreis hergestellt werden, muss festgehalten werden, dass die Wahrscheinlichkeit für einen Kaufpreis innerhalb der Konfidenzregionen sehr gering ausfällt. Als Schluss lässt sich festhalten, dass die Schätzwerte zwar eine sehr geringe Unsicherheit aufweisen, jedoch keinen unmittelbaren Bezug zum realen Marktgeschehen aufweisen.

Unter diesem Aspekt lassen sich auch die Unsicherheiten in den geschätzten Zielgrößen der Bayesschen Kollokationsansätze *K.1* und *K.2* beurteilen. Wie aus den Abbildungen des Anhangs C.2 (S. 221 ff.) zu entnehmen ist, ergeben sich größere Unsicherheitsspannen für die Schätzwerte als für die vergleichbaren HPDIs der Bayesschen Regressionsansätze *R.1* und *R.2*. Da die Trendkoeffizienten der Kollokationsansätze zu den Trendkoeffizienten der Bayesschen Regressionen vergleichbare Unsicherheitsspannen aufweisen, ist dies wie bereits angedeutet auf die Einflüsse der Signalkomponente zurückzuführen, die als zusätzlicher Modellparameter mit eigenem Unsicherheitshaushalt auf die Schätzungen von $\bar{y}$ wirkt. Im Vergleich der informativen und nichtinformativen Lösung in der Bayesschen Kollokation ist keine durchgängige Verbesserung durch die Integration von Priori-Information zu erkennen: In den einzelnen Epochen ergeben sich teilweise geringere Unsicherheiten in den Schätzwerten der Posteriori-Zielgrößen. Bestätigt wird dies durch die Abbildung 5.45 (S. 147), in der die Ergebnisse der Einzelepochen zusammengefasst werden. Hier liegen die entsprechenden Medianwerte der HPDIs des Ansatzes *K.1* in den Epochen 2005 sowie 2010 bis 2012 unterhalb der Ergebnisse des nichtinformativen Ansatzes *K.2*, während in den übrigen Epochen eine Verbesserung der Unsicherheiten zu beobachten ist. Zusammenfassend lässt sich festhalten, dass durch die Fortpflanzung der Unsicherheiten über den gesamten Auswertezeitraum eine kontinuierliche

Entwicklung in den Unsicherheitsmaßen erreicht werden kann, die überwiegend eine Verbesserung gegenüber den größeren Variationen des nichtinformativen Ansatzes ergibt.

Bezüglich der funktionalen Approximation konnte in den Modellen *K.1* und *K.2* bereits eine deutliche Verbesserung zu den jeweiligen Regressionsansätzen festgestellt werden; die Unsicherheitsspannen haben sich im Vergleich hingegen erwartungsgemäß vergrößert und bewirken daher die Schätzung von Vergleichswerten $\bar{y}$ mit größeren Wahrscheinlichkeitsbereichen. Wie aus den Abbildungen des Anhangs C.2 (S. 221 ff.) ebenfalls zu entnehmen ist, liegen die Kaufpreise in sämtlichen Auswertepochen größtenteils innerhalb der HPDIs. Unter diesem Aspekt lässt sich der Schluss ziehen, dass die rein mathematische Modellformulierung in einem Kollokationsansatz zwar unsicherere Werte mit sich bringt, jedoch einen größeren Bezug zum realen Marktgeschehen aufweist. An dieser Stelle erweist sich die Unsicherheit in der mathematischen Modellierung im Gegensatz zu den Ergebnissen der Bayesschen Regressionen als Vorteil für den unmittelbaren Marktbezug. Zudem liegen die numerischen Werte für die Spannen der Konfidenzregionen in einem für die Verkehrswertermittlung realistischen Rahmen.

Modell	R.1	R.2	K.1	K.2
R.1		0	0	0
R.2	2		0	0
K.1	2	2		0
K.2	2	2	2	
Summe	6	4	2	0
Rang	1.	2.	3.	4.

Legende:

0 - schlechter, 1 - neutral, 2 - besser

Abbildung 6.16: Paarvergleichsmatrix für die Unsicherheit der Schätzwerte der Posteriori-Zielgröße $\bar{y}$

Die Erkenntnisse über die Unsicherheitsmodellierung in den Schätzwerten der Posteriori-Zielgröße lassen sich in der Paarvergleichsmatrix nach Abbildung 6.16 zusammenfassen. Werden zunächst die Modelle *R.1* und *R.2* miteinander verglichen, wird durch die Berücksichtigung von Priori-Informationen eine Verringerung in den Spannen der HPDIs erreicht. Die Spannen der Konfidenzregionen des Modells *R.2* liegen durchgängig für jede Auswertepochen überhalb derer des Modells *R.1*. Aus diesem Grund erweist sich die informative Bayessche Regression als der Ansatz mit der größten Präzision in der Schätzung von Vergleichswerten. Dementsprechend wird im Paarvergleich der Wert 2 zu Gunsten des Modells *R.1* gesetzt. Die gleiche Wertung lässt sich im Vergleich zu den Modellen *K.1* und *K.2* vornehmen, die aufgrund der zusätzlichen Unsicherheit der Signalkomponente über größere HPDIs verfügen. Wird das Modell *R.2* den Modellen *K.1* und *K.2* gegenübergestellt, zeigt sich auch hier eine geringere Unsicherheit in den geschätzten Vergleichswerten; wiederum wird der Einfluss der Unsicherheit der Signalkomponente deutlich. Das Modell *R.2* ist daher gegenüber den Kollokationsansätzen mit einem Wert von 2 zu beurteilen. Werden abschließend die Modelle *K.1* und *K.2* miteinander verglichen, kann zwar keine klare Präferenz für einen der Ansätze erkannt werden. Dennoch erweisen sich die Größenordnungen der HPDIs im Ansatz *K.1* durch den Prozess der Rekursion als stabil, während sich die Unsicherheitsspannen im Ansatz *K.2* als schwankend erweisen. Insofern wird die Fortpflanzung der Unsicherheit in der informativen Bayesschen Kollokation als Vorteil gewertet hinsichtlich der Stabilität über den Auswertezeitraum gewertet. Die teilweise kleineren Unsicherheitsspannen des nichtinformativen Modells *K.2* werden der zufälligen Zusammenstellung der Stichprobendaten zugeschrieben. Für das Modell *K.1* wird gegenüber dem Modell *K.2* der Wert 2 gesetzt. Insgesamt ergibt sich so die in Abbildung 6.16 dargestellte Rangfolge. Die informative Bayessche Regression weist die geringsten Unsicherheitsspannen auf und wird entsprechend mit der Rangstufe 1 bewertet. An zweiter Stelle folgt das Modell der nichtinformativen Bayesschen Regression. Auf dem dritten Rang findet sich die informative Bayessche Kollokation, während die nichtinformativen Bayessche Kollokation mit der Rangstufe 4 die vergleichsweise größten HPDIs aufweist.

Abschließend für die Diskussion der Ergebnisse soll im folgenden Abschnitt die Beurteilung der Unsicherheiten für die Prädiktion thematisiert werden.

6.2.2.2. Diskussion der Unsicherheit in der Schätzung der prädizierten Posteriori-Zielgröße

Im Kontext der Anwendung der rekursiven Auswertestrategie konnte bereits ein Mehrwert informativer Ansätze für die Unsicherheitsmodellierung der Modellparameter und der Schätzwerte der Posteriori-Zielgröße $\bar{y}$ nachgewiesen werden. Für eine Aussage zur Prädiktion von Vergleichswerten sind im Folgenden die Unsicherheiten in der prädizierten Posteriori-Schätzung der Zielgröße gegenüberzustellen. Wie im Zusammenhang mit der funktionalen Modellierung erläutert, wurde zu diesem Zweck die Methode der Kreuzvalidierung genutzt: Durch die Elimination eines ausgewählten Kauffalls aus der jeweils betrachteten Stichprobe ergeben sich lediglich geringfügige Änderungen in den numerischen Ausprägungen der Modellparameter. Gleichmaßen sind für die Unsicherheitsmodellierung der Modellparameter in den reduzierten Modellen ebenfalls nur geringe Änderungen in den HPDIs zu erwarten; auf eine Diskussion der Modellparameter für die Modelle, die in der Kreuzvalidierung geschätzt werden, wird daher an dieser Stelle verzichtet. Im Fokus stehen die Unsicherheiten für die Prädiktion von Vergleichswerten.

Für eine grafische Veranschaulichung der *Ergebnisse der Bayesschen Regressionsansätze R.1 und R.2* wird auf die exemplarischen Epochen 2005, 2009 und 2014 in den Abbildungen 5.46 bis 5.48 (S. 150) bzw. die Darstellungen des gesamten Auswertzeitraumes in Anhang B.3 (S. 215 ff.) verwiesen. Im Vergleich zu den HPDIs der Schätzungen der Posteriori-Zielgröße $\bar{y}$ ergeben sich im Zuge der Prädiktion ungleich größere Wertespannen. Die festgestellte geringe Unsicherheit in den Modellparametern spiegelt sich insofern nicht in den reinen Schätzergebnissen für $\bar{y}_p$ wider, was sowohl für den informativen Fall *R.1* als auch den nichtinformativen Fall *R.2* zutrifft.

Neben den Einzeldarstellungen der Auswertepochen können für eine allgemeine Einordnung die Abbildungen 5.21 (S. 133) und 5.50 (S. 151) gegenübergestellt werden. Im direkten Vergleich der Größenordnung der Medianwerte der HPDIs ist zu sehen, dass diese sich über den Auswertzeitraum 2005 bis 2014 um den durchschnittlichen Faktor 7 im informativen Fall und den Faktor 8 im nichtinformativen Fall erhöht haben und so in den Prädiktionen ein größerer Wahrscheinlichkeitsbereich abgedeckt wird.

Trotz der Vergrößerung der Konfidenzregionen kann auch hier der Vorteil der rekursiven Auswertestrategie festgestellt werden: Die Unsicherheiten der informativen Modellschätzungen *R.1* für die prädizierten Vergleichswerte $\bar{y}_p$ liegen für sämtliche Auswertepochen unterhalb derer der nichtinformativen Ansätze. Die Verläufe, die sich bereits bei der Schätzung von $\bar{y}$ gezeigt haben, bestätigen sich. Als Vorteil der informativen Lösung im Vergleich zur nichtinformativen Lösung ist zu erkennen, dass neben den verringerten Wahrscheinlichkeitsbereichen durch die Rekursion wiederum eine kontinuierliche Entwicklung in den Spannen erreicht wird. Demgegenüber stehen die HPDIs des nichtinformativen Modells *R.2*, welche wie im Zusammenhang mit der Schätzung der Unsicherheiten von $\bar{y}$ stärker variieren und ausschließlich durch die Stichprobendaten der aktuellen Auswertepoche beeinflusst sind.

Die Wahrscheinlichkeitsbereiche decken insgesamt einen größeren Bereich ab und werden so der Tatsache gerecht, dass sich ohne Kaufpreisinformation und lediglich basierend auf wertbeeinflussenden Merkmalen ein Vergleichswert vornehmlich durch sachverständige Einschätzung ergibt. Diese Unsicherheit in der realen Entscheidungsfindung wird auch durch die rein mathematischen Zusammenhänge abgebildet, sodass die Prädiktion in den Bayesschen Regressionsmodellen *R.1* und *R.2* als zutreffend zu beurteilen ist. Anhand der Abbildungen in Anhang B.3 (S. 215 ff.) ist zu erkennen, dass die Kaufpreise in jeder Epoche unabhängig von der Schätzung in Modell *R.1* oder *R.2* mit wenigen Ausnahmen innerhalb der Konfidenzregionen liegen. Eine Systematik für die Fälle, in denen die Kaufpreise außerhalb liegen, ist nicht zu erkennen und scheint damit dem Zufall zu unterliegen. Für die Unsicherheitsbetrachtung ergibt sich der wesentliche Schluss, dass sich zum einen bedingt durch die Prädiktion und damit verbundene mangelnde Kaufpreisinformation, große Unsicherheiten in der Schätzung der Vergleichswerte ergeben. Zum anderen ist dieser Aspekt im Kontext der ökonometrischer Modellbildung und speziell der Verkehrswertermittlung als Vorteil zu werten, da Vergleichswerte definitionsgemäß größere Abweichungen zwischen Kaufpreisen und Funktionswerten abzufangen haben als beispielsweise rein ingenieurgeodätisch basierte Modelle der Messtechnik, an die ungleich höhere Forderungen gestellt werden. Im Rahmen der Prädiktion lassen sich so Bereiche vorgeben, innerhalb derer die Kaufpreise mit vorgegebener Wahrscheinlichkeit vermutet werden. Für die Ableitung von Vergleichsfaktoren ist die Angabe der HPDIs basierend auf der Prädiktion als geeignet zu beurteilen, da sie einen größeren Raum für den gutachterlichen Sachverstand lassen, der letztlich dann benötigt wird, um von einem Vergleichswert zum Verkehrswert zu gelangen (vgl. Abbildung 2.6, S. 42).

Eine Verringerung der Unsicherheiten in den Prädiktionsergebnissen lässt sich durch die *Bayesschen Kollokationsansätze K.1* und *K.2* erreichen. Wie den Ergebnissen nach den Darstellungen in den Abbildungen 5.54 bis 5.56 (S. 153) für die exemplarischen Epochen 2005, 2009 und 2014 zu entnehmen ist und durch die Darstellung sämtlicher Auswertepochen bestätigt wird (vgl. Anhang C.4, S. 234 ff.), ergeben sich im Vergleich zu den HPDIs der Ansätze *R.1* und *R.2* Konfidenzregionen mit geringeren Wertespanssen. Im Vergleich der Spannen informativer Bayesscher Regression zu Bayesscher Kollokation liegen diese um den durchschnittlichen Faktor 3,5 niedriger, bezüglich der nichtinformativen Bayesschen Regression zu Bayesscher Kollokation um den Faktor 4,5. Die Verringerung in beiden Modellansätzen ist auf das erweiterte stochastische Modell zurückzuführen, welche in den Kollokationsansätzen berücksichtigt werden. Durch die zusätzliche Ausschöpfung von Restinformationen und deren stochastische Modellierung in der Signalkomponente kann im Gegensatz zur Prädiktion in der Bayesschen Regression eine Verringerung der Unsicherheiten erreicht werden, wie anhand der Ergebnisse deutlich wird. Im Vergleich der geschätzten HPDIs für die Schätzung der Posteriori-Zielgröße $\bar{y}$ der Bayesschen Kollokation zu den Unsicherheiten der prädizierten Posteriori-Zielgrößen $\bar{y}_p$ ergibt sich eine Vergrößerung der Wertespanssen mit dem Faktor 2,2 im informativen Kollokationsmodell *K.1* und dem Faktor 2,3 im nichtinformativen Kollokationsmodell *K.2*. Die Unterschiede sind damit wesentlich kleiner als die in den Lösungen der Bayesschen Regressionen. Somit kann festgehalten werden, dass sich wie im Rahmen der Diskussion zu den funktionalen Modellen in der Prädiktion durch einen Kollokationsansatz zwar keine nennenswerte Steigerung in der Approximationsgüte, durch die zusätzliche stochastische Information jedoch eine verbesserte Unsicherheitsmodellierung in den Prädiktionen erreichen lässt.

Weiterhin ist bezüglich der Unsicherheitsspannen in den informativen und nichtinformativen Bayesschen Kollokationsansätzen auffällig, dass sich in einigen Epochen größere prädizierte Unsicherheitsspannen für die informative Lösung im Vergleich zur nichtinformativen Lösung ergeben (vgl. Abbildung 5.59, S. 156). Konkret ist dies in den Epochen 2005 sowie 2010 bis 2012 der Fall und entspricht damit den Erkenntnissen, die bereits in Verbindung mit den Unsicherheitsspannen der Signalkomponente $\bar{s}$ und deren Einfluss auf die Schätzung von $\bar{y}$ formuliert wurden. Für die prädizierte Signalkomponente $\bar{s}_p$ ist gemäß Abbildung 5.58 (S. 155) ein ähnliches Muster erkennbar.

Äquivalente Ergebnisse haben sich auch bereits im Zusammenhang mit den entsprechenden Medianwerten der Schätzung von $\bar{y}$ sowie der zugehörigen Schätzung des Varianzfaktors gezeigt. In sich werden die numerischen Resultate der Prädiktion damit zwar bestätigt, sprechen jedoch nicht für einen pauschalen Mehrwert der Integration von Priori-Wissen für Prädiktionszwecke. Die Ursachen finden sich aufgrund der ermittelten Unsicherheiten in den Modellparametern der Trendkoeffizienten und des Varianzfaktors in der zusätzlichen Berücksichtigung des Signals. Da die Berechnung der Signalkomponente keine Vorinformation über das Signal selbst berücksichtigt, sondern ausschließlich aus den übrigen Modellparametern Vorwissen bezieht, scheinen die Effekte des Signals das Priori-Wissen über die übrigen Modellparameter zu überlagern – in den Auswertepochen, in denen die Streuung in den Daten gering ausfällt, ergeben sich im Modell *K.2* für die Unsicherheiten der prädizierten Schätzwerte $\bar{y}_p$ geringere Spannen als in der informativen Lösung *K.1*. Liegt eine hohe Streuung in den Daten vor, erweist sich die informative Lösung *K.1* durch die Berücksichtigung von Vorinformationen als geringer in den Wertespanssen der HPDIs, da die stärkere Datenheterogenität durch die Ergebnisse vorheriger Auswertepochen abgefangen wird. Daraus ergibt sich ein genereller Vorteil für die informative Bayessche Kollokation, die Unregelmäßigkeiten im Datenmaterial geeignet berücksichtigen kann.

Im informativen Fall des Modells *K.1* sowie im nichtinformativen Fall des Modells *K.2* ist in den Abbildungen des Anhangs C.4 (S. 234 ff.) zu sehen, dass trotz der im Vergleich zu den Bayesschen Regressionen verringerten Unsicherheitsspannen die originären Kaufpreise in sämtlichen Auswertepochen größtenteils innerhalb der HPDIs liegen. Weiterhin ist zu erkennen, dass die informative Lösung *K.1* dabei eine kleinere Spanne umfasst. Analog zu den vorstehenden Ausführungen zu den Regressionsansätzen *R.1* und *R.2* lässt sich auch hier der Schluss ziehen, dass die mathematischen Modellierungen im Prädiktionsansatz der Bayesschen Kollokation zwar generell mit größeren Unsicherheitsspannen verbunden sind, diese sich jedoch im Hinblick auf die Eigenschaften ökonomischer Modelle als Vorteil in der praktischen Anwendung der indirekten Vergleichswertermittlung erweisen. Zusätzlich ergibt sich der Vorteil, dass die Modelle der Bayesschen Kollokation zwar im Vergleich zu den Bayesschen Regressionsmodellen über eine geringere Unsicherheit in der Prädiktion verfügen, die Spannen jedoch ausreichend

groß sind, um die Kaufpreise mit der gewählten Wahrscheinlichkeit zu überdecken. Die Unsicherheit in den Vergleichswerten kann daher sowohl durch die Verwendung eines informativen als auch nichtinformativen Bayesschen Kollokationsansatzes verringert werden, ohne den Bezug zum realen Marktgeschehen einzuschränken.

Werden die Ergebnisse im Paarvergleich gegenübergestellt, ergibt sich die Paarvergleichsmatrix für die Unsicherheit der prädizierten Schätzwerte der Posteriori-Zielgröße nach Abbildung 6.17. Die Resultate des Modells *R.1* haben eine Präferenz gegenüber dem Modell *R.2* gezeigt; die Unsicherheit in den Schätzwerten $\bar{y}_p$ ist geringer einzustufen, was zum Wert 2 für den unmittelbaren Paarvergleich führt. Im Vergleich zu den Ansätzen der Bayesschen Kollokation haben sich die HPDIs des Modells *R.1* hingegen als größer erwiesen. Aus diesem Grund werden die Kollokationsansätze jeweils gegenüber dem Modell *R.1* mit dem Wert 2 beurteilt. Das Modell *R.2* weist im Gesamtvergleich die größten Unsicherheitspannen in den prädizierten Schätzwerten auf, so dass die Bayesschen Kollokationsmodelle ergänzend zum Vergleich mit *R.1* ebenfalls eine höhere Präferenz mit dem Wert 2 erhalten. Der direkte Vergleich der informativen und nichtinformativen Bayesschen Kollokation ist ähnlich wie die Unsicherheit in der Schätzung der Posteriori-Zielgröße $\bar{y}$ zu werten: Zwar zeigen sich in den Epochen 2005 sowie 2010 bis 2012 für das nichtinformative Modell *K.2* geringere Spannen der HPDIs im Vergleich zum Modell *K.1*, jedoch variieren diese über den gesamten Auswertzeitraum stärker und liegen überwiegend über den Spannen der informativen Lösung. Da zudem eine kontinuierliche Entwicklung und damit verbunden eine Stabilität der HPDIs durch die rekursive Schätzung erreichbar ist, wird die informative Bayessche Kollokation präferiert. Entsprechend ist im direkten Paarvergleich der Wert 2 zu setzen.

Modell	R.1	R.2	K.1	K.2
R.1		0	2	2
R.2	2		2	2
K.1	0	0		0
K.2	0	0	2	
Summe	2	0	6	4
Rang	3.	4.	1.	2.

Legende:
0 - schlechter, 1 - neutral, 2 - besser

Abbildung 6.17: Paarvergleichsmatrix für die Unsicherheit der prädizierten Schätzwerte der Posteriori-Zielgröße $\bar{y}_p$

Die Rangfolgen ergeben sich aus dem Paarvergleich wie in Abbildung 6.17 dargestellt. Im Gesamtvergleich zeigt sich die informative Bayessche Kollokation als das Modell mit den geringsten Unsicherheiten in der Schätzung der prädizierten Posteriori-Zielgröße. Auf dem zweiten Rang findet sich die nichtinformativ Bayessche Kollokation, gefolgt von der informativen und der nichtinformativen Bayesschen Regression auf dem dritten und vierten Rang. Das Ergebnis lässt keine pauschale Bevorzugung von Modellen mit Integration von Priori-Information erkennen. Für die Prädiktion spielt insbesondere die Ausnutzung der zusätzlichen Information der prädizierten Signalkomponenten eine Rolle. Die Vorteile der Bayes-Strategie zeigen sich erst im direkten Vergleich der Modelle *K.1* und *K.2*, da sich im Modell *K.1* die Einflüsse der Priori-Information positiv im Sinne verringerter Unsicherheitspannen auswirken.

Basierend auf den Ergebnissen und Darstellungen wird im folgenden Abschnitt ein Fazit zum zweiten Forschungsziel gezogen.

6.2.3. Fazit zum zweiten Forschungsziel

Für das zweite Forschungsziel lässt sich auf Grundlage der gewonnen Erkenntnisse das folgende Fazit ziehen:

Durch die Einführung der Bayes-Strategie in die Verkehrswertermittlung wird eine erweiterte statistische Behandlung von Unsicherheiten im Bewertungsprozess ermöglicht. Im Grundlagenteil nach Kapitel 2 wurde erläutert, welche hervorgehobene Bedeutung Kaufpreise in der Verkehrswertermittlung als Grundlage jeder Marktanalyse aufweisen: Eine Verknüpfung zwischen Kaufpreisen und theoretischen, ökonometrischen Modellen stellt sich als unverzichtbar dar. Werden die Unsicherheitskomponenten der Vagheit und der Unschärfe explizit ausgeschlossen, lässt sich durch die Formulierung des indirekten Vergleichsverfahrens in den vorgestellten Bayesschen Modellen die zufällige, kaufpreisinduzierte Variabilität umfassend abbilden. Mit der in der Arbeit dargelegten Methodik wird

der Sichtweise Rechnung getragen, dass Kaufpreise grundsätzlich als Zufallswerte anzusehen sind, die damit dem Bereich der Wahrscheinlichkeit zugeordnet werden. Aufgrund der nicht beliebig oft wiederholbaren Ermittlung von Kaufpreisen – welche ausschließlich durch wiederholte Veräußerung ein- und desselben Objekts durch unterschiedliche Vertragsparteien erreicht werden könnte – wird der Zufallscharakter deutlich. In Verbindung mit der zunächst rein sachlogisch begründeten kausalen Verbindung zu wertbeeinflussenden Merkmalen lässt sich die Unsicherheit in Schätzergebnissen nicht vermeiden. Da die Bayes-Strategie im Sinne der Plausibilität von Schätzergebnissen interpretiert wird, lässt sich im Hinblick auf das komplexe Zusammenspiel von unsicheren Zusammenhängen sowie zufallsbedingten Kaufpreisen festhalten, dass die resultierenden Unsicherheiten für die Modellparameter und die Schätzung von Vergleichswerten einen signifikanten Mehrwert für Interpretationen in der Verkehrswertermittlung mit sich bringen. In diesem Kontext wird das zweite formulierte Forschungsziel erreicht:

1. **Die praktischen Untersuchungen anhand exemplarischer Stichproben zeigen, dass durch Ansätze der informativen Bayesschen Regression und Bayesschen Kollokation eine signifikante Verringerung der Unsicherheiten in der Schätzung der Modellparameter sowie der Zielgrößen erreicht wird.** Ermöglicht wird diese Aussage durch die Gegenüberstellung mit den Lösungen entsprechender nichtinformativer Ansätze, in denen keine Vorinformationen vorliegen und die in den Schätzalgorithmen den klassisch-frequentistischen Ansätzen ähneln. Einschränkend ergeben sich für den Modellparameter des Signals im Kollokationsansatz nicht in jeder Auswertepoche des Beispieldatensatzes signifikante Verringerungen. Eine Erklärung lässt sich darin finden, dass für dessen Berechnung keine Vorinformationen über die Signalstruktur der vorherigen Auswertepochen verwendet wird, diese vielmehr eine Modifikation durch die Verwendung von Vorinformationen über Trendkoeffizienten und deren Unsicherheit erfahren. Größtenteils lässt sich jedoch auch hier ein Mehrwert durch die Integration der Vorinformation nachweisen.
2. **Für die Prädiktion lässt sich ebenfalls ein Mehrwert in Form verringerter Unsicherheitsspannen für die Schätzungen der prädizierten Posteriori-Zielgröße erreichen.** Sowohl für die informative Bayessche Regression als auch die nichtinformativ Bayessche Kollokation wird im direkten Vergleich zu den nichtinformativen Varianten eine Verringerung in den Unsicherheiten erzielt.
3. **Für das Bayessche Vergleichsverfahren wurde eine Auswertestrategie entwickelt, die auf die Anforderungen der heterogenen Datengrundlage zugeschnitten ist und die in ihrer rekursiven Anwendung die Fortpflanzung von Unsicherheiten in den Eingangsdaten und den Modellparametern erlaubt.** Für die Zwecke der Vergleichswertermittlung erlaubt die Auswertestrategie eine Fortpflanzung der Unsicherheiten im Bewertungsprozess. Zudem ermöglicht die Auswertestrategie eine kontinuierliche Darstellung des gewählten räumlichen und sachlichen Teilmarktes, die in der Lage ist, kurzfristige Änderung am Immobilienmarkt zu dämpfen und rein numerisch bedingte Streuungen in den Daten abzufangen.

Zusammenfassend lässt sich festhalten, dass das zweite Forschungsziel erreicht wurde. Im Rahmen der Untersuchungen hat sich gezeigt, dass in der Gesamtzusammenstellung die informative Bayessche Kollokation und die informative Bayessche Regression die Ergebnisse mit den geringsten Unsicherheitsspannen der vier Modelle aufweisen. Die Integration von Priori-Information und die entwickelte Auswertestrategie lassen sich somit als genereller Mehrwert der Bayes-Strategie für die Vergleichswertermittlung beurteilen. Der Beitrag rekursiver Bayesscher Schätzverfahren zur Verkehrswertermittlung wird in diesem Kontext eingehender im folgenden Abschnitt diskutiert.

6.3. Beitrag rekursiver Bayesscher Schätzverfahren zur Verkehrswertermittlung

Basierend auf der exemplarischen Stichprobe der Ein- und Zweifamilienhäuser in der Wohnungsmarktregion Osnabrück wurde in den vorherigen Abschnitten nachgewiesen, dass sich die Bayes-Strategie erfolgreich auf die Vergleichswertermittlung übertragen lässt. In der funktionalen Modellierung haben sich die Vorteile gegenüber den bisherigen, klassischen Ansätzen nicht gezeigt, in der Unsicherheitsmodellierung ist durch die rekursive Auswertung jedoch ein deutlicher Mehrwert zu erkennen. Bislang wurden die Forschungsziele lediglich im Hinblick auf die konkret ausgewerteten Stichproben untersucht. Von Interesse ist daher, wie die Ergebnisse in Bezug auf die generelle Vergleichswertermittlung zu werten sind – es stellt sich die Frage, wo die Stärken und Schwächen der einzelnen

Modelle liegen. In Abschnitt 6.3.1 wird das Forschungsergebnis daher im Kontext der Vergleichswertermittlung gewertet.

Multivariate statistische Analyseverfahren werden darüber hinaus nicht nur in der Vergleichswertermittlung genutzt, sondern tragen durch die Anwendung in weiteren Aufgabenstellungen der Verkehrswertermittlung zur Analyse des Grundstücksmarktes bei. Im Fokus stehen die Ableitung von Kenngrößen und Daten für die Verkehrswertermittlung, die einen Beitrag zur Grundstücksmarkttransparenz leisten. Teilweise werden diese auf Grundlage der gesetzlichen Vorgaben des BauGB abgeleitet, teilweise handelt es sich um Anwendungen der Statistik, die als Hilfsmittel für eine zusätzliche mathematisch-statistisch begründete Abbildung des Marktgeschehens genutzt werden. Als Beispiel dient der niedersächsische Immobilien-Preis-Kalkulator (IPK), der insbesondere von nicht sachverständigen Personen genutzt werden kann, um das Preisniveau einer Immobilie zu ermitteln. Das Preisniveau darf zwar nicht mit dem Verkehrswert gleichgesetzt werden, trägt aber dennoch zur Grundstücksmarkttransparenz bei. Abschnitt 6.3.2 ist aus diesen Gründen der Wertung der Forschungsergebnisse im Kontext der Verkehrswertermittlung gewidmet. Anhand typischer Aufgabenfelder der Verkehrswertermittlung wird die Übertragungsmöglichkeit der Forschungsergebnisse diskutiert und so der Beitrag der Arbeit in den übergeordneten Rahmen eingeordnet. Die Ausführungen werden dabei als Ausblick aufgefasst; detaillierte Ausführungen zu spezifischen Vorgehensweisen in der Umsetzung sind nicht beabsichtigt und bleiben künftigen Forschungsaufgaben überlassen.

6.3.1. Wertung der Ergebnisse im Kontext der Vergleichswertermittlung

Die Wertung der Ergebnisse im Kontext der Vergleichswertermittlung lässt sich in zwei Aspekte differenzieren: Zum einen die Übertragbarkeit der Auswertestrategie, zum anderen die Eignung der Bayesschen Modelle für die Schätzung von Vergleichswerten in Bezug auf ihre Stärken und Schwächen.

Für die Vergleichswertermittlung nach den gesetzlichen Grundlagen und insbesondere den Vorgaben der VW-RL ist es erforderlich, eine objektive und in den Ergebnissen reproduzierbare Ableitung zu gewährleisten. Die entwickelte *Auswertestrategie* ist damit eine notwendige Voraussetzung für eine marktgerechte Analyse von Daten des Grundstücksmarktes und grundlegender Baustein in Auswertungen der Bayes-Strategie. Dabei handelt es sich um lediglich eine mögliche Realisierung für eine Strategie in der Vergleichswertermittlung; auf abweichende Möglichkeiten der Formulierung wird hingewiesen. Für die praktischen Untersuchungen der Arbeit bildet die Auswertestrategie damit den Rahmen für eine nach den fachlichen Standards der VW-RL als geeignet zu beurteilende Vergleichswertermittlung. Eine direkte Übertragungsmöglichkeit auf die Auswertung weiterer sachlicher und räumlicher Teilmärkte ist ohne die Notwendigkeit von Modifikationen zu erwarten. Dabei ist zu beachten, dass Bayessche Regression und Bayessche Kollokation nicht zwingend als gemeinsame Auswertung durchgeführt werden müssen: Die Bayessche Kollokation ist als Erweiterung der funktionalen Modellierung in der Regression aufzufassen. Die formulierte Auswertestrategie kann insofern auch für eine alleinige Auswertung mittels Bayesscher Regression verwendet werden. In den Abbildungen 4.4 (S. 92) und 4.8 (S. 104) sind in diesem Fall lediglich die Komponenten der Bayesschen Kollokation zu streichen. Eine ausschließliche Anwendung der Bayesschen Kollokation ist hingegen nicht möglich, da diese auf den Residuen der Bayesschen Regression als Eingangsgrößen beruht. Die Auswertestrategie erweist sich damit als flexibel anwendbar, wenn die Berechnung eines Kollokationsansatzes als zu aufwendig oder für eine konkrete Aufgabenstellung nicht zielführend beurteilt wird.

Werden geringfügige Modifikationen vorgenommen, ist die Auswertestrategie unabhängig von der angewendeten Methodik: Während in der vorliegenden Arbeit explizit konjugierte Modelle im Vordergrund stehen, lassen sich basierend auf dem Grundprinzip der Auswerteschritte auch abweichende Modellformulierungen vornehmen. Als Beispiel dienen die nichtkonjugierten Modelle, die durch das Hinzufügen des Auswerteschritt der Monte-Carlo-Integration nach Abschnitt 3.4 (S. 73 ff.) ebenfalls für die Vergleichswertermittlung herangezogen werden können. Auf diese Weise lassen sich u. a. Lösungen der Parameterschätzung im Fall von abweichenden Verteilungsannahmen herleiten. Ein Mehrwert für die Vergleichswertermittlung ist aufgrund der flexiblen Anpassung an die Verteilung von Stichprobendaten zu erwarten.

Die *Eignung der Bayesschen Modelle* für die Schätzung von Vergleichswerten ist im Zusammenhang mit den thematischen Schwerpunkten der Arbeit zu sehen. Diese wurden in den Ausführungen des Abschnitts 1.1 (S. 15 ff.) zum Stand der Forschung mit der Aufgabenstellung der funktionalen Modellierung und der Aufgabenstellung der

Unsicherheitsmodellierung definiert. Auf Basis praktischer Untersuchungen wurden für beide Aspekte Paarvergleiche der formulierten Bayesschen Modelle durchgeführt, die in den vorherigen Abschnitten beschrieben wurden. Werden die Ergebnisse der Paarvergleiche für eine Beurteilung des Gesamtbeitrags zur Vergleichswertermittlung zusammengefasst, ergeben sich die in Abbildung 6.18 dargestellten Rangfolgen der einzelnen Modelle. Die Rangfolgen ergeben sich aus der Summe der Häufigkeiten, mit denen ein Modell in den Paarvergleichen der funktionalen Modellierung oder der Unsicherheitsmodellierung die Rangstufe 1 belegt hat. Auf diese Weise wird eine Aussage ermöglicht, welches der vier eingeführten Modelle sich in den meisten Fällen für eine praktische Anwendung eignet. Gleichwohl wird keine weitere Abstufung hinsichtlich der Häufigkeiten zweiter, dritter oder vierter Ränge vorgenommen, sodass ausschließlich ein grundsätzlicher Trend zu erkennen ist.

Modell	R.1	R.2	K.1	K.2
Summe	3	1	5	2
Rang	2.	4.	1.	3.

Abbildung 6.18: Zusammenfassung der Ergebnisse der Paarvergleiche

Die Abbildung zeigt, dass das Modell der informativen Bayesschen Kollokation am häufigsten den ersten Rang belegt, gefolgt von der informativen Bayesschen Regression. Auf Rang 3 folgt die nichtinformativ Bayessche Kollokation, abschließend auf dem vierten Rang die nichtinformativ Bayessche Regression. Aus der Rangfolge lässt sich erkennen, dass durch informative Bayessche Modelle in der vorgenommenen Vergleichswertermittlung ein Mehrwert gegeben ist. Dabei ist zu beachten, dass die Ergebnisse lediglich auf der Auswertung eines einzelnen räumlichen und sachlichen Teilmarktes und unterschiedliche Wertungen in den Paarvergleichen teilweise nur auf geringen Differenzen in den Ergebnissen beruhen.

Wird zunächst die Aufgabenstellung der funktionalen Modellierung betrachtet, erweisen sich die vier Modelle gleichermaßen als geeignet für die Vergleichswertermittlung. Bedingt durch die Auswertestrategie kommen die geschätzten Vergleichswerte den normativen Anforderungen nach und bilden das Marktgeschehen – unter Verweis auf die natürlichen Differenzen zwischen Kaufpreis und Wert – damit zutreffend ab. Insbesondere in der Prädiktion von Vergleichswerten als wichtigste Anwendung sind nur geringfügige Unterschiede in den Schätzwerten zu erkennen. Es ist zu erwarten, dass sich diese Aussage auch bei der Übertragung auf die Analyse weiterer sachlicher und räumlicher Teilmärkte treffen lässt. Der generelle Mehrwert für die Vergleichswertermittlung wird in den informativen Modellen durch eine höhere Anzahl von Stichprobendaten gegeben, die in die Auswertungen einfließen. In Form des Priori-Wissens stehen so mehr Informationen zur Verfügung, als in den nichtinformativen Modellen berücksichtigt werden können. Aufgrund dieser Tatsache sind die Auswirkung in der Verringerung von Unsicherheiten zu erwarten, nicht in der funktionalen Modellierung. Vorteile in der funktionalen Modellierung werden sich im Umkehrschluss dann ergeben, wenn lediglich wenige Kauffallinformationen zur Verfügung stehen. Durch die Integration von Priori-Information werden in diesen Fällen die Schätzungen der Modellparameter gestützt.

Der wichtigste Beitrag der Forschungsergebnisse zur Vergleichswertermittlung ist in den Unsicherheiten der Schätzwerte zu sehen. Wie in Abschnitt 1.1 (S. 15 ff.) erläutert, ist für die Angabe von Unsicherheiten in der Praxis der Verkehrswertermittlung ein Defizit zu erkennen. Durch die Entwicklung der (informativen) Bayesschen Modelle wird ein Ansatz gegeben, der durch die Angabe der Plausibilität von Schätzungsergebnissen einen signifikanten Mehrwert leistet. Sowohl für die informative Bayessche Regression als auch Bayessche Kollokation kann so der Forderung der VW-RL nach Angabe und Veröffentlichung von Unsicherheiten nachgekommen werden, die unter dem nicht näher definierten Begriff der „Aussagen zur Qualität“ subsummiert werden (vgl. VW-RL, Anlage 4). Durch die Ableitung von Unsicherheitsspannen nicht nur für die Vergleichswerte, sondern auch für die zugrunde liegenden Modellparameter wird die Forderung erfüllt. Die Bayes-Strategie gibt dabei eine in der Vergleichs- und Verkehrswertermittlung innovative Sichtweise durch die Interpretation der HPDIs als Grad der Plausibilität für Schätzwerte. Der festgestellte Mehrwert wird auch in den Rangfolgen nach 6.18 deutlich: Diese bildet sich insbesondere durch die Paarvergleiche in Bezug auf die Unsicherheitsmodellierung, wie ein Vergleich mit den Einzelergebnissen nach den Abschnitten 6.1 und 6.2 zeigt. Die Stärken der informativen Modelle gegenüber den nichtinformativen liegen damit in der Fortpflanzung von Unsicherheiten: Durch die Rekursion werden kurz- und mittelfristige Schwankungen am

Immobilienmarkt ausgeglichen. Nicht relevante Einflüsse werden so gedämpft. Gleichzeitig erweist sich diese Eigenschaft als Schwäche, wenn sich Änderungen am Immobilienmarkt ergeben, die zu langfristigen Änderungen führen. Durch die Vorinformation werden lediglich langsame Anpassungsprozesse in den Modellschätzungen ermöglicht. Reale Strukturbrüche in der Modellformulierung können daher nur verzögert aufgedeckt werden.

Gemäß Abbildung 6.18 ist die informative Bayessche Kollokation in Bezug auf die gewählten Vergleichskriterien vor den übrigen Modellen zu präferieren. Die Stärken der Unsicherheitsmodellierung in den Modellparametern verbinden sich hier mit plausiblen Unsicherheitsspannen für die Prädiktion von Vergleichswerten. Gleichwohl hat sich herausgestellt, dass eine hohe Abhängigkeit von der Signalkomponente besteht. Deren Ableitung ist bislang nicht optimal gelöst, sodass bei der Lösung des Problems eine weitere Steigerung der Prädiktionsgüte zu erwarten ist. Weiterhin erfordert die Anwendung der Bayesschen Kollokation ein im Vergleich zur Bayesschen Regression größeres Maß an statistischen Vorkenntnissen: Die dem Signal zugrunde liegenden Autokorrelationsfunktionen sind jeweils aufgabenspezifisch zu formulieren und können in abweichenden räumlichen und sachlichen Teilmärkten von den in der Arbeit gezeigten Ansätzen abweichen. Nur unter sachverständiger und sorgfältiger Ableitung sind Verbesserungen in der Schätzung von Vergleichswerten zu erwarten. Die praktische Relevanz der Bayesschen Regression ist durch die Bedeutung der klassischen linearen Regressionsanalyse in der derzeitigen Vergleichswertermittlung höher einzustufen. Trotzdem verbleibt als wesentliche Schwäche, dass die Unsicherheiten in den Prädiktionswerten deutlich höher ausfallen als bei der Bayesschen Kollokation.

Zusammenfassend ist die praktische Anwendung der informativen Bayesschen Modelle für die Ableitung von Vergleichsfaktoren sorgfältig abzuwägen. Als geeignet erweisen sich die Modelle dann, wenn keine Steigerung der funktionalen Approximationsgüte, sondern die Ableitung von Vergleichswerten mit geringerer Unsicherheit im Vordergrund steht. Von besonderem Interesse sind die Modelle zudem in Fällen, in denen wenige Kauffalldaten für die Ableitung von Vergleichswerten zur Verfügung stehen, da durch die Rekursion eine vergleichsweise stabile Schätzung – sowohl hinsichtlich Schwankungen am Immobilienmarkt als auch in der numerischen Lösung der Algorithmen – erreicht wird.

Bislang wurden die Ergebnisse ausschließlich auf das Vergleichswertverfahren bezogen. Anhand des gezeigten Mehrwerts für die Angabe von Unsicherheitsmaßen stellt sich die Frage, welchen Beitrag die Forschungsergebnisse zusätzlich auf weitere Aufgabenstellungen der Verkehrswertermittlung zu leisten vermögen. Im folgenden Abschnitt werden die Übertragungsmöglichkeiten behandelt.

6.3.2. Wertung der Ergebnisse im Kontext der Verkehrswertermittlung

Die praktische Bedeutung der Forschungsergebnisse bemisst sich an den Aufgaben der Verkehrswertermittlung, die sowohl durch Sachverständige als auch die nach § 193 BauGB institutionalisierten Gutachterausschüsse wahrgenommen werden. Dabei kommt insbesondere der Arbeit der Gutachterausschüsse eine hohe Bedeutung zu, da als wesentliche Kernaufgabe die Analyse des Grundstücksmarktes mit dem Ziel der Markttransparenz identifiziert werden kann. Zu diesem Zweck werden durch das BauGB Aufgaben vorgegeben, die die Gutachterausschüsse zu erfüllen haben. Im Einzelnen sind dies nach § 193 BauGB vornehmlich die

1. Erstellung von Verkehrswertgutachten,
2. Ermittlung von Bodenrichtwerten sowie
3. Ableitung sonstiger zur Wertermittlung erforderlicher Daten.

Da die Ergebnisse sowohl für die amtliche als auch die gewerbliche und private Verkehrswertermittlung genutzt werden, wird die hohe Bedeutung der Aufgabenerledigung schnell deutlich. Die Übertragung einer erweiterten Unsicherheitsmodellierung bietet sich daher insbesondere für die statistischen Analysen der Gutachterausschüsse an. Von den Ergebnissen profitieren in der Folge sowohl private Sachverständige als auch gewerbliche Institutionen, die sich mit der Verkehrswertermittlung oder einer Wertermittlung, basierend auf den wesentlichen Daten, befassen. Die Übertragbarkeit der Forschungsergebnisse auf praktische Anwendungsmöglichkeiten soll daher anhand der genannten Aufgaben der Gutachterausschüsse gezeigt werden. Abschließend erfolgt ein Überblick über sonstige Anwendungsmöglichkeiten.

6.3.2.1. Erstellung von Verkehrswertgutachten

Für Wertermittlungsfälle, in denen sich der Verkehrswert im Vergleichswertverfahren bewerten lässt, ergibt sich der Mehrwert direkt anhand der Forschungsergebnisse: Die Zielgröße der praktischen Untersuchungen in Kapitel 5 wurde als Kaufpreis zu Wohnfläche formuliert und entspricht damit per Definition einem Vergleichs- bzw. Gebäudefaktor. Für die Nutzung in einem Verkehrswertgutachten ist für die Ableitung eines Vergleichswertes lediglich die Bezugsgröße (hier: Wohnfläche) mit dem Gebäudefaktor zu multiplizieren. Daher lässt sich die dargestellte Auswertung direkt auf die Ermittlung von Verkehrswerten im Vergleichswertverfahren nach § 15 ImmoWertV übertragen und so bei der Erstellung eines Gutachtens über ein Einzelobjekt nutzen.

Der Vorteil der informativen Bayesschen Modelle liegt darin, durch die Unsicherheitsspannen des Prädiktionswertes erste Anhaltspunkte für die sachverständige Einschätzung des Vergleichswertes geben zu können. Bis zu diesem Punkt handelt es sich beim Vergleichswert ausschließlich um einen Funktionswert; der Marktbezug wird erst im Anschluss beurteilt. Wie die Untersuchungen gezeigt haben, liefern die informativen Modelle Unsicherheitsspannen der Prädiktion, die in den überwiegenden Fällen die Kaufpreise mit der gewählten Wahrscheinlichkeit beinhalten und im Gegensatz zu den Ergebnissen der nichtinformativen Modelle geringer sind. Gutachter können das zu bewertende Objekt so bereits grob einschätzen und erkennen, in welchen Spannen sich der Verkehrswert bewegt. Als zusätzliche Einschätzung dienen die Unsicherheiten der Modellparameter, die angeben, wie einzelne Einflussgrößen auf das Ergebnis wirken und wie sicher sie bestimmt sind.

Wie durch die VW-RL gefordert, erfolgt auf Basis des Vergleichswertes die Einschätzung der Marktkonformität und die Berücksichtigung besonderer objektspezifischer Grundstücksmerkmale (vgl. VW-RL, Punkt 7 (4)), die sich in Form von Zu- und Abschlägen auf den Vergleichswert niederschlagen und so vom Funktionswert zum Verkehrswert führen. Durch Berücksichtigung der Unsicherheitsspannen lassen sich die Zu- und Abschläge sachlich begründen: Es wird deutlich, dass es sich lediglich um einen Vergleichswert handelt und sich in Bezug auf das reale Marktgeschehen Schwankungen ergeben können. Durch den erweiterten Unsicherheitshaushalt lassen sich im Rahmen von Verkehrswertgutachten die resultierenden Verkehrswerte auf Basis der prädizierten Vergleichswerte mit einer zusätzlichen, realitätsnahen Unsicherheitsspanne angeben. Durch die Angabe eines Wahrscheinlichkeitsbereichs für den Verkehrswert kann eine zusätzliche Informationssteigerung für die Markttransparenz erreicht werden, sodass auch nicht in den Grundlagen der Verkehrswertermittlung kundige Personen eine Einschätzung der Plausibilität vornehmen können. Da bisherige Verkehrswertgutachten lediglich einen spitz berechneten Wert angeben, wird an dieser Stelle stets eine hohe Rechengenauigkeit suggeriert. Durch die Angabe eines zusätzlichen Wahrscheinlichkeitsbereiches werden die Unsicherheiten im Bewertungsprozess geeignet abgebildet.

Insgesamt ist anhand der Forschungsergebnisse die Nutzung eines Bayesschen Vergleichswertverfahrens zu empfehlen, welches auf einer rekursiven Anwendung beruht und kontinuierlich durch neue Marktdaten aktualisiert wird. Nicht nur ergeben sich geringere Unsicherheitsspannen als im klassischen Fall, auch wird den kontinuierlichen Entwicklungen am Immobilienmarkt Rechnung getragen. In nichtinformativen und damit nicht rekursiven Modellen können kurzfristige Schwankungen hingegen nicht aufgefangen werden und bilden sich direkt im resultierenden Verkehrswert eines Gutachtens ab. Die Nutzung als Vergleichswertverfahren in Gutachten erfordert für private und gewerbliche Sachverständige die Veröffentlichung der Modellparameter seitens der Gutachterausschüsse, die aufgrund der Kaufpreissammlung exklusiv über die zugrunde liegenden Kauffalldaten verfügen und daher die statistischen Auswertungen durchführen. Zwingende Voraussetzungen für die Anwendung sind statistische Grundkenntnisse für die Interpretation der Daten.

Werden die informativen Bayesschen Modelle im Einzelnen betrachtet, bietet sich für die Erstellung von Verkehrswertgutachten sowohl die Bayessche Regression als auch die Bayessche Kollokation an. Zu beachten ist, dass die Bayessche Kollokation zwar keine unmittelbaren Vorteile in der funktionalen Modellierung der Prädiktionswerte aufweist, jedoch über realitätsnahe Unsicherheitsspannen der prädizierten Zielgröße verfügt. Im Hinblick auf die erforderlichen statistischen Kenntnisse kann sich die Bayessche Regression jedoch als vorteilhafter erweisen, da sie auf der bislang genutzten klassischen linearen Regressionsanalyse beruht. Kollokationsmethoden erfordern ungleich höhere Vorkenntnisse und können aufgrund der objektspezifischen Berechnung einer prädizierten Signalkomponente ausschließlich durch einzelfallbezogene Auswertungen einer Kaufpreissammlung genutzt werden. Die Anwendung ist somit für private und gewerbliche Sachverständige nicht möglich.

6.3.2.2. Ableitung von Bodenrichtwerten

Als weiteres Anwendungsgebiet für die Bayesschen Modelle lässt sich die Ableitung von Bodenrichtwerten nennen. Die Ermittlung der durchschnittlichen Lagewerte obliegt den Gutachterausschüssen und wird regelmäßig in einem Abstand von maximal zwei Jahren durchgeführt. Die Ableitungsmethodik ist grundsätzlich nicht vorgeschrieben, die Nutzung multivariater Analyseverfahren lässt sich damit auch hier übertragen. Bodenrichtwerte können ebenso wie Kaufpreise als Zielgröße in den Bayesschen Modellen formuliert werden. Unter der Voraussetzung, dass zutreffende wertbeeinflussende Merkmale in die Ansätze eingeführt werden, können die prädierten Posteriori-Schätzwerte im Hinblick auf die rekursive Auswertung von Jahr zu Jahr fortgeschrieben werden. Wie die praktischen Untersuchungen gezeigt haben, ist als wesentlicher Vorteil der informativen Bayesschen Modelle eine erkennbare Kontinuität in den Schätzergebnissen zu erreichen. Hinsichtlich der Entwicklung von Bodenrichtwerten, welche i. d. R. einen kontinuierlich steigenden Verlauf aufweisen, erweist sich dies als Vorteil. Sprünge in den Schätzungen, die nicht unmittelbar sachlogisch zu begründen sind, werden vermieden. Im Umkehrschluss bedeutet dies, dass beispielsweise der Einfluss aktueller Kauffälle, die eine größere Abweichung zu vergleichbaren Kauffällen aufweisen, in ihrer Auswirkung auf den Bodenrichtwert gedämpft werden. Ein weiterer wesentlicher Vorteil ist die generelle Angabe von Unsicherheitsspannen. Da Bodenrichtwerte einen für eine fest definierte Zone durchschnittlichen Lagewert mit durchschnittlichen Eigenschaften darstellen, können durch die Angabe von Wahrscheinlichkeitsbereichen Abweichungen in einer konkreten Verkehrswertermittlung plausibilisiert werden. Bislang werden in praktischen Anwendungen durch Gutachter sachverständig Zu- und Abschläge angebracht – ohne dass die generelle Unsicherheit des Bodenrichtwertes als zusätzliche Argumentationsgrundlage verwendet werden könnte. Zudem ist trotz der zonalen Abgrenzungen nicht immer die geforderte Homogenität in den durchschnittlichen Eigenschaften innerhalb einer Bodenrichtwertzone zu erreichen, sodass die Angabe von Wahrscheinlichkeitsbereichen unverzichtbar ist. Offensichtlich wird diese Problematik an Zonengrenzen, für die stark voneinander abweichende Bodenrichtwerte beschlossen sind.

Für die Ableitung von Bodenrichtwerten bieten sich sowohl die Modelle der Bayesschen Regression als auch der Bayesschen Kollokation an. Insbesondere die Anwendung der Bayesschen Kollokation kann sich durch die Ausnutzung räumlicher Informationen in der Signalberechnung als Vorteil erweisen. Da die Bodenrichtwerte ausschließlich von den Gutachterausschüssen abgeleitet werden, liegen die für die Berechnung einer Kollokation notwendigen Voraussetzungen vor, insbesondere die Datenverfügbarkeit für die individuelle Berechnung der Signalkomponente. Zudem sind die Unsicherheitsmaße für die prädierten Werte als realitätsnah einzustufen und könnten daher einen generellen Mehrwert bei der Veröffentlichung von Bodenrichtwerten bieten.

6.3.2.3. Ableitung sonstiger zur Wertermittlung erforderlicher Daten

Neben der Ableitung von Bodenrichtwerten ist nach § 193 Abs. 5 BauGB eine wesentliche Aufgabe der Gutachterausschüsse, die Ableitung sonstiger zur Wertermittlung erforderlicher Daten vorzunehmen. Unter dem Begriff werden im Einzelnen

- Indexreihen,
- Umrechnungskoeffizienten,
- Vergleichsfaktoren,
- Sachwertfaktoren sowie
- Kapitalisierungs- bzw. Liegenschaftszinssätze

subsummiert. Die Ergebnisse werden regelmäßig in den Grundstücksmarktberichten veröffentlicht und den Anwendern für die Verkehrswertermittlung zur Verfügung gestellt.

Indexreihen werden nach § 11 ImmoWertV abgeleitet, um allgemeine Änderungen von Wertverhältnissen auf dem Grundstücksmarkt zu erfassen. Eine Einschränkung in Bezug auf die Objektart ist nicht gegeben; in der Praxis werden bspw. für individuelles Wohnbauland, Grün- und Ackerflächen, Eigentumswohnungen oder Ein- und

Zweifamilienhäuser jeweils eigene Indexreihen berechnet. Neben den sachlichen Teilmärkten ist auch keine Einschränkung der räumlichen Teilmärkte gegeben. In Abhängigkeit der Datenlage lassen sich lokale, regionale oder globale Indexreihen angeben; ein Beispiel für lokale Indexreihen wurde mit den Abbildungen 5.3 (S. 111) gegeben. Die Methodik zur Ableitung ist nicht einheitlich vorgeschrieben, Grundlage sind jedoch stets Kaufpreise, die das reale Marktgeschehen widerspiegeln. Durch eine geeignete Formulierung lassen sich Indexreihen mit Hilfe von multivariaten statistischen Analyseverfahren berechnen. Diese Vorgehensweise wird u. a. durch die niedersächsischen Gutachterausschüsse genutzt, insofern sind die entwickelten Modelle der vorliegenden Arbeit direkt übertragbar. Die Ableitung von Unsicherheitsspannen einzelner Indexwerte tragen zur Markttransparenz bei, indem die Unsicherheiten der Eingangsdaten und der Modellformulierung direkt ablesbar sind. Die Indexwerte werden so nicht als absolute Größen gegeben, sondern bei Veröffentlichung der Unsicherheiten gleichzeitig eine Einschätzung über die Güte der mathematischen Modellierung gegeben. Dies ist insofern relevant, als dass die Datengrundlage bei Indexwerten oftmals sehr heterogen ist. In der Praxis werden häufig größere Selektionsbereiche vorgegeben, um eine ausreichende Datenbasis zu gewährleisten. Deutlich wird dies am Beispiel des sogenannten Häuserpreisindex: Das statistische Bundesamt Destatis leitet bundesweite Häuserpreisindizes ab, wie sie durch die Europäische Union (EU) im Rahmen einer übergeordneten Rechtsverordnung verlangt werden. Da es sich um eine bundesweite Auswertung handelt – die wesentlich auf den Daten der Gutachterausschüsse beruhen – und somit eine hohe Aggregationsstufe notwendig ist, werden entsprechend weitgefaste Selektionsgrenzen für die Stichproben gegeben (vgl. KASTE (2013)). Durch die Datenlage wird eine entsprechende Unsicherheit gegeben, die mittels der informativen Bayesschen Modelle nicht nur angegeben, sondern auch in den quartalsweisen Neuberechnungen rekursiv fortgepflanzt werden kann. Der Häuserpreisindex stellt lediglich ein Beispiel dar, der Grundgedanke lässt sich jedoch auf jede weitere Indexreihe, unabhängig vom räumlichen oder sachlichen Teilmarkt übertragen. Gemeinsam ist jeder Indexreihe die Darstellung von kontinuierlichen Marktänderungen, sodass sich die Nutzung rekursiver Schätzungen anbietet. Die in den praktischen Untersuchungen der Arbeit nachgewiesene Kontinuität in den Schätzungen der Modellparameter und Zielgrößen erweist sich als Vorteil. Ein potentieller Mehrwert kann sich sowohl durch die informative Bayessche Regression als auch die Bayessche Kollokation ergeben. Als Nachteil könnte sich in diesem Zusammenhang erweisen, dass die informativen Bayesschen Modelle eine dämpfende Wirkung auf größere Schwankungen im Marktgeschehen aufweisen. Für plötzliche Brüche in den allgemeinen Wertentwicklungen, bspw. ausgelöst durch die Wirtschafts- und Finanzkrise, erweist sich diese Eigenschaft u. U. als zu unflexibel. Für eine abschließende Wertung sind an dieser Stelle Detailuntersuchungen erforderlich.

Nach § 12 ImmoWertV sind durch die Gutachterausschüsse weiterhin *Umrechnungskoeffizienten* abzuleiten, mit denen Wertunterschiede von Grundstücken erfasst werden sollen, die sich aus Abweichungen bestimmter Grundstücksmerkmale sonst gleichartiger Grundstücke ergeben. Insbesondere sind dies Abweichungen aus dem unterschiedlichen Maß der baulichen Nutzung oder der Grundstücksgröße und -tiefe. Auch in diesem Fall werden in der Praxis multivariate statistische Verfahren genutzt. Der Arbeitskreis der Gutachterausschüsse und Oberen Gutachterausschüsse in der Bundesrepublik Deutschland (AK OGA) hat in diesem Zusammenhang bundesweit anwendbare Umrechnungskoeffizienten abgeleitet, die auf regressionsanalytischen Verfahren basieren; im Speziellen handelt es sich um eine doppelt-logarithmische Regression, deren Modellparameter im Abschlussbericht der Untersuchungen veröffentlicht wurden (LIEBIG & ACHE, 2013). Auch einzelne Gutachterausschüsse nutzen die Möglichkeiten, die die klassische Regression bietet. Eine Übertragbarkeit ist somit auch in dieser Anwendung gegeben. Bei den Umrechnungskoeffizienten handelt es sich um Werte, die i. d. R. einen größeren Bereich umfassen (für Gutachterausschüsse bspw. den lokalen Markt im Amtsbezirk): Entsprechend viele Daten fließen in die Auswertungen ein, die als sehr heterogen zu beurteilen sind. Da sich die Umrechnungskoeffizienten auf ein durchschnittliches Musterobjekt beziehen und Grundstücke gemäß der Ausführungen in Kapitel 2 durch Individualität geprägt sind, lassen sich Unsicherheiten in den Ergebnissen nicht vermeiden. Hier bietet sich in erster Linie die Nutzung der informativen Bayesschen Kollokation an, da in die Auswertungen räumliche Informationen mit einfließen und so ein stärkerer Bezug zur geografischen Lage von Objekten berücksichtigt wird. Die Unsicherheiten in den Prädiktionen fallen geringer aus, was als Vorteil gegenüber der informativen Bayesschen Regression zu werten ist. Da sich jedoch auch die informative Bayessche Regression als Erweiterung der bisherigen Vorgehensweise empfiehlt, sind Vor- und Nachteile anwendungsfallbezogen abzuwägen. Durch die Fortpflanzung der Unsicherheiten im Rekursionsprozess werden

gleichermaßen Vorteile sichtbar, insbesondere im Hinblick auf die Kontinuität in den Schätzungen; kurzfristige Schwankungen und Änderungen sind in Umrechnungskoeffizienten generell nicht zu erwarten.

Die Eignung der Auswertestrategie für die Prädiktion von Vergleichswerten wurde in Kapitel 5 anhand von Gebäudefaktor als Zielgröße gezeigt. Die Anwendbarkeit für eine Ableitung von allgemeinen *Vergleichsfaktoren* nach § 13 ImmoWertV ist damit sofort erkennbar und wurde hinsichtlich der Vorteile für die Unsicherheitsangaben bereits diskutiert: Die kontinuierliche Auswertung von Vergleichsfaktoren im rekursiven Ansatz ermöglicht es, in Veröffentlichungen wie den Grundstücksmarktberichten der Gutachterausschüsse zusätzliche Unsicherheitsspannen zu geben, die für die Ermittlung von Verkehrswerten auf Basis der Vergleichsfaktoren unverzichtbar sind. Diese dienen insbesondere der Angabe von Unsicherheiten des Verkehrswertes und schließen damit den Kreis zur Erstellung von Verkehrswertgutachten, die auf den Analysen der Gutachterausschüsse basieren. Aktuelle Darstellungen in den Grundstücksmarktberichten verzichten derzeit sowohl auf eine Angabe der Unsicherheiten als auch auf die Veröffentlichung der zugrunde liegenden Modellparameter, wie sie nach VW-RL, Anlage 4 gefordert wird. Ein genereller Mehrwert ist somit auch an dieser Stelle zu erkennen.

Gleiches gilt für die Ableitung von *Sachwertfaktoren* nach § 14 ImmoWertV, die sich durch die Formulierung der Zielgröße Kaufpreis zu Sachwert ebenfalls in den Bayesschen Ansätzen modellieren lassen. Die Angabe von Unsicherheitsspannen bietet sich insbesondere an, da dem Sachwertfaktor im Sachwertverfahren die Aufgabe zukommt, den Marktbezug eines theoretischen Modellwertes abzubilden. Unsicherheitsspannen können in der Argumentation für die letztliche Festlegung seitens eines Gutachters eine sinnvolle Erweiterung der bisherigen Vorgehensweise darstellen. Zudem sind in den praktischen Ableitungen die Eingangsdaten als sehr unsicher einzustufen: Die für die Berechnungen benötigten Angaben werden in Kaufverträgen i. d. R. nicht vollständig gegeben. Dies trifft bspw. auf den Ausstattungsstandard oder Baumängel- und Bauschäden zu. Hierbei handelt es sich um Faktoren, die auch durch die obligatorischen Eigentümerbefragungen nicht zufriedenstellend abgebildet werden können. Letztlich werden die Daten durch die auswertende Person auf Basis von Fotos und Fragebögen in der Kaufpreissammlung erfasst und beruhen daher auf sachverständigem aber auch subjektiv geprägten Einschätzungen. Die Unsicherheit in den Eingangsdaten einer statistischen Analyse wird deutlich; Abweichungen zu den realen Verhältnissen sind kaum zu vermeiden und spiegeln sich dann in nicht erklärbaren Differenzen von Kaufpreisen für Objekte mit theoretisch vergleichbaren wertbeeinflussenden Merkmalen wieder. Die Unsicherheit im Sachwertfaktor wird in der Praxis derzeit berücksichtigt, indem durch sachverständiges Ermessen Zu- und Abschläge an einem für ein konkretes Objekt nach statistischen Gesichtspunkten abgeleiteten Sachwertfaktor angebracht werden. Die Zu- und Abschläge beruhen ausschließlich auf gewonnen Erfahrungen und Kenntnissen der Sachverständigen über den betrachteten Teilmarkt. Durch die in dieser Arbeit entwickelten Ansätze lassen sich Sachwertfaktoren ableiten, die mit einer zusätzlichen Angabe für deren Plausibilität veröffentlicht und genutzt werden können. Dem sachverständigen Ermessen wird damit eine zusätzliche, transparente Argumentationsgrundlage für Abweichungen vom Funktionswert gegeben. Angaben in den eingangs thematisierten Verkehrswertgutachten sorgen für eine zusätzliche Transparenz und Nachvollziehbarkeit. Da die Sachwertfaktoren ausschließlich durch die Gutachterausschüsse abgeleitet werden, bieten sich als Hilfsmittel sowohl die informative Bayessche Regression und Bayessche Kollokation an und würden damit eine signifikante Erweiterung der bisherigen Vorgehensweise darstellen.

Werden abschließend für die sonstigen zur Wertermittlung erforderlichen Daten die *Liegenschaftszinssätze* nach § 14 ImmoWertV betrachtet, ergibt sich ein Bedarf an der Ermittlung von Konfidenzregionen. Die Datenlage für die Ableitung von Liegenschaftszinssätzen ist als sehr heterogen zu beurteilen. Insbesondere für spezielle Immobilienarten wie Bürogebäude ergibt sich u. U. in einem lokalen Bereich kein ausreichender Datensatz, um gesicherte Liegenschaftszinssätze ableiten zu können. Die Angabe von Unsicherheitsspannen stellt sich somit als unverzichtbar dar, jedoch werden diese nicht nach den in Kapitel 3 gegebenen Modellansätzen abgebildet. Liegenschaftszinssätze werden durch eine rekursive Anwendung des Ertragswertverfahrens berechnet, sodass sich eine abweichende Modellbildung ergibt. Die in dieser Arbeit thematisierten Modellansätze sind auf diese Anwendung bislang nicht ausgelegt und müssten neu formuliert werden. Generell ist die Anwendung Bayesscher Ansätze auch hier denkbar und würde hinsichtlich der Unsicherheitsmodellierung einen Mehrwert bei der Veröffentlichung der Zinssätze mit sich bringen. Es ist jedoch zu erwarten, dass sich aufgrund der Ertragswertformel nichtlineare Zusammenhänge ergeben, die nicht mehr auf Basis konjugierter Dichten gelöst werden können. Eine analytische Lösung wäre in diesem

Zuge nicht anwendbar, das Modell müsste mit Hilfe einer in Abschnitt 3.4 vorgestellten Monte-Carlo-Simulation gelöst werden. Die Übertragung der Auswertestrategie auf die Abweichung von analytisch lösbaren Modellen muss an dieser Stelle künftigen Forschungsvorhaben überlassen werden.

Als abschließende Wertung lässt sich festhalten, dass der Beitrag rekursiver Bayesscher Schätzverfahren zur Vergleichswertermittlung theoretisch auf unterschiedliche Aufgabenstellungen der Verkehrswertermittlung übertragbar ist. Marktanalysen beruhen heute in den meisten Fällen auf der Anwendung statistischer Methoden, weisen jedoch bislang einen Mangel in der Unsicherheitsmodellierung auf. Angaben zu Spannen der Ermittlungsergebnisse werden weder in Verkehrswertgutachten noch in den Grundstücksmarktdaten selbst gegeben. In sämtlichen Aufgabenbereichen der Gutachterausschüsse, in denen als Hilfsmittel multivariate statistische Analyseverfahren genutzt werden, ist daher ein Mehrwert durch die erweiterte Modellierung in Bayesschen Ansätzen zu erwarten. Darüber hinaus sind Anwendungen auch dann denkbar, wenn es sich nicht um gesetzlich vorgeschriebene Aufgaben handelt. Als Beispiele dienen der IPK Niedersachsen oder die Ermittlung sanierungsbedingter Bodenwerterhöhungen im Modell Niedersachsen. Insbesondere werden Mehrwerte anhand der kontinuierlichen Entwicklungen am Immobilienmarkt deutlich. Aktuelle Auswerteprinzipien beruhen meist darauf, einzelne Epochen auszuwerten und so die Grundstücksmarktdaten fortzuschreiben. Einflüsse zurückliegender Auswertungen werden auf diese Weise nicht integriert, jede Auswerteepoche ist als Einzelauswertung zu interpretieren. Durch die dynamischen Entwicklungen des Immobilienmarktes wird jedoch deutlich, dass eine rekursive Auswertung – unabhängig von der Wahl des Zeitraums einer Auswerteepoche – durch die Fortpflanzung von Schätzungsergebnissen und deren Unsicherheiten auf nachfolgende Epochen ein zutreffenderes Bild des Immobilienmarktes geben kann. Die Forschungsergebnisse sind daher als Beitrag aufzufassen, eine kontinuierliche Auswertung von Marktdaten zu ermöglichen und zugleich durch adäquate Unsicherheitsbehandlung und -fortpflanzung einen Mehrwert für die Markttransparenz zu leisten.

7. Fazit und Ausblick

Die Verkehrswertermittlung als wesentlicher Bestandteil der Analyse des Immobilienmarktes ist von jeher mit hohen Anforderungen verbunden, die sich insbesondere in normativen Vorgaben zur klar abgegrenzten Definition des Verkehrswertes sowie den Methoden zu seiner Bestimmung widerspiegeln. Grundvoraussetzung ist die Objektivität in der Wertfindung, die in dem angewendeten Verfahren erkennbar sein muss. Erst durch den Ausschluss jeglicher subjektiver Einflüsse und spekulativer Momente kann ein zutreffendes Abbild des betrachteten Marktes erreicht werden. Mit der Voraussetzung der Objektivität geht gleichermaßen die Forderung nach einer jederzeitigen Reproduzierbarkeit der Wertermittlungsergebnisse einher; eine enge Wechselwirkung ist offensichtlich. In diesem Kontext konnte als Kern der Arbeit die Bayes-Theorie erfolgreich genutzt werden, um eine Neuformulierung der indirekten Vergleichswertermittlung vorzunehmen. Für die Basis der Modellierung als indirekte Vergleichswertverfahren wurden die bewährten Ansätze der multiplen linearen Regressionsanalyse und der einfachen Kollokation gewählt. Diese ließen sich unter den in der Arbeit getroffenen Annahmen für die Wahrscheinlichkeitsverteilungen der Parameter und Beobachtungen als lineare konjugierte Modelle formulieren. Als wesentlicher Vorteil hat sich unter diesen Umständen die einfach handhabbare analytische Lösung der resultierenden Gleichungen ergeben, die sich leicht in einen rekursiven Auswerteprozess einbinden lässt.

Neben der Neuformulierung der Modelle als Bayessche indirekte Vergleichswertverfahren ist aufgrund der speziellen Anforderungen der Verkehrswertermittlung grundsätzlich eine sorgfältige Vor- und Aufbereitung der Eingangsdaten notwendig. Zu diesem Zweck wurde eine Auswertestrategie entwickelt, deren wesentliche Schritte in der Modelldefinition, der Aufbereitung einer Datenstichprobe eines betrachteten Teilmarkts sowie der eigentlichen Auswertung in den Bayesschen Modellen bestehen. Die Einzelschritte bauen aufeinander auf und sind teilweise als iterativer Prozess zu durchlaufen. Wie anhand der praktischen Ergebnisse gezeigt werden konnte, ermöglicht die Auswertestrategie die Analyse des Teilmarkts in Bezug auf die Schätzung von unbekannten Parametern, ausgeglichenen Schätzgrößen sowie insbesondere der Angabe von zugehörigen Unsicherheitsmaßen. Darüber hinaus ist die Auswertestrategie rekursiv anwendbar, sodass durch die fortlaufende Integration von empirischem Vorwissen die Fortpflanzung von Unsicherheiten über mehrere Auswertepochen ermöglicht wurde und damit eine innovative Vorgehensweise in einer kontinuierlichen Abbildung des Immobilienmarktes darstellt.

Wird die reine Ebene der Modellbildung betrachtet, hat die Integration von Priori-Informationen in den Auswerteprozess gezeigt, dass sich insbesondere in der Unsicherheitsmodellierung durch die rekursiven Bayesschen Schätzungen ein Mehrwert für die Verkehrswertermittlung ergibt. In der funktionalen Modellierung konnten für das gewählte Fallbeispiel keine größeren Unterschiede festgestellt werden, die nennenswert für Anwendungen in der Verkehrswertermittlung wären; dies hat dazu geführt, dass das erste Forschungsziel nur bedingt erreicht werden konnte. Die Stärken einer Bayesschen Modellierung mit der Integration von Vorinformationen sind jedoch klar auf eine Steigerung in der Unsicherheitsmodellierung ausgelegt, was durch die Erreichung des zweiten Forschungsziels belegt werden konnte. In Fällen, in denen die Stichprobe einen vergleichsweise großen Datenumfang enthält – wie es auch im Rahmen der Arbeit der Fall ist –, liefern informative und nichtinformative Bayessche Ansätze Modelle mit annähernd gleicher Approximationsgüte. Reduziert sich jedoch der Datensatz der Stichprobe erheblich, werden die eigentlichen Vorteile der Bayesschen Modelle deutlich: Die nichtinformativen Modelle können dann u. U. nicht mehr gelöst werden, wenn die Mindestanforderungen an den Stichprobenumfang nicht mehr erfüllt werden. Liegen Vorinformationen vor, werden die informativen Modelle hingegen durch die zusätzlichen Daten gestützt und lassen sich weiterhin lösen. Die Stabilität der Schätzergebnisse der Modellparameter, die in der vorliegenden Arbeit nachgewiesen werden konnte, erweist sich hier als Vorteil. Relevant für Anwendungen in der Verkehrswertermittlung ist diese Tatsache insbesondere für Teilmärkte, in denen kaum Kauffälle und damit Stichprobendaten vorliegen und sich ggf. auch als sehr heterogen erweisen. Die Auswertungen in diesen sogenannten kaufpreisarmen Lagen sind in nichtinformativen Modellen – und damit auch den klassisch-frequentistischen Ansätzen – nicht möglich. Die Integration von (zutreffender) Priori-Information ermöglicht auch in diesen Fällen eine Modellschätzung.

Zusammenfassend erweitert die vorliegende Arbeit zum Themenkomplex multivariater statistischer Verfahren in der Verkehrswertermittlung die bisherigen Ansätze um die folgenden Aspekte:

- a) Gewinnung und Integration von *Vorinformation* aus empirischen Daten
 - Methodik zur Gewinnung von Vorinformation aus empirischen Daten
 - Integration von Vorinformation in den Bewertungsprozess mit dem Ziel der Verbesserung der Unsicherheit
- b) Untersuchung der *funktionalen Modellierung*
 - Adaption der multiplen linearen Regression und der einfachen Kollokation als indirekte Vergleichswertverfahren im Bayesischen Ansatz
 - Verbesserung der mittleren quadratischen Abweichung in der Schätzung von Vergleichswerten
- c) Reduzierung der *Unsicherheiten* basierend auf zufälliger Variabilität
 - Reduzierung der Unsicherheiten der unbekannten Modellparameter
 - Reduzierung der Unsicherheiten in den Schätzergebnissen der Vergleichswerte
- d) Entwicklung einer *Auswertestrategie*
 - Formulierung eines rekursiv anwendbaren Auswerteschemas
 - Entwicklung von Ausreißersuche und Hypothesentests basierend auf der Bayes-Theorie

Im Rahmen der hier vorgestellten Forschungsergebnisse kann lediglich ein Teilbereich des Anwendungsgebietes betrachtet werden, den die Integration der Bayes-Strategie in die Verkehrswertermittlung eröffnet. Künftige Ansätze für offene Forschungsfragen beziehen sich zum einen auf die Weiterentwicklung der hier vorgestellten Ansätze und die Auswertestrategie, zum andern auf die Verknüpfung mit weiteren zu untersuchenden Themenkomplexen. Die wesentlichen Aspekte werden im Folgenden kurz beschrieben:

- *Übertragung der rekursiven Auswertestrategie*
Die in dieser Arbeit dargestellten Forschungsergebnisse beruhen auf der Untersuchung eines eingeschränkten räumlichen und sachlichen Teilmarkts. Teilaspekte wurden in diversen Veröffentlichungen bereits auf abweichende Teilmärkte übertragen (vgl. ZADDACH & ALKHATIB (2012a), ZADDACH & ALKHATIB (2013a), ZADDACH & ALKHATIB (2013b), ZADDACH & ALKHATIB (2014a), ZADDACH & ALKHATIB (2014b)); die Anwendung der vollständigen rekursiven Auswertestrategie ist jedoch bislang nicht erfolgt. Für die Auswertung weiterer Stichproben ist demnach zu prüfen, ob sich die in der vorliegenden Arbeit entwickelte Grundlagenmethodik mit ähnlichen Ergebnissen auf unterschiedliche Datensätze übertragen lässt.
- *Übertragung auf weitere Wertermittlungsverfahren*
Bislang wurde ausschließlich das Vergleichswertverfahren im Bayesschen Ansatz modelliert. Damit stellt sich die Frage nach der Übertragbarkeit der Bayes-Strategie auf das Sach- und Ertragswertverfahren (und die nicht normierten Verfahren) im Hinblick auf eine generelle Anwendungsmöglichkeit. Entsprechend ist eine modifizierte Auswertestrategie zu entwickeln, die mit den zu erwartenden nicht analytisch lösbaren Modellen auf Basis von Monte-Carlo-Simulationen zu formulieren wäre.
- *Modelloptimierung der Bayesschen Kollokation*
Hinsichtlich der Bayesschen Kollokation verbleiben offene Forschungsansätze, die zu Gunsten einer grundlegenden Methodenentwicklung im Rahmen der Arbeit nicht umfassend behandelt werden konnten. Wie bereits erwähnt und im Laufe der Abhandlungen thematisiert, zeigen sich die ermittelten Wertabstände als nicht optimal für die Ermittlung von Korrelationen in der Berechnung der Signalkomponenten. Inwiefern sich diese Problematik rein numerisch aus den Stichprobendaten ergibt oder ob sich ein abweichendes zu definierendes Ähnlichkeitsmaß für die Vergleichbarkeit von Immobilien als günstiger für die Berechnung von Autokorrelationsfunktionen erweist, ist zu prüfen. Weiterhin werden im Kollokationsansatz bislang keine Gewichte für das

Verhältnis Signal- zu Rauschanteil berücksichtigt. Eine signifikante Modellerweiterung ergibt sich daher durch die Einführung Bayesscher Varianzkomponenten, die sich u. U. auch hoch korreliert mit der Problematik der Wertabstände zeigt. Inwiefern sich eine Berücksichtigung auf die Prädiktionsgüte auswirkt, ist durch einen erweiterten Modellansatz zu prüfen.

- *Modellerweiterung auf die Behandlung impräziser Daten*

Die Untersuchungen im Kontext der Arbeit befassen sich ausschließlich mit der zufälligen Variabilität der Eingangsdaten. Wie in ZADDACH (2014c) gezeigt, lässt sich für die Behandlung impräziser Einflüsse die Fuzzy-Theorie in ein kombiniertes Fuzzy-Bayes-Modell als indirektes Vergleichswertverfahren integrieren. Die umfassende Formulierung unter dem Aspekt der Auswertestrategie und der Mehrwert für praktische Anwendung ist eine bislang offene Forschungsfrage.

- *Nachweis eines Mehrwerts für kaufpreisarme Lagen*

Der Mehrwert in der Unsicherheitsmodellierung konnte im Zuge der Arbeit nachgewiesen werden. Der Mehrwert der funktionalen Modellierung in informativen Bayesschen Modellen kann für die Auswertung kaufpreis- armer Lagen erwartet werden. Erste Studien zu diesem Thema bestätigen die generelle Annahme (vgl. WEITKAMP & ALKHATIB (2014)). Dabei basieren die Auswertungen vornehmlich auf Simulationen von kaufpreis- armen Lagen, die durch systematische Reduzierung von Stichprobendaten erreicht werden. Lösungsansätze werden durch robuste Bayessche Modelle gegeben. Die Verknüpfung der entwickelten rekursiven Auswerte- strategie sowie der aufgezeigten Methodik für die Generierung von empirischen Priori-Informationen mit der Untersuchung von kaufpreisarmen Lagen ist abschließend als weitere Forschungsaufgabe zu nennen.

Literatur

- AITKIN, M. (1991): *Posterior Bayes factor (with discussion)*. In: Journal of the Royal Statistical Society, , B53: 111–142.
- ALEFELD, G. & HERZBERGER, J. (1983): *Introduction to Interval Computations*. Academic Press, New York.
- ALKHATIB, H. (2008): *On Monte Carlo Methods with Applications to the Current Satellite Gravity Missions*. Dissertation, Landwirtschaftliche Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn.
- ALKHATIB, H. & WEITKAMP, A. (2012): *Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 1*. In: zvf - Zeitschrift für Geodäsie, Geoinformation und Landmanagement, **137**, 2: 93–102.
- ALKHATIB, H. & WEITKAMP, A. (2013): *Robust Bayesian regression approach for areas with small numbers of purchases*. In: Proceedings of the RICS Cobra Research Conference, New Delhi, India.
- ALKHATIB, H., WEITKAMP, A., ZADDACH, S. & NEUMANN, I. (2015): *Realistic Uncertainty Estimation of the Market Value Based on a Fuzzy-Bayesian Sales Comparison Approach*. In: Proceedings of the FIG Working Week 2015, Sofia, Bulgaria, 17 - 21 May 2015.
- ANDO, T. (2010): *Bayesian Model Selection and Statistical Modeling*. Chapman & Hall/CRC and Taylor & Francis Group, Boca Raton and FL.
- BANDEMER, H. (1997): *Ratschläge zum mathematischen Umgang mit Ungewissheit: Reasonable Computing*. B. G. Teubner Verlagsgesellschaft Leipzig, Stuttgart and Leibzig.
- BANDEMER, H. (2005): *Mathematik und Ungewißheit: Drei Essays zu Problemen der Anwendung*. Edition am Gutenbergplatz, Leibzig.
- BANDEMER, H.-W. (2006): *Mathematics of Uncertainty: Ideas, Methods, Application Problems*. Springer-Verlag, Berlin Heidelberg New York.
- BARNETT, V. & LEWIS, T. (1984): *Outliers in Statistical Data*. 2. Aufl. Wiley, Chichester.
- BAUCH, H., JAHN, K.-U., OELSCHLÄGEL, D., SÜSSE, H. & WIEBIGKE, V. (1987): *Intervallmathematik - Theorie und Anwendungen*. B. G. Teubner Verlagsgesellschaft Leipzig, Leipzig.
- BECKMAN, R. J. & COOK, R. D. (1983): *Outlier.....s*. In: Technometrics, , 25: 119–163.
- BERGER, J. O. & PERICCHI, L. R. (1996): *The intrinsic Bayes factor for linear models*. In: BERNARDO, J. M., BERGER, J. O., DAWID, A. P. & SMITH, A. F. M. [Hrsg.]: *Bayesian Statistics 5*, 25–44. Oxford University Press, Oxford.
- BIEWER, B. (1997): *Fuzzy-Methoden: Praxisrelevante Rechenmodelle und Fuzzy-Programmiersprachen*. Springer Verlag, Berlin Heidelberg New York.
- BISCHOFF, B. (2007): *Grundstückswertermittlung*. In: SCHMOLL GENANNT EISENWERTH, FRITZ [Hrsg.]: *Basiswissen Immobilienwirtschaft*, 443–622. Grundeigentum-Verlag, Berlin-Schöneberg.
- BORTZ, J. (2005): *Statistik für Human- und Sozialwissenschaftler*. Sechste, vollständig überarbeitete und aktualisierte auflage Aufl. Springer Medizin Verlag, Heidelberg.
- BRAUER, K.-U. (2013): *Grundlagen der Immobilienwirtschaft: Recht - Steuern - Marketing - Finanzierung - Bestandsmanagement - Projektentwicklung*. 8. Aufl. Springer-Gabler, Wiesbaden.
- BROCKHAUS (2005): *Brockhaus - Die Enzyklopädie in 30 Bänden*. 21. Aufl. Online Ausgabe, F. A. Brockhaus, Leipzig and Mannheim.
- BRUSCO, M. J. & STAHL, S. (2005): *Branch-and Bound Applications in Combinatorial Data Analysis*. Statistics and Computing. Springer Science+Business Media, Inc., New York.
- BUNDESMINISTERIUM FÜR UMWELT, NATURSCHUTZ, BAU UND REAKTORSICHERHEIT (2014): *Richtlinie zur Ermittlung des Vergleichswerts und des Bodenwerts: Vergleichswerttrichtlinie - VW-RL*.
- CASPARY, W. & WICHMANN, K. (2007): *Auswertung von Messdaten: Statistische Methoden für Geo- und Ingenieurwissenschaften*. Oldenbourg Wissenschaftsverlag GmbH, München.
- CHALONER, K. & ROLLIN, B. (1988): *A Bayesian Approach to Outlier Detection and Residual Analysis*. In: Biometrika, **75**, 4: 651–659.
- CHATTERJEE, S. & PRICE, B. (1995): *Praxis der Regressionsanalyse*. 2. auflage Aufl. R. Oldenbourg Verlag GmbH, München and Wien.
- CHRISTENSEN, R. (2011): *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. Chapman & Hall/CRC Press, Boca Raton London New York.
- COX, R. T. (1946): *Probability, frequency and reasonable expectation*. In: American Journal of Physics, , 14: 1–13.
- CRESSIE, N. A. C. (1993): *Statistics for Spatial Data: Revised Edition*. John Wiley & Sons, Inc., New York and Chichester and Toronto and Brisbane and Singapore.
- DEGROOT, M. H. (1970): *Optimal Statistical Decisions*. McGraw-Hill, New York.
- DOUCET, A. & SMITH, A. (2001): *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- ENGEL, R. (2008): *Das Problem der Unschärfe in der Wertermittlung: Überlegungen zu marktgerechten Bewertungsansätzen in der Ertragswertberechnung*. In: GuG - Grundstücksmarkt und Grundstückswert, , 05: 269–276.
- FAHRMEIR, L., KNEIB, T. & LANG, S. (2009): *Regression: Modelle, Methoden und Anwendungen*. 2. auflage Aufl. Springer Verlag, Heidelberg and Dordrecht and London and New York.
- FISCHER, R. (2007): *Werttheorie*. In: SANDNER, S. & WEBER, U. [Hrsg.]: *Lexikon der Immobilienwertermittlung*. Bundesanzeiger Verlagsges.mbh, Köln.
- FREEMAN, P. R. (1985): *On the number of outliers in data from a linear model*. In: BERNARDO, J. M., DEGROOT, M. H., LINDLEY, D. V. & SMITH, A. F. M. [Hrsg.]: *Bayesian Statistics*, 349–365. University Press, Valencia.
- FURNIVAL, G. M. & WILSON, R. W. (1974): *Regressions by Leaps and Bounds*. In: Technometrics, **16**, 4: 499–511.
- GATU, C. & KONTOGHIOGHES, E. J. (2006): *Branch-and-Bound Algorithms for Computing the Best-Subset Regression Models*. In: Journal of Computational and Graphical Statistics, **15**, 1: 139–156.
- GEISSER, S. (1975): *The predicitive sample reuse method with applications*. In: Journal of the American Statistical Association, , 70: 320–328.

- GELMAN, A., CARLIN, J. B., STERN, H. S. & RUBIN, D. B. (2004): *Bayesian Data Analysis: Second Edition*. Chapman & Hall/CRC Press LLC, Florida.
- GERARDY, T., MÖCKEL, R., TROFF, H. & BISCHOFF, B. [Hrsg.] (2014): *Praxis der Grundstücksbewertung*. OLZOG Verlag GmbH, München.
- GONDRING, H. (2009): *Immobilienwirtschaft*. 2. Aufl. Verlag Franz Vahlen GmbH, München.
- GRAFAREND, E. W. & SCHAFFRIN, B. (1993): *Ausgleichsrechnung in linearen Modellen*. BI Wissenschaftsverlag, Mannheim Leipzig.
- GRÜNING, R. & KÜHN, R. (2013): *Entscheidungsverfahren für komplexe Probleme: Ein heuristischer Ansatz*. 4. auflage Aufl. Springer Gabler, Berlin.
- GUM (2008): *Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement*. ISO/IEC Guide 98-3:2008, Genf.
- HAACK, B. (2006): *Sensitivitätsanalyse zur Verkehrswertermittlung von Grundstücken*. Dissertation, Schriftenreihe der Professur für Städtebau und BODenordnung der Universität Bonn.
- HARTUNG, J. & ELPELT, B. (2007): *Multivariate Statistik: Lehr- und Handbuch der angewandten Statistik*. 7 Aufl. Oldenbourg Wissenschaftsverlag GmbH, München.
- HASTINGS, W. K. (1970): *Monte-carlo sampling methods using Markov chains and their applications*. In: *Biometrika*, , 57: 97–109.
- HAWERK, W. (1992): *Zur Anwendung kinematischer Modelle in der Grundstücksbewertung*. Dissertation, Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Leibniz Universität Hannover, Nr. 178, Hannover.
- HAWKINS, D. M. (1980): *Identification of Outliers*. Chapman and Hall, London.
- HELD, L. (2008): *Methoden der statistischen Inferenz: Likelihood und Bayes*. Spektrum Akademischer Verlag, Heidelberg.
- HELD, L. & BOVÉ, D. S. (2014): *Applied Statistical Inference: Likelihood and Bayes*. Springer-Verlag, Berlin Heidelberg.
- HELMERT, F. R. (1872): *Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. Teubner, Leipzig.
- HEUNECKE, O., KUHLMANN, H., WELSCH, W., EICHHORN, A. & NEUNER, H. (2013): *Handbuch Ingenieurgeodäsie: Auswertung geodätischer Überwachungsmessungen*. 2., neu bearbeitete und erweiterte auflage Aufl. Wichmann, Berlin.
- HILDEBRANDT, H. (1999): *Variationsbreite bei Grundstückswertermittlungen*. In: *VR - Vermessung und Raumordnung*, **61**, 6+7: 320–328.
- HÖPCKE, W. (1978): *Strategie der Grundstückswertermittlung durch Preisvergleich*. In: BRÜCKNER, R. [Hrsg.]: *Anwenderseminar "Mathematische Statistik bei der Ermittlung von Grundstückswerten"*, *Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der Technischen Universität Hannover*, Bd. 65a. Hannover.
- HOWIND, J. (2005): *Analyse des stochastischen Modells von GPS-Trägerphasenbeobachtungen*. Dissertation, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften.
- HÜBLER, O. (1989): *Ökonometrie*. Gustav Fischer Verlag, Stuttgart and New York.
- HUO, X. & NI, X. (2007): *When do stepwise algorithms meet subset selection criteria?* In: *The Annals of Statistics*, **35**, 2: 870–887.
- JANSSEN, O. (2002): *Monte-Carlo-Simulationen verbessern die Bewertungsqualität von Immobilien*. In: *GuG - Grundstücksmarkt und Grundstückswert*, , 01: 37–43.
- JAYNES, E. T. (2003): *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge.
- JEFFREYS, H. (1935): *Some Tests of Significance, Treated by the Theory and the Simplification of Models*. In: *Proceedings of the Cambridge Philosophy Society*, , 31: 203–222.
- JEFFREYS, H. (1961): *Theory of Probability*. Clarendon, Oxford.
- JESTER, S. & ROESCH, G. (2006): *Der Verkehrswert als Näherungswert*. In: *GuG - Grundstücksmarkt und Grundstückswert*, , 03: 157–161.
- KANNGIESER, E., KERTSCHER, D. & DEICHSEL, C. (1994): *Modellierung der Lageabhängigkeit von Bodenrichtwerten*. In: *ZfV - Zeitschrift für Vermessungswesen*, 527–539.
- KANNGIESER, E., KERTSCHER, D. & VOLLMER, H. (1990): *Untersuchung des Lageinflusses auf den Verkehrswert unbebauter Grundstücke*. In: *AVN (Allgemeine Vermessungs-Nachrichten)*, **97**, 10: 358–365.
- KASS, R. E. & RAFTERY, A. E. (1995): *Bayes Factors*. In: *Journal of the American Statistical Association*, **90**, 430: 773–795.
- KASTE, J. (2013): *Statistische Erfassung der Preisentwicklung für Wohneigentum in Niedersachsen mittels Häuserpreisindex nach Destatis*. Dissertation, Gottfried Wilhelm Leibniz Universität Hannover, Hannover.
- KAUKO, T. & D'AMATO, M. [Hrsg.] (2008): *Mass Appraisal Methods: An international perspective for property valuers*. John Wiley & Sons Ltd., The Atrium and Southern Gate and Chichester and West Sussex and United Kingdom.
- KENDALL, M. G. & STUART, A. (1969): *The Advanced Theory of Statistics*, Bd. 1. Third edition Aufl. Charles Griffin & Company Limited, London.
- KLEIBER, W. (2007): *Verkehrswertermittlung von Grundstücken: Kommentar und Handbuch zur Ermittlung von Marktwerten (Verkehrswerten), Versicherungs- und Beleihungswerten unter Berücksichtigung von WertV und BelWertV*. 5. auflage Aufl. Bundesanzeiger Verlagsges., Köln.
- KLEIBER, W. (2010): *Verkehrswertermittlung von Grundstücken: Kommentar und Handbuch zur Ermittlung von Marktwerten (Verkehrswerten), Versicherungs- und Beleihungswerten unter Berücksichtigung der ImmoWertV*. 6., vollständig neu bearbeitete auflage Aufl. Bundesanzeiger Verlagsges., Köln.
- KLEIBER, W. (2014): *Verkehrswertermittlung von Grundstücken: Kommentar und Handbuch zur Ermittlung von Marktwerten (Verkehrswerten), Versicherungs- und Beleihungswerten unter Berücksichtigung der ImmoWertV*. 7., vollständig neu bearbeitete auflage Aufl. Bundesanzeiger Verlagsges., Köln.
- KOCH, K.-R. (1987): *Statistische Methoden zur Analyse von Grundstückspreisen*. In: *ZfV - Zeitschrift für Vermessungswesen*, **112**, 12: 617–621.
- KOCH, K.-R. (1990): *Bayesian Inference with Geodetic Applications*. Springer, Berlin.
- KOCH, K.-R. (1994): *Bayessche Inferenz für die Prädiktion und Filterung*. In: *ZfV - Zeitschrift für Vermessungswesen*, **119**, 9: 464–470.
- KOCH, K.-R. (1999): *Parameter Estimation and Hypothesis Testing in Linear Models*. Second, updated and enlarged edition Aufl. Springer-Verlag, Berlin Heidelberg New York.
- KOCH, K.-R. (2000): *Einführung in die Bayes-Statistik*. Springer Verlag, Berlin Heidelberg New York.

- KOCH, K.-R. (2004): *Parameterschätzung und Hypothesentests in linearen Modellen*. Vierte, bearbeitete Auflage. Ehemals Ferd. Dummlers Verlag, Bonn.
- KOCH, K.-R. (2007): *Introduction to Bayesian Statistics*. Second, updated and enlarged edition. Springer-Verlag, Berlin Heidelberg New York.
- KOCH, K.-R. & SCHMIDT, M. (1994): *Deterministische und stochastische Signale mit Anwendungen in der digitalen Bildverarbeitung*. Ferd. Dummlers Verlag, Bonn.
- KRARUP, T. (1969): *A contribution to the mathematical foundation of physical geodesy*, Bd. Meddelelse No.44. Geodætisk Institut, Kopenhagen.
- KRUSE, R., GEBHARDT, J. & KLAWONN, F. (1995): *Fuzzy-Systeme*. Wiley, Chichester.
- KUTTERER, H. (1994): *Intervallmathematische Behandlung endlicher Unschärfen linearer Ausgleichungsmodelle: Dissertation*, DGK C, Bd. 423. München.
- KUTTERER, H. (1999): *On the sensitivity of the results of least-squares adjustments concerning the stochastic model*. In: Journal of Geodesy, , 73: 350–361.
- KUTTERER, H. (2002): *Zum Umgang mit Ungewissheit in der Geodäsie: Bausteine für eine neue Fehlertheorie, Reihe C*, Bd. Heft Nr. 553. Verlag der Bayerischen Akademie der Wissenschaften in Kommission beim Verlag C. H. Beck, München.
- LAUER, M. (2004): *Entwicklung eines Monte-Carlo-Verfahrens zum selbständigen Lernen von Gauß-Mischverteilungen*. Dissertation, Universität Osnabrück, Osnabrück.
- LAUX, H., GILLENKIRCH, R. M. & SCHENK-MATHES, H. Y. (2014): *Entscheidungstheorie*. 9., vollständig überarbeitete Auflage. Springer Gabler, Berlin.
- LEE, P. M. (2012): *Bayesian Statistics: An Introduction*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex.
- LGLN (2015): *Handbuch für die Führung und Auswertung der Kaufpreissammlung (Handbuch-AKS)*. Stand: 04.03.2015 Aufl. Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN) - Landesvermessung und Geobasisinformation, Fachgebiet 223 Wertermittlung, Hannover.
- LIEBIG, S. & ACHE, P. (2013): *Ableitung von bundesweit anwendbaren Umrechnungskoeffizienten: Abschlussbericht*. Geschäftsstelle des Oberen Gutachterausschusses für Grundstückswerte Niedersachsen, Oldenburg.
- MEISSL, P. (1976): *Hilbert spaces and their applications to geodetic least squares problems*. In: Bollettino di Geodesia e Scienze Affini, , 35: 49–80.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953): *Equations of state calculations by fast computing machines*. In: Journal of Chemical Physics, , 21: 1087–1091.
- METZGER, B. (2008): *Wertermittlung von Immobilien und Grundstücken*. 3. Aufl. Haufe Verlag, München.
- MI (2015): *Beschreibung der Elemente der Kaufpreissammlung 2008*. Stand: 21.06.2015 Aufl. Niedersächsisches Ministerium für Inneres und Sport - Referat 43, Hannover.
- MORITZ, H. (1973a): *Neuere Ausgleichungs- und Prädiktionsverfahren*. In: ZfV - Zeitschrift für Vermessungswesen, **98**, 4: 137–146.
- MORITZ, H. (1973b): *Least-squares collocation*. Dissertation, Deutsche Geodätische Kommission, München.
- MORITZ, H. (1980): *Advanced Physical Geodesy, Sammlung Wichmann*, Bd. 13. Herbert Wichmann Verlag, Karlsruhe.
- MÖSER, M., MÜLLER, G., SCHLEMMER, H. & WERNER, H. [Hrsg.] (2000): *Handbuch Ingenieurgeodäsie: Auswertung geodätischer Überwachungsmessungen*. Herbert Wichmann Verlag, Heidelberg.
- MUNIZZO, M. A. & VIRRUSO MUSIAL, L. (2009): *General Sales Comparison Approach*. Cengage Learning, Mason and USA.
- MÜRLE, M. (2006): *Aufbau eines Wertermittlungsinformationssystems*. Dissertation, Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften der Universität Fridericiana zu Karlsruhe.
- NEUMANN, I. (2009): *Zur Modellierung eines erweiterten Unsicherheitsrahmens in Parameterschätzung und Hypothesentests, Reihe C*, Bd. Heft Nr. 634. Verlag der Bayerischen Akademie der Wissenschaften in Kommission beim Verlag C. H. Beck, München.
- NIEMEIER, W. (2008): *Ausgleichungsrechnung: Statistische Auswertemethoden*. 2. Auflage. Walter de Gruyter, Berlin and New York.
- OBERMAIER, R. & SALIGER, E. (2013): *Betriebswirtschaftliche Entscheidungstheorie: Einführung in die Logik individueller und kollektiver Entscheidungen*. 6., grundlegend überarbeitete Auflage. Oldenbourg, München.
- O'HAGAN, A. (1995): *Fractional Bayes factors for model comparison (with discussion)*. In: Journal of the Royal Statistical Society, , B57: 99–138.
- PAPULA, L. (2008): *Mathematik für Ingenieure und Naturwissenschaftler Band 3: Vektoranalysis, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Fehler- und Ausgleichungsrechnung*. 5., verbesserte und erweiterte Auflage. Vieweg+Teubner Verlag, GWV Fachverlage GmbH, Wiesbaden.
- PATTON, C. V. & SAWICKI, D. S. (1993): *Basic Methods of Policy Analysis and Planning*. 2nd ed. Aufl. Prentice-Hall, Englewood Cliffs and New Jersey.
- PELZER, H. (1976): *Mathematische Modelle für die Wertermittlung*. In: BRÜCKNER, R. [Hrsg.]: *Mathematische Statistik bei der Ermittlung von Grundstückswerten*, Bd. 65, 208–218. Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der Technischen Universität Hannover.
- PELZER, H. (1978): *Ein indirektes Vergleichsverfahren unter Anwendung statistischer Methoden*. In: ZfV - Zeitschrift für Vermessungswesen, **Jg. 103**, 6: 245–254.
- PETERSEN, H. (2007): *Was können wir von der Verkehrswertermittlung in der Bundesrepublik an Ergebnissicherheit erwarten?* In: GuG - Grundstücksmarkt und Grundstückswert, , 04: 203–208.
- PETTIT, L. I. & SMITH, A. (1985): *Outliers and influential observations in linear models*. In: BERNARDO, J. M., DEGROOT, M. H., LINDLEY, D. V. & SMITH, A. F. M. [Hrsg.]: *Bayesian Statistics*, 473–494. University Press, Valencia.
- REUTER, F. (1989): *Zur Umsetzung des Verkehrswertbegriffes in Wertermittlungsmethoden*. In: VR - Vermessung und Raumordnung, **51**, 6+7: 377–386.
- ROMMELFANGER, H. (1994): *Fuzzy Decision Support-Systeme: Entscheiden bei Unschärfe*. 2. Aufl. Berlin Heidelberg.
- RÖSSLER, R., LANGNER, JOHANNES, SIMON, J., KLEIBER, W., JOERIS, DAGMAR & SIMON, T. (2005): *Schätzung und Ermittlung von Grundstückswerten*. 8., überarbeitete und erweiterte Auflage. Wolters Kluwer Deutschland GmbH, München/Unterschleißheim.
- RUMMEL, R. (1976): *A model comparison in least squares collocation*. In: Bulletin Géodésique, , 50: 181–192.

- SANDNER, S. & WEBER, U. [Hrsg.] (2007): *Lexikon der Immobilienwertermittlung*. 2., erweiterte und überarbeitete auflage Aufl. Bundesanzeiger Verlagsges.mmbH, Köln.
- SCHLITTEGEN, R. (1998): *Einführung in die Statistik: Analyse und Modellierung von Daten*. 8 Aufl. Oldenbourg Verlag, München and Wien.
- SCHÖN, S. (2003): *Analyse und Optimierung geodätischer Messanordnungen unter besonderer Berücksichtigung des Intervallansatzes, Reihe C*, Bd. Heft Nr. 567. Verlag der Bayerischen Akademie der Wissenschaften in Kommission beim Verlag C. H. Beck, München.
- SCHÖNFELD, P. (1971): *Methoden der Ökonometrie: Band II - Stochastische Regressoren und simultane Gleichungen*. Verlag Franz Vahlen, München.
- SCHÖNWANDT, W. & MÜLLER-HERBERS, S. (2007): *Methoden zur Beurteilung von Varianten*. Arbeitspapier, 4. auflage Aufl. Fakultät Architektur und Stadtplanung - Institut für Grundlagen der Planung - Universität Stuttgart, Stuttgart.
- SCHULTE, K.-W. [Hrsg.] (2008): *Immobilienökonomie: Band IV Volkswirtschaftliche Grundlagen*. Oldenbourg Wissenschaftsverlag GmbH, München.
- SCHULTE, K.-W., STURM, V. & WIFFLER, M. (2008): *Volkswirtschaftslehre und Immobilienökonomie*. In: SCHULTE, K.-W. [Hrsg.]: *Immobilienökonomie*, 2–26. Oldenbourg Wissenschaftsverlag GmbH, München.
- SEELE, W. (1982): *Wertermittlung bei der Preisprüfung und Ungewißheit des Verkehrswertes*. In: VR - Vermessung und Raumordnung, **44**, 3: 105–121.
- SIMON, T. (2004): *Verbessert die Monte-Carlo-Simulation die Grundstückswertermittlung?* In: GuG - Grundstücksmarkt und Grundstückswert, , 02: 93–101.
- SOMMER, G. (2000): *Das Monte-Carlo Verfahren in der Ertragswertermittlung*. In: Der Sachverständige, , 12: 27–31.
- STONE, M. (1974): *Cross-validatory choice and assessment of statistical predictions*. In: Journal of the Royal Statistical Society B, , 36: 111–133.
- STÖWE, H. (1977): *Ökonometrie*. Verlag Anton Hain Meisenheim GmbH, Meisenheim.
- STREICH, J.-W. (2003): *Die Ortsübliche Vergleichsmiete - ein nicht exakt bestimmbarer Rechtsbegriff*. In: GuG - Grundstücksmarkt und Grundstückswert, , 01: 1–7.
- STROTKAMP, H.-P. & SAUERBORN, C. (2002): *Welche Genauigkeiten können bei der Anwendung des Ertrags- und Sachwertverfahrens der Wertermittlungsverordnung erwartet werden?* In: WFA - WertermittlungsForum Aktuell, , 04: 157–163.
- TAPPE, S. (2013): *Einführung in die Wahrscheinlichkeitstheorie*. Springer Spektrum, Berlin Heidelberg.
- TEUSCH, A. (2006): *Einführung in die Spektral- und Zeitreihenanalyse*. Dissertation, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften.
- TSCHERING, C. C. (1978): *Collocation and least squares methods as a tool for handling gravity field dependent data obtained through space research techniques techniques*. In: HIEBER, S. & GUYENNE, T. D. [Hrsg.]: *European Workshop on Space Oceanography, Navigation and Geodynamics*, 141–149. Paris.
- TSCHIRK, W. (2014): *Statistik: Klassisch oder Bayes*. Springer-Verlag, Berlin Heidelberg.
- UHDE, C. (1982): *Mathematische Modelle zur Analyse von Grundstücksmärkten*, Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der Technischen Universität Hannover, Bd. 118.
- URBAN, D. & MAYERL, J. (2006): *Regressionsanalyse: Theorie, Technik und Anwendung: Lehrbuch - Studienskripten zur Soziologie*. VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH, Wiesbaden.
- VIERTL, R. (2003): *Einführung in die Stochastik: Mit Elementen der Bayes-Statistik und der Analyse unscharfer Informationen*. 3., überarbeitete und erweiterte auflage Aufl. Springer-Verlag, Wien.
- VIERTL, R. (2011): *Statistical Methods for Fuzzy Data*. John Wiley Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex.
- WANG, K. & WOLVERTON, M. L. [Hrsg.] (2002): *Real Estate Valuation Theory*. Springer Science + Business Media, New York.
- WEITKAMP, A. & ALKHATIB, H. (2012): *Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 2*. In: zvf - Zeitschrift für Geodäsie, Geoinformation und Landmanagement, **137**, 2: 103–114.
- WEITKAMP, A. & ALKHATIB, H. (2014): *Die Bewertung kaufpreisarmer Lagen mit multivariaten statistischen Verfahren - Möglichkeiten und Grenzen robuster Methoden bei der Auswertung weniger Kauffälle*. In: AVN (Allgemeine Vermessungs-Nachrichten), , 1: 3–12.
- WERNECKE, M. & ROTTKE, N. (2001): *Immobilienfinanzierung - Alle reden davon, doch keiner tut's: Antizyklisch investieren*. In: Immobilien Zeitung, , 23: 10.
- WOLF, H. (1977): *Zur Grundlegung der Kollokationsmethode*. In: ZfV - Zeitschrift für Vermessungswesen, **102**, 6: 237–239.
- WÜNSCHE, A. & NÄTHER, W. (2002): *Least-Squares fuzzy regression with fuzzy random variables*. In: Fuzzy Sets and Systems, , 130: 43–50.
- ZADDACH, S. (2014c): *Kombination von Fuzzy-Bayes-Ansätzen für die Ermittlung von Verkehrswerten*. DFG - Deutsche Forschungsgemeinschaft, Projektabschlussbericht (unveröffentlicht).
- ZADDACH, S. & ALKHATIB, H. (2012a): *Use of Collocation Method as an Improved Sales Comparison Approach*. In: Proceedings of the FIG Working Week 2012, Rome, Italy, 6 - 10 May 2014.
- ZADDACH, S. & ALKHATIB, H. (2012b): *Enhancement of the Classical Regression Analysis by Means of Least Squares Collocation Method*. In: Proceedings of the FIG Commission 9 Workshop - Mass Appraisal Techniques, Paphos, Cyprus, 14 - 16 September 2012.
- ZADDACH, S. & ALKHATIB, H. (2013a): *Anwendung der Kollokation als erweitertes Vergleichswertverfahren in der Immobilienwertermittlung*. In: zvf - Zeitschrift für Geodäsie, Geoinformation und Landmanagement, **138**, 2: 144–153.
- ZADDACH, S. & ALKHATIB, H. (2013b): *Quantifying the impact of uncertainty to the market value by introducing a Bayesian sales comparison approach*. In: Proceedings of the RICS Cobra Research Conference, New Delhi, India, 10 - 12 September 2013.
- ZADDACH, S. & ALKHATIB, H. (2014a): *Least Squares Collocation as an Enhancement to Multiple Regression Analysis in Mass Appraisal Applications*. In: Journal of Property Tax Assessment & Administration, **11**, 1: 47–66.
- ZADDACH, S. & ALKHATIB, H. (2014b): *Propagating the uncertainty of the market value by the use of a Bayesian regression approach*. In: Proceedings of the XXV FIG Congress, Kuala Lumpur, Malaysia, 16 - 21 June 2014.

- ZADDACH, S., WEITKAMP, A. & ALKHATIB, H. (2010): *New method to deduce regional factors of German cost approach*. In: Polish Real Estate Scientific Society, , 4: 95 ff.
- ZADEH, L. A. (1965): *Fuzzy sets*. In: Information Control, , 8: 338–353.
- ZEISSLER, M. (2012): *Zur Ermittlung von Bodenrichtwerten bei fehlenden Kaufpreisen unbebauter Grundstücke*. Dissertation, Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover, Nr. 301.
- ZELLNER, A. & MOULTON, B. R. (1985): *Bayesian regression diagnostics with applications to international consumption and income data*. In: Econometrics, , 29: 187–211.
- ZIEGENBEIN, W. (1976): *Anwendung der multiplen Regression*. In: BRÜCKNER, R. [Hrsg.]: *Mathematische Statistik bei der Ermittlung von Grundstückswerten*, Bd. 65, 219–235. Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der Technischen Universität Hannover.
- ZIEGENBEIN, W. (1977): *Zur Anwendung multivariater Verfahren der mathematischen Statistik in der Grundstückswertermittlung*. Dissertation, Technische Universität Hannover.
- ZIEGENBEIN, W. (1986): *Zur Analyse der automatisiert geführten Kaufpreissammlungen für Grundstücke in Niedersachsen*. In: Nachrichten der Niedersächsischen Vermessungs- und Katasterverwaltung, **36**, 3: 195–218.
- ZIEGENBEIN, W. (1995): *Programmgesteuerte Regressionsanalyse und Vergleichswertermittlung im Programmsystem AKS*. In: Nachrichten der Niedersächsischen Vermessungs- und Katasterverwaltung, **45**, 4: 321–335.
- ZIEGENBEIN, W. (1999): *Wertermittlungsinformationssystem Niedersachsen*. In: Nachrichten der Niedersächsischen Vermessungs- und Katasterverwaltung, **49**, 3: 121–126.
- ZIEGENBEIN, W. (2010): *Immobilienwertermittlung*. In: KUMMER, K. & FRANKENBERGER, J. [Hrsg.]: *Das deutsche Vermessungs- und Geoinformationswesen*, 421–468. Wichmann, Verlagsgruppe Hüthig Jehle Rehm GmbH, Heidelberg and München and Landsberg and Frechen and Hamburg.
- ZIEGENBEIN, W. & HAWERK, W. (1978): *Erfahrungen bei der Prädiktion von Grundstückswerten*. In: ZfV - Zeitschrift für Vermessungswesen, **Jg. 103**, 6: 254–261.

Abbildungsverzeichnis

1.1. Aufbau der Arbeit	21
2.1. Wesentliche Eigenschaften von Immobilien und Immobilienmärkten	24
2.2. Prinzip der Preisbildung am Grundstücksmarkt	27
2.3. Grundstruktur ökonometrischer Modelle	32
2.4. Prinzip von Genauigkeit und Präzision.	36
2.5. Übersicht der Wertermittlungsverfahren als Preisvergleiche	41
2.6. Arten des Vergleichswertverfahrens	42
2.7. Prinzipskizze der Kollokation im zweidimensionalen Fall.	45
4.1. Allgemeiner Ablauf der Auswertestrategie	86
4.2. Modelldefinition	87
4.3. Ablauf der Stichprobenaufbereitung	88
4.4. Auswertestrategie für das indirekte Vergleichswertverfahren basierend auf der Bayes-Theorie	92
4.5. Grundprinzip der Rückwärtsstrategie für die Modellwahl im Branch-and-Bound Algorithmus	101
4.6. Grundprinzip der Vorwärtsstrategie für die Modellwahl im Branch-and-Bound Algorithmus	101
4.7. Beispiel des iterativen Branch-and-Bound Algorithmus für die Modellwahl	102
4.8. Schema der rekursiven Bayesschen Schätzung	104
4.9. Beispiel einer Paarvergleichsmatrix	107
4.10. Vergleichskriterien für die Wertung der Bayesschen Modelle	108
5.1. Modellübersicht in der Auswertestrategie	109
5.2. Räumlicher Teilmarkt für die praktischen Untersuchungen: Stadt Osnabrück und Stadtrandgemeinden (Kartengrundlage: Amtliche Geobasisdaten des LGLN, 2015)	110
5.3. Indexreihen für freistehende Ein- und Zweifamilienhäuser (links) und Bodenpreisindexreihen für individuelles Wohnbauland in Osnabrück (rechts)	111
5.4. Histogramme der Zielgröße KPW der Epoche 2004 (links: Originaldaten, rechts: Aufbereitete Daten)	116
5.5. Histogramme der Zielgröße KPW der Epoche 2005 (links: Originaldaten, rechts: Aufbereitete Daten)	116
5.6. Posteriori-Schätzung des Varianzfaktors σ^2 der Epochen 2005 bis 2014, Bayessche Regression	126
5.7. Posteriori-Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2005 (oben); Differenzplots der HPDIs (unten)	128
5.8. Posteriori-Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2009 (oben); Differenzplots der HPDIs (unten)	128
5.9. Posteriori-Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2014 (oben); Differenzplots der HPDIs (unten)	128
5.10. Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2005	129
5.11. Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2009	129
5.12. Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2014	129
5.13. Medianwerte der Abweichungen zwischen Kaufpreisen und Posteriori-Schätzwerten $\bar{\mathbf{y}}$ der Epochen 2005 bis 2014, Bayessche Regression	130
5.14. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_0$ der Epochen 2005 bis 2014, Bayessche Regression	130
5.15. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_1$ der Epochen 2005 bis 2014, Bayessche Regression	130
5.16. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_2$ der Epochen 2005 bis 2014, Bayessche Regression	130
5.17. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_3$ der Epochen 2005 bis 2014, Bayessche Regression	131

5.18. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_1$ der Epochen 2005 bis 2014, Bayessche Regression	131
5.19. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_1$ der Epochen 2005 bis 2014, Bayessche Regression	131
5.20. Varianz $V_{\bar{\sigma}^2}$ des Varianzfaktors $\bar{\sigma}^2$ der Epochen 2005 bis 2014, Bayessche Regression	132
5.21. Medianwerte der HPDI-Spannen von $\bar{\mathbf{y}}$ der Epochen 2005 bis 2014, Bayessche Regression	133
5.22. Theoretische Verläufe von Funktionen des exponentiellen und des Gaußschen Typs	134
5.23. Empirische und analytische Autokorrelationsfunktionen der Epoche 2005	136
5.24. Empirische und analytische Autokorrelationsfunktionen der Epoche 2009	136
5.25. Empirische und analytische Autokorrelationsfunktionen der Epoche 2014	136
5.26. Posteriori-Schätzung des Varianzfaktors $\bar{\sigma}^2$ der Epochen 2005 bis 2014, Bayessche Kollokation	139
5.27. Posteriori-Schätzwerte $\bar{\mathbf{s}}$, Epoche 2005 (oben); Differenzplots der HPDIs (unten)	140
5.28. Posteriori-Schätzwerte $\bar{\mathbf{s}}$, Epoche 2009 (oben); Differenzplots der HPDIs (unten)	140
5.29. Posteriori-Schätzwerte $\bar{\mathbf{s}}$, Epoche 2014 (oben); Differenzplots der HPDIs (unten)	140
5.30. Posteriori-Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Kollokation, Epoche 2005 (oben); Differenzplots der HPDIs (unten)	141
5.31. Posteriori-Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Kollokation, Epoche 2009 (oben); Differenzplots der HPDIs (unten)	141
5.32. Posteriori-Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Kollokation, Epoche 2014 (oben); Differenzplots der HPDIs (unten)	141
5.33. Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Kollokation, Epoche 2005	142
5.34. Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Kollokation, Epoche 2009	142
5.35. Abweichungen der Kaufpreise zu Posteriori-Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Kollokation, Epoche 2014	142
5.36. Medianwerte der prozentualen Abweichungen zwischen Kaufpreisen und Posteriori-Schätzwerten $\bar{\mathbf{y}}$ der Epochen 2005 bis 2014, Bayessche Kollokation	142
5.37. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_0$ der Epochen 2005 bis 2014, Bayessche Kollokation	144
5.38. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_1$ der Epochen 2005 bis 2014, Bayessche Kollokation	144
5.39. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_2$ der Epochen 2005 bis 2014, Bayessche Kollokation	144
5.40. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_3$ der Epochen 2005 bis 2014, Bayessche Kollokation	145
5.41. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_4$ der Epochen 2005 bis 2014, Bayessche Kollokation	145
5.42. HPDI des Posteriori-Trendkoeffizienten $\bar{\beta}_5$ der Epochen 2005 bis 2014, Bayessche Kollokation	145
5.43. Varianz $V_{\bar{\sigma}^2}$ des Varianzfaktors $\bar{\sigma}^2$ der Epochen 2005 bis 2014, basierend auf Bayesscher Regression .	146
5.44. Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{\mathbf{s}}$ der Epochen 2005 bis 2014	147
5.45. Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{\mathbf{y}}$ der Epochen 2005 bis 2014, Bayessche Kollokation	147
5.46. Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Bayesschen Regression, Epoche 2005 (oben); Differenzplots der HPDIs (unten)	150
5.47. Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Bayesschen Regression, Epoche 2009 (oben); Differenzplots der HPDIs (unten)	150
5.48. Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Bayesschen Regression, Epoche 2014 (oben); Differenzplots der HPDIs (unten)	150
5.49. Prozentuale Abweichungen zwischen Kaufpreisen und prädizierten Posteriori-Schätzwerten $\bar{\mathbf{y}}_p$ in den Epochen 2005 bis 2014, Bayessche Regression	151
5.50. Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ in den Epochen 2005 bis 2014, Bayessche Regression	151
5.51. Posteriori-Schätzwerte $\bar{\mathbf{s}}_p$, Epoche 2005 (oben); Differenzplots der HPDIs (unten)	152
5.52. Posteriori-Schätzwerte $\bar{\mathbf{s}}_p$, Epoche 2009 (oben); Differenzplots der HPDIs (unten)	152
5.53. Posteriori-Schätzwerte $\bar{\mathbf{s}}_p$, Epoche 2014 (oben); Differenzplots der HPDIs (unten)	152
5.54. Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Bayesschen Kollokation, Epoche 2005 (oben); Differenzplots der HPDIs (unten)	153

5.55. Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Bayesschen Kollokation, Epoche 2009 (oben); Differenzplots der HPDIs (unten)	153
5.56. Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Bayesschen Kollokation, Epoche 2014 (oben); Differenzplots der HPDIs (unten)	153
5.57. Prozentuale Abweichungen zwischen Kaufpreisen und prädizierten Posteriori-Schätzwerten $\bar{\mathbf{y}}_p$ in den Epochen 2005 bis 2014, Bayessche Kollokation	154
5.58. Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{\mathbf{s}}_p$ der Epochen 2005 bis 2014	155
5.59. Medianwerte der HPDI-Spannen der Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Epochen 2005 bis 2014, Bayessche Kollokation	156
6.1. Konzept der Ergebnisdiskussion	157
6.2. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2005	161
6.3. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2009	161
6.4. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2014	161
6.5. Mittlerer quadratischer Fehler der Posteriori-Schätzung $\bar{\mathbf{y}}$ der Epochen 2005 bis 2014, Bayessche Regression	162
6.6. Mittlere quadratische Fehler der Posteriori-Schätzung $\bar{\mathbf{y}}$ der Epochen 2005 bis 2014, Bayessche Kollokation	163
6.7. Paarvergleichsmatrix für die Schätzung der Posteriori-Zielgröße $\bar{\mathbf{y}}$	164
6.8. Absolute Residuen der prädizierten Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Epoche 2005	165
6.9. Absolute Residuen der prädizierten Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Epoche 2009	165
6.10. Absolute Residuen der prädizierten Posteriori-Schätzwerte $\bar{\mathbf{y}}_p$ der Epoche 2014	165
6.11. Mittlere quadratische Fehler der prädizierten Posteriori-Schätzung $\bar{\mathbf{y}}_p$ der Epochen 2005 bis 2014, Bayessche Regression	166
6.12. Mittlere quadratische Fehler der prädizierten Posteriori-Schätzung $\bar{\mathbf{y}}_p$ der Epochen 2005 bis 2014, Bayessche Kollokation	166
6.13. Paarvergleichsmatrix für die funktionale Modellierung der prädizierten Posteriori-Zielgröße $\bar{\mathbf{y}}_p$	168
6.14. Paarvergleichsmatrix für die Unsicherheitsmodellierung der Trendkoeffizienten $\bar{\beta}$	170
6.15. Paarvergleichsmatrix für die Unsicherheit der Varianz $\bar{\sigma}^2$	171
6.16. Paarvergleichsmatrix für die Unsicherheit der Schätzwerte der Posteriori-Zielgröße $\bar{\mathbf{y}}$	174
6.17. Paarvergleichsmatrix für die Unsicherheit der prädizierten Schätzwerte der Posteriori-Zielgröße $\bar{\mathbf{y}}_p$	177
6.18. Zusammenfassung der Ergebnisse der Paarvergleiche	180
B.1. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2005 (oben); Differenzplots der HPDIs (unten)	209
B.2. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2006 (oben); Differenzplots der HPDIs (unten)	210
B.3. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2007 (oben); Differenzplots der HPDIs (unten)	210
B.4. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2008 (oben); Differenzplots der HPDIs (unten)	210
B.5. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2009 (oben); Differenzplots der HPDIs (unten)	211
B.6. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2010 (oben); Differenzplots der HPDIs (unten)	211
B.7. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2011 (oben); Differenzplots der HPDIs (unten)	211
B.8. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2012 (oben); Differenzplots der HPDIs (unten)	212
B.9. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2013 (oben); Differenzplots der HPDIs (unten)	212
B.10. Schätzwerte $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2014 (oben); Differenzplots der HPDIs (unten)	212
B.11. Abweichungen der Kaufpreise zu Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2005	213
B.12. Abweichungen der Kaufpreise zu Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2006	213
B.13. Abweichungen der Kaufpreise zu Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2007	213
B.14. Abweichungen der Kaufpreise zu Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2008	213
B.15. Abweichungen der Kaufpreise zu Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2009	214
B.16. Abweichungen der Kaufpreise zu Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2010	214
B.17. Abweichungen der Kaufpreise zu Schätzwerten $\bar{\mathbf{y}}$ der Bayesschen Regression, Epoche 2011	214

C.37. Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2011	233
C.38. Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2012	233
C.39. Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2013	234
C.40. Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2014	234
C.41. Schätzwerte $\bar{s}_p$, Epoche 2005 (oben); Differenzplots der HPDIs (unten)	235
C.42. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2005, Bayesische Kollokation)	235
C.43. Schätzwerte $\bar{s}_p$, Epoche 2006 (oben); Differenzplots der HPDIs (unten)	236
C.44. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2006, Bayesische Kollokation)	236
C.45. Schätzwerte $\bar{s}_p$, Epoche 2007 (oben); Differenzplots der HPDIs (unten)	237
C.46. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2007, Bayesische Kollokation)	237
C.47. Schätzwerte $\bar{s}_p$, Epoche 2008 (oben); Differenzplots der HPDIs (unten)	238
C.48. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2008, Bayesische Kollokation)	238
C.49. Schätzwerte $\bar{s}_p$, Epoche 2009 (oben); Differenzplots der HPDIs (unten)	239
C.50. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2009, Bayesische Kollokation)	239
C.51. Schätzwerte $\bar{s}_p$, Epoche 2010 (oben); Differenzplots der HPDIs (unten)	240
C.52. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2010, Bayesische Kollokation)	240
C.53. Schätzwerte $\bar{s}_p$, Epoche 2011 (oben); Differenzplots der HPDIs (unten)	241
C.54. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2011, Bayesische Kollokation)	241
C.55. Schätzwerte $\bar{s}_p$, Epoche 2012 (oben); Differenzplots der HPDIs (unten)	242
C.56. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2012, Bayesische Kollokation)	242
C.57. Schätzwerte $\bar{s}_p$, Epoche 2013 (oben); Differenzplots der HPDIs (unten)	243
C.58. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2013, Bayesische Kollokation)	243
C.59. Schätzwerte $\bar{s}_p$, Epoche 2014 (oben); Differenzplots der HPDIs (unten)	244
C.60. Prädizierte Schätzwerte $\bar{y}_p$ und Differenzplots HPDI (2014, Bayesische Kollokation)	244
C.61. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2005 .	245
C.62. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2006 .	245
C.63. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2007 .	246
C.64. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2008 .	246
C.65. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2009 .	246
C.66. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2010 .	247
C.67. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2011 .	247
C.68. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2012 .	247
C.69. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2013 .	248
C.70. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2014 .	248
C.71. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2005 (Prädiktion)	249
C.72. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2006 (Prädiktion)	249
C.73. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2007 (Prädiktion)	250
C.74. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2008 (Prädiktion)	250
C.75. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2009 (Prädiktion)	250
C.76. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2010 (Prädiktion)	251
C.77. Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2011 (Prädiktion)	251

C.78.Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2012 (Prädiktion)	251
C.79.Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2013 (Prädiktion)	252
C.80.Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2014 (Prädiktion)	252

Tabellenverzeichnis

2.1. Potenziell wertbeeinflussende Grundstücksmerkmale nach Nr. 4.2 bis 4.4 VW-RL	41
4.1. Multivariate statistische Verfahren in der Verkehrswertermittlung	79
4.2. Exemplarische Zielgrößen in der Verkehrswertermittlung	81
4.3. Übersicht über die Berechnungen von Korrelationen in der Bayesschen Auswertestrategie	92
4.4. Parameter k in der Bayesschen Ausreißersuche für die Priori-Wahrscheinlichkeit 95 %	94
4.5. Skala für die Evidenz von Bayes-Faktoren nach JEFFREYS (1961)	100
5.1. Vertrags- und Ordnungsmerkmale für die Stichprobenerhebung	112
5.2. Selektionsgrenzen für die Gesamtstichprobe der Ein- und Zweifamilienhäuser (Stadt und Stadtrand- gemeinden Osnabrück)	114
5.3. Deskriptive Statistik der Epochen 2004 und 2005	115
5.4. Iterationen in der Aufbereitung der Zielgröße der Initialepochen	117
5.5. Häufigkeiten der Klassenbelegung der Einflussgröße AST	118
5.6. Iterationen in der Aufbereitung der stetigen Einflussgrößen in den Epochen 2004 und 2005	119
5.7. Korrelationen zwischen den Einflussgrößen (Epoche 2004)	119
5.8. Korrelationen zwischen den Einflussgrößen (Epoche 2005)	119
5.9. Einheiten der Zielgröße und der Einflussgrößen in den praktischen Untersuchungen	120
5.10. Ergebnisse für die Schätzung des Gesamtmodells M_g^{2005}	121
5.11. Übersicht über die Kofaktormatrizen des Gesamtmodells $M_g^{g,2005}$	121
5.12. Übersicht über die Submodelle auf Basis des Gesamtmodells $M_g^{g,2005}$ (1. Iteration)	121
5.13. Ergebnisse für die Schätzung der Submodelle	122
5.14. Übersicht über die Kofaktormatrizen der Submodelle $M_1^{s,2005}$ bis $M_5^{s,2005}$	122
5.15. Bayes-Faktoren für den Vergleich der Submodelle $M_k^{s,2005}$ mit dem Gesamtmodell $M_g^{g,2005}$	123
5.16. Stichprobenaufbereitung der Epochen 2006 bis 2014	124
5.17. Bayes-Faktoren der Epochen 2006 bis 2014 unter rekursiver Auswertung (Submodelle)	124
5.18. Modellparameter der Epochen 2005 bis 2014, basierend auf Bayesscher Regression	125
5.19. Geschätzte Parameter $\hat{c}$ der analytischen Autokorrelationsfunktionen der Epochen 2005 bis 2014	136
5.20. Modellparameter der Epochen 2005 bis 2014, basierend auf Bayesscher Kollokation	138
6.1. Medianwerte der absoluten Residuen in den Auswertepochen 2005 bis 2014	166
A.1. Deskriptive Statistiken der Epochen 2004 bis 2010	205
A.2. Deskriptive Statistiken der Einzelstichproben (2010 bis 2014)	206
A.3. Korrelationen zwischen den Einflussgrößen (Epoche 2004)	206
A.4. Korrelationen zwischen den Einflussgrößen (Epoche 2005)	206
A.5. Korrelationen zwischen den Einflussgrößen (Epoche 2006)	207
A.6. Korrelationen zwischen den Einflussgrößen (Epoche 2007)	207
A.7. Korrelationen zwischen den Einflussgrößen (Epoche 2008)	207
A.8. Korrelationen zwischen den Einflussgrößen (Epoche 2009)	207
A.9. Korrelationen zwischen den Einflussgrößen (Epoche 2010)	207
A.10. Korrelationen zwischen den Einflussgrößen (Epoche 2011)	207
A.11. Korrelationen zwischen den Einflussgrößen (Epoche 2012)	208
A.12. Korrelationen zwischen den Einflussgrößen (Epoche 2013)	208
A.13. Korrelationen zwischen den Einflussgrößen (Epoche 2014)	208

A. Anhang – Stichprobenbeschreibungen

Die folgenden Abschnitte geben durch die Stichprobenbeschreibungen einen Überblick über die Kauffalldaten des Fallbeispiels.

A.1. Deskriptive Statistiken der Stichproben 2006 bis 2014

Die Tabellen A.1 und A.2 enthalten die deskriptiven Statistiken der Einzelstichproben der Auswertepochen 2004 bis 2014, die im Zuge des nach Kapitel 5 (S. 109 ff.) beschriebenen Fallbeispiels erhoben werden.

Tabelle A.1.: Deskriptive Statistiken der Epochen 2004 bis 2010

Epoche	Merkmal	DAT	KPW	BRW	GRF	WOF	AST	KEL	ALT
	Einheit	[-]	[€/m ²]	[€/m ²]	[m ²]	[m ²]	[-]	[%]	[Jahre]
2004	Anzahl	94	94	94	94	94	87	94	94
	Minimum	02.01.2004	769	65	305	90	1	0	4
	Maximum	28.12.2004	2.867	340	1.499	265	5	100	54
	Mittelwert	-	1.458	148	774	161	3	77	30
	Median	-	1.394	140	726	150	3	100	30
2005	Anzahl	105	105	105	105	105	100	105	105
	Minimum	05.01.2005	813	75	267	92	1	0	3
	Maximum	30.12.2005	2.667	360	1.414	280	5	100	55
	Mittelwert	-	1.481	164	780	162	3	75	32
	Median	-	1.436	145	761	155	3	100	33
2006	Anzahl	84	84	84	84	84	82	84	84
	Minimum	06.01.2006	627	60	287	85	1	0	3
	Maximum	29.12.2006	3.138	400	1.256	310	5	100	56
	Mittelwert	-	1.411	155	757	157	3	81	34
	Median	-	1.333	135	723	150	3	100	36
2007	Anzahl	95	95	95	95	95	84	95	95
	Minimum	05.01.2007	600	75	327	85	1	0	3
	Maximum	28.12.2007	3.471	340	1.430	320	5	100	57
	Mittelwert	-	1.333	154	795	167	3	77	33
	Median	-	1.293	140	742	160	3	100	34
2008	Anzahl	94	94	94	94	94	85	94	94
	Minimum	07.01.2008	575	75	300	102	1	0	3
	Maximum	22.12.2008	2.370	400	1.292	325	5	100	55
	Mittelwert	-	1.313	167	722	167	3	75	36
	Median	-	1.285	150	693	160	3	100	38
2009	Anzahl	107	107	107	107	107	92	107	107
	Minimum	03.01.2009	588	75	417	74	1	0	4
	Maximum	23.12.2009	2.800	340	1.409	315	5	100	59
	Mittelwert	-	1.296	158	773	167	3	76	38
	Median	-	1.255	140	752	160	3	100	39
2010	Anzahl	128	128	128	128	128	114	128	128
	Minimum	11.01.2010	667	70	273	60	1	0	1
	Maximum	28.12.2010	3.107	340	1.371	341	5	100	59
	Mittelwert	-	1.309	147	777	161	3	67	37
	Median	-	1.238	130	772	152	3	100	40

Tabelle A.2.: Deskriptive Statistiken der Einzelstichproben (2010 bis 2014)

Epoche	Merkmal	DAT	KPW	BRW	GRF	WOF	AST	KEL	ALT
	Einheit	[-]	[€]	[€/m ²]	[m ²]	[m ²]	[-]	[%]	[Jahre]
2011	Anzahl	112	112	112	112	112	104	112	112
	Minimum	21.01.2011	522	60	327	93	1	0	2
	Maximum	28.12.2011	2.400	340	1.368	360	5	100	61
	Mittelwert	-	1.326	147	761	158	3	78	37
	Median	-	1.243	135	734	150	3	100	38
2012	Anzahl	134	134	134	134	134	121	134	134
	Minimum	05.01.2012	667	60	290	95	1	0	5
	Maximum	21.12.2012	2.938	340	1.486	300	5	100	61
	Mittelwert	-	1.347	143	805	157	3	79	39
	Median	-	1.285	135	774	152	3	100	41
2013	Anzahl	129	129	129	129	129	107	129	129
	Minimum	03.01.2013	734	70	281	80	1	0	3
	Maximum	30.12.2013	3.750	360	1.401	370	5	100	63
	Mittelwert	-	1.456	144	752	163	3	73	36
	Median	-	1.377	135	727	158	3	100	39
2014	Anzahl	76	76	76	76	76	55	76	76
	Minimum	17.01.2014	703	65	360	100	1	0	3
	Maximum	16.12.2014	3.064	360	1.476	283	5	100	60
	Mittelwert	-	1.486	148	772	160	3	78	38
	Median	-	1.377	130	721	156	3	100	42

A.2. Korrelationen der Einflussgrößen in den Stichproben 2004 bis 2014

Die Tabellen A.3 bis A.13 zeigen die Korrelationen, die zwischen den Einflussgrößen der jeweiligen Stichproben der Auswertepochen 2004 bis 2014 nach den theoretischen Ausführungen in Abschnitt 4.2.2.2 (S. 90 ff.) ermittelt werden können. Die Korrelationen basieren ausschließlich auf den Berechnungen von Spearmanschen Korrelationskoeffizienten, da sich lediglich Einflussgrößen vom stetigen Datentyp im Auswertansatz befinden.

Tabelle A.3.: Korrelationen zwischen den Einflussgrößen (Epoche 2004)

Einflussgröße	BRW	GRF	ALT	WOF	KEL
BRW	1,00	0,00	0,07	0,16	-0,03
GRF	0,00	1,00	0,31	0,32	0,12
ALT	0,07	0,31	1,00	-0,08	0,29
WOF	0,16	0,32	-0,08	1,00	-0,02
KEL	-0,03	0,12	0,29	-0,02	1,00

Tabelle A.4.: Korrelationen zwischen den Einflussgrößen (Epoche 2005)

Einflussgröße	BRW	GRF	ALT	WOF	KEL
BRW	1,00	0,02	0,19	0,11	0,20
GRF	0,02	1,00	0,32	0,09	-0,04
ALT	0,19	0,32	1,00	-0,30	0,42
WOF	0,15	0,09	-0,30	1,00	-0,08
KEL	0,20	-0,04	0,42	-0,08	1,00

Tabelle A.5.: Korrelationen zwischen den Einflussgrößen (Epoche 2006)

Einflussgröße	BRW	GRF	ALT	WOF	KEL
BRW	1,00	0,20	0,07	0,20	0,18
GRF	0,20	1,00	0,38	0,25	0,28
ALT	0,07	0,38	1,00	-0,25	0,36
WOF	0,20	0,25	-0,25	1,00	0,04
KEL	0,18	0,28	0,36	0,04	1,00

Tabelle A.6.: Korrelationen zwischen den Einflussgrößen (Epoche 2007)

Einflussgröße	BRW	GRF	ALT	WOF	KEL
BRW	1,00	0,09	0,09	0,29	-0,07
GRF	0,09	1,00	0,28	0,24	0,16
ALT	0,09	0,28	1,00	-0,09	0,18
WOF	0,29	0,24	-0,09	1,00	0,07
KEL	-0,07	0,16	0,18	0,07	1,00

Tabelle A.7.: Korrelationen zwischen den Einflussgrößen (Epoche 2008)

Einflussgröße	BRW	GRF	ALT	WOF	KEL
BRW	1,00	0,03	0,19	-0,14	0,13
GRF	0,03	1,00	0,32	0,35	0,11
ALT	0,19	0,32	1,00	0,06	0,37
WOF	-0,14	0,35	0,06	1,00	-0,07
KEL	0,13	0,11	0,37	-0,07	1,00

Tabelle A.8.: Korrelationen zwischen den Einflussgrößen (Epoche 2009)

Einflussgröße	BRW	GRF	ALT	WOF	
BRW	1,00	-0,21	0,10	-0,04	-0,01
GRF	-0,21	1,00	0,03	0,29	-0,03
ALT	0,10	0,03	1,00	-0,33	0,32
WOF	-0,04	0,29	-0,33	1,00	-0,04
KEL	-0,01	-0,03	0,32	-0,04	1,00

Tabelle A.9.: Korrelationen zwischen den Einflussgrößen (Epoche 2010)

Einflussgröße	BRW	GRF	ALT	WOF	KEL
BRW	1,00	-0,14	0,09	0,18	0,10
GRF	-0,14	1,00	0,31	0,21	0,17
ALT	0,09	0,31	1,00	-0,28	0,30
WOF	0,18	0,21	-0,28	1,00	0,12
KEL	0,10	0,17	0,30	0,12	1,00

Tabelle A.10.: Korrelationen zwischen den Einflussgrößen (Epoche 2011)

Einflussgröße	BRW	GRF	ALT	WOF	
BRW	1,00	-0,03	0,13	0,14	-0,03
GRF	-0,03	1,00	0,33	0,30	0,22
ALT	0,13	0,33	1,00	-0,16	0,33
WOF	0,14	0,30	-0,16	1,00	0,05
KEL	-0,03	0,22	0,33	0,05	1,00

Tabelle A.11.: Korrelationen zwischen den Einflussgrößen (Epoche 2012)

Einflussgröße	BRW	GRF	ALT	WOF	
BRW	1,00	-0,07	0,12	0,17	0,12
GRF	-0,07	1,00	0,39	0,10	0,07
ALT	0,12	0,39	1,00	-0,06	0,25
WOF	0,17	0,10	-0,06	1,00	0,05
KEL	0,12	0,07	0,25	0,05	1,00

Tabelle A.12.: Korrelationen zwischen den Einflussgrößen (Epoche 2013)

Einflussgröße	BRW	GRF	ALT	WOF	
BRW	1,00	-0,15	0,13	0,14	0,09
GRF	-0,15	1,00	0,39	0,20	0,23
ALT	0,13	0,39	1,00	-0,15	0,42
WOF	0,14	0,20	-0,15	1,00	-0,06
KEL	0,09	0,23	0,42	-0,06	1,00

Tabelle A.13.: Korrelationen zwischen den Einflussgrößen (Epoche 2014)

Einflussgröße	BRW	GRF	ALT	WOF	
BRW	1,00	-0,38	-0,00	0,13	0,17
GRF	-0,38	1,00	0,30	0,28	0,12
ALT	-0,00	0,30	1,00	-0,17	0,40
WOF	0,13	0,28	-0,17	1,00	-0,14
KEL	0,17	0,12	0,40	-0,14	1,00

B. Anhang – Ergebnisse der Bayesschen Regression

Die folgenden Grafiken fassen die Ergebnisse für das nach Kapitel 5 (S. 109 ff.) beschriebene Fallbeispiel zusammen. Dargestellt werden in diesem Anhang ausschließlich die Ergebnisse der Bayesschen Regressionen über die Auswertepochen 2005 bis 2014. Für weitere Interpretationen zu den Grafiken wird entsprechend auf die textlichen Ausführungen verwiesen. In den Unterabschnitten werden die Posteriori-Schätzungen der Zielgröße $\bar{y}$ mit den zugehörigen HPDIs, die prozentualen Abweichungen zwischen den Kaufpreisbeobachtungen und den Schätzwerten sowie die Ergebnisse der Prädiktion und deren Unsicherheit aufgeführt.

B.1. Posteriori-Schätzung der Zielgröße $\bar{y}$ und HPDIs

Die Grafiken B.1 bis B.10 zeigen die Posteriori-Schätzungen der Zielgröße $\bar{y}$ für die einzelnen Auswertepochen sowie die zugehörigen HPDIs der Schätzwerte. In den oberen Abbildungen werden jeweils die Ergebnisse der informativen Bayesschen Regression *R.1* (links) und der nichtinformativen Bayesschen Regression *R.2* (rechts) gezeigt. Die Darstellung orientiert sich – wie auch im Zusammenhang mit dem Fallbeispiel nach Kapitel 5 beschrieben – an der nach ihrer Größe aufsteigenden Sortierung der realen Kaufpreisbeobachtungen und der entsprechenden Zuordnung der Schätzwerte $\bar{y}$. In der unteren Abbildung werden die Differenzen der HPDIs dargestellt, die sich aus dem Verhältnis der nichtinformativen Lösung *R.2* zur informativen Lösung *R.1* ergeben.

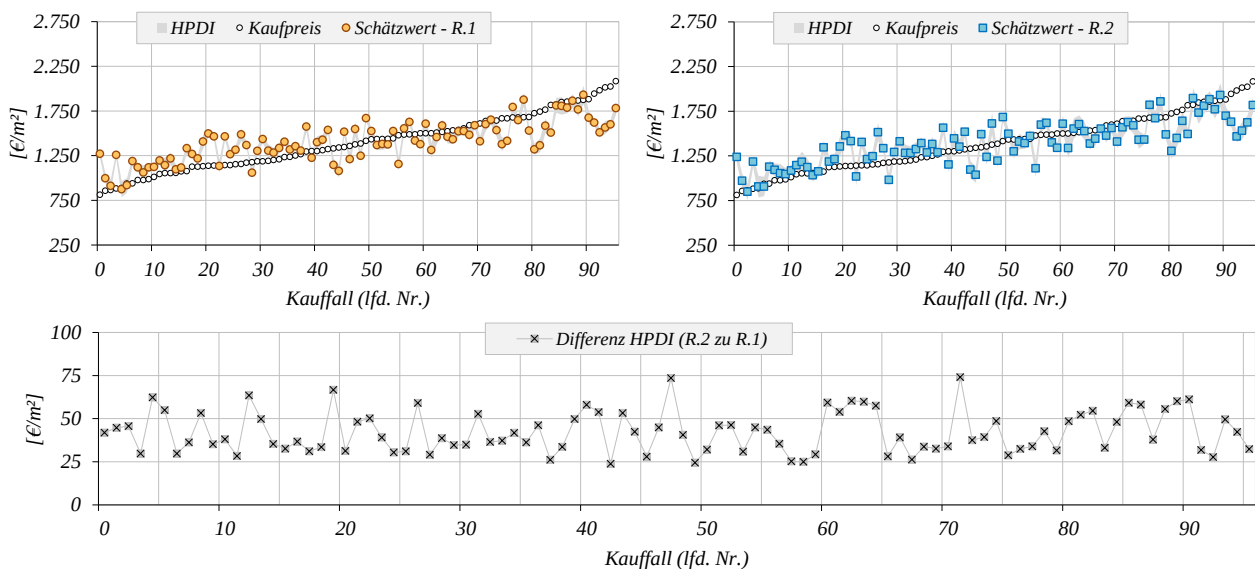


Abbildung B.1: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

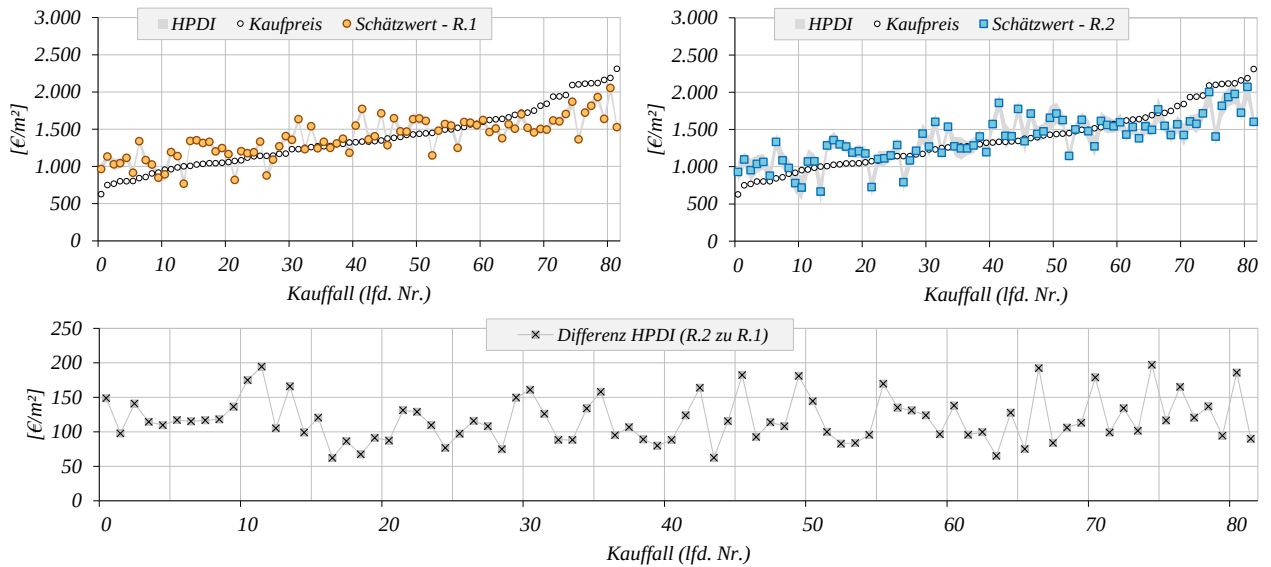


Abbildung B.2: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2006 (oben); Differenzplots der HPDIs (unten)

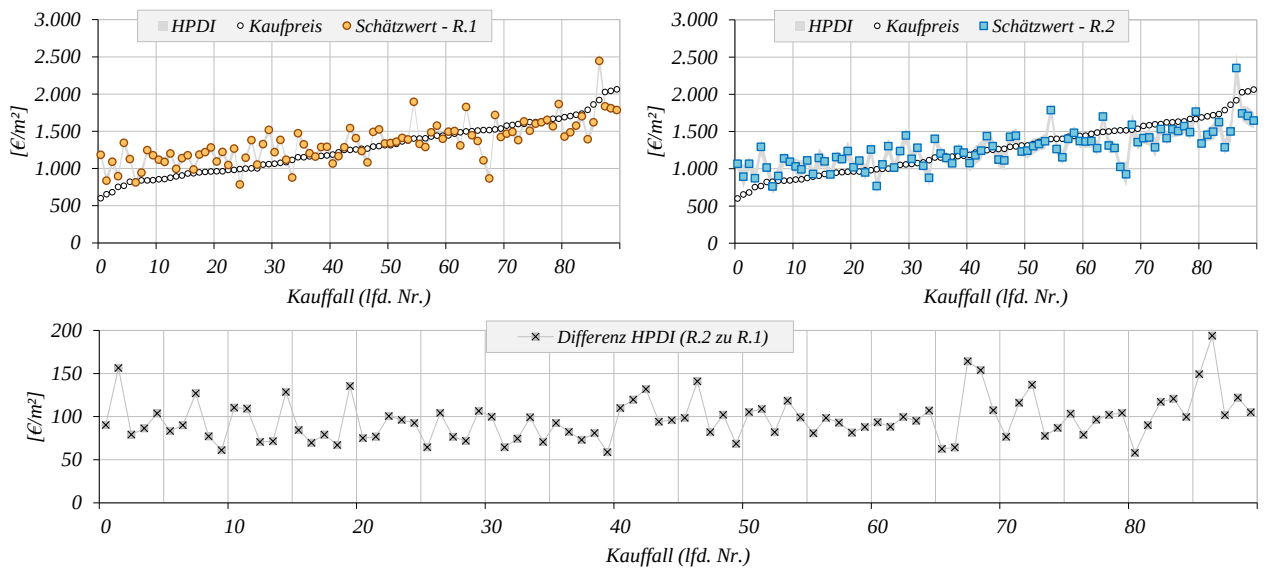


Abbildung B.3: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2007 (oben); Differenzplots der HPDIs (unten)

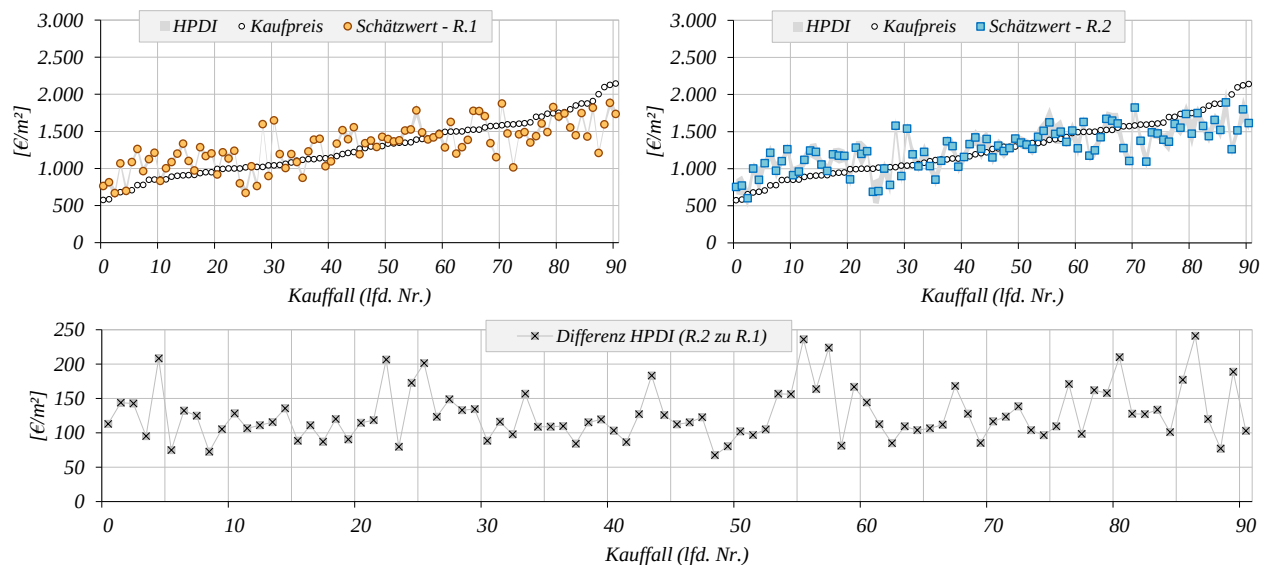
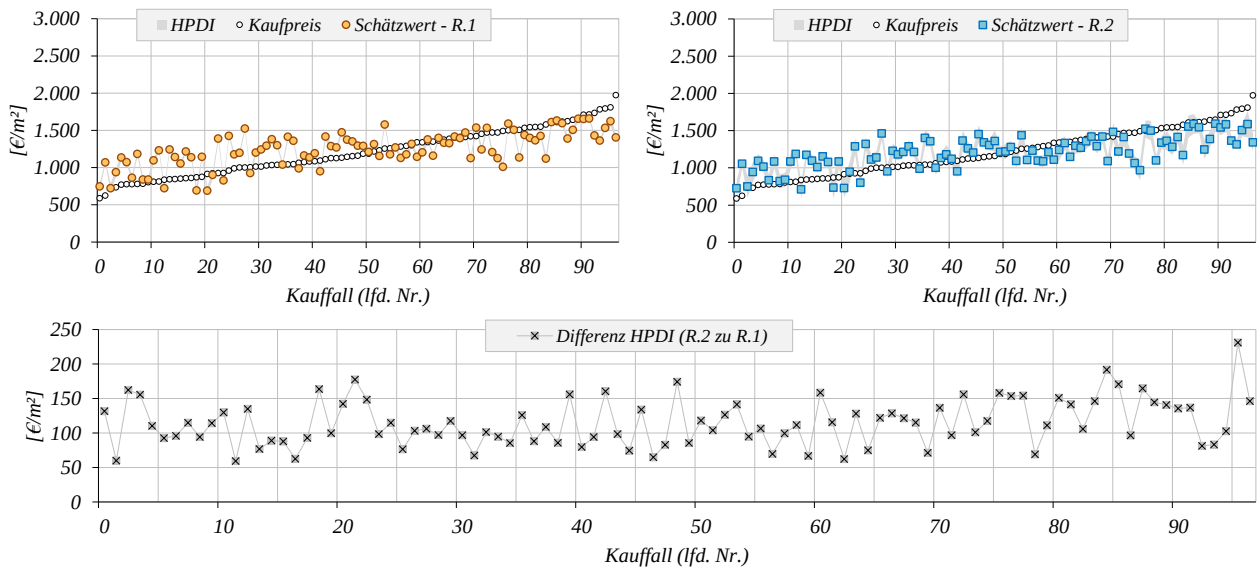
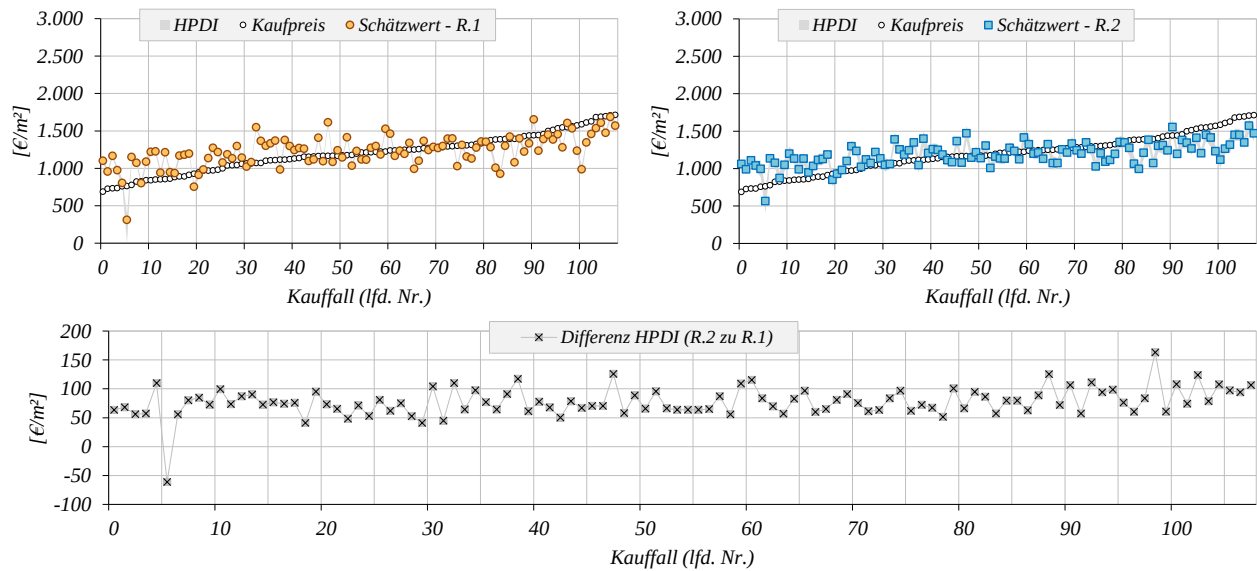
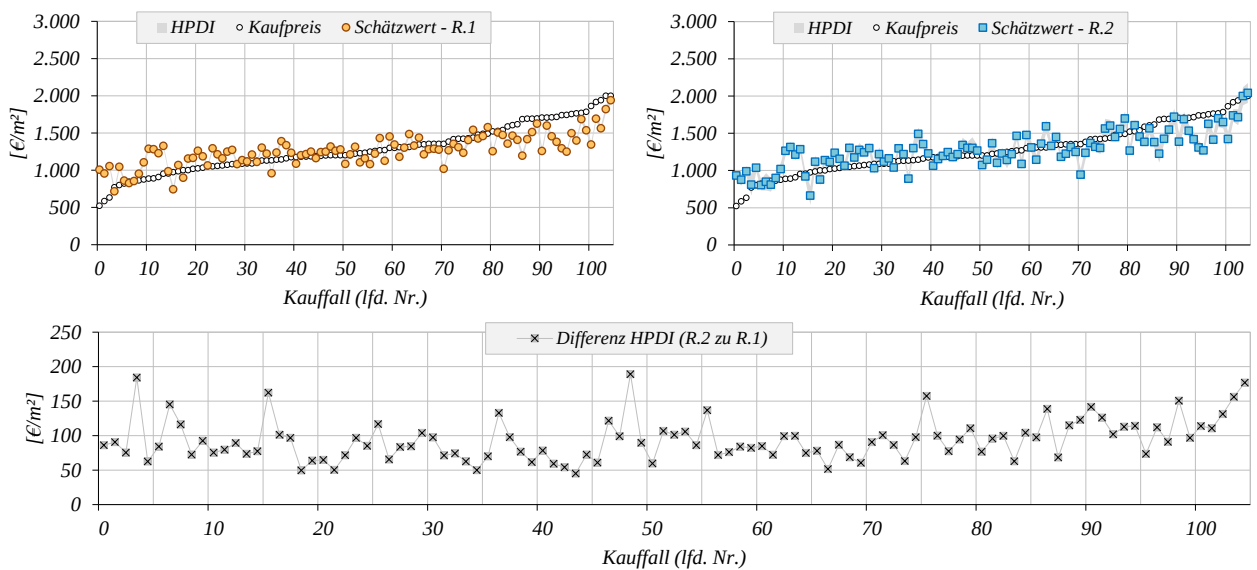


Abbildung B.4: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2008 (oben); Differenzplots der HPDIs (unten)

Abbildung B.5: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2009 (oben); Differenzplots der HPDIs (unten)Abbildung B.6: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2010 (oben); Differenzplots der HPDIs (unten)Abbildung B.7: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2011 (oben); Differenzplots der HPDIs (unten)

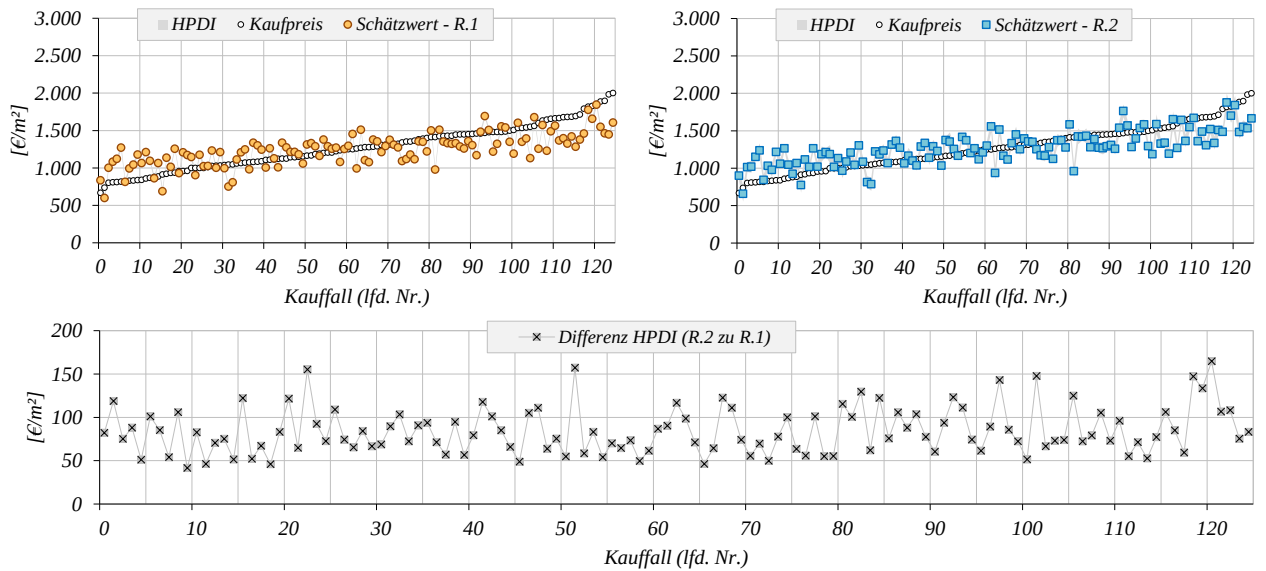


Abbildung B.8: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2012 (oben); Differenzplots der HPDIs (unten)

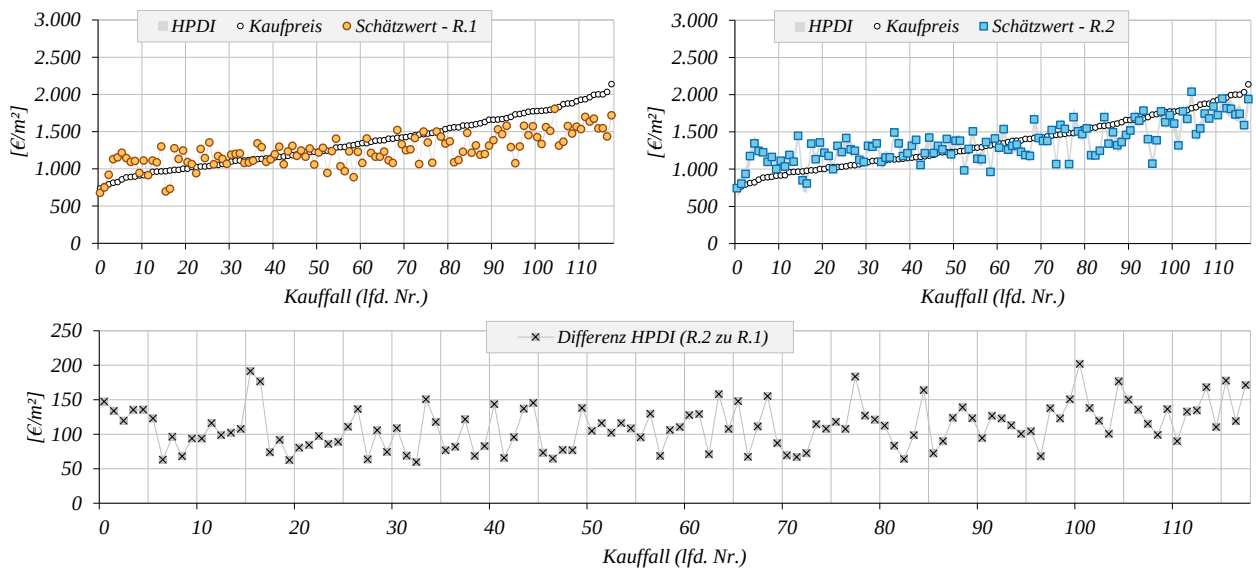


Abbildung B.9: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2013 (oben); Differenzplots der HPDIs (unten)

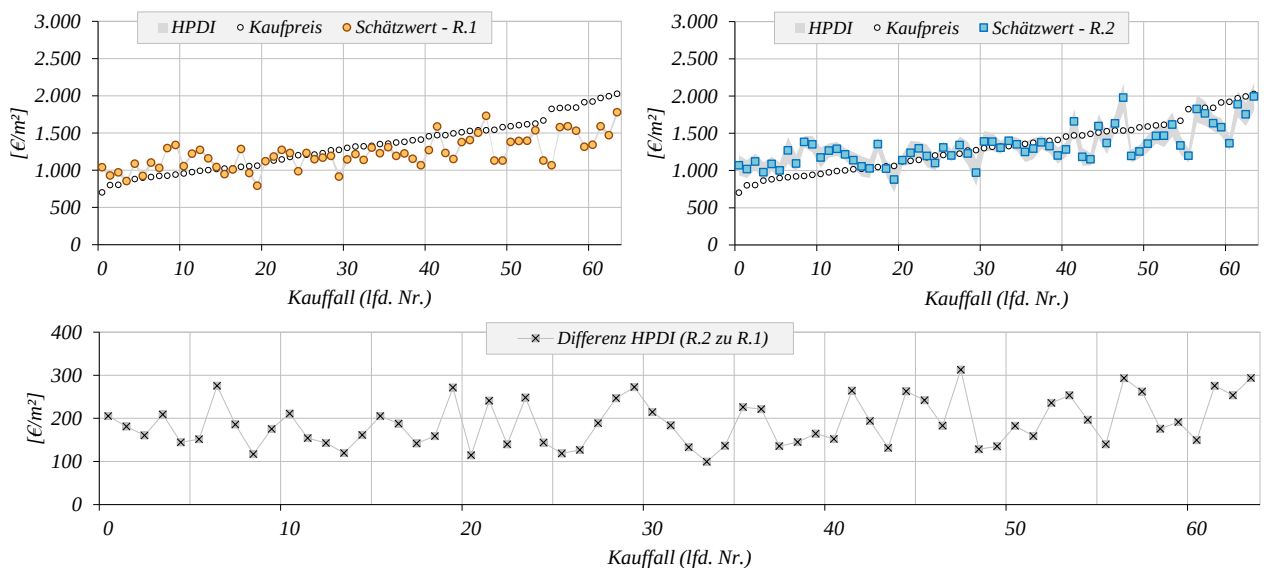


Abbildung B.10: Schätzwerte $\bar{y}$ der Bayesschen Regression, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

B.2. Abweichungen zwischen Kaufpreisen und Schätzwerten $\bar{y}$

In den Abbildungen B.11 bis B.20 werden die prozentualen Abweichungen zwischen den originären Kaufpreisen und deren Schätzwerten der Bayesschen Regression gegenübergestellt. In der jeweils linken Abbildung finden sich die Ergebnisse der informativen Lösungen *R.1*, in den jeweils rechten Abbildungen die der nichtinformativen Lösungen *R.2*.

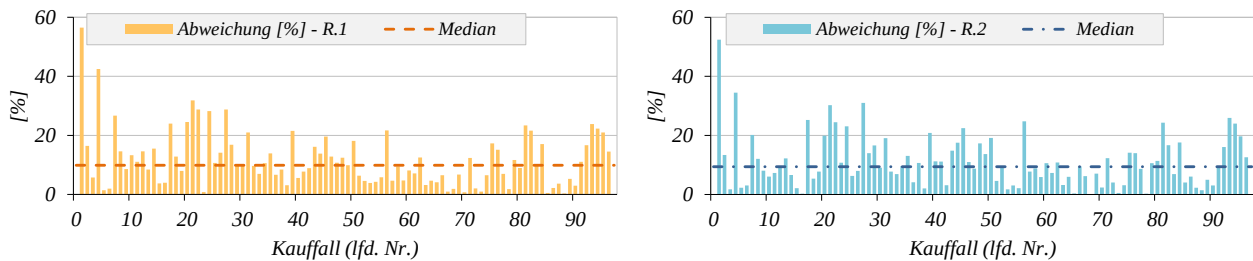


Abbildung B.11: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2005

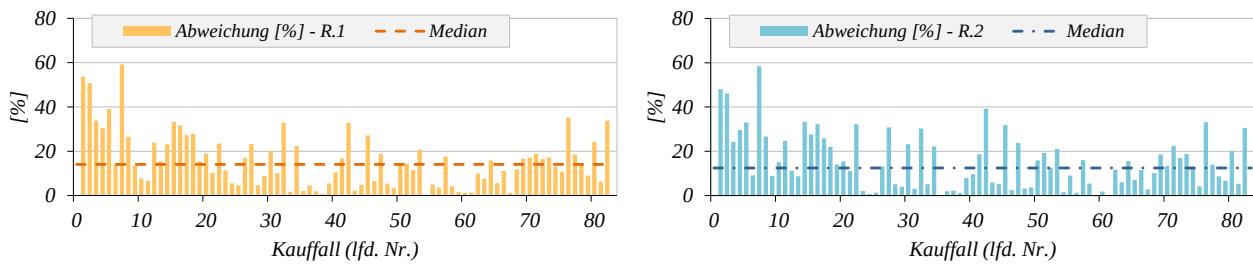


Abbildung B.12: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2006

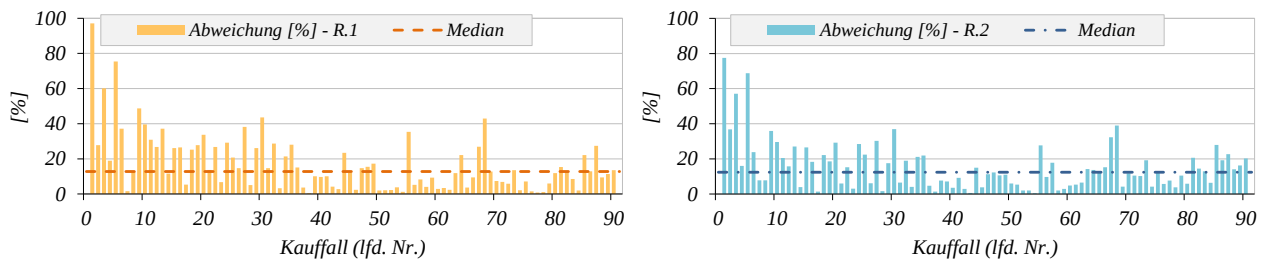


Abbildung B.13: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2007

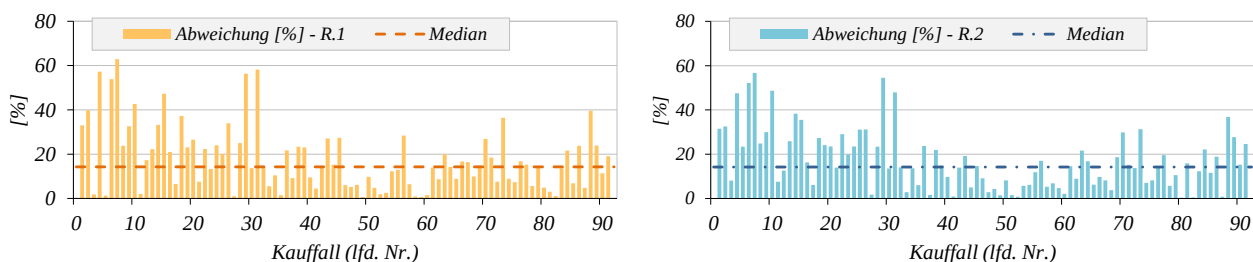


Abbildung B.14: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2008

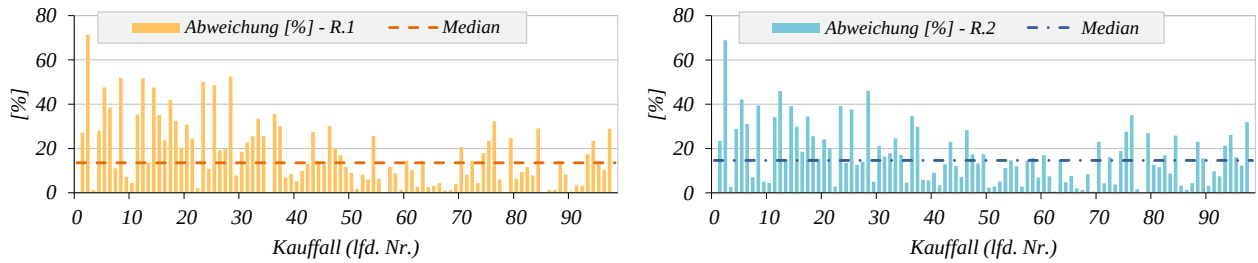


Abbildung B.15: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2009

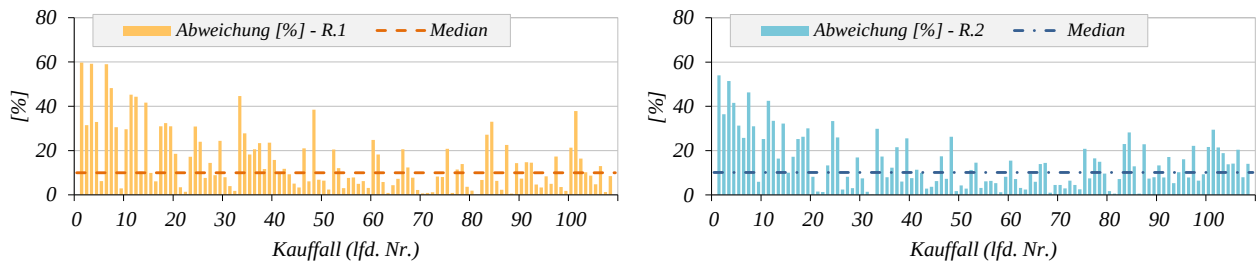


Abbildung B.16: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2010

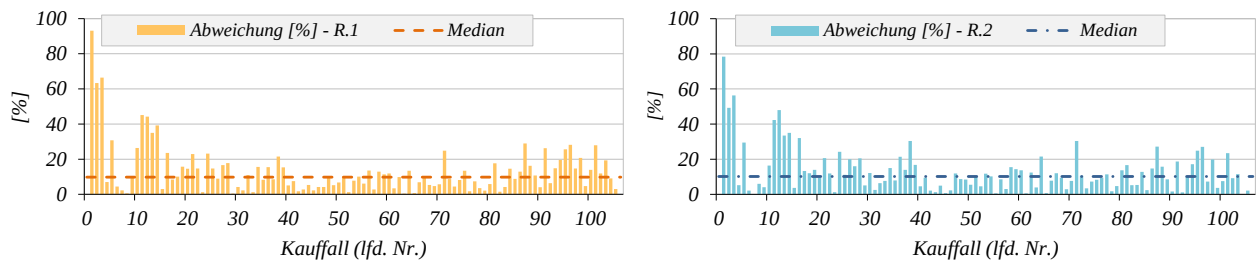


Abbildung B.17: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2011

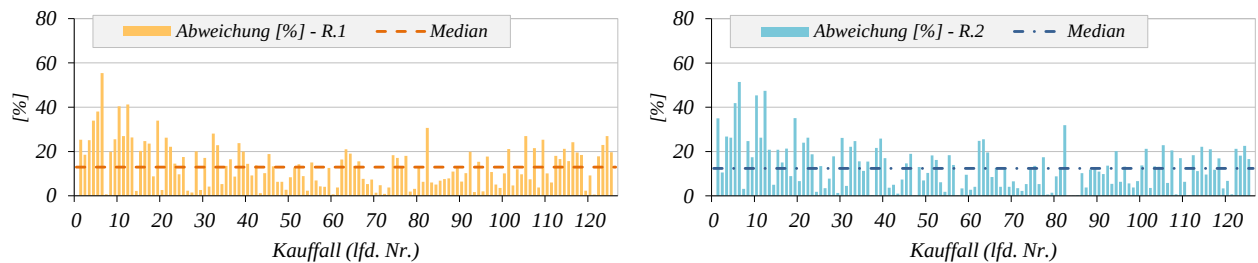
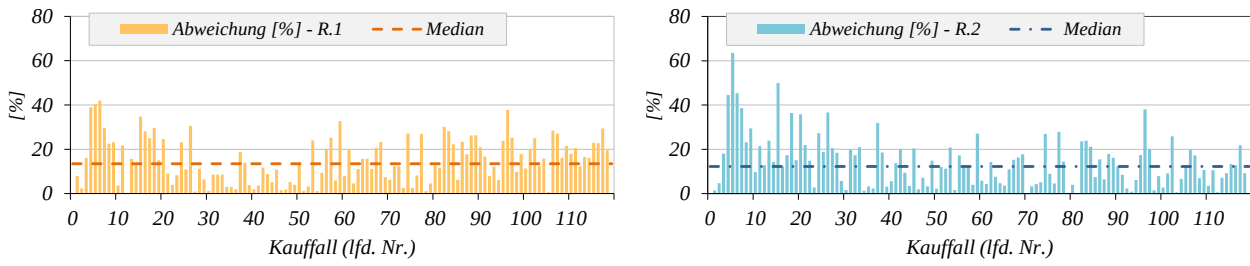
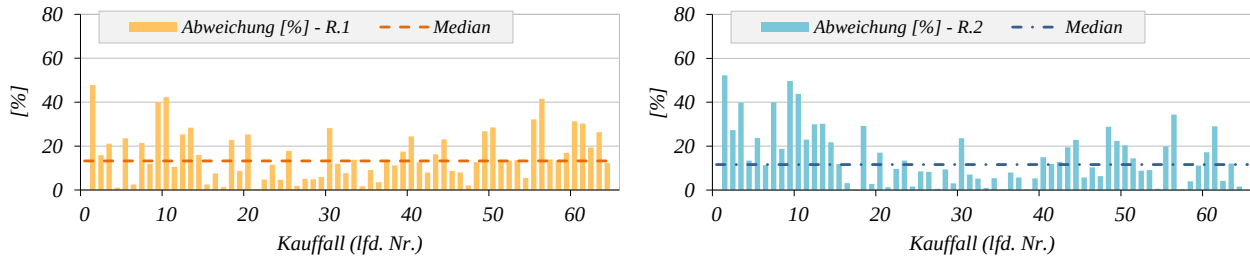
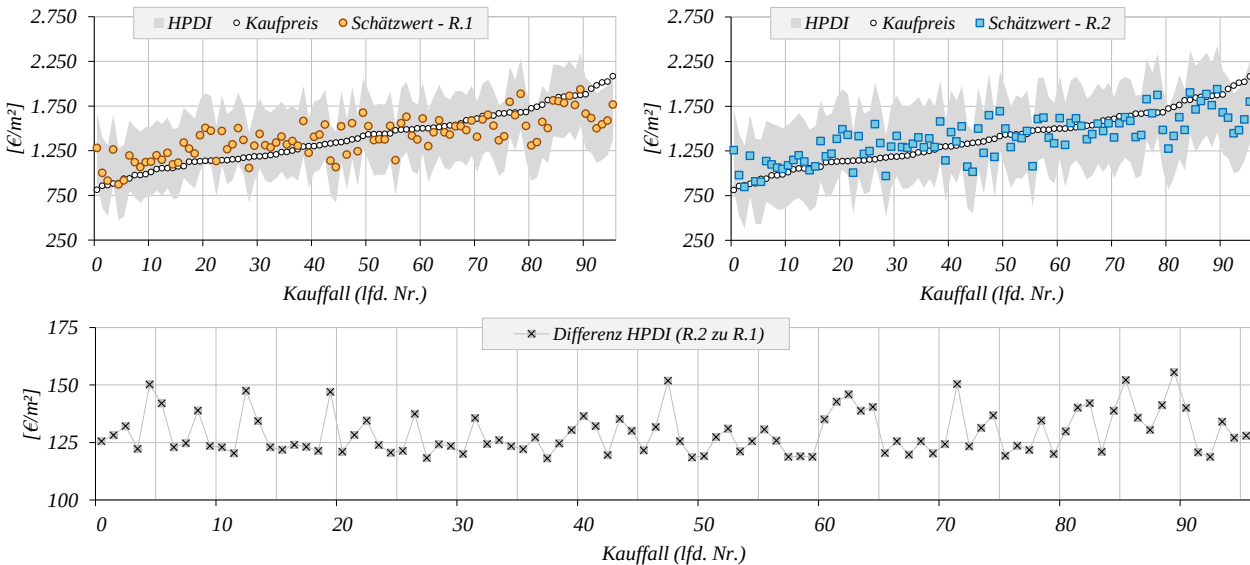


Abbildung B.18: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2012

Abbildung B.19: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2013Abbildung B.20: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Regression, Epoche 2014

B.3. Prädizierte Posteriori-Schätzung der Zielgröße $\bar{y}_p$ und HPDIs

Die Grafiken B.21 bis B.30 zeigen die Ergebnisse der Prädiktion von Vergleichswerten im Ansatz der Kreuzvalidierung und basierend auf den Daten des im textlichen Teil behandelten Fallbeispiels. Wie im Kontext der Darstellung der geschätzten Vergleichswerte $\bar{y}$ erläutert, werden in der jeweiligen oberen linken Abbildung die Ergebnisse der informativen Bayesschen Regression (*R.1*) dargestellt, in der oberen entsprechend die der nichtinformativen Bayesschen Regression (*R.2*). Wiederum erfolgt die Sortierung nach aufsteigenden Kaufpreisen. Unterhalb der genannten Grafiken werden die Differenzplots für die Unsicherheitsspannen zwischen informativer und nichtinformativer Lösung gegeben.

Abbildung B.21: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

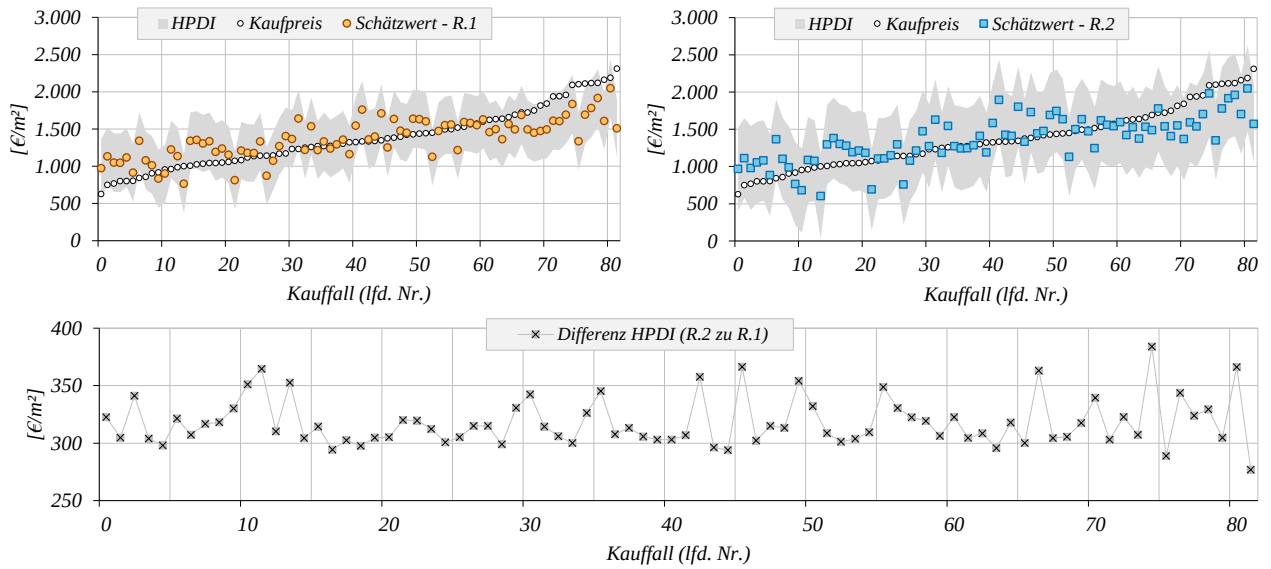


Abbildung B.22: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2006 (oben); Differenzplots der HPDIs (unten)

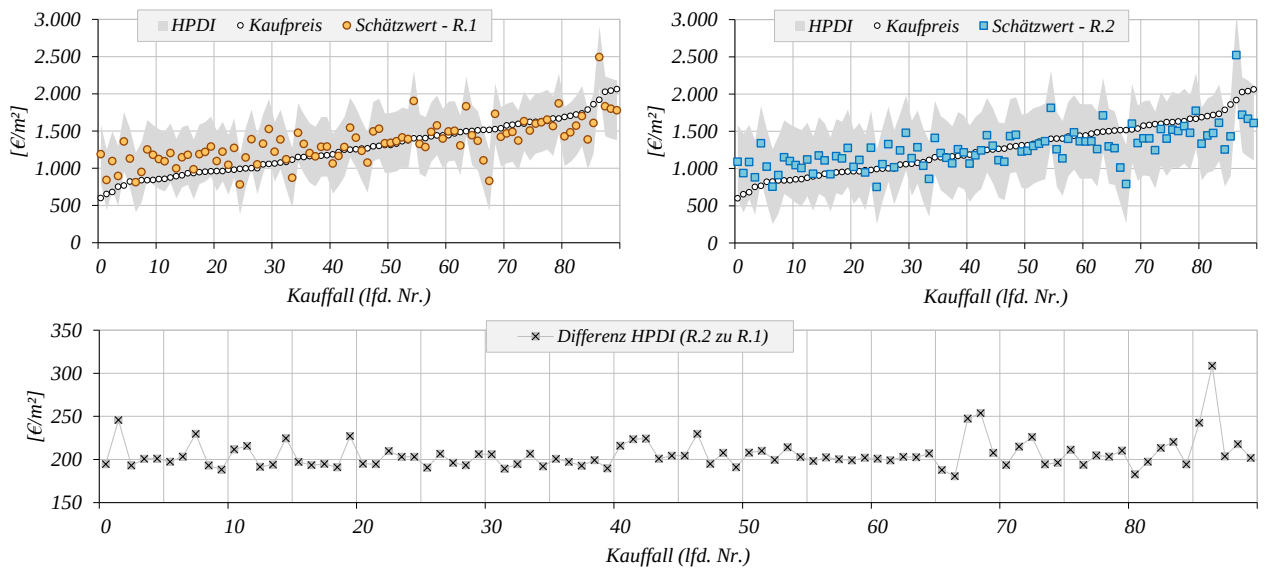


Abbildung B.23: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2007 (oben); Differenzplots der HPDIs (unten)

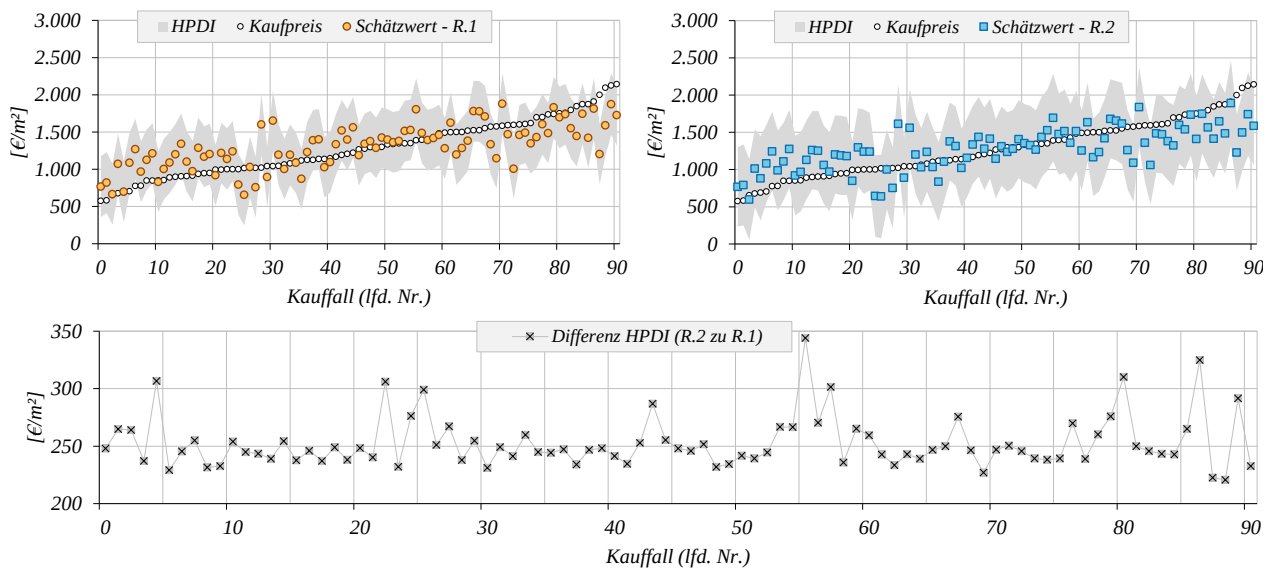
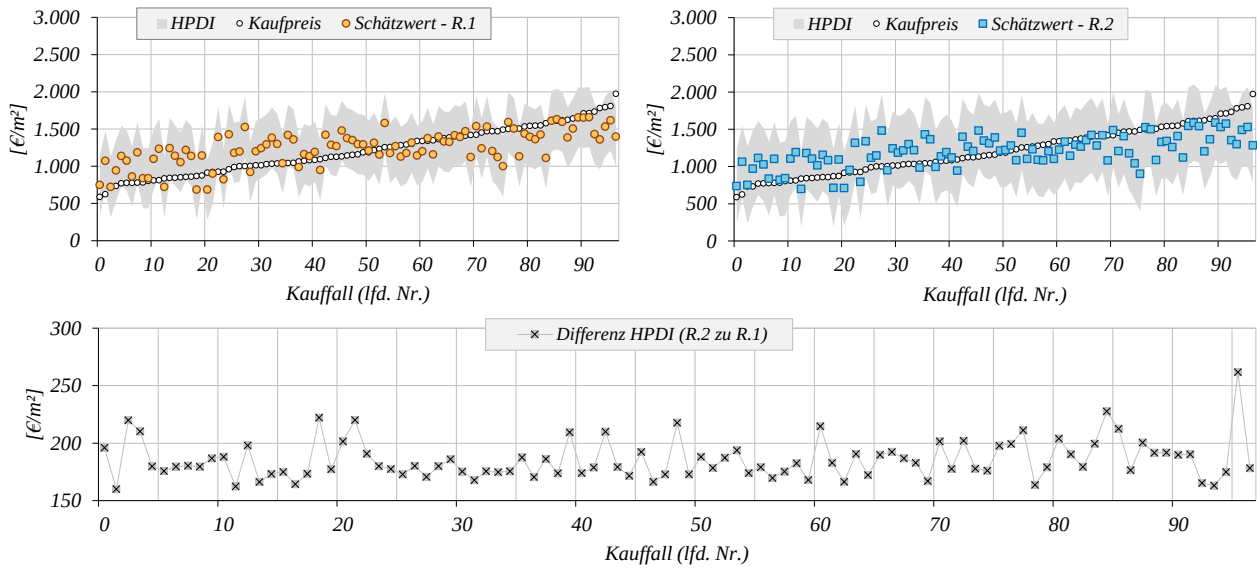
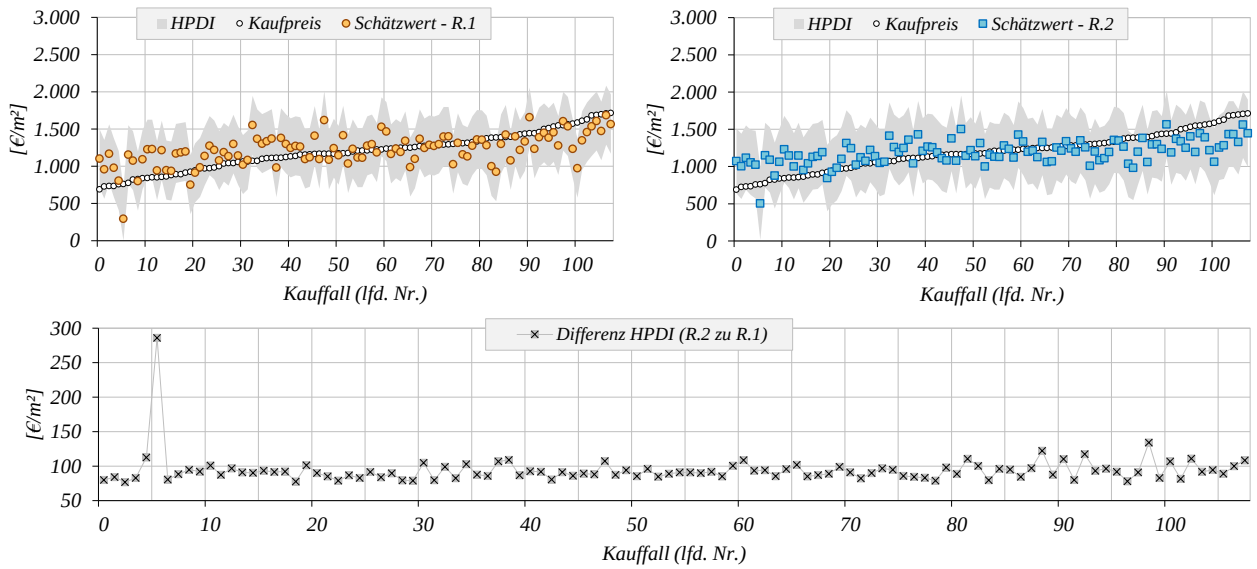
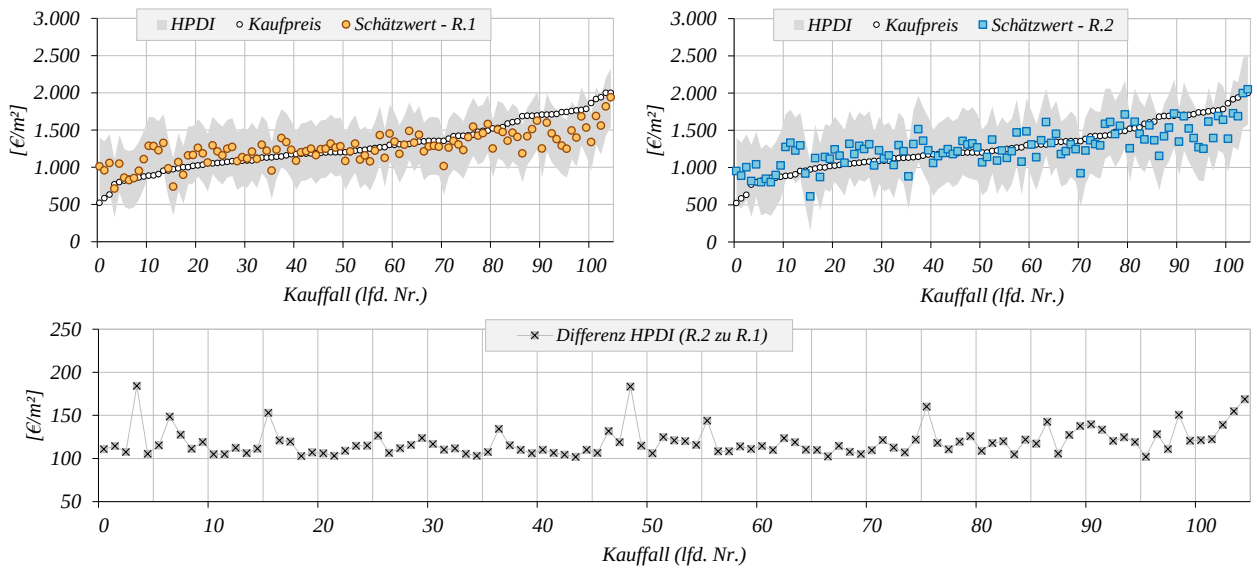


Abbildung B.24: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2008 (oben); Differenzplots der HPDIs (unten)

Abbildung B.25: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2009 (oben); Differenzplots der HPDIs (unten)Abbildung B.26: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2010 (oben); Differenzplots der HPDIs (unten)Abbildung B.27: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2011 (oben); Differenzplots der HPDIs (unten)

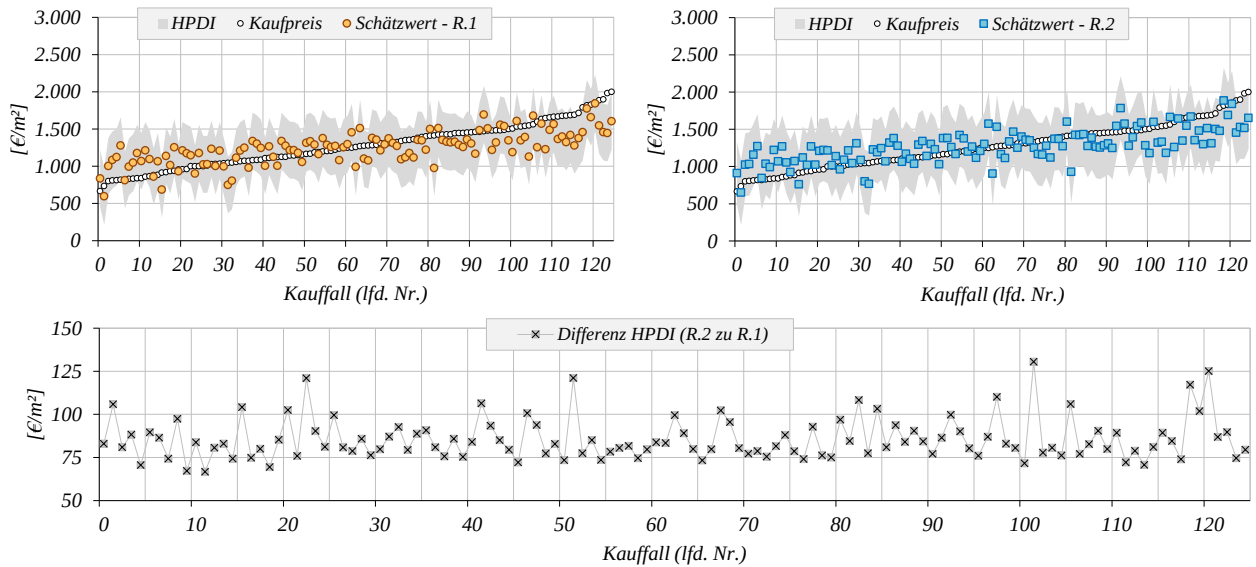


Abbildung B.28: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2012 (oben); Differenzplots der HPDIs (unten)

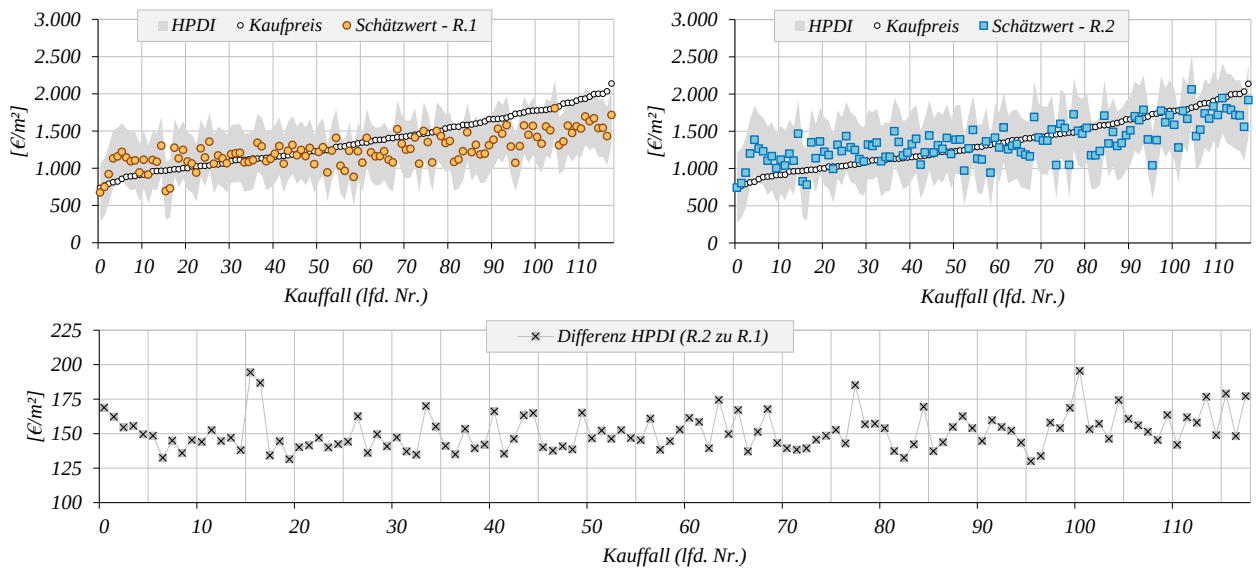


Abbildung B.29: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2013 (oben); Differenzplots der HPDIs (unten)

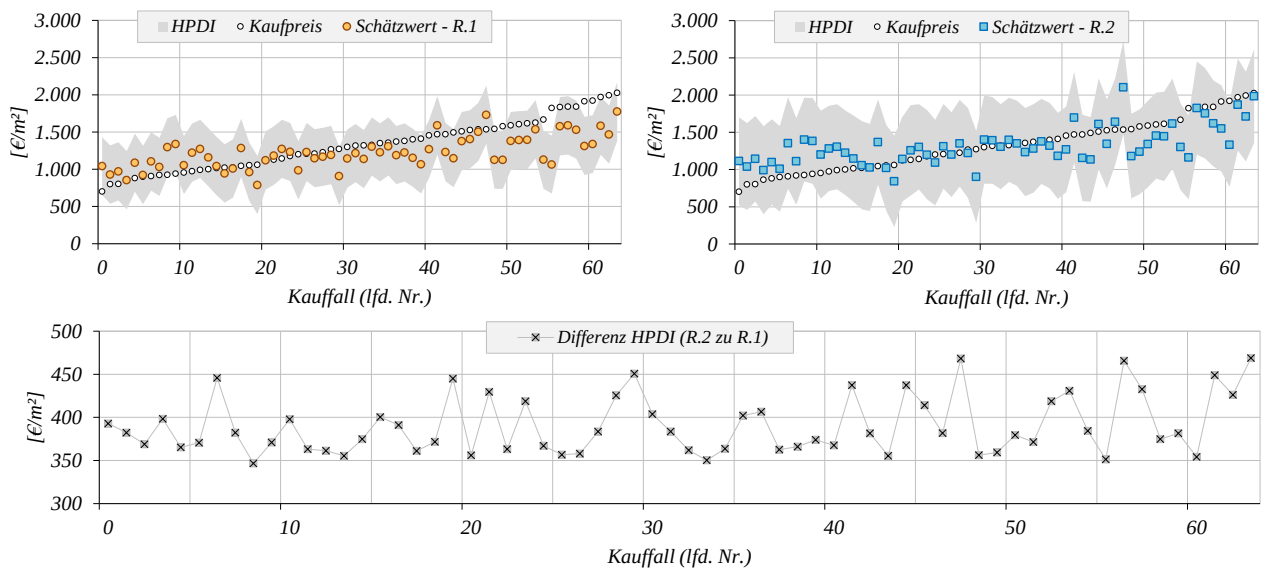


Abbildung B.30: Schätzwerte $\bar{y}_p$ der Bayesschen Regression, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

C. Anhang – Ergebnisse der Bayesschen Kollokation

Die Inhalte des folgenden Anhangs zeigen die Ergebnisse, die für das nach Kapitel 5 (S. 109 ff.) beschriebene Fallbeispiel erhalten werden. Wiederum wird für Beschreibungen und Interpretationen der Grafiken auf die Ausführungen im textlichen Teil der Arbeit verwiesen. Die Unterabschnitte des Anhangs zu den Ergebnissen der Bayesschen Kollokation enthalten die Abbildungen zur Schätzung der Autokorrelationsfunktionen, den Posteriori-Schätzungen der Signalkomponente $\bar{s}$, der Zielgröße $\bar{y}$, den jeweiligen Prädiktionsergebnisse $\bar{s}_p$ und $\bar{s}_p$ sowie die zugehörigen HPDIs. Zudem werden die prozentualen Abweichungen zwischen den Kaufpreisbeobachtungen und deren Schätzwerten gegeben. Abschließend erfolgt eine Darstellung der absoluten Residuen der Bayesschen Modellansätze für Schätzung und Prädiktion.

C.1. Schätzung der Autokorrelationsfunktionen

Entsprechend den Ausführungen des textlichen Teils in Kapitel 5 (S. 109 ff.) zeigen die folgenden Abbildungen eine Gegenüberstellung von empirischen und analytischen Autokorrelationsfunktionen, die auf Grundlage der Daten des Fallbeispiels erhalten werden. Die jeweils linken Grafiken zeigen die Ergebnisse basierend auf informativen Lösungen (K.1), die rechten der nichtinformativen Lösungen (K.2).

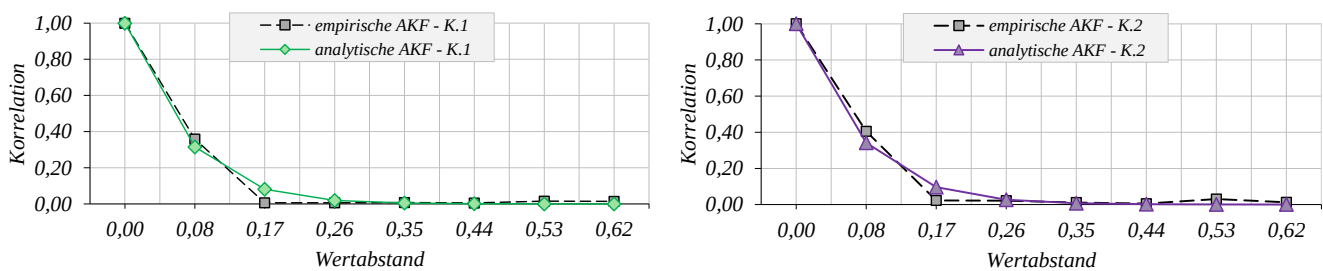


Abbildung C.1: Empirische und analytische Autokorrelationsfunktionen der Epoche 2005

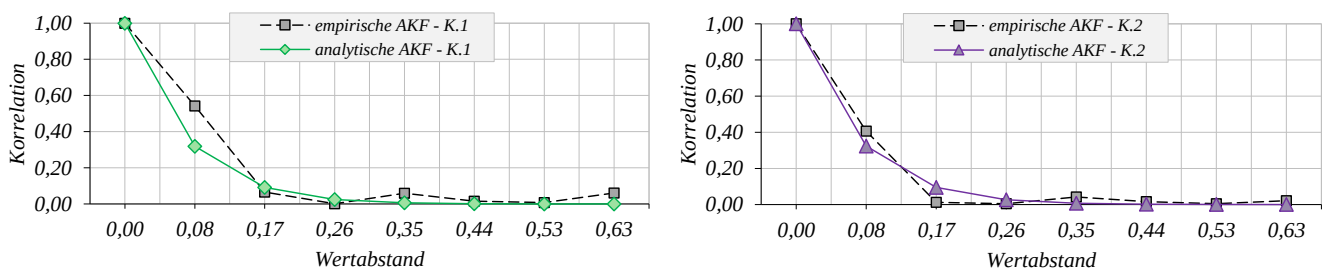


Abbildung C.2: Empirische und analytische Autokorrelationsfunktionen der Epoche 2006

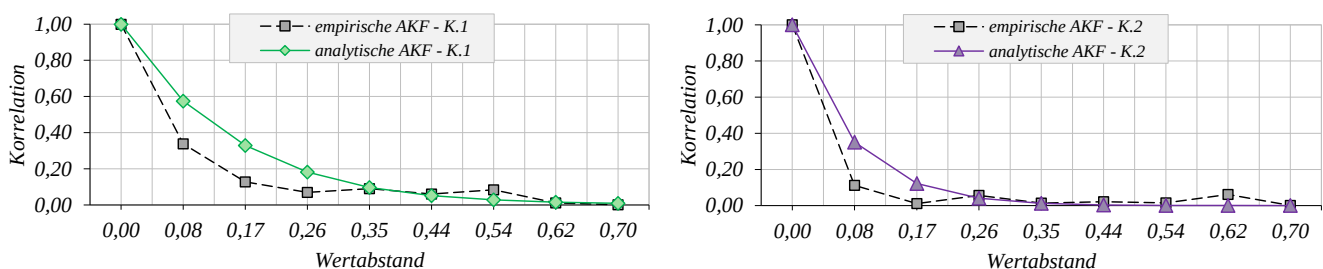


Abbildung C.3: Empirische und analytische Autokorrelationsfunktionen der Epoche 2007

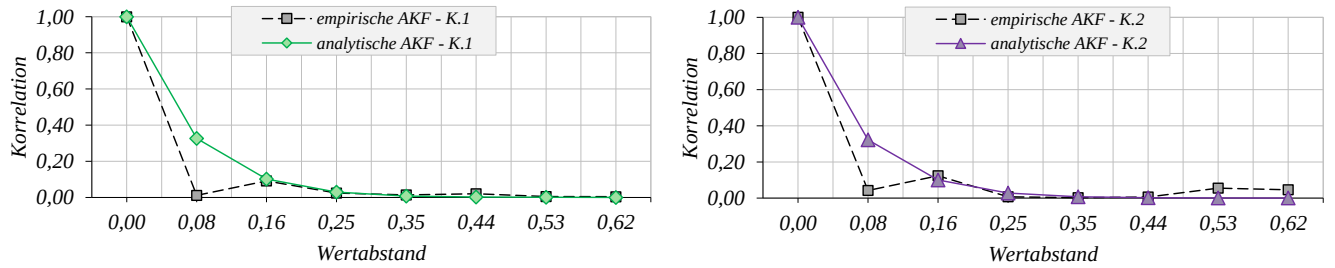


Abbildung C.4: Empirische und analytische Autokorrelationsfunktionen der Epoche 2008

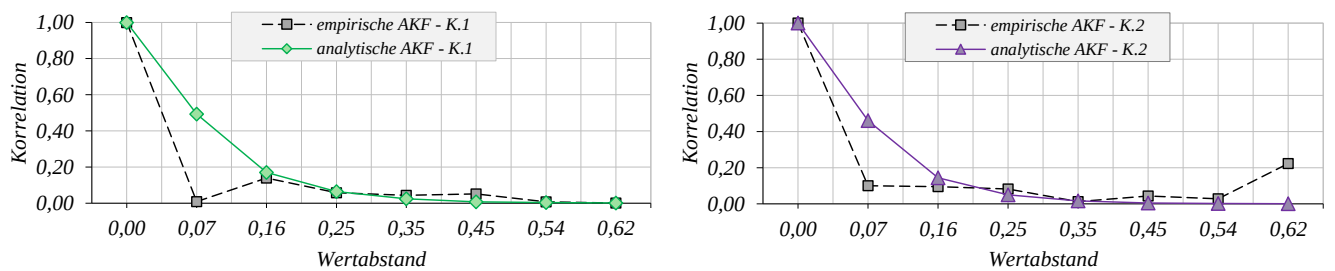


Abbildung C.5: Empirische und analytische Autokorrelationsfunktionen der Epoche 2009

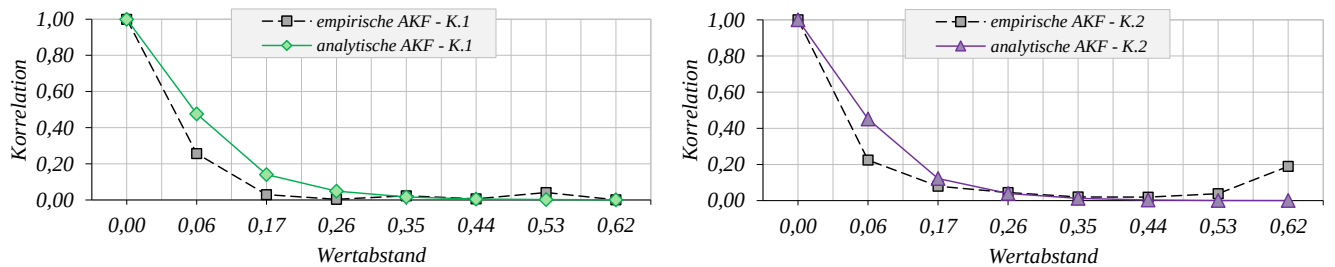


Abbildung C.6: Empirische und analytische Autokorrelationsfunktionen der Epoche 2010

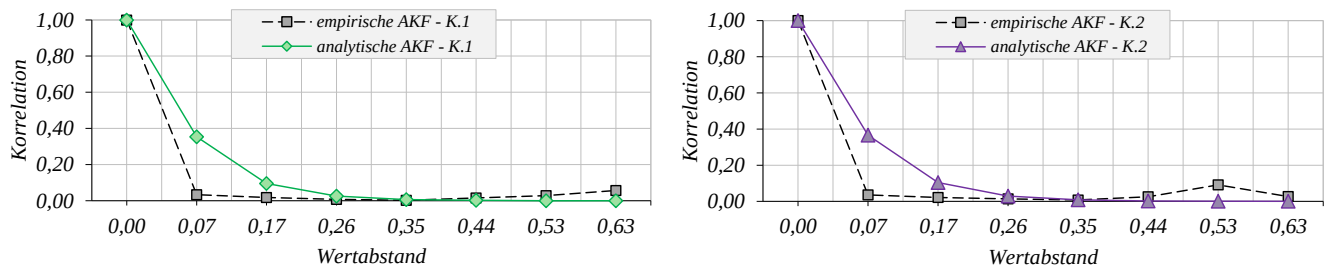


Abbildung C.7: Empirische und analytische Autokorrelationsfunktionen der Epoche 2011

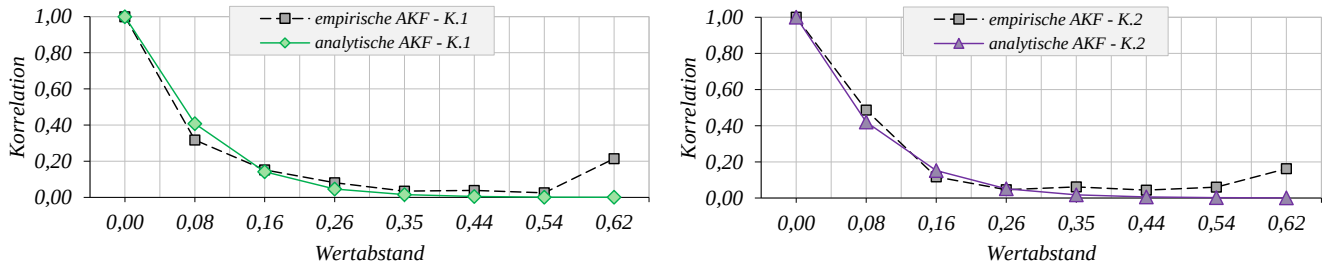


Abbildung C.8: Empirische und analytische Autokorrelationsfunktionen der Epoche 2012

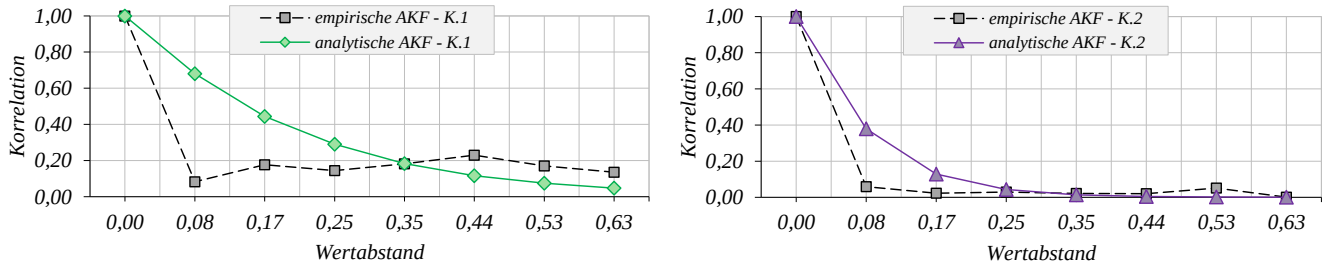


Abbildung C.9: Empirische und analytische Autokorrelationsfunktionen der Epoche 2013

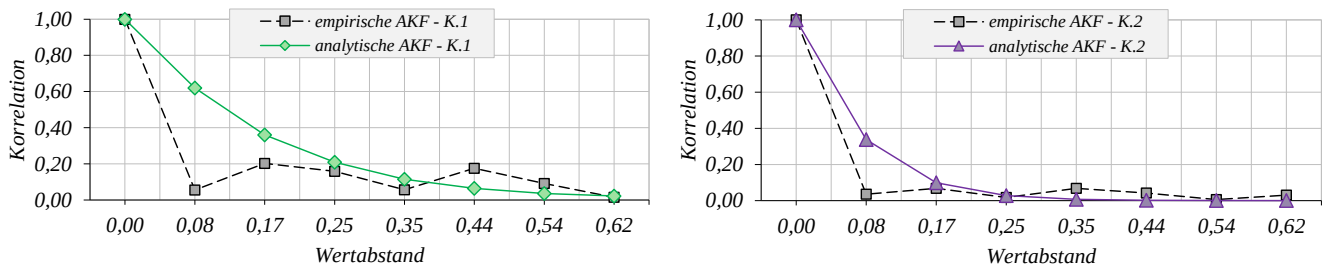


Abbildung C.10: Empirische und analytische Autokorrelationsfunktionen der Epoche 2014

C.2. Posteriori-Schätzung der Signalkomponente $\bar{s}$, der Zielgröße $\bar{y}$ und HPDIs

Die folgenden Abbildungen C.11 bis C.30 zeigen jeweils die Posteriori-Schätzung der Signalkomponente $\bar{s}$, der Zielgröße $\bar{y}$ sowie deren zugehörige HPDIs und die Differenzplots zwischen den Ergebnissen der unterschiedlichen Lösungsansätze. Die Ergebnisdarstellung erfolgt dabei epochenweise: In der jeweils ersten Abbildung (bspw. Abbildung C.11) werden die geschätzten Signalkomponenten dargestellt. Den oberen linken Grafiken sind die Ergebnisse zu entnehmen, die auf Grundlage der informativen Bayesschen Kollokation (*K.1*) erhalten werden. Dabei sind die Werte in der Aufreihung so dargestellt, wie sie sich aus der Sortierung der realen Kaufpreisbeobachtungen nach aufsteigender Größe ergeben. Eine direkte Verknüpfung mit der Darstellung der Posteriori-Schätzwerte $\bar{y}$ ist so direkt gegeben. Die oberen rechten Grafiken enthalten entsprechend die Ergebnisse der Schätzung im nichtinformativen Ansatz *K.2*; die Sortierung erfolgt analog. Die untere Grafik zeigt die Differenzplots der HPDIs von nichtinformativer zu informativer Lösung.

Die Abbildungen, die auf die Darstellung der geschätzten Signalkomponenten folgen (bspw. Abbildung C.12), zeigen die Ergebnisse für die Schätzung der Vergleichswerte $\bar{y}$. Wie auch im Zusammenhang mit der Darstellung der Daten der Bayesschen Regression werden in der jeweils oberen linken Grafik die Ergebnisse des informativen Ansatzes *K.1* gezeigt, in der oberen rechten die des nichtinformativen Ansatzes *K.2*. Unterhalb der beiden Grafiken findet sich die Darstellung der Differenzen zwischen den Unsicherheitsspannen für die einzelnen Schätzwerte.

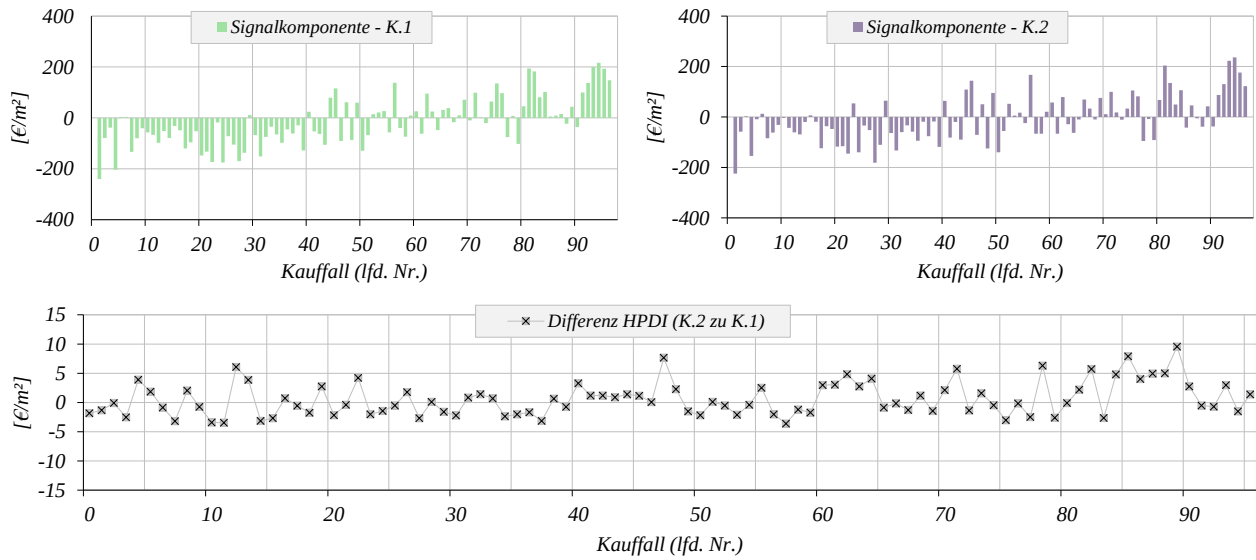


Abbildung C.11: Schätzwerte $\bar{s}$, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

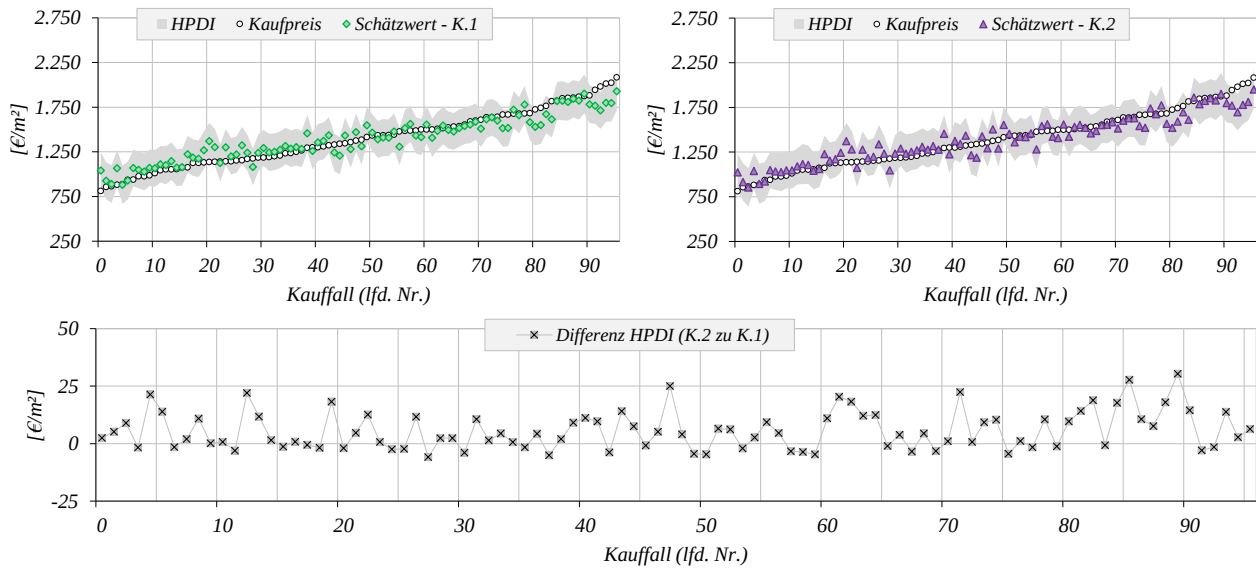
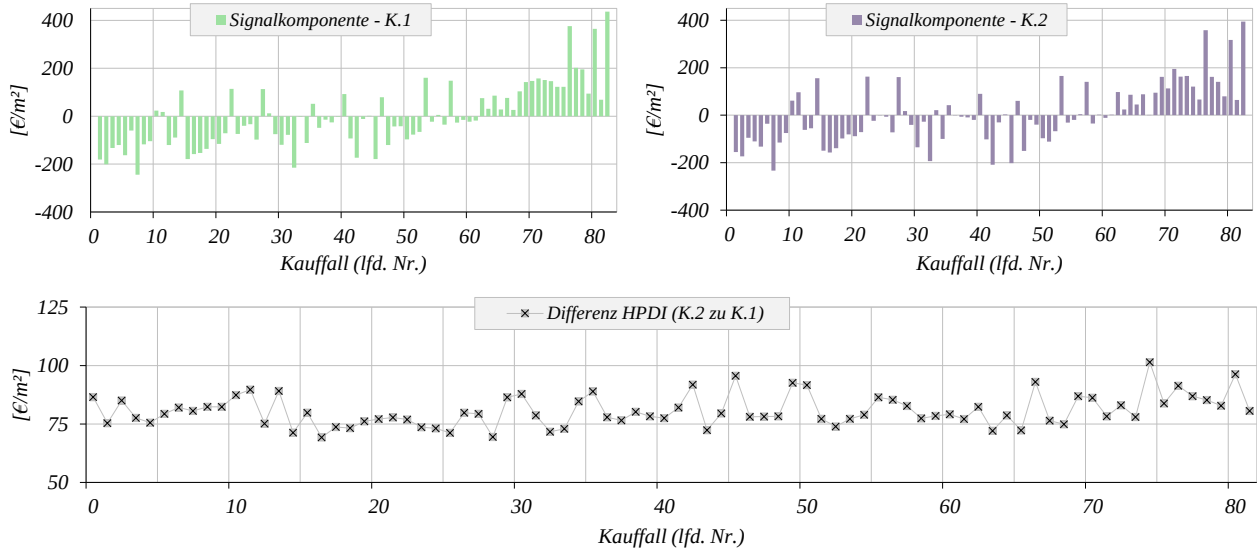
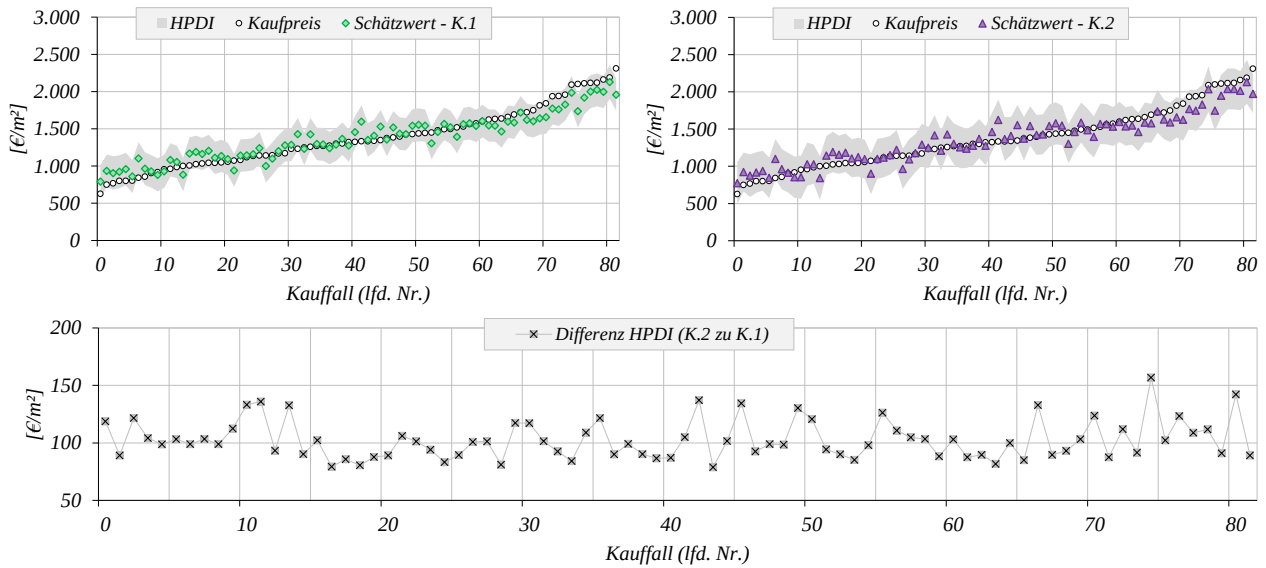


Abbildung C.12: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

Abbildung C.13: Schätzwerte $\bar{s}$, Epoche 2006 (oben); Differenzplots der HPDIs (unten)Abbildung C.14: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2006 (oben); Differenzplots der HPDIs (unten)

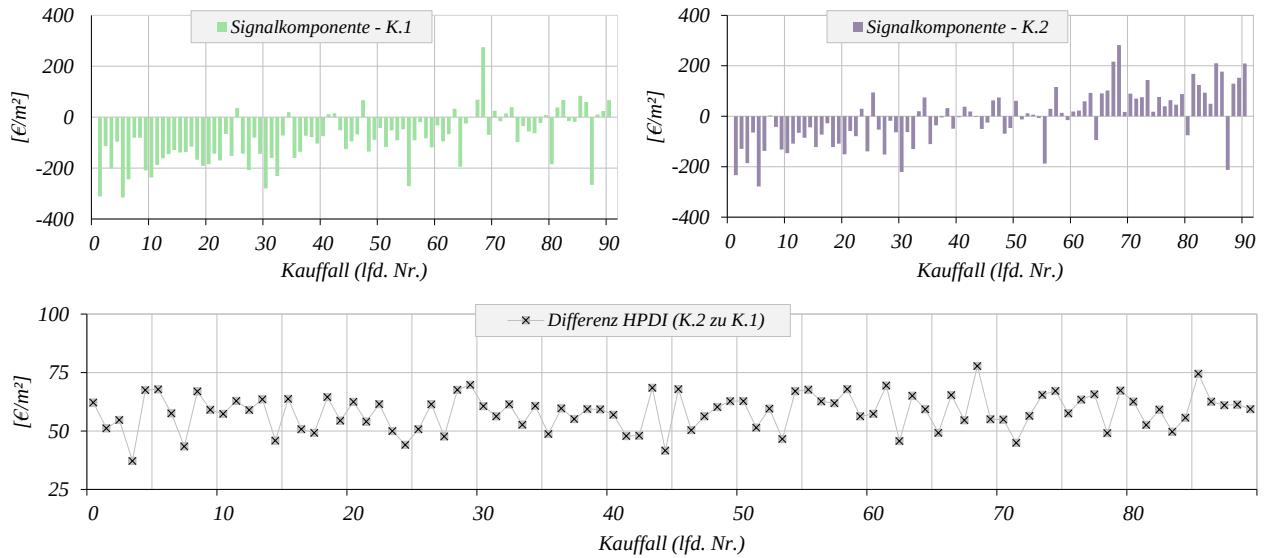


Abbildung C.15: Schätzwerte $\bar{s}$, Epoche 2007 (oben); Differenzplots der HPDIs (unten)

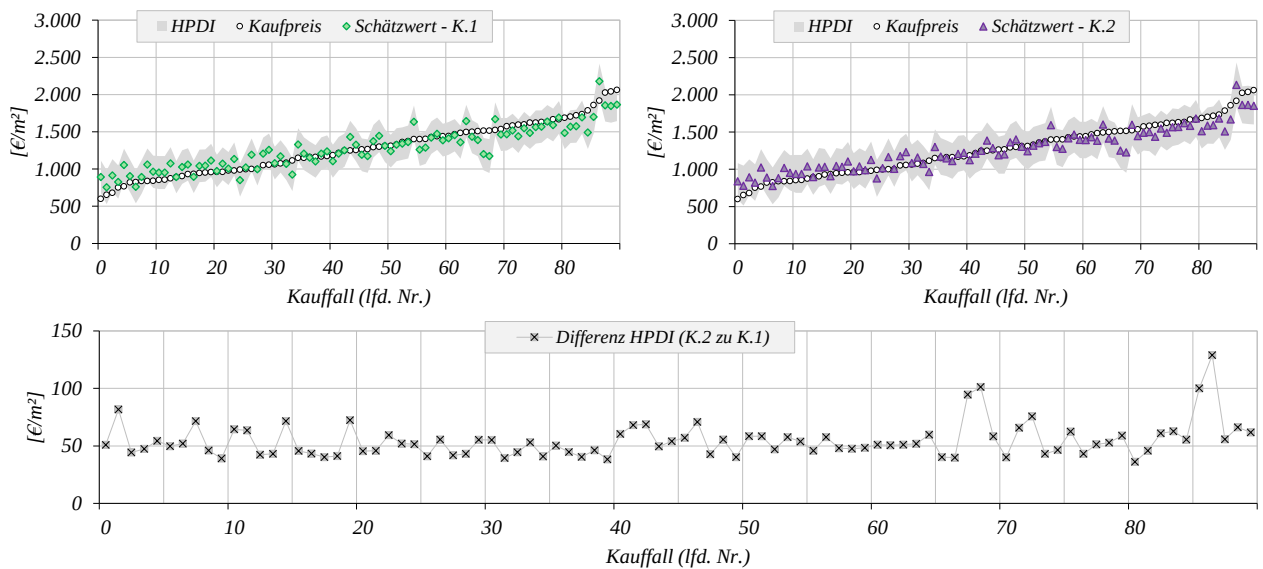


Abbildung C.16: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2007 (oben); Differenzplots der HPDIs (unten)

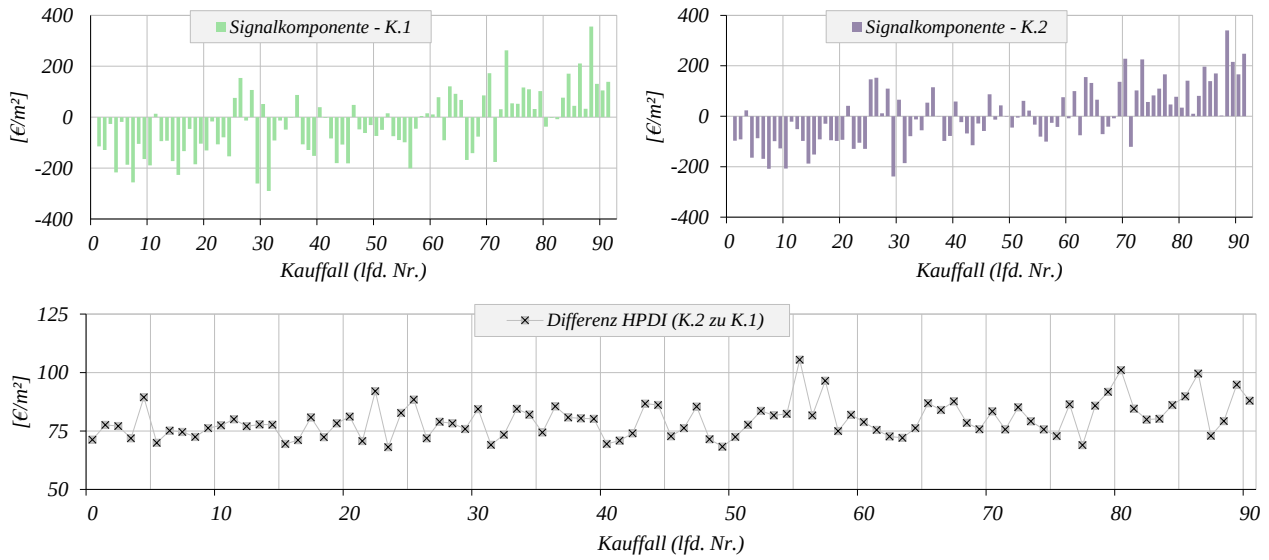


Abbildung C.17: Schätzwerte $\bar{s}$, Epoche 2008 (oben); Differenzplots der HPDIs (unten)

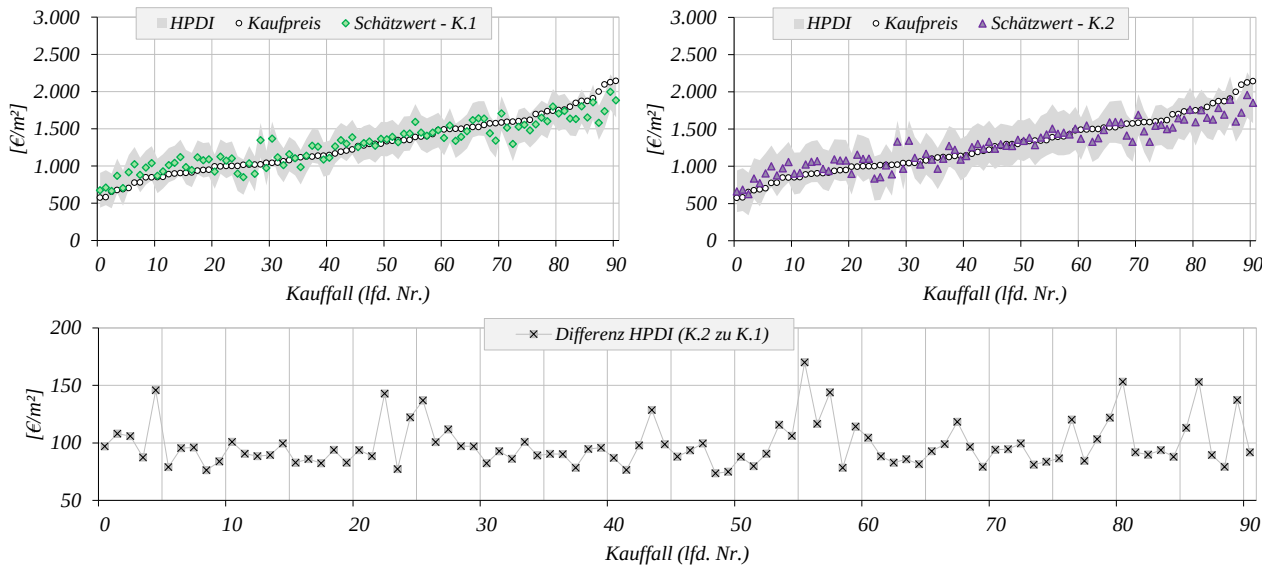


Abbildung C.18: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2008 (oben); Differenzplots der HPDIs (unten)

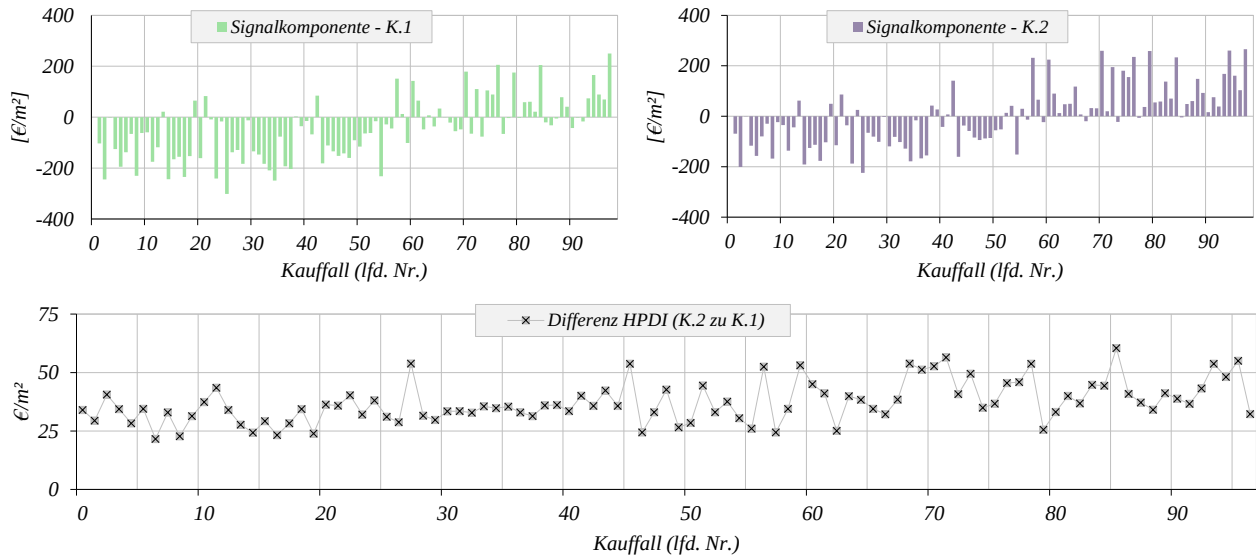


Abbildung C.19: Schätzwerte $\bar{s}$, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

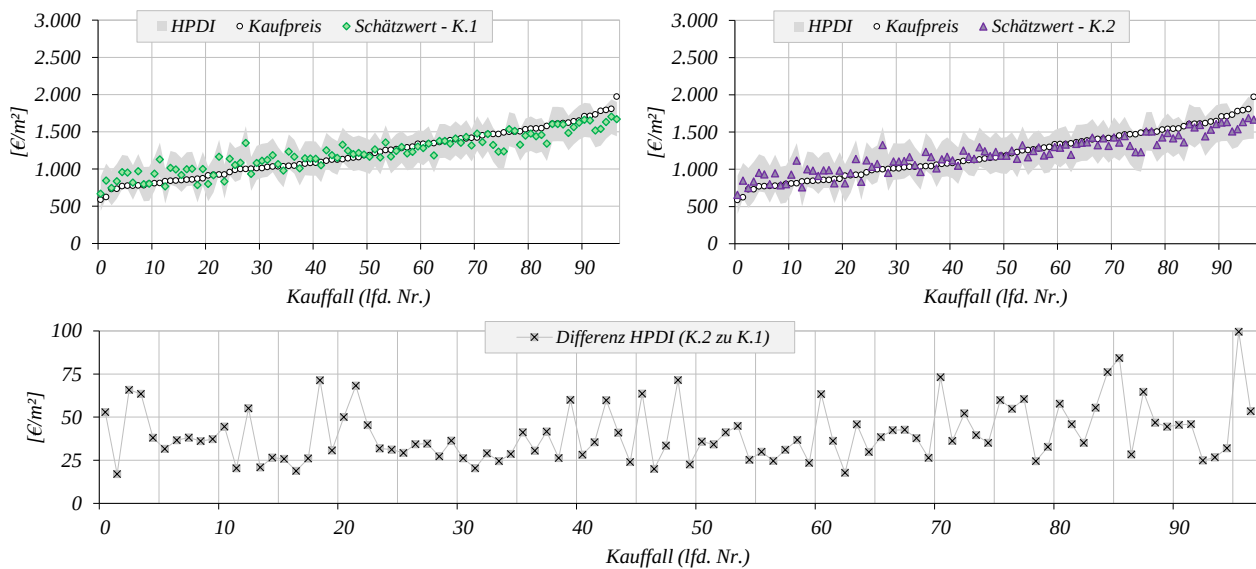
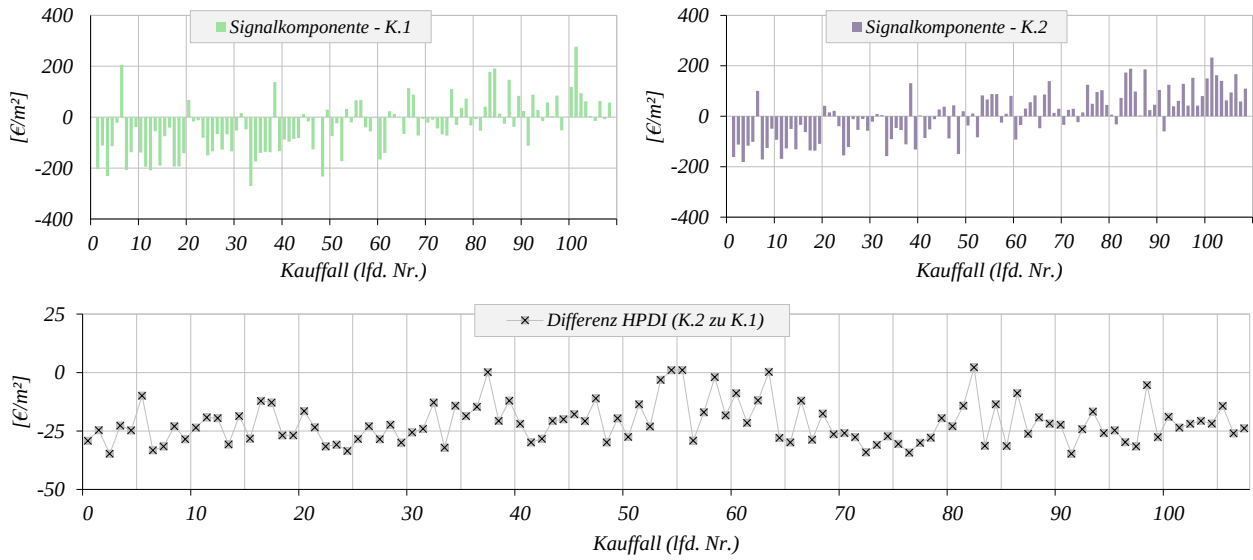
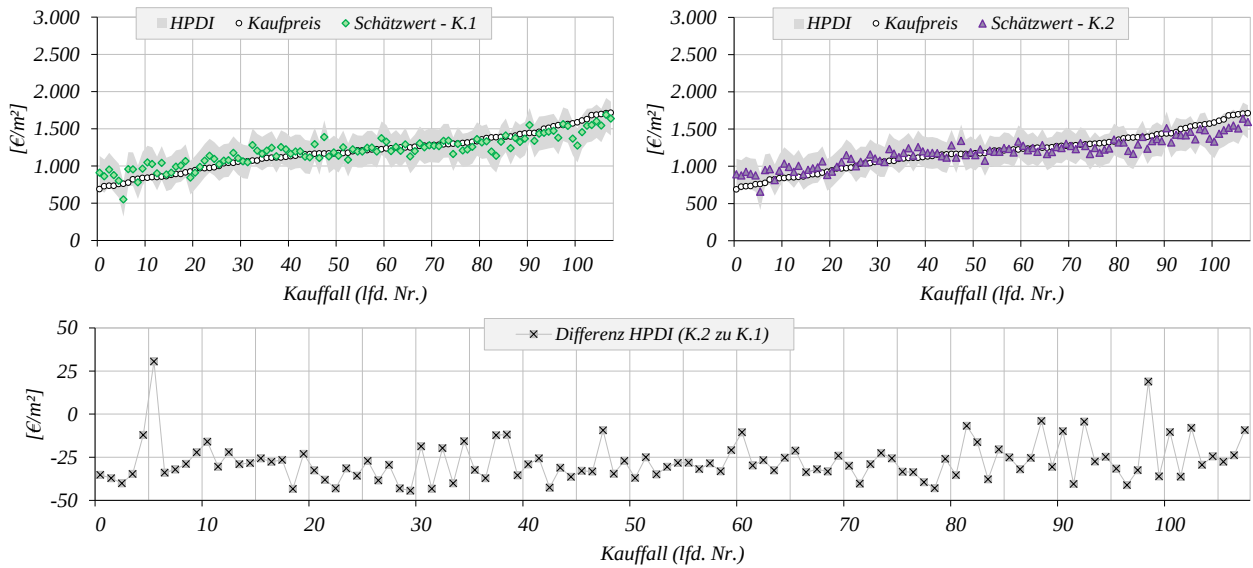


Abbildung C.20: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

Abbildung C.21: Schätzwerte $\bar{s}$, Epoche 2010 (oben); Differenzplots der HPDIs (unten)Abbildung C.22: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2010 (oben); Differenzplots der HPDIs (unten)

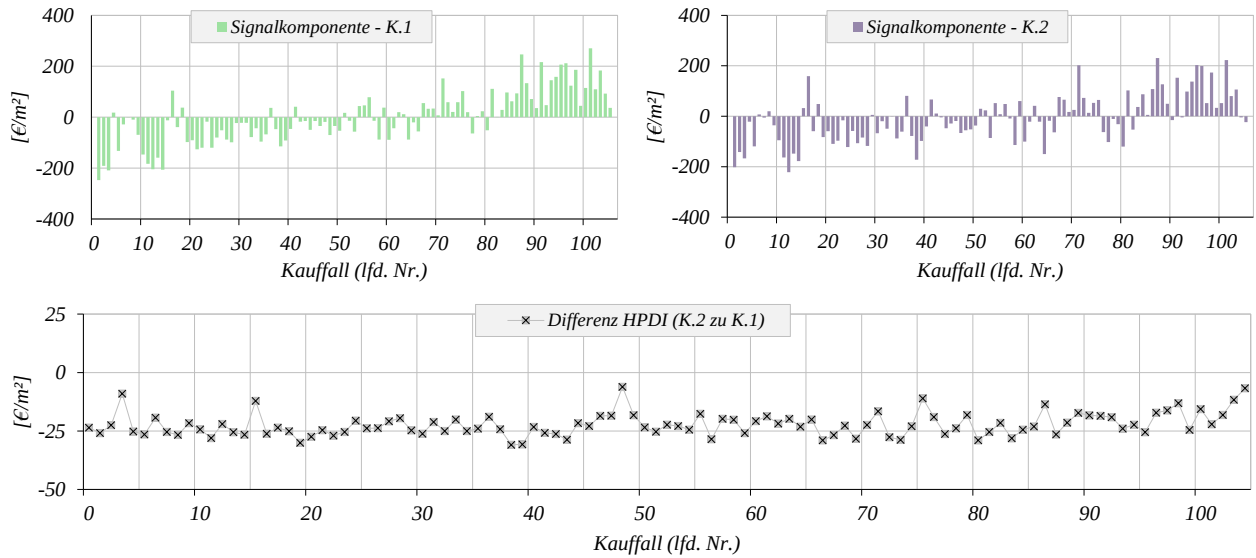


Abbildung C.23: Schätzwerte $\bar{s}$, Epoche 2011 (oben); Differenzplots der HPDIs (unten)

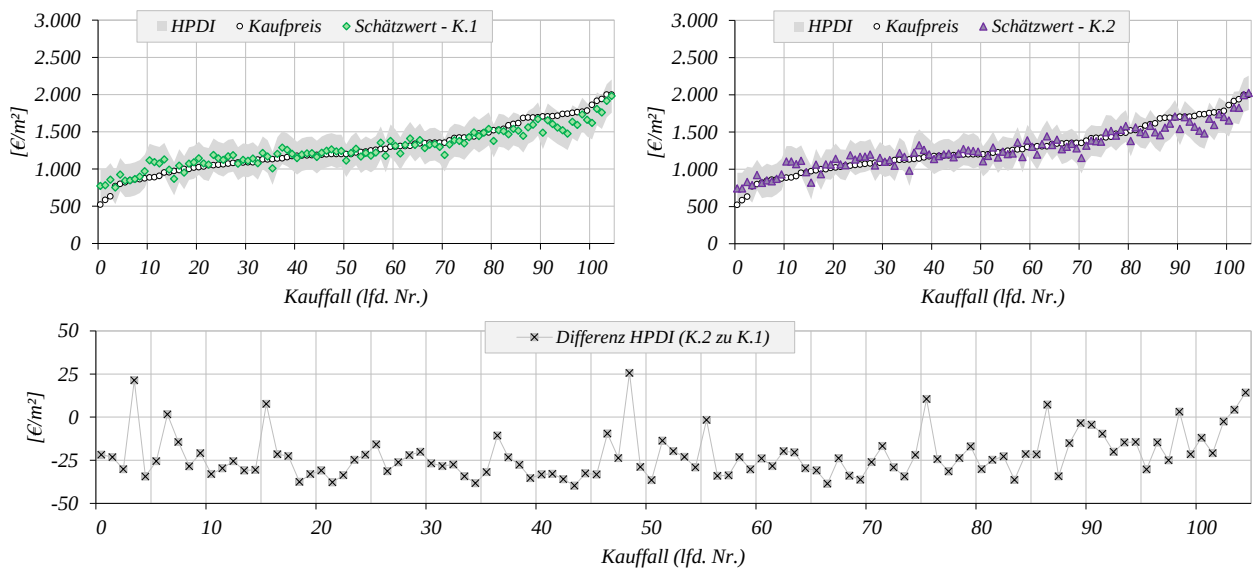


Abbildung C.24: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2011 (oben); Differenzplots der HPDIs (unten)

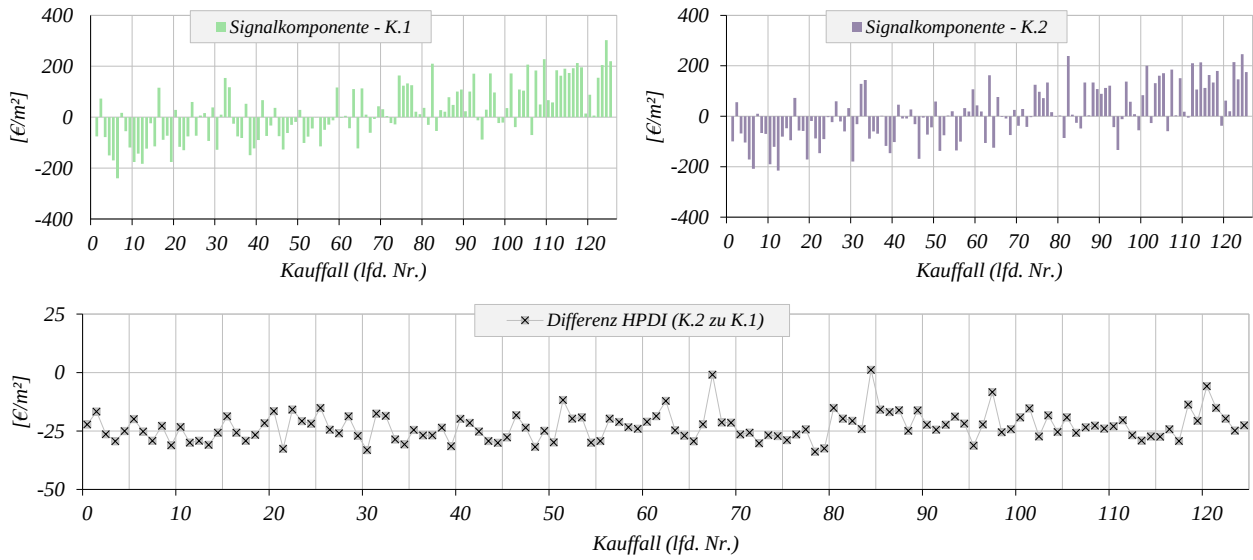


Abbildung C.25: Schätzwerte $\bar{s}$, Epoche 2012 (oben); Differenzplots der HPDIs (unten)

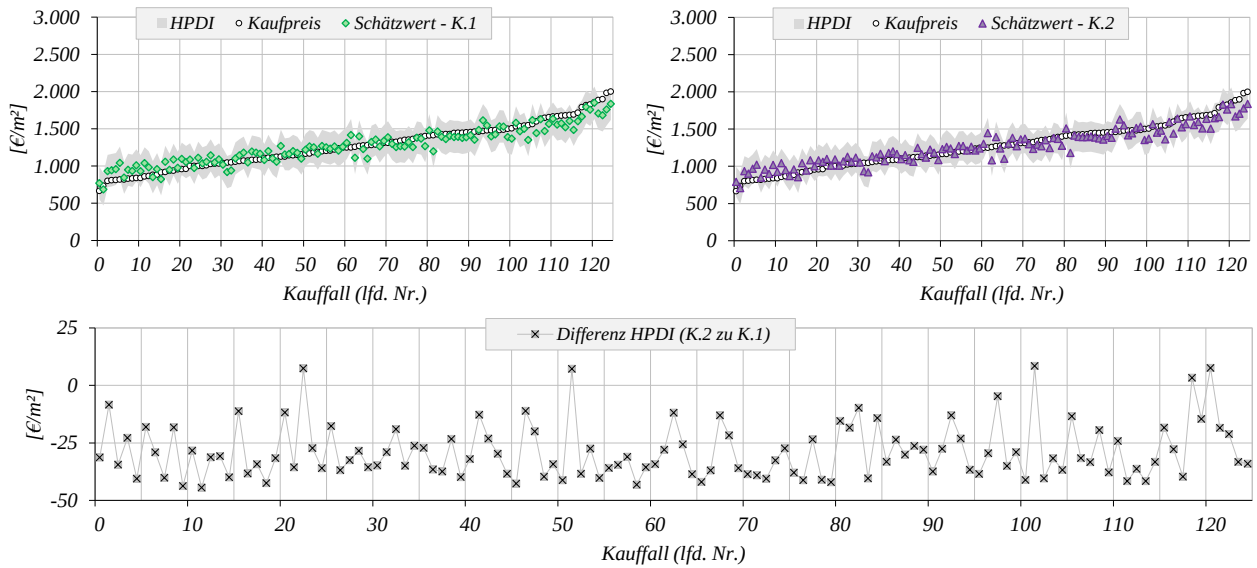


Abbildung C.26: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2012 (oben); Differenzplots der HPDIs (unten)

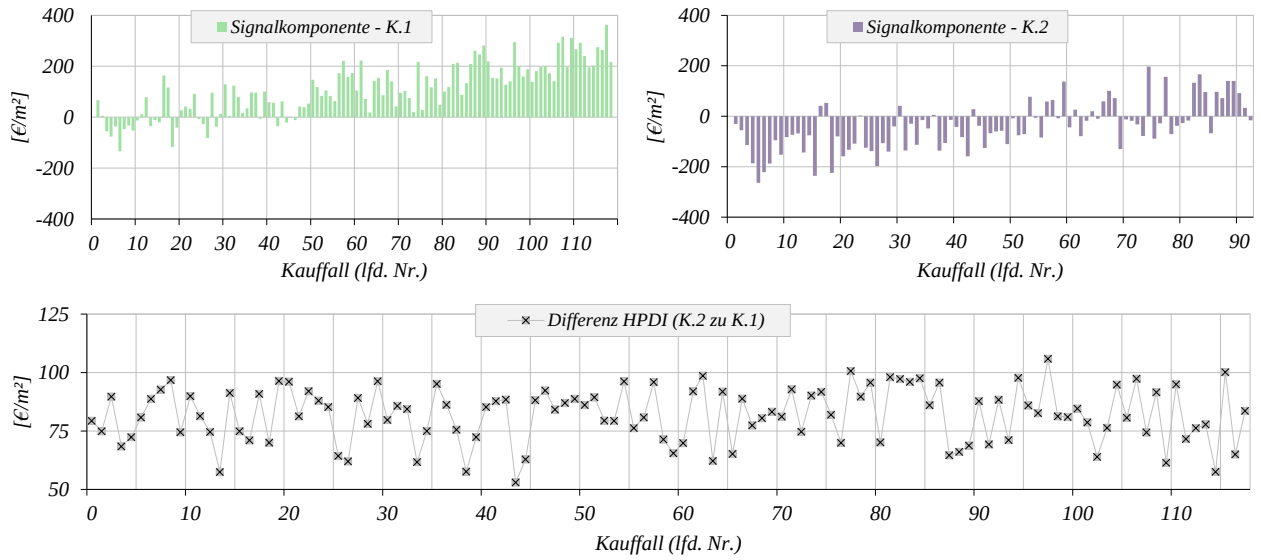


Abbildung C.27: Schätzwerte $\bar{s}$, Epoche 2013 (oben); Differenzplots der HPDIs (unten)

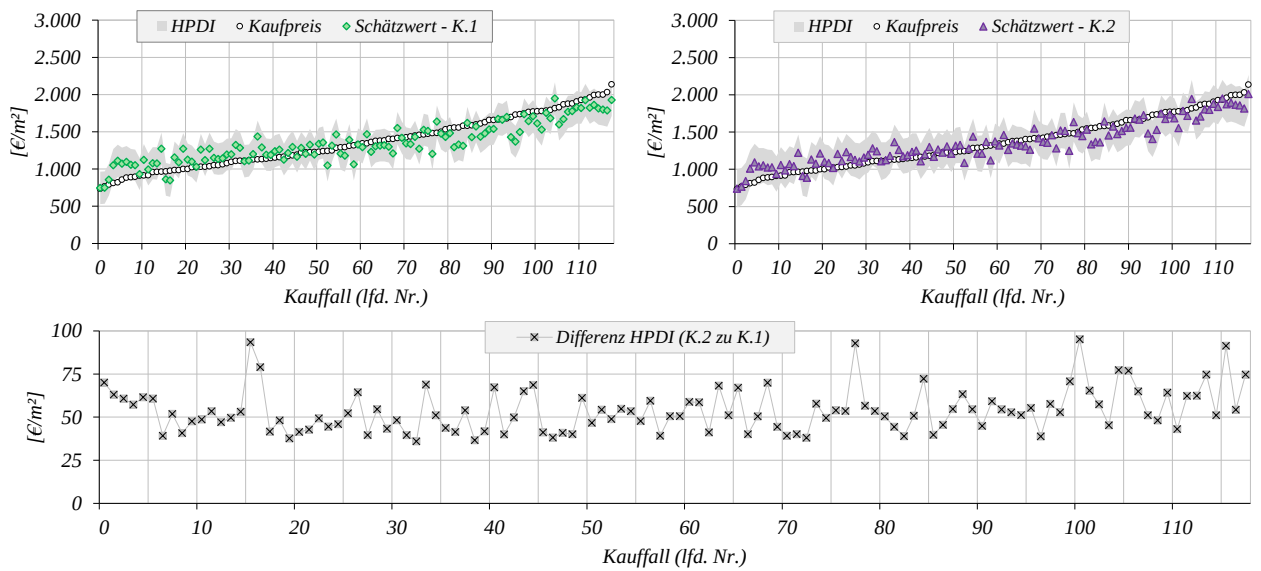


Abbildung C.28: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2013 (oben); Differenzplots der HPDIs (unten)

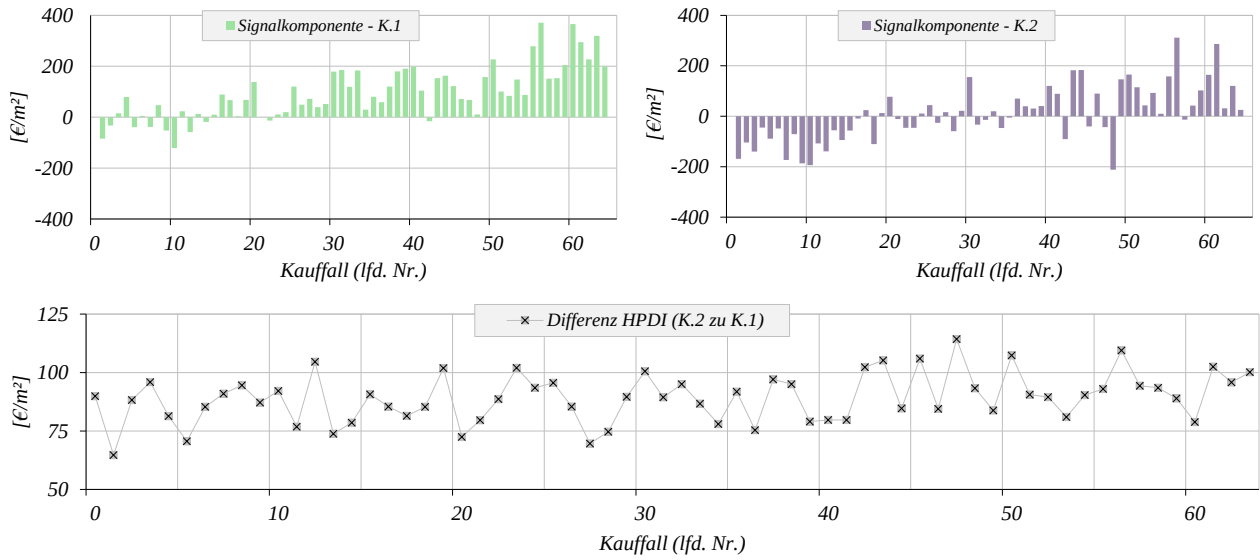


Abbildung C.29: Schätzwerte $\bar{s}$, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

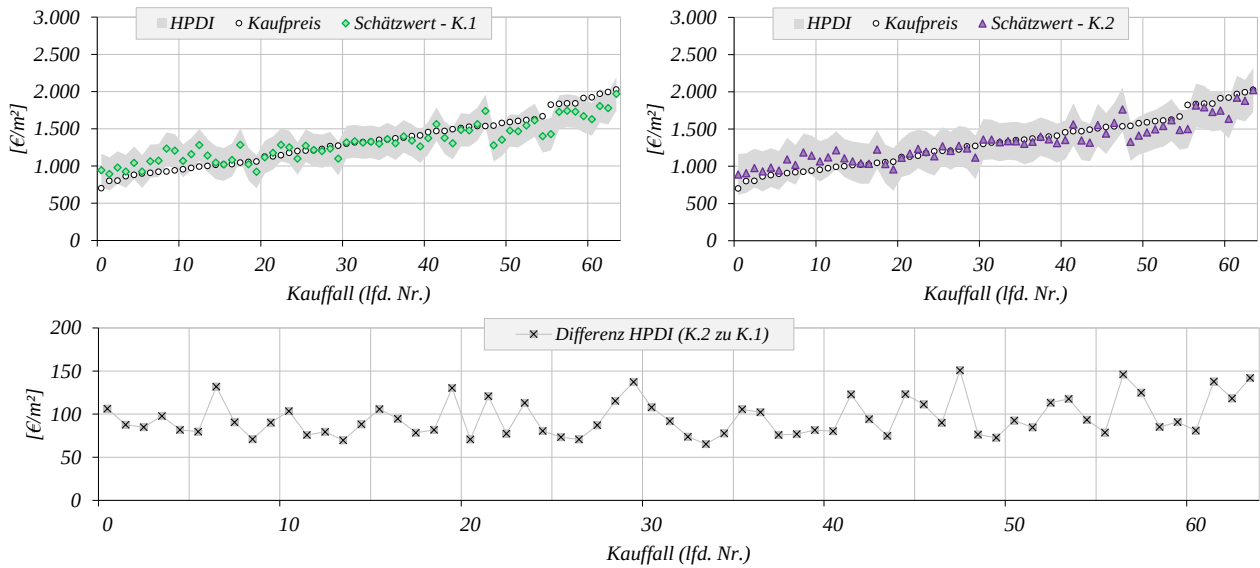


Abbildung C.30: Schätzwerte $\bar{y}$ der Bayesschen Kollokation, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

C.3. Abweichungen zwischen Kaufpreisen und Schätzwerten $\bar{y}$

Die Abbildungen C.31 bis C.40 zeigen die prozentualen Abweichungen zwischen den originären Kaufpreisen und den jeweiligen Schätzwerten $\bar{y}$ der Bayesschen Kollokation. In den linken Grafiken finden sich die Ergebnisse der informativen Lösungen *K.1*, in den rechten Grafiken die der nichtinformativen Lösungen *K.2*.

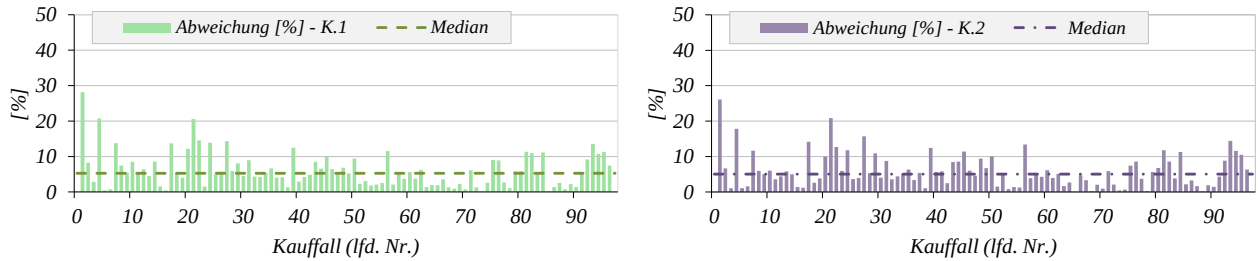


Abbildung C.31: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2005

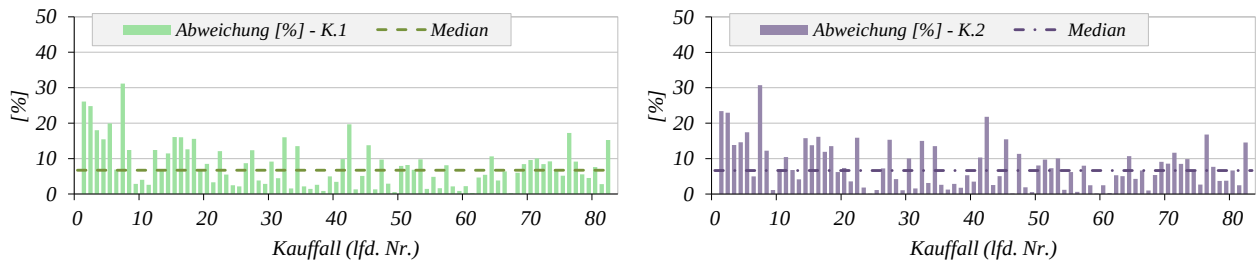


Abbildung C.32: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2006

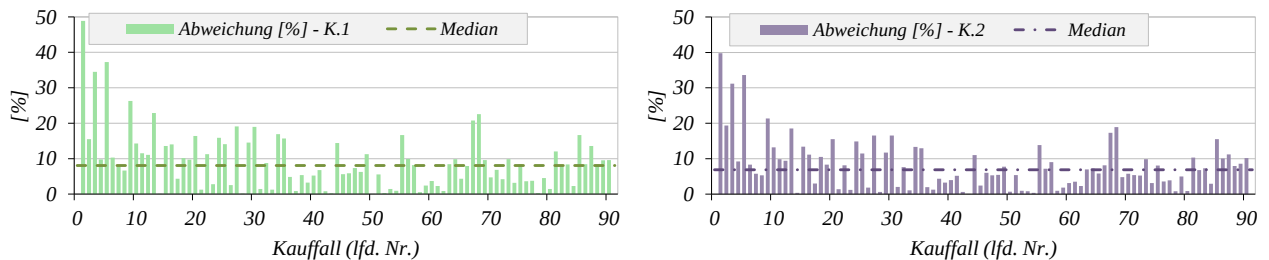


Abbildung C.33: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2007

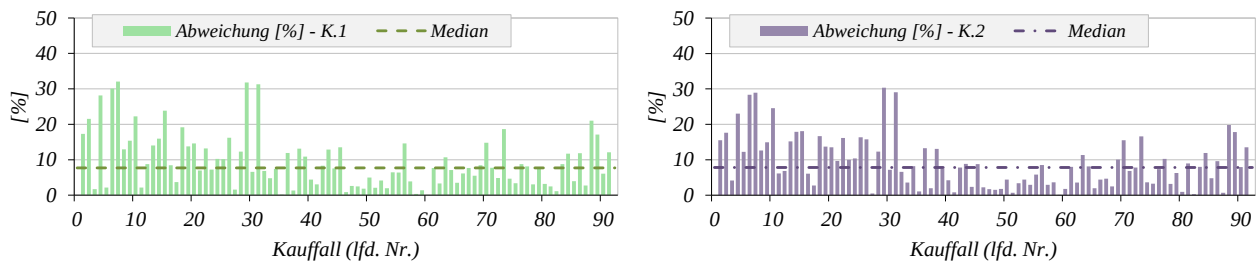
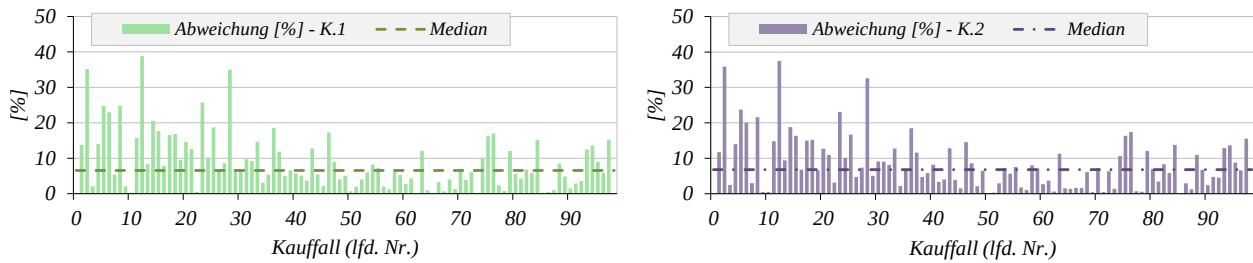
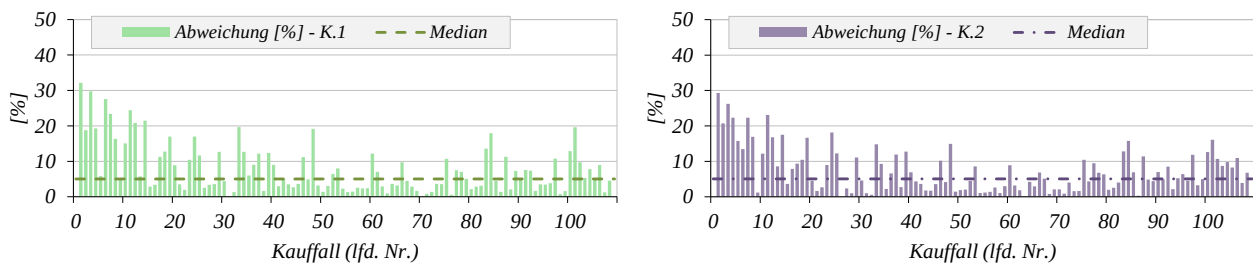
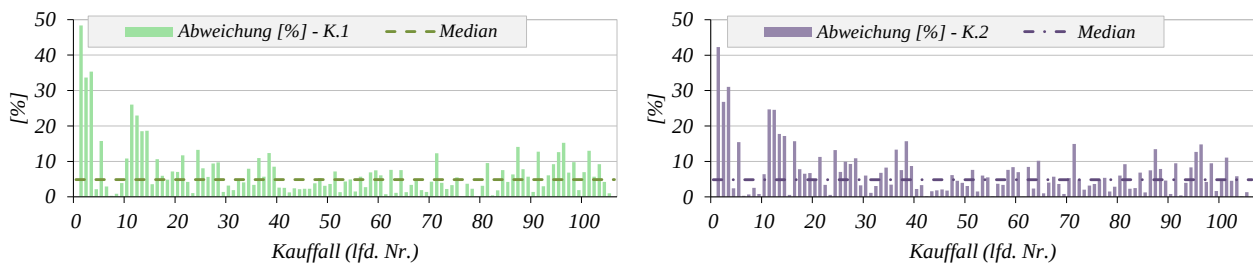
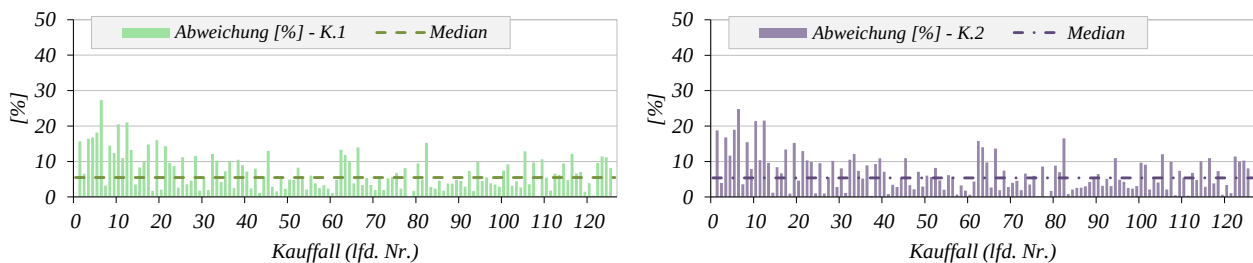


Abbildung C.34: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2008

Abbildung C.35: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2009Abbildung C.36: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2010Abbildung C.37: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2011Abbildung C.38: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2012

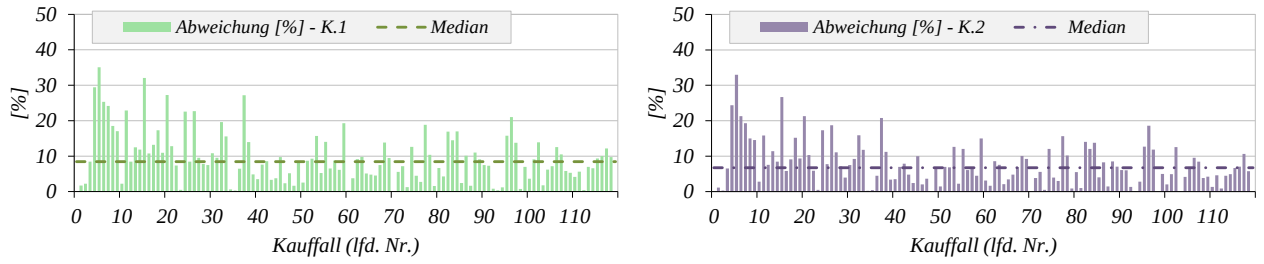


Abbildung C.39: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2013

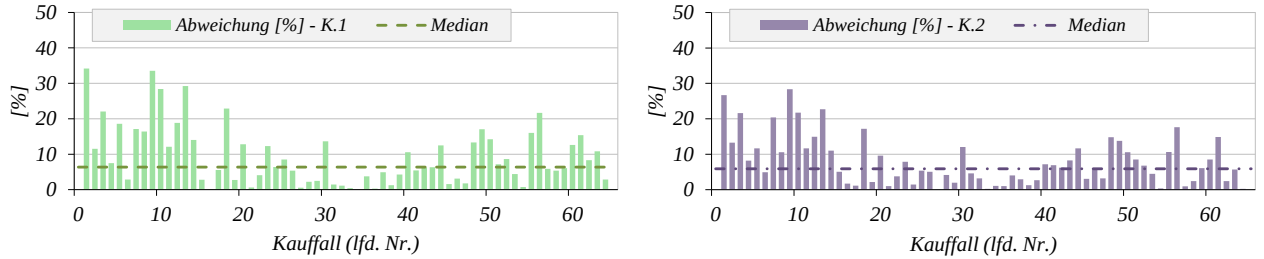


Abbildung C.40: Abweichungen der Kaufpreise zu Schätzwerten $\bar{y}$ der Bayesschen Kollokation, Epoche 2014

C.4. Posteriori-Schätzung der Signalkomponente $\bar{s}_p$ und der Zielgröße $\bar{y}_p$ und HPDIs

Die Abbildungen C.41 bis C.60 zeigen die Ergebnisse der Prädiktion im Ansatz der Bayesschen Kollokation. Wie im Fall der Darstellung der Ergebnisse für die Schätzung der Vergleichswerte $\bar{y}$ und derer Signalkomponenten $\bar{s}$ sind die Abbildungen epochenweise sortiert. Die ungerade nummerierten Abbildungen (bspw. Abbildung C.41) geben einen Überblick über die geschätzten Signalkomponenten $\bar{s}_p$; entsprechend finden sich in den Grafiken oben links die Ergebnisse des informativen Ansatzes *K.1*, oben rechts des nichtinformativen Ansatzes *K.2*. Unterhalb der beiden Grafiken sind die Differenzplots der HPDIs gegeben. Die Sortierung erfolgt wiederum nach den zugehörigen, aufsteigenden Kaufpreisbeobachtungen. Gemäß der bisherigen Ausführungen werden in den folgenden, gerade nummerierten Abbildungen (bspw. Abbildung C.42) die Ergebnisse für die Posteriori-Schätzungen der prädizierten Vergleichswerte $\bar{y}_p$ gezeigt. Die Aufteilung der Abbildungen orientiert sich entsprechend an den Darstellungen für die Signalkomponente.

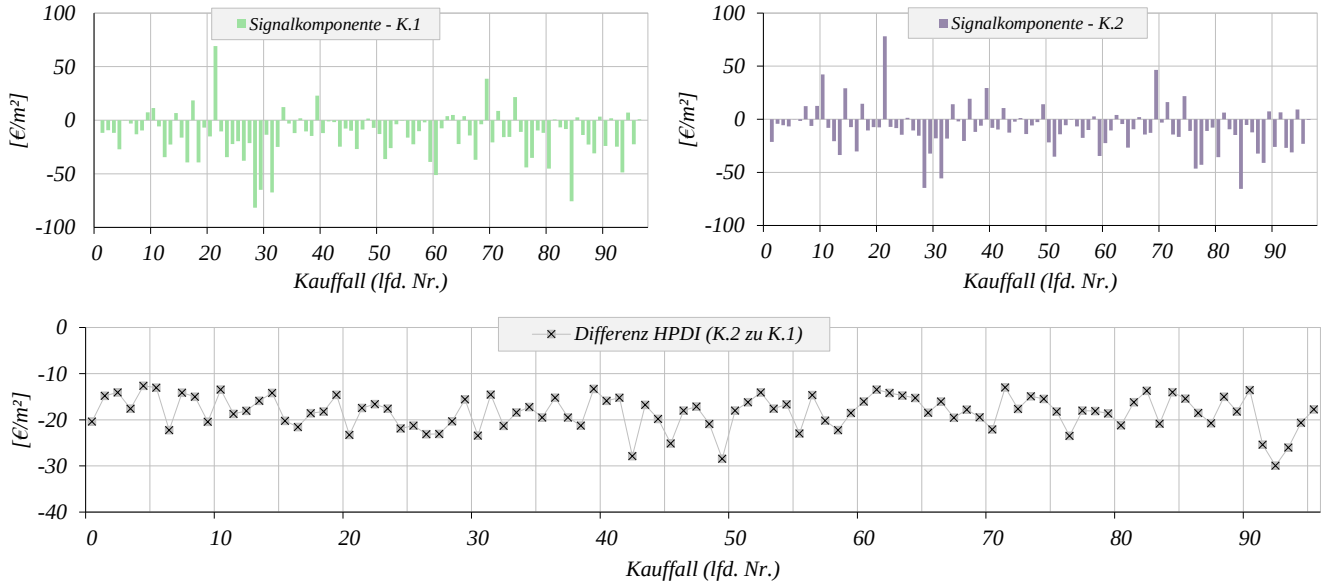


Abbildung C.41: Schätzwerte $\bar{s}_p$, Epoche 2005 (oben); Differenzplots der HPDIs (unten)

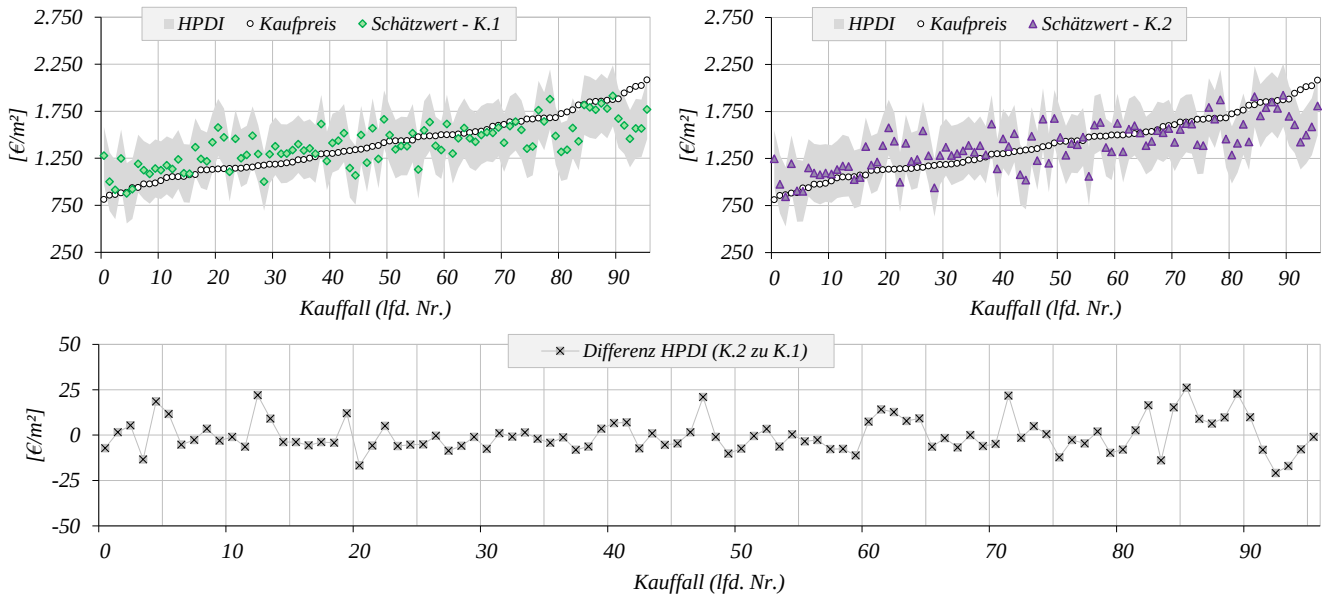


Abbildung C.42: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2005, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

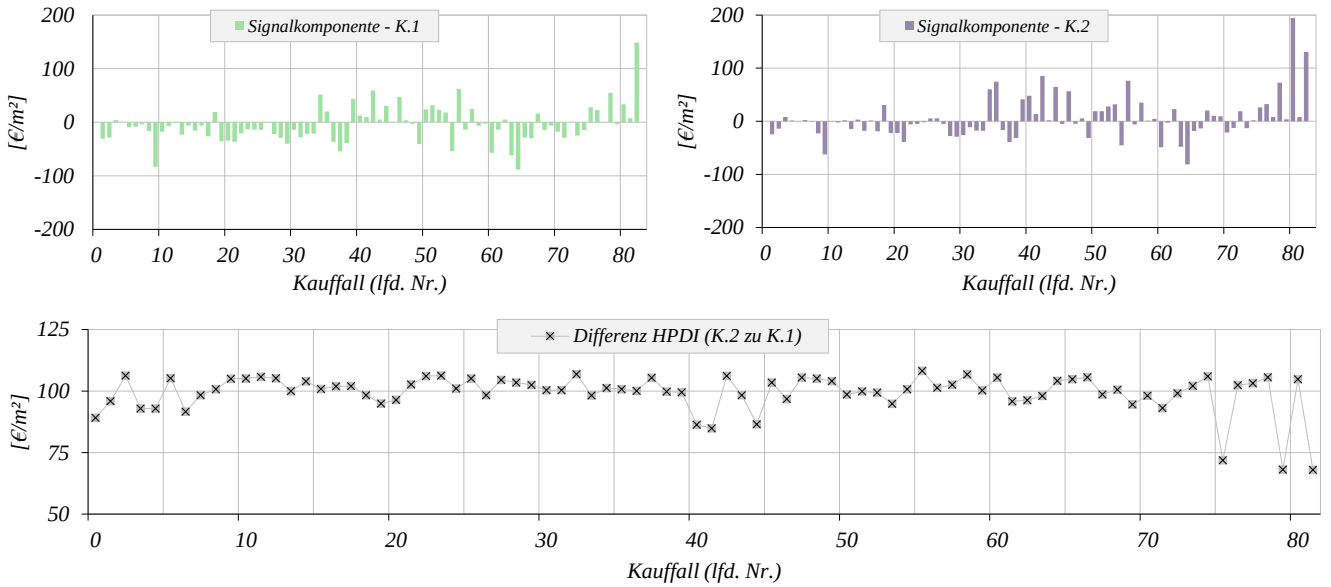


Abbildung C.43: Schätzwerte $\bar{s}_p$, Epoche 2006 (oben); Differenzplots der HPDIs (unten)

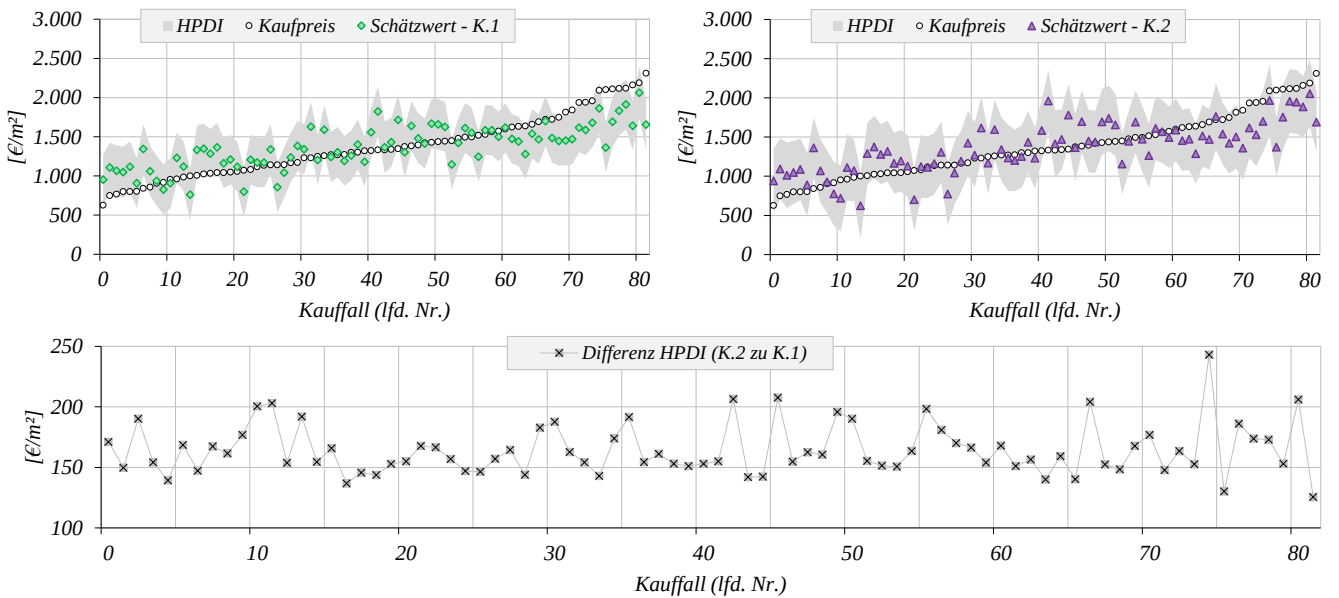


Abbildung C.44: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2006, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

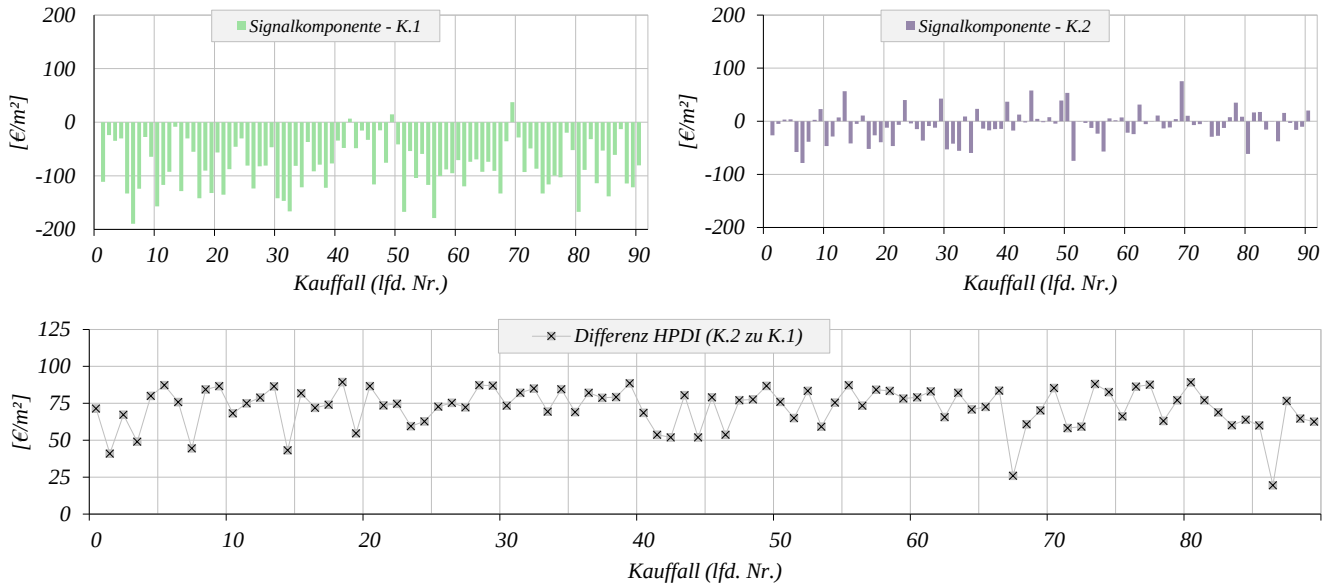


Abbildung C.45: Schätzwerte $\bar{s}_p$, Epoche 2007 (oben); Differenzplots der HPDIs (unten)

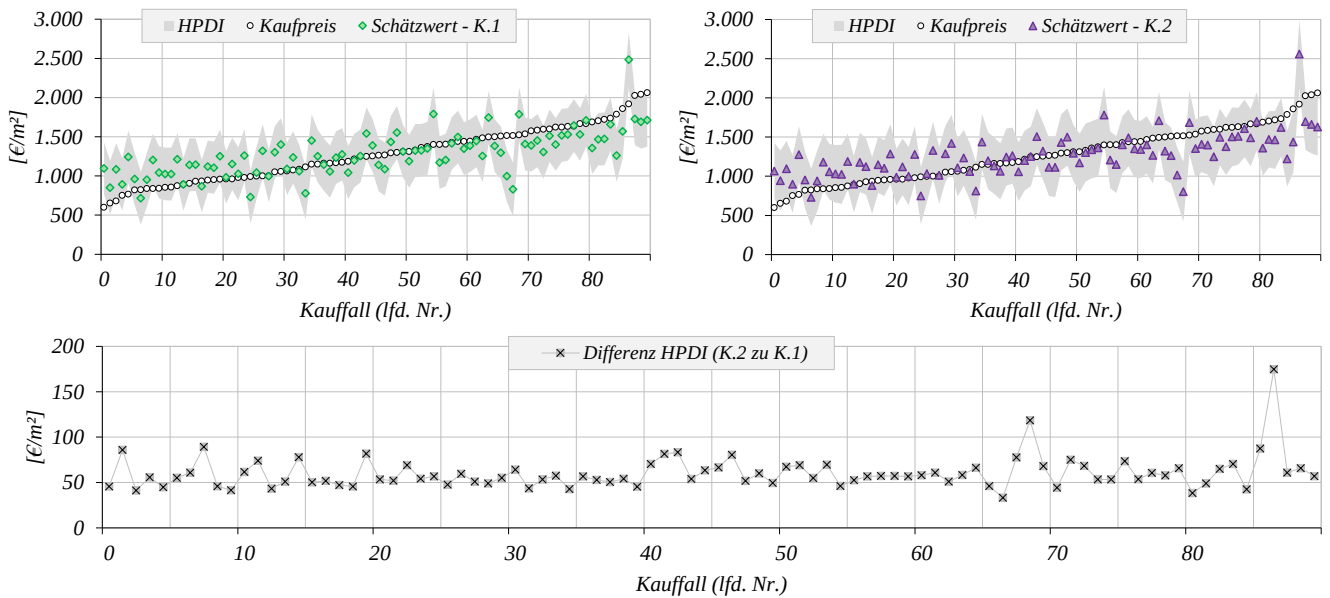


Abbildung C.46: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2007, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

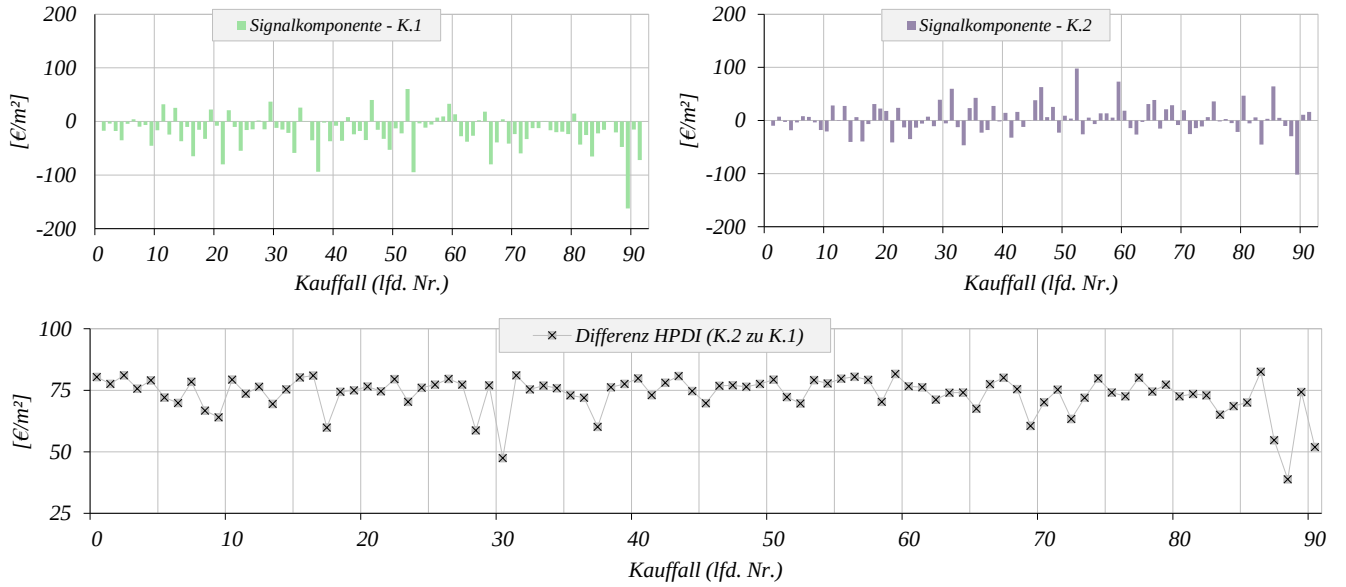


Abbildung C.47: Schätzwerte $\bar{s}_p$, Epoche 2008 (oben); Differenzplots der HPDIs (unten)

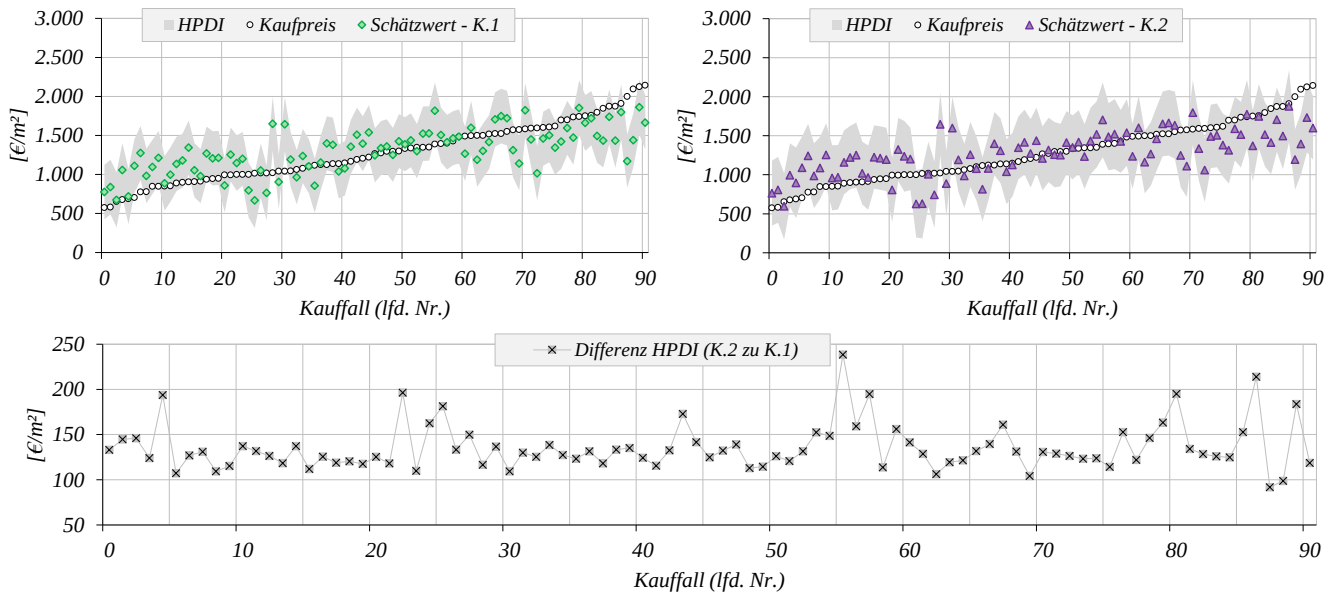


Abbildung C.48: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2008, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

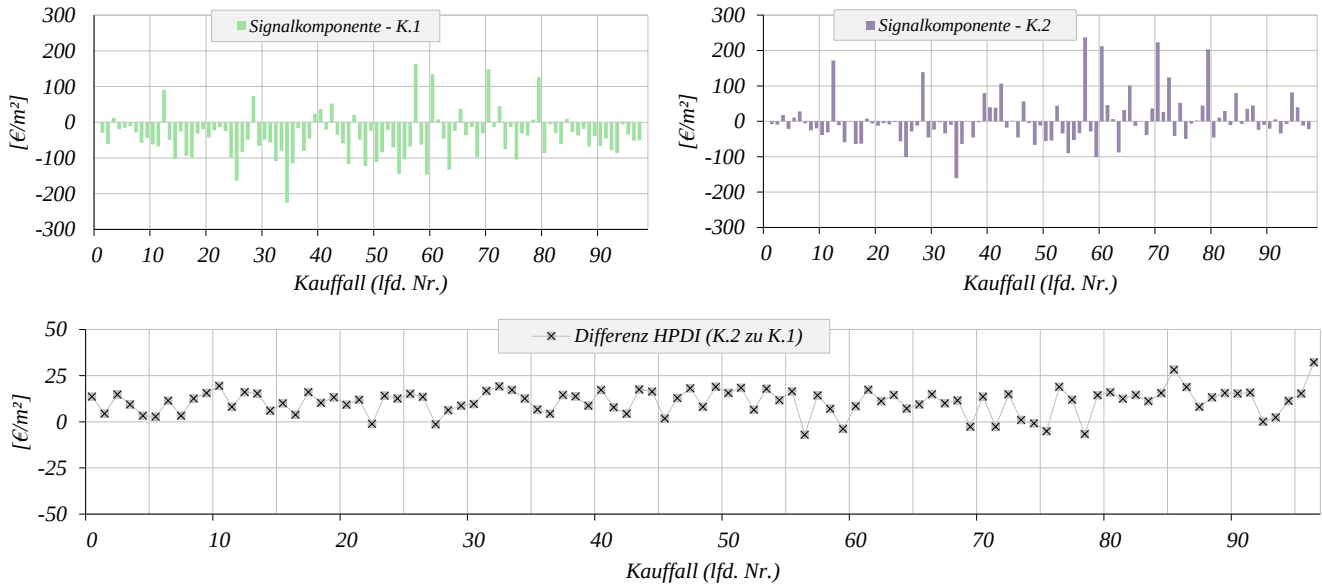


Abbildung C.49: Schätzwerte $\bar{s}_p$, Epoche 2009 (oben); Differenzplots der HPDIs (unten)

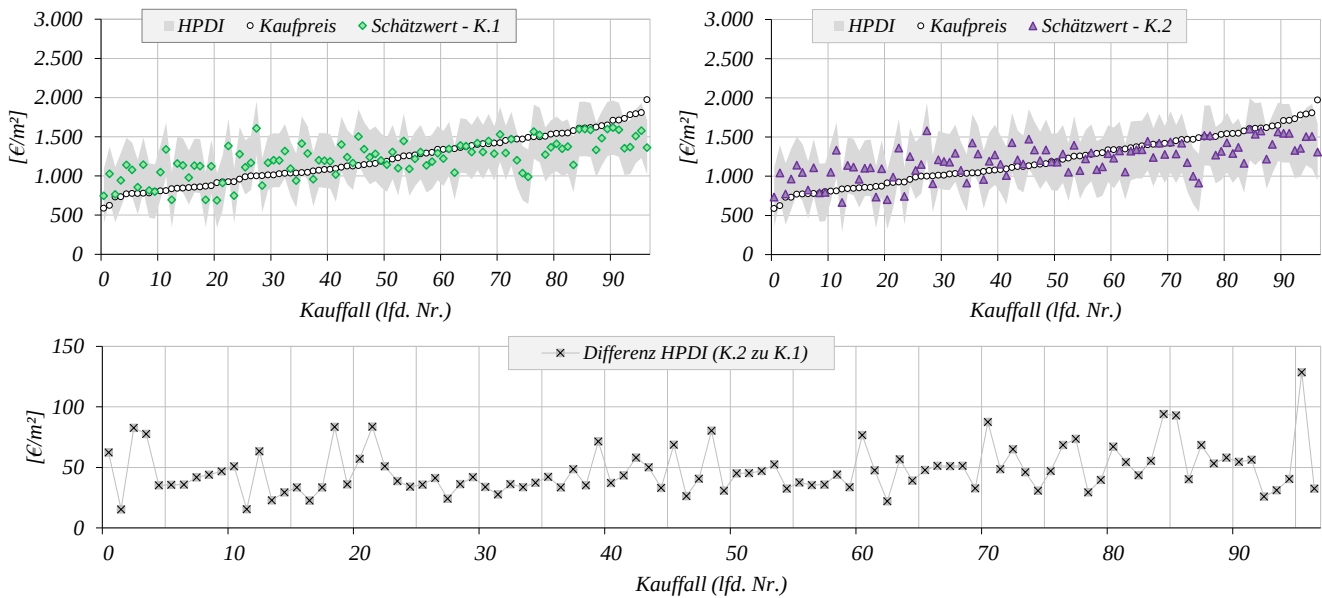


Abbildung C.50: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2009, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

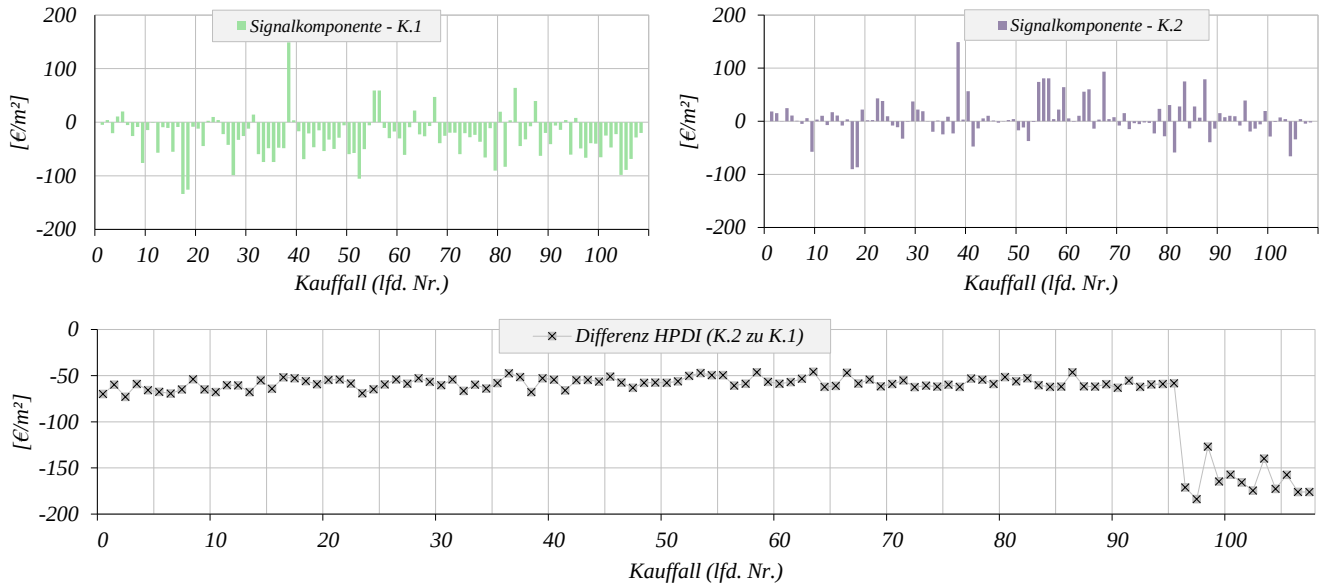


Abbildung C.51: Schätzwerte $\bar{s}_p$, Epoche 2010 (oben); Differenzplots der HPDIs (unten)

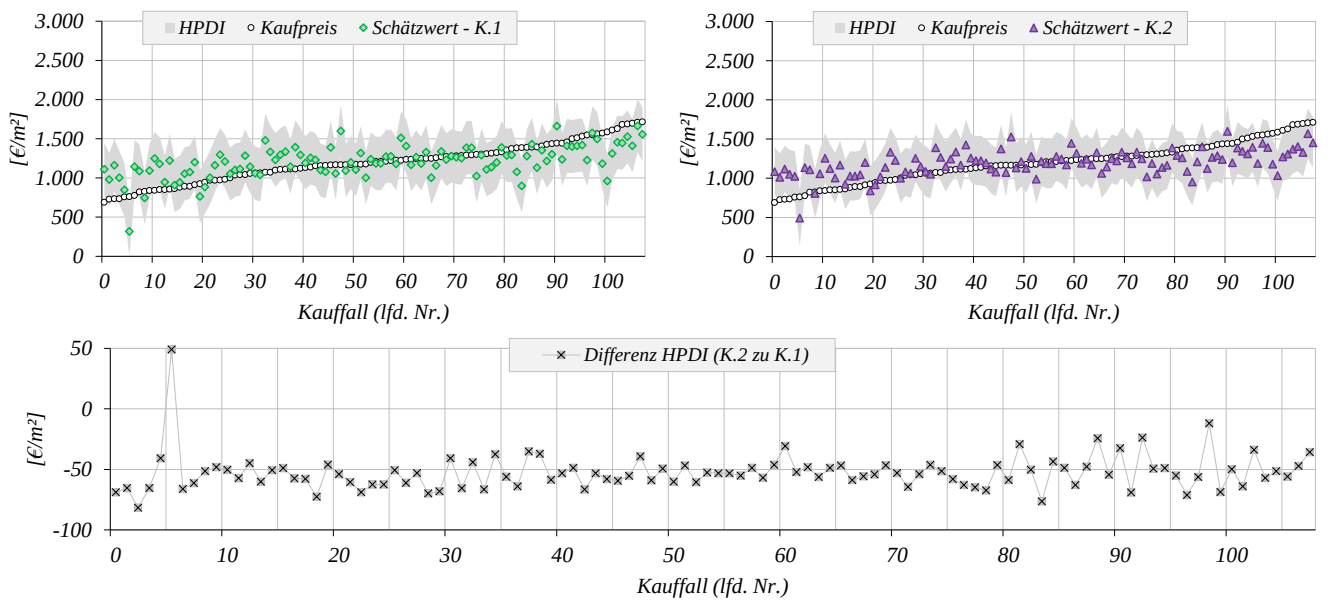
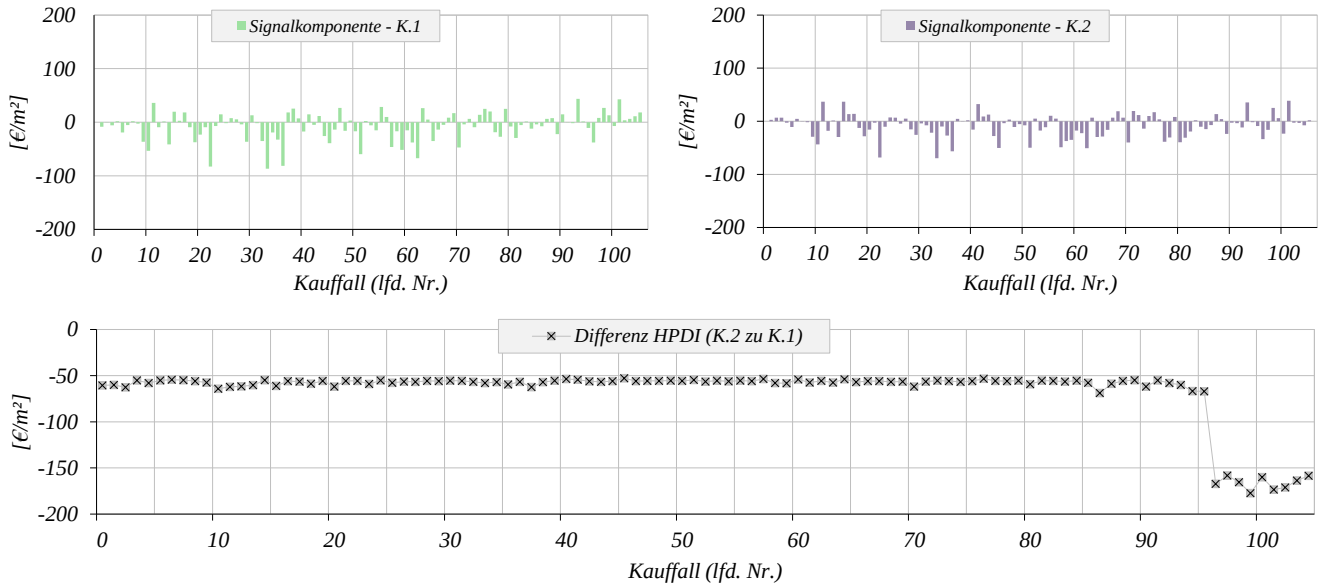
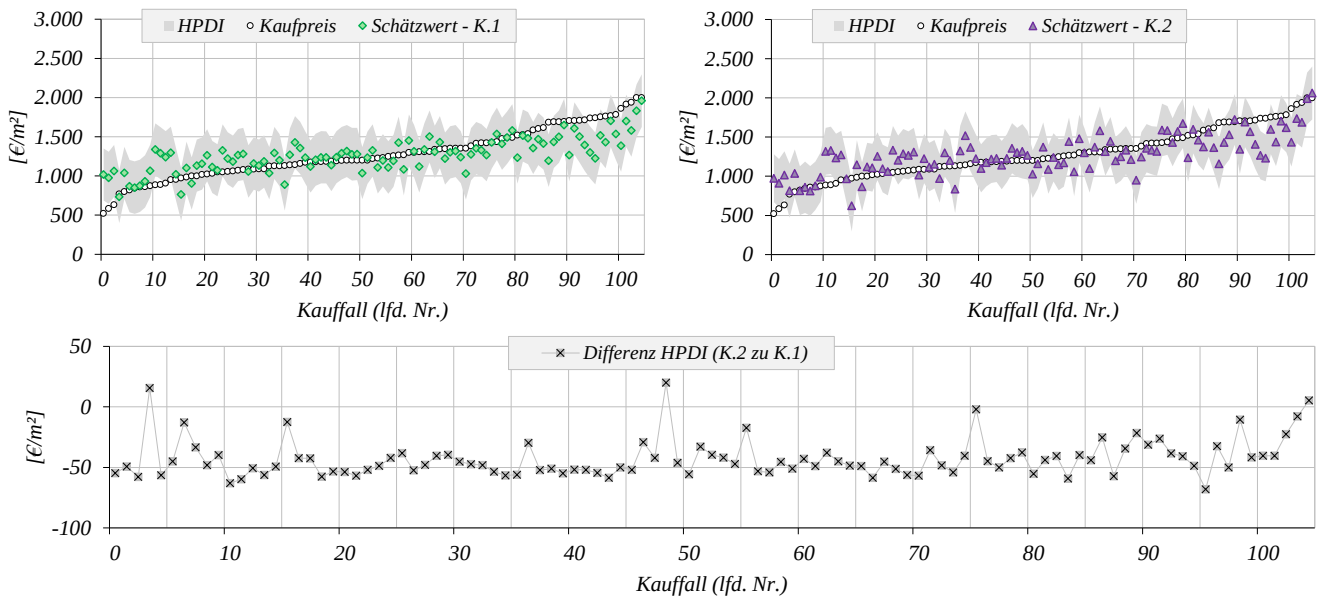


Abbildung C.52: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2010, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

Abbildung C.53: Schätzwerte $\bar{s}_p$, Epoche 2011 (oben); Differenzplots der HPDIs (unten)Abbildung C.54: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2011, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

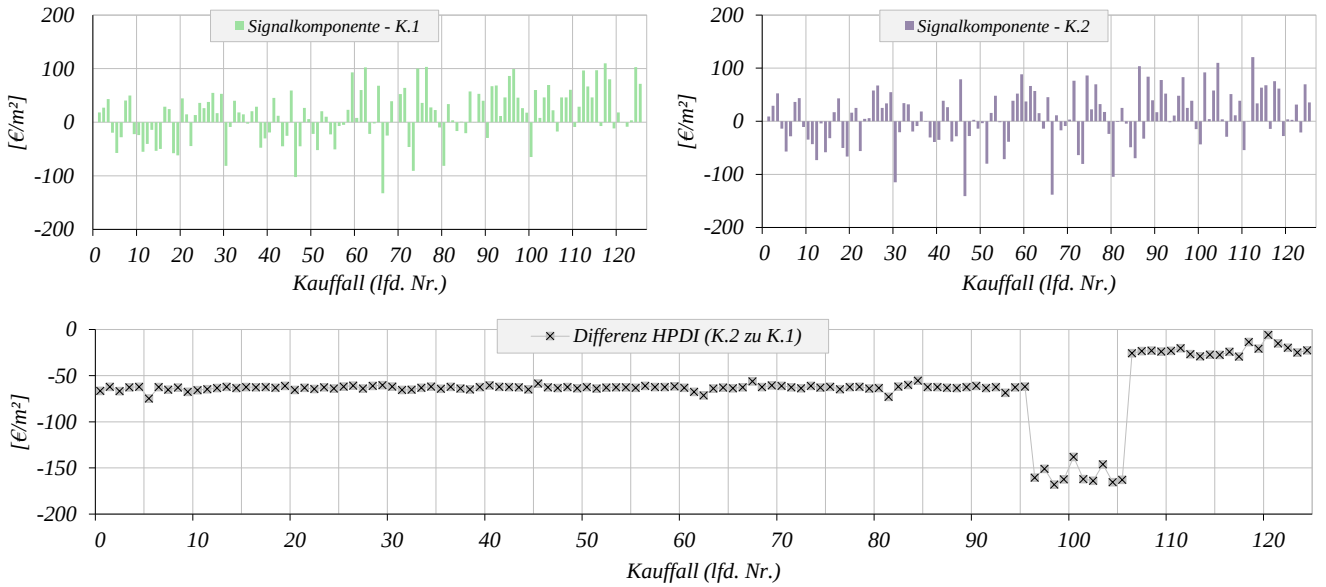


Abbildung C.55: Schätzwerte $\bar{s}_p$, Epoche 2012 (oben); Differenzplots der HPDIs (unten)

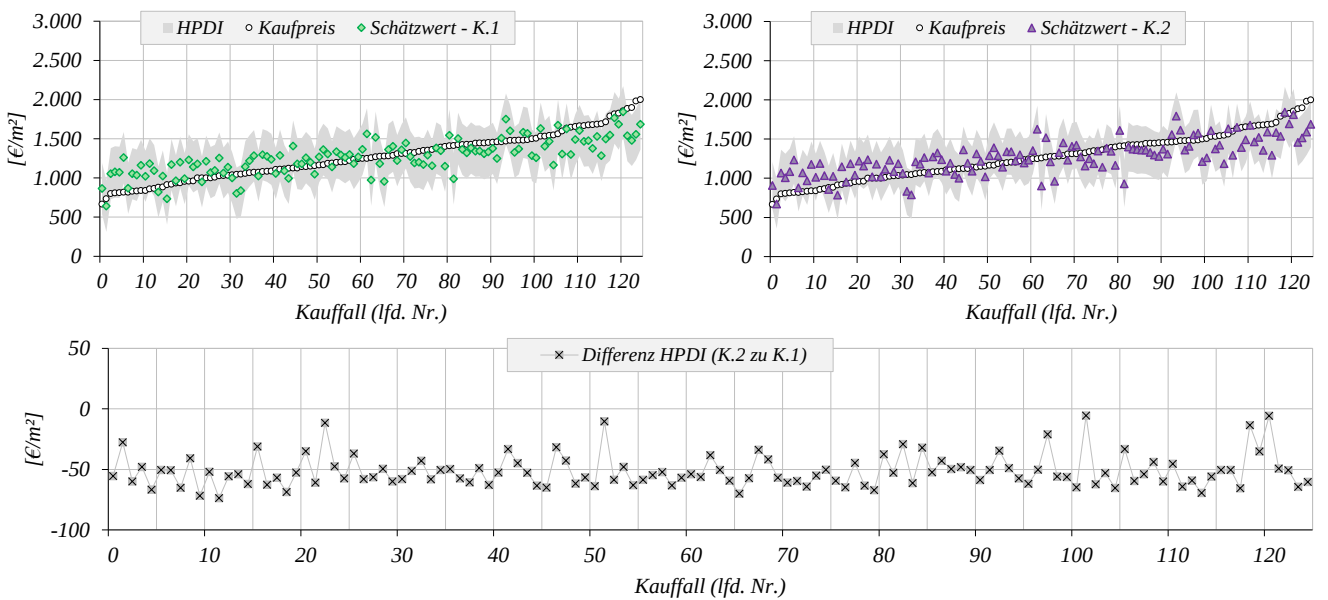


Abbildung C.56: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2012, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

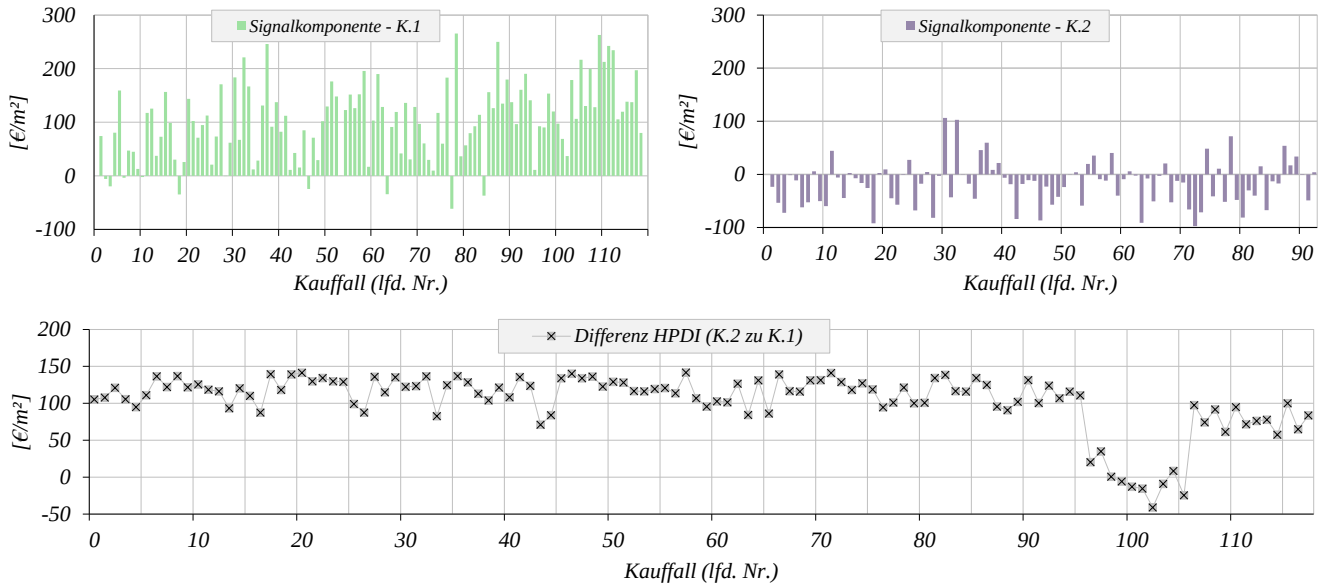


Abbildung C.57: Schätzwerte $\bar{s}_p$, Epoche 2013 (oben); Differenzplots der HPDIs (unten)

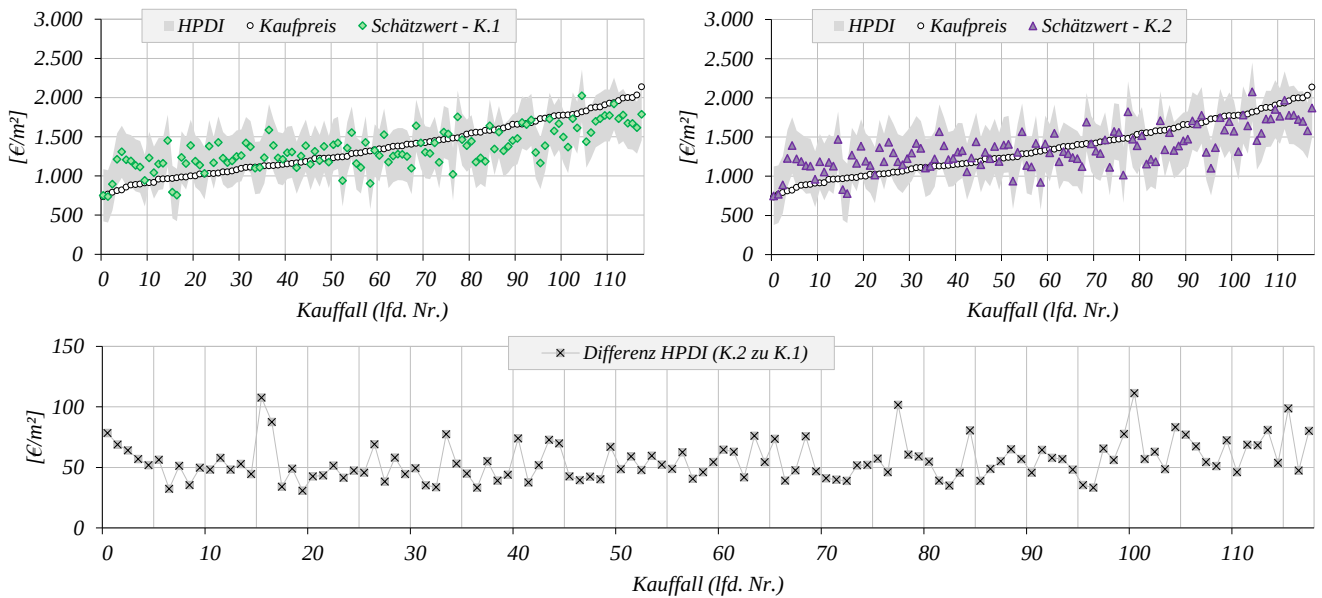


Abbildung C.58: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2013, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

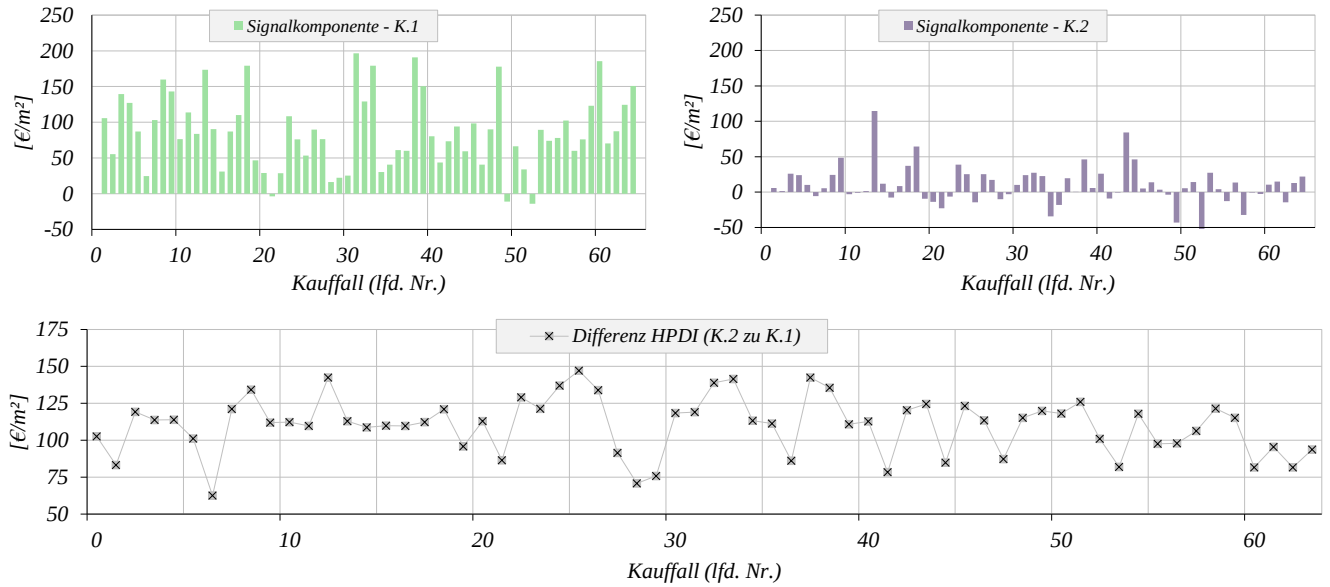


Abbildung C.59: Schätzwerte $\bar{s}_p$, Epoche 2014 (oben); Differenzplots der HPDIs (unten)

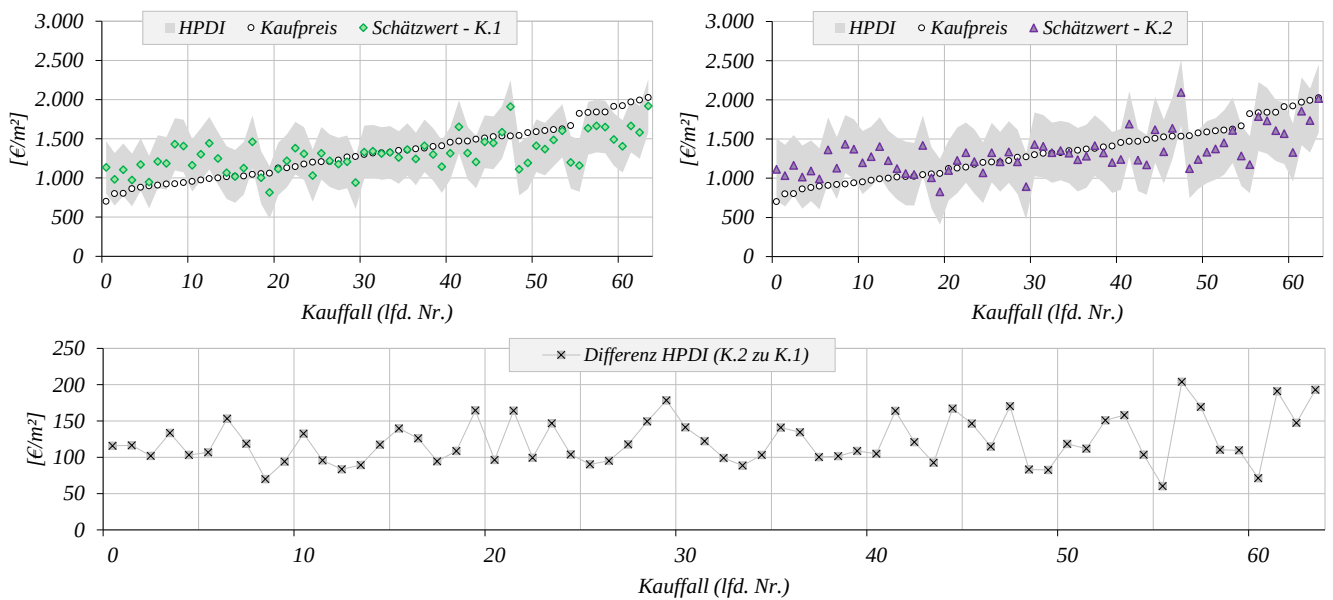


Abbildung C.60: Prädizierte Schätzwerte $\bar{y}_p$ der Epoche 2014, basierend auf Bayesischer Kollokation (oben); Differenzplots der HPDIs (unten)

C.5. Absolute Residuen der Bayesschen Modellansätze

Die nachfolgenden Abbildungen C.61 bis C.70 zeigen die absoluten Residuen, die aus den Schätzungen der Bayesschen Regressions- und Kollokationsmodelle aus den Gleichungen

$$\begin{aligned} |\bar{e}_{R.1}| &= |\mathbf{y} - \bar{\mathbf{y}}_{R.1}| = |\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}_{R.1}|, \\ |\bar{e}_{R.2}| &= |\mathbf{y} - \bar{\mathbf{y}}_{R.2}| = |\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}_{R.2}|, \\ |\bar{e}_{K.1}| &= |\mathbf{y} - \bar{\mathbf{y}}_{K.1}| = |\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}_{K.1} - \bar{s}_{K.1}|, \\ |\bar{e}_{K.2}| &= |\mathbf{y} - \bar{\mathbf{y}}_{K.2}| = |\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}_{K.2} - \bar{s}_{K.2}|, \end{aligned}$$

resultieren, wobei die Indizes sich jeweils auf die Modellvariante nach den Ausführungen der Kapitel 5 (S. 109 ff.) und 6 (S. 157 ff.) beziehen. In den Grafiken werden jeweils in der oberen Zeile die absoluten Residuen der informativen Bayesschen Regression und der informativen Bayesschen Kollokation (Modellvarianten $R.1$ und $K.1$) gegenübergestellt, in der jeweils zweiten Zeile die der nichtinformativen Lösungen $R.2$ und $K.2$. Die Sortierung bezieht sich auf die laufenden Kauffallnummern, die sich bereits aus den Darstellungen der Abbildungen B.1 bis B.10 (S. 209 ff.) für die Schätzungen der Bayesschen Regressionen und der Abbildungen C.12 bis C.30 (S. 222 ff.) für die Bayesschen Kollokationen ergeben und auf der aufsteigenden Höhe der Kaufpreise basieren. Zusätzlich werden für sämtliche Modellvarianten die Medianwerte der absoluten Residuen angegeben.

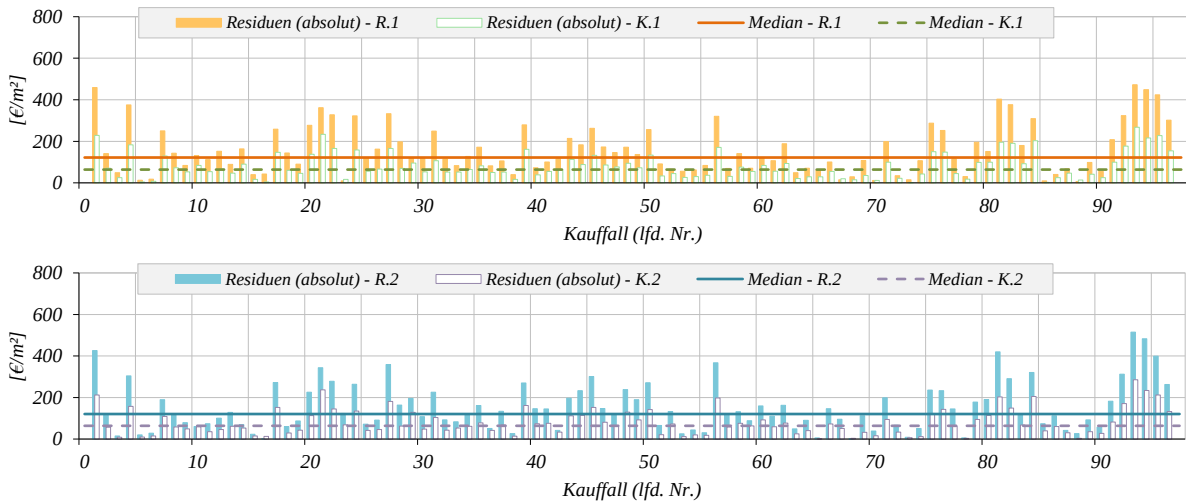


Abbildung C.61: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2005

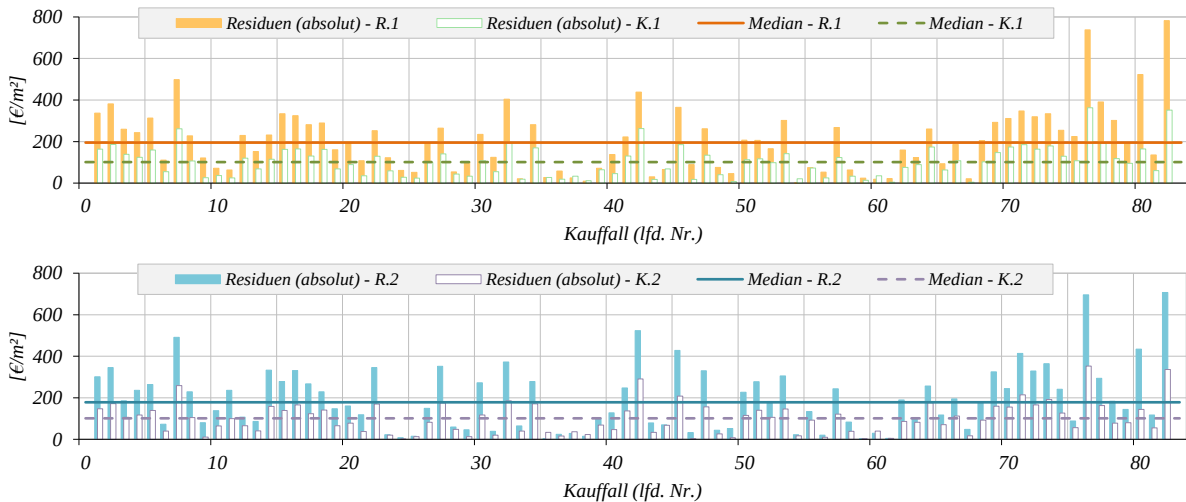


Abbildung C.62: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2006

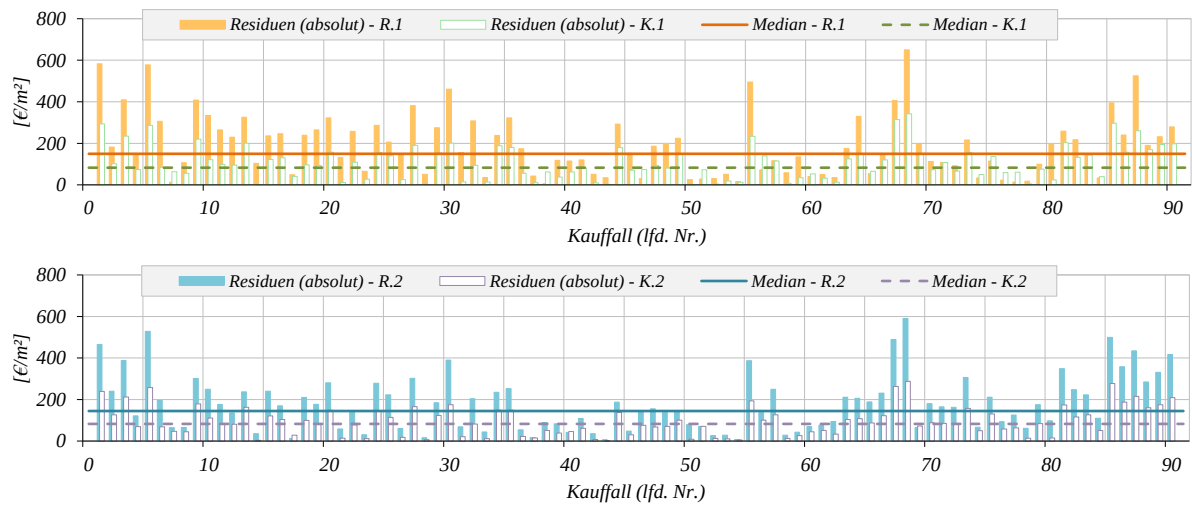


Abbildung C.63: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2007

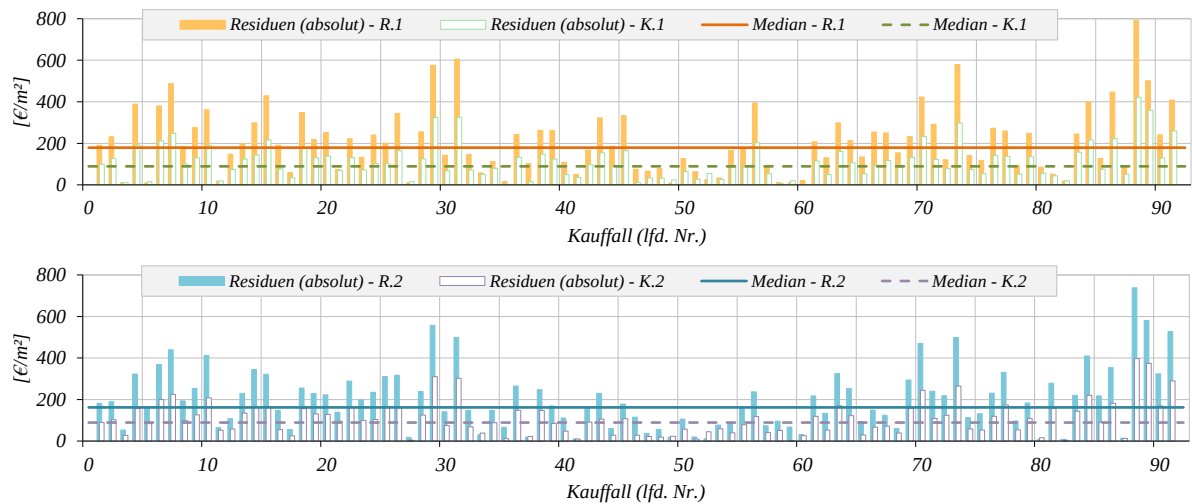


Abbildung C.64: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2008

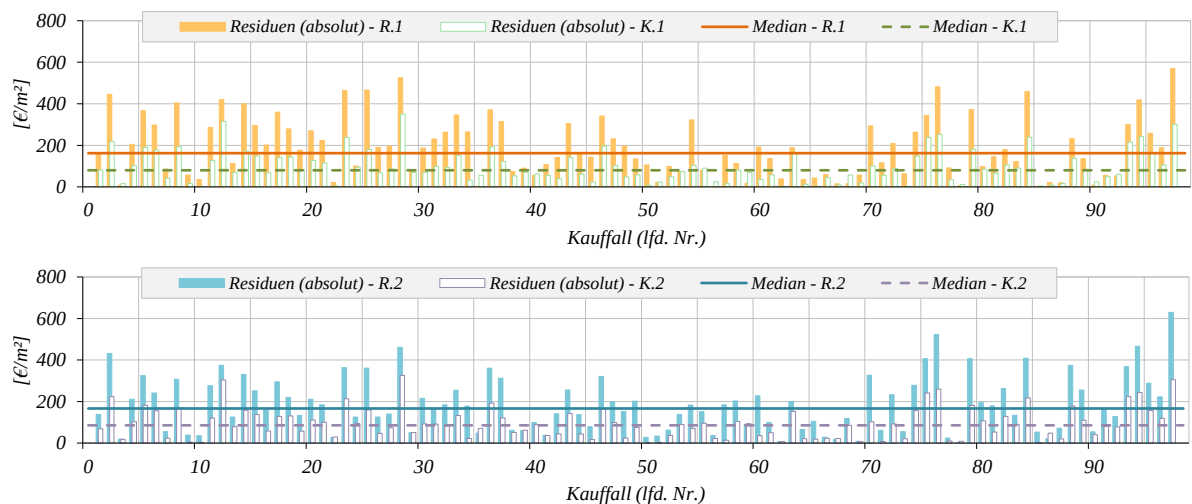


Abbildung C.65: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2009

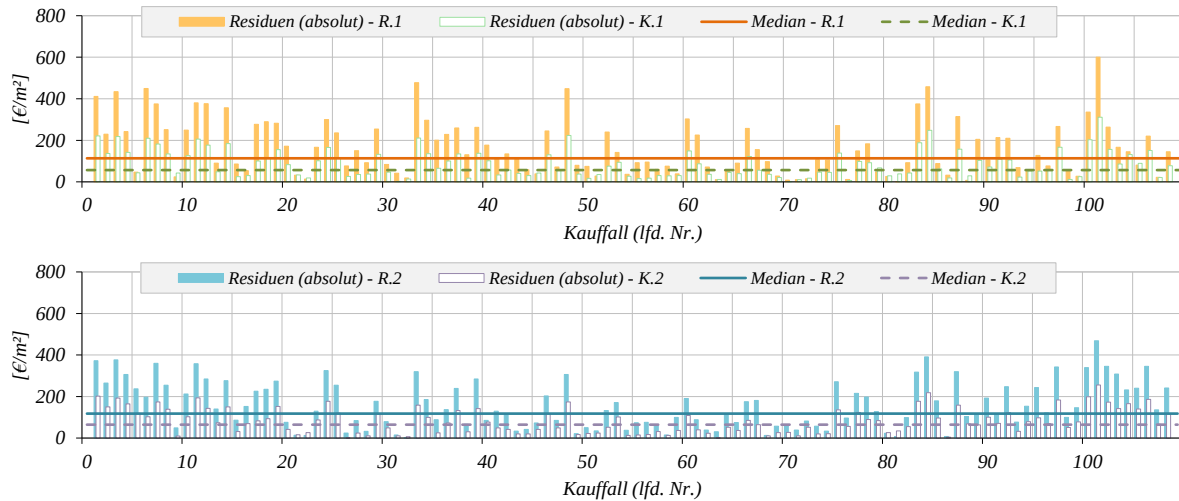


Abbildung C.66: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2010

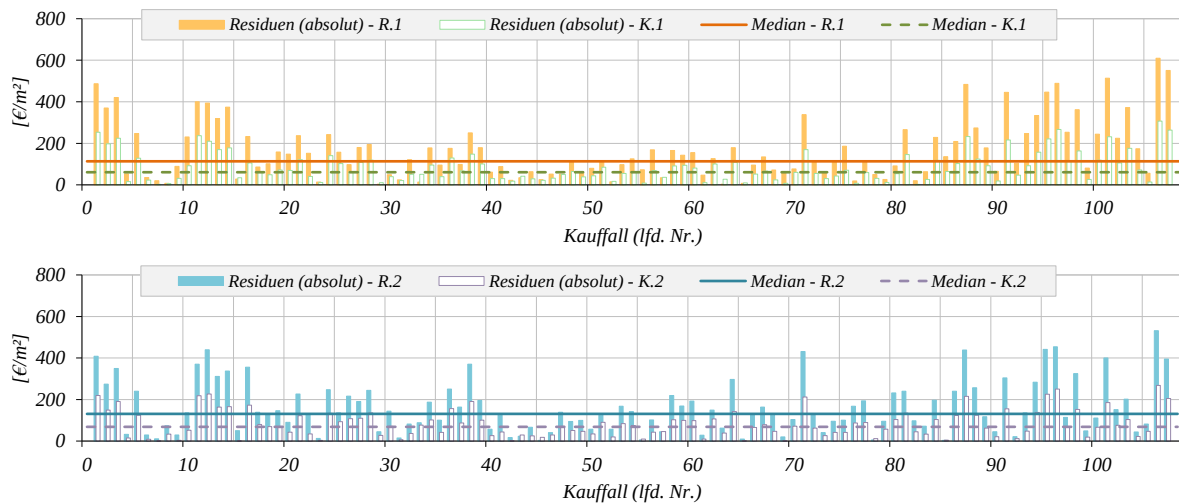


Abbildung C.67: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2011

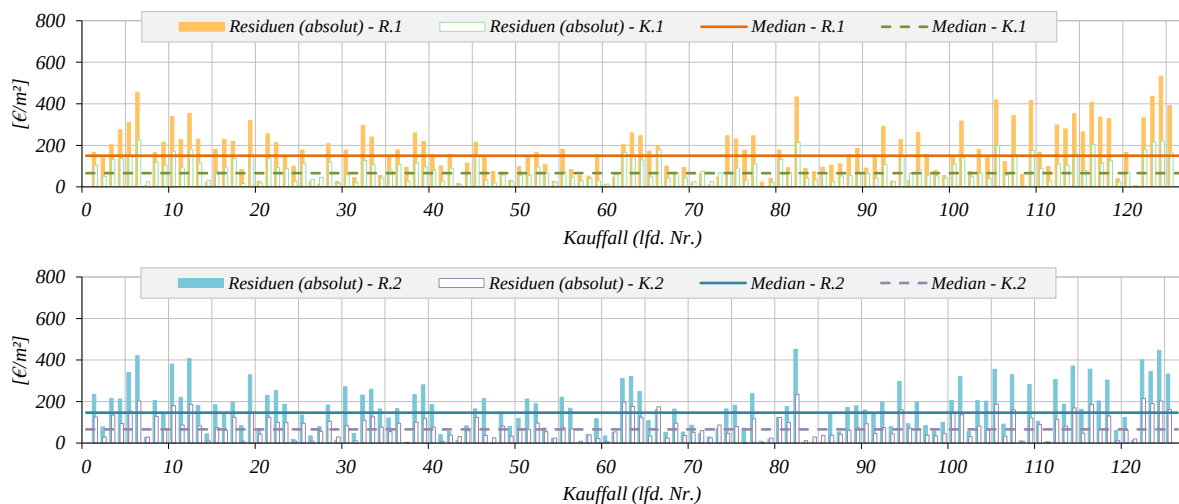


Abbildung C.68: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2012

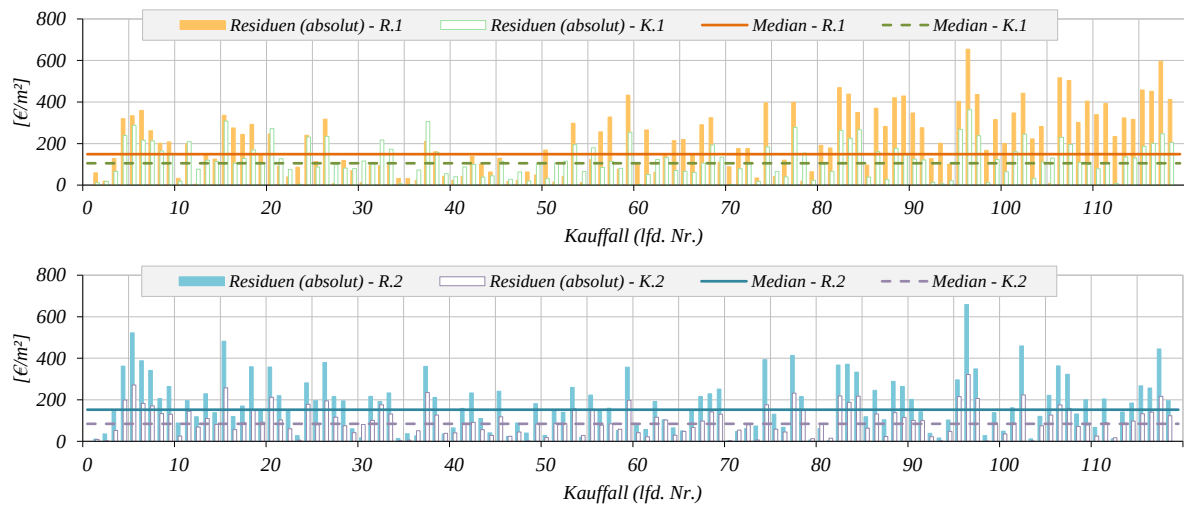


Abbildung C.69: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2013

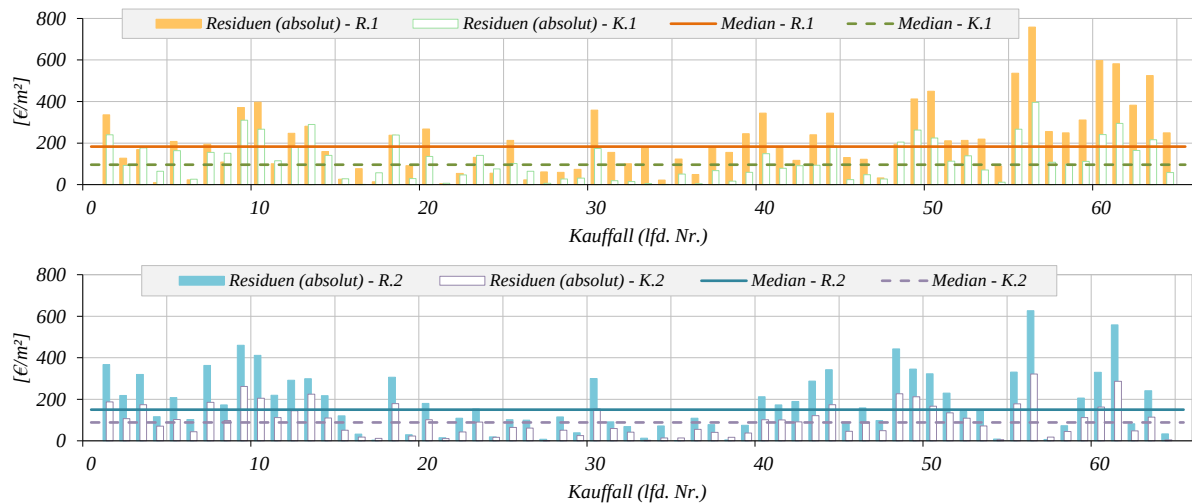


Abbildung C.70: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2014

C.6. Absolute Residuen der Bayesschen Modellansätze (Prädiktion)

In den folgenden Abbildungen C.71 bis C.80 wird eine Übersicht über die absoluten Residuen gegeben, die sich aus der Prädiktion im Bayesschen Vergleichswertverfahren ergeben und die anhand der Gleichungen

$$\begin{aligned} |\bar{e}_{p,R.1}| &= |\mathbf{y} - \bar{\mathbf{y}}_{p,R.1}| = |\mathbf{y} - \mathbf{X}_p \bar{\boldsymbol{\beta}}_{R.1}|, \\ |\bar{e}_{p,R.2}| &= |\mathbf{y} - \bar{\mathbf{y}}_{p,R.2}| = |\mathbf{y} - \mathbf{X}_p \bar{\boldsymbol{\beta}}_{R.2}|, \\ |\bar{e}_{p,K.1}| &= |\mathbf{y} - \bar{\mathbf{y}}_{p,K.1}| = |\mathbf{y} - \mathbf{X}_p \bar{\boldsymbol{\beta}}_{K.1} - \bar{s}_{p,K.1}|, \\ |\bar{e}_{p,K.2}| &= |\mathbf{y} - \bar{\mathbf{y}}_{p,K.2}| = |\mathbf{y} - \mathbf{X}_p \bar{\boldsymbol{\beta}}_{K.2} - \bar{s}_{p,K.2}|, \end{aligned}$$

erhalten werden. Die Prädiktionen werden anhand der Kreuzvalidierung erhalten, die in Abschnitt 4.4.1 (S. 105 f.) theoretisch behandelt und in Abschnitt 5.4 (S. 148 ff.) praktisch auf das gewählte Fallbeispiel angewendet wird. In den jeweils ersten Zeilen der Grafiken werden die absoluten Residuen der informativen Bayesschen Regression und der informativen Bayesschen Kollokation (*R.1* und *K.1*) gegenübergestellt, entsprechend in den jeweils zweiten Zeilen die Lösungen der nichtinformativen Modelle *R.2* und *K.2*. Die Sortierung bezieht sich auf die laufenden Kauffallnummern, die sich aus den Darstellungen der Abbildungen B.21 bis B.30 (S. 215 ff.) und C.42 bis C.60 (S. 235 ff.) ergeben. Für die absoluten Residuen werden ergänzend die Medianwerte angegeben.

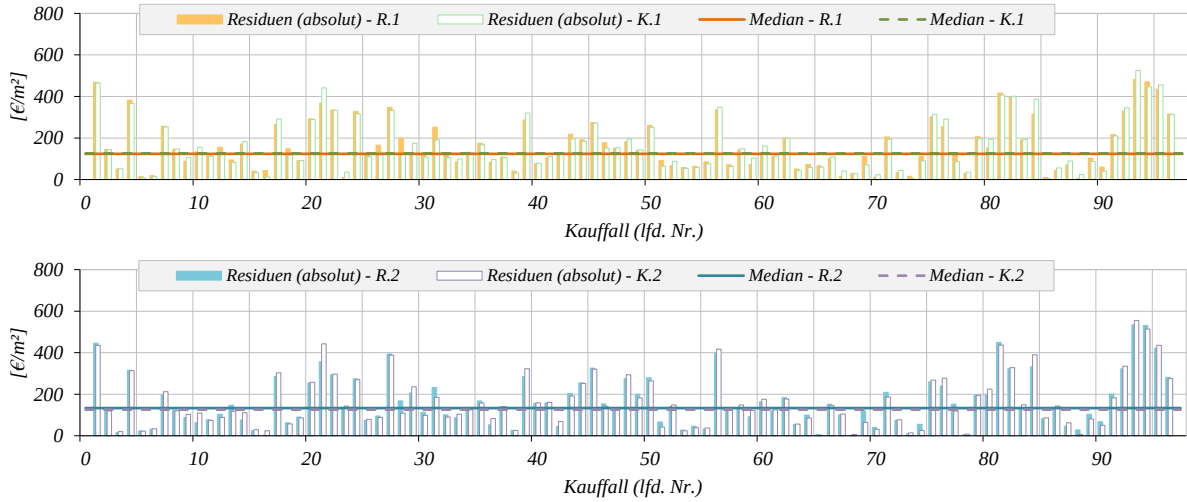


Abbildung C.71: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2005 (Prädiktion)

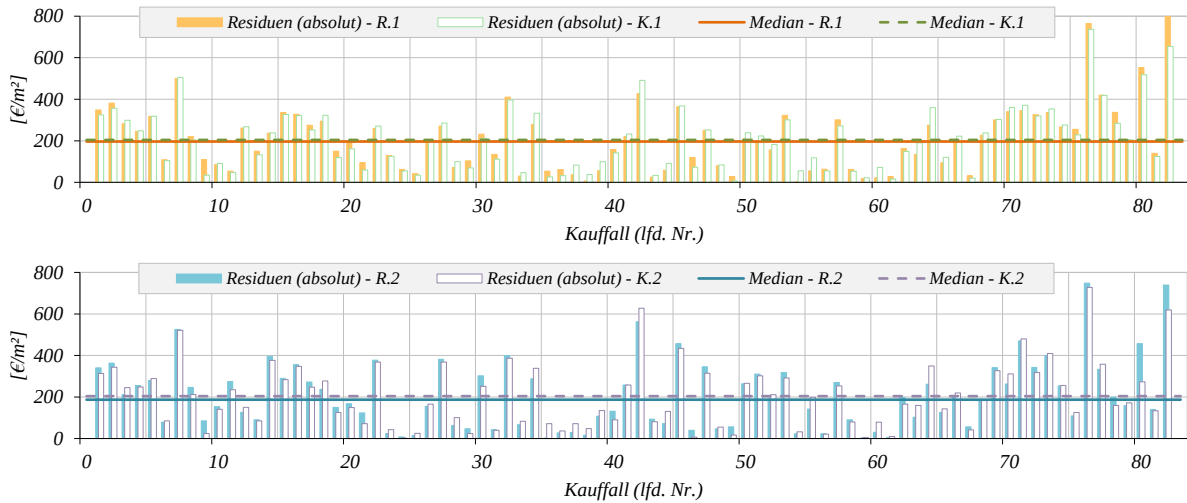


Abbildung C.72: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2006 (Prädiktion)

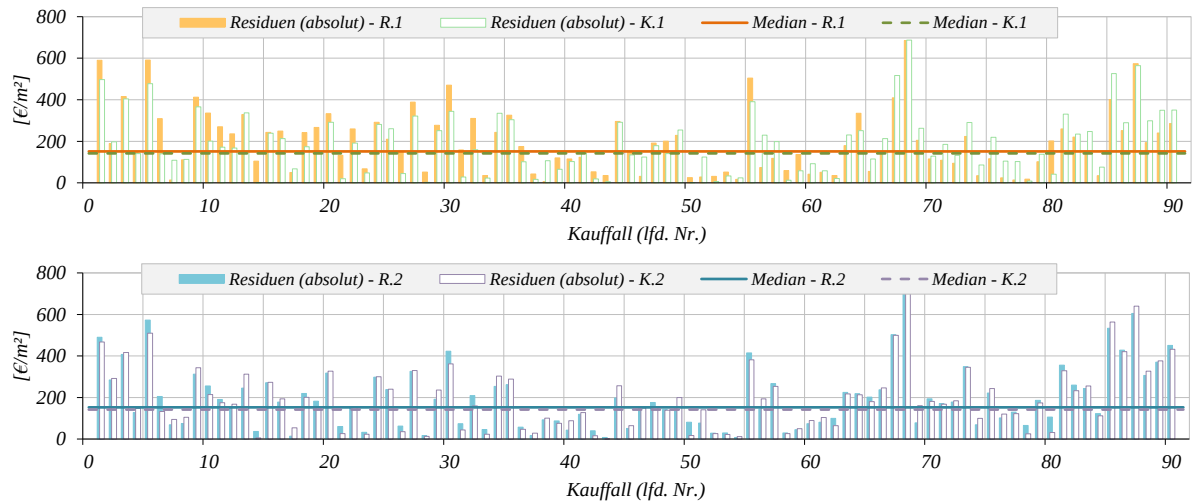


Abbildung C.73: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2007 (Prädiktion)

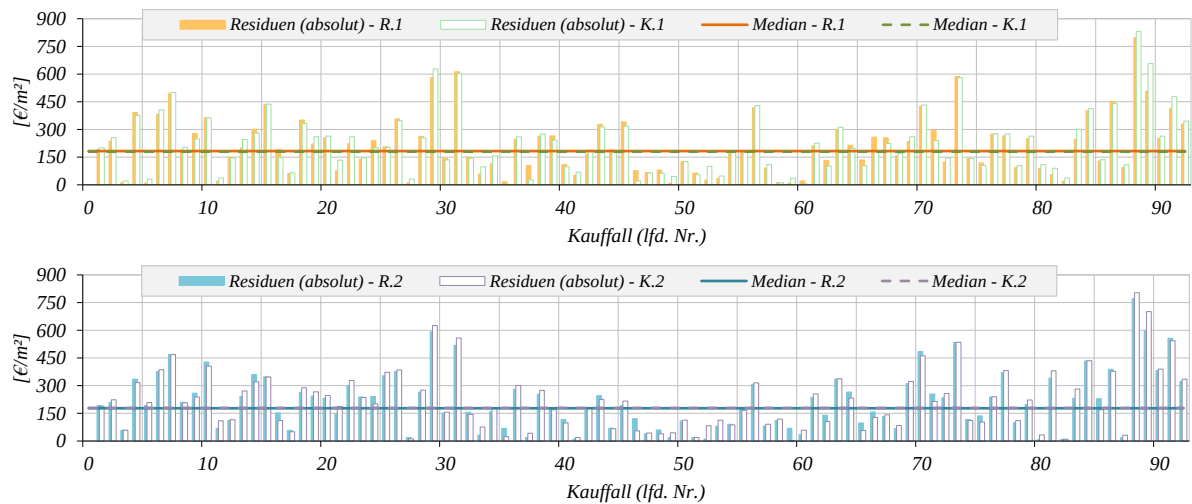


Abbildung C.74: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2008 (Prädiktion)

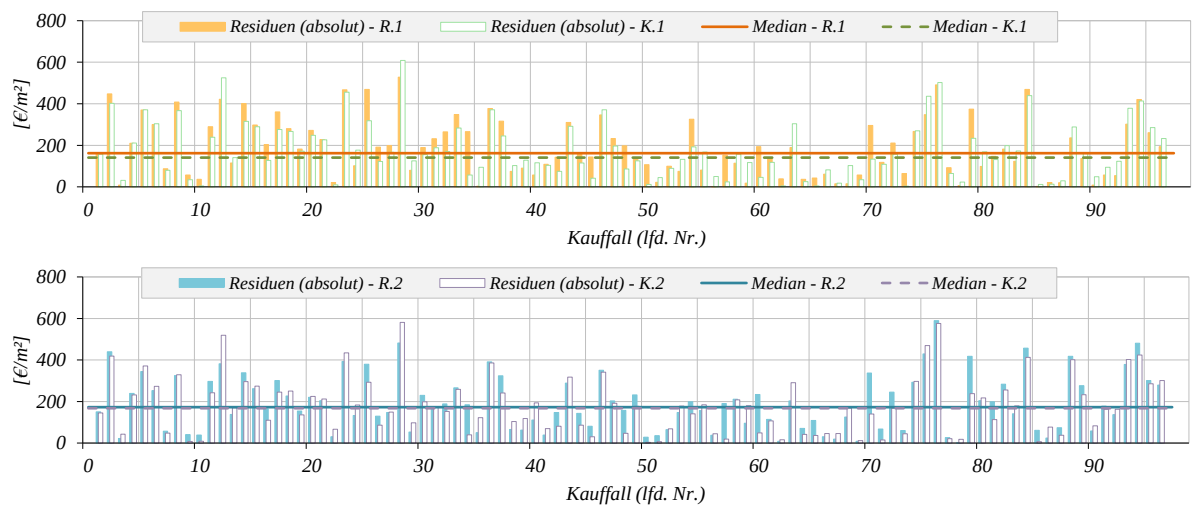


Abbildung C.75: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2009 (Prädiktion)

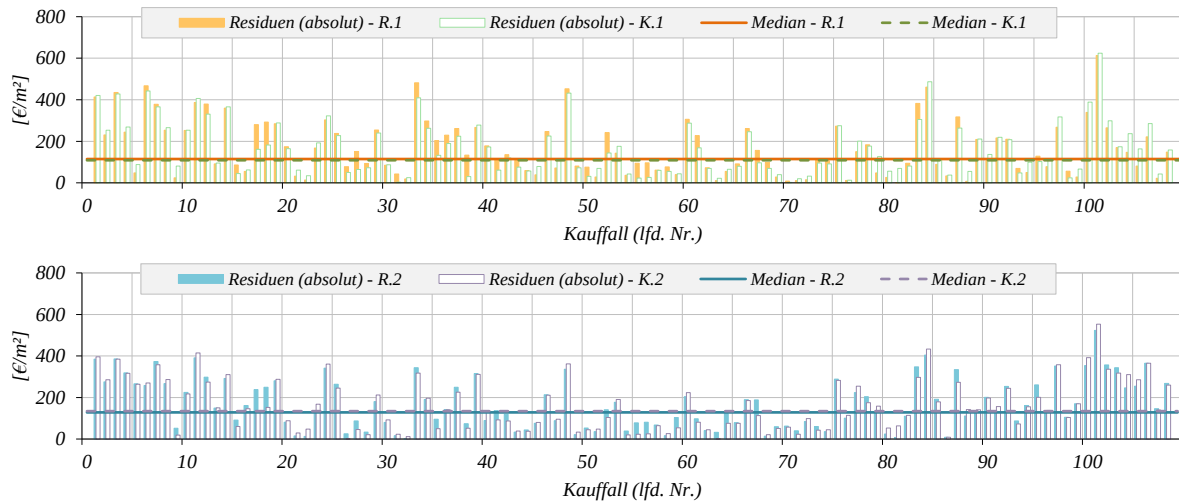


Abbildung C.76: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2010 (Prädiktion)

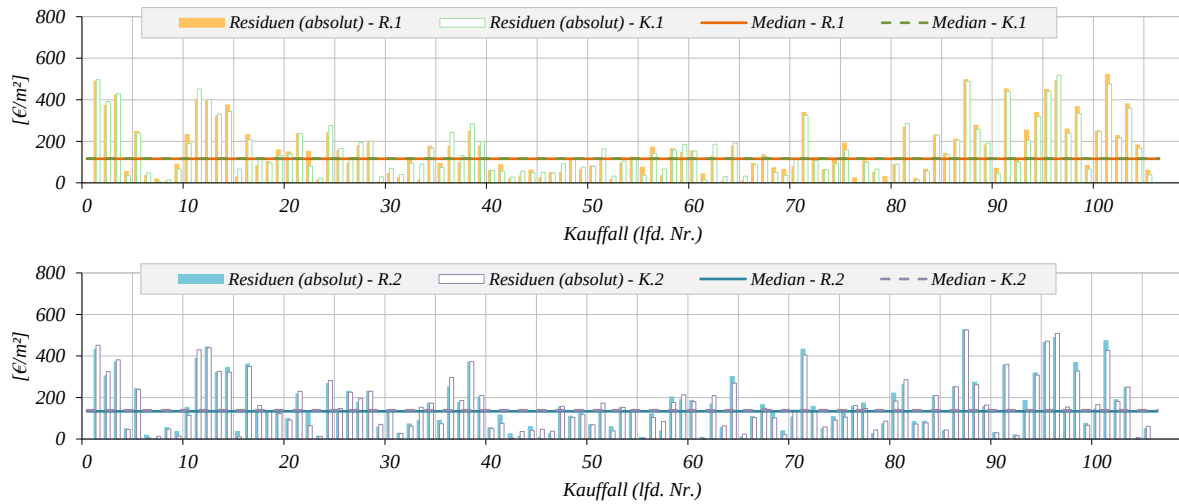


Abbildung C.77: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2011 (Prädiktion)

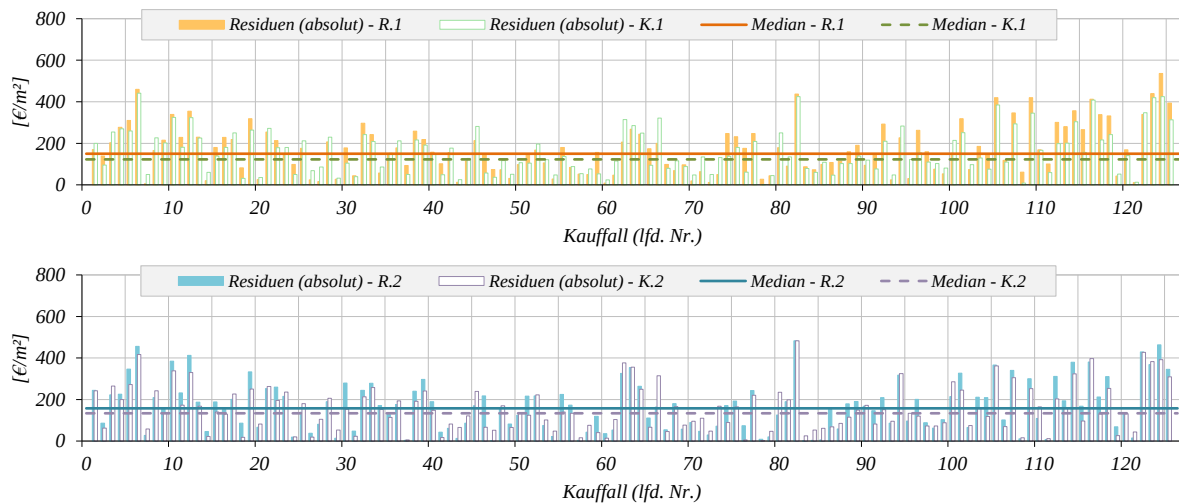


Abbildung C.78: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2012 (Prädiktion)

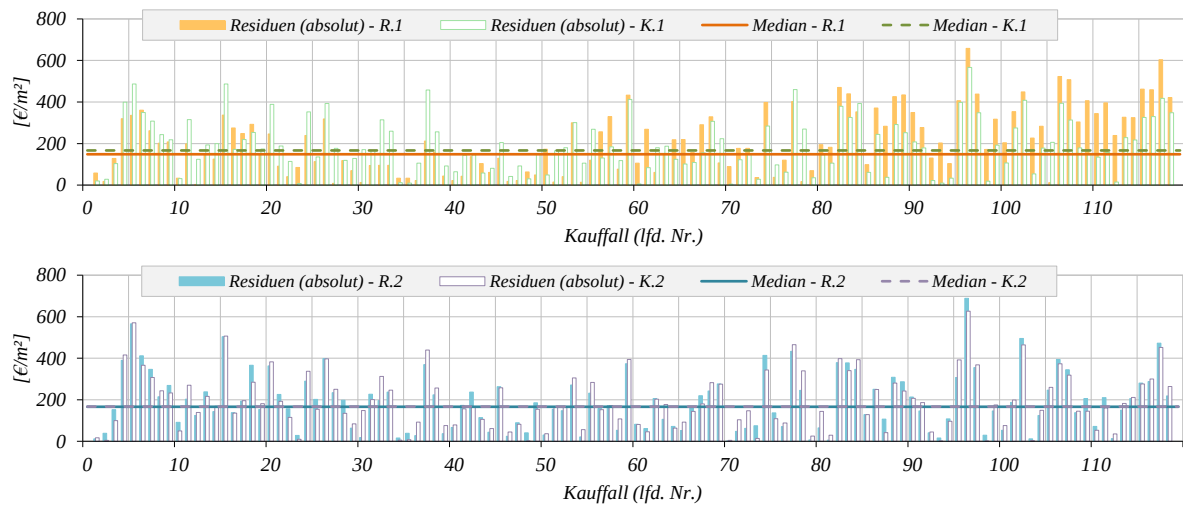


Abbildung C.79: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2013 (Prädiktion)

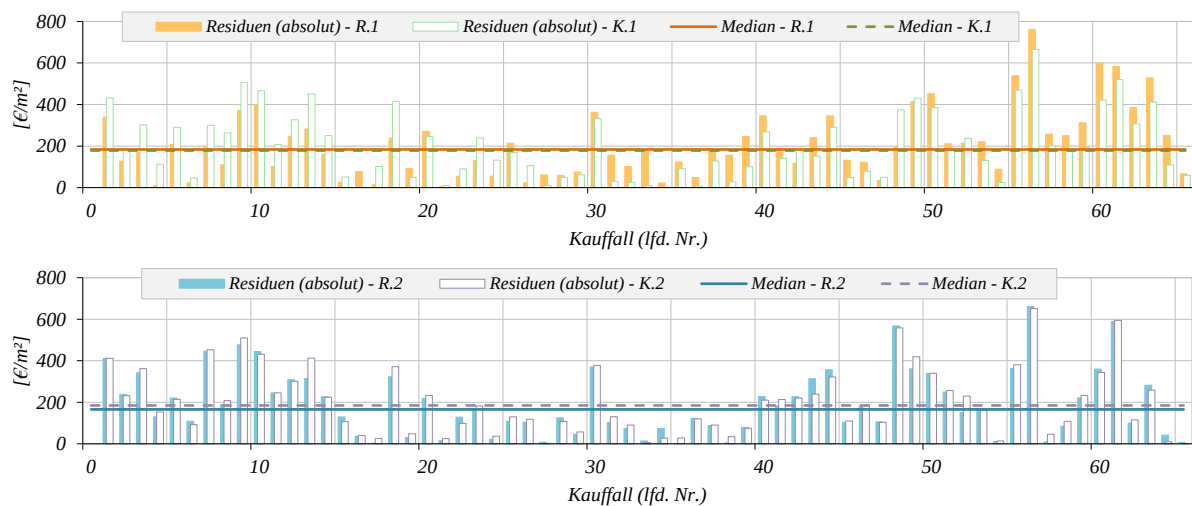


Abbildung C.80: Absolute Residuen im Vergleich der informativen und nichtinformativen Lösungen der Epoche 2014 (Prädiktion)

Curriculum vitae

Persönliche Informationen

Name	Sebastian Zaddach
Geburtsdatum	10. März 1981
Geburtsort	Osnabrück, Deutschland

Berufliche Tätigkeiten

04/2014 – aktuell	Landesamt für Geoinformation und Landesvermessung Niedersachsen Regionaldirektion Osnabrück-Meppen Dezernatsleiter
01/2011 – 03/2014	Gottfried Wilhelm Leibniz Universität Hannover Geodätisches Institut – Flächen- und Immobilienmanagement Projektmitarbeiter
10/2010 – 12/2010	Gottfried Wilhelm Leibniz Universität Hannover Geodätisches Institut – Ingenieurgeodäsie und geodätische Auswertemethoden Projektmitarbeiter
07/2003	Ingenieurplanung - Lubenow, Witschel und Partner GbR, Wallenhorst Messgehilfe

Hochschulstudium

10/2003 – 12/2007	Gottfried Wilhelm Leibniz Universität Hannover Fakultät für Bauingenieurwesen und Geodäsie Fachrichtung Geodäsie und Geoinformatik Abschluss: Diplom-Ingenieur
-------------------	---

Ausbildung

05/2008 – 06/2010	Landesvermessung und Geobasisinformation Niedersachsen Referendariat in der Laufbahn des höheren technischen Verwaltungsdienstes Abschluss: 2. Staatsexamen, Assessor des Vermessungs- und Liegenschaftswesens
08/2001 – 06/2003	Ingenieurplanung - Lubenow, Witschel und Partner GbR, Wallenhorst Berufsausbildung zum Bauzeichner (Tief-, Straßen- und Landschaftsbau) Abschluss: Bauzeichner

Wehr-/Zivildienst

07/2000 – 04/2001	allgemeine Wehrpflicht
-------------------	------------------------

Schulausbildung

08/1993 – 06/2000	Graf-Stauffenberg-Gymnasium, Osnabrück Abschluss: Allgemeine Hochschulreife
-------------------	--

Danksagung

An dieser Stelle möchte ich mich bei denen bedanken, die mich auf meinem langen Weg bis zur Fertigstellung der Arbeit begleitet und unterstützt haben.

Mein ganz besonderer Dank gilt Herrn Prof. Dr.-Ing. Winrich Voß. Bereits vor meiner beruflichen Tätigkeit am Geodätischen Institut Hannover hat er mir die Möglichkeit eröffnet, mich als externer Doktorand mit dem Promotionsvorhaben zu beschäftigen. Während meiner Zeit als wissenschaftlicher Mitarbeiter hatte ich die Gelegenheit, die ersten Ansätze und Gedanken zu vertiefen, weiterzuführen und schließlich zu einem gelungenen Abschluss zu führen. Für die vielen intensiven Diskussionen, wertvollen Anregungen zur Fokussierung der Inhalte und die persönliche Unterstützung während dieser Zeit möchte ich mich bei Herrn Voß herzlich bedanken.

Ein weiterer Dank gilt Herrn Prof. Dr.-Ing. Hansjörg Kutterer für die Übernahme eines Koreferats, die konstruktiven Diskussionen insbesondere zum Projektstart sowie die hilfreichen Anregungen zur inhaltlichen und formalen Strukturierung der Arbeit. Herrn Prof. Dr.-Ing. Hans Joachim Linke danke ich ebenfalls für die Übernahme eines Koreferats und die Unterstützung in der Endphase der Arbeit.

Einen sehr großen Anteil am erfolgreichen Abschluss tragen Prof. Dr.-Ing. Alexandra Weitkamp und Dr.-Ing. Hamza Alkhatib. Durch die Initiierung des DFG-Forschungsprojektes haben sie mir den Weg bereitet, mein Promotionsvorhaben am Geodätischen Institut Hannover umsetzen zu können. Ich möchte mich herzlich für die zahllosen intensiven Diskussionen, die vielen Anregungen und Hilfestellungen sowie die gemeinsame erfolgreiche Arbeit an Forschungs- und Institutsaufgaben bedanken. Danke für die gemeinsame Zeit am Institut und darüber hinaus.

Die Arbeit wäre zudem ohne die Unterstützung von Herrn Dr.-Ing. Jens-André Paffenholz nicht erfolgreich abgeschlossen worden. Ich danke ihm für die vielen Gespräche und Hilfen, insbesondere in der Endphase, ohne die ich noch heute über diverse LaTeX-Fallstricke stolpern würde und vermutlich am Ende eine handschriftliche Arbeit abgegeben hätte. Weiterhin danke ich allen Kolleginnen und Kollegen, die mich über eine mal kürzere, mal längere Zeit während meiner Tätigkeit am Geodätischen Institut begleitet haben und so ebenfalls ein Teil einer erfolgreichen Arbeit geworden sind. Für die Kollegialität, Freundschaften und Unterstützung bedanke ich mich herzlich.

Der größte Dank gilt jedoch meiner Familie: Für die vielen, vielen Jahre (und Jahrzehnte) der Unterstützung. Meine Eltern Gabriele und Michael Zaddach haben es mir ermöglicht, meinen Weg zu gehen und dabei immer auf ihren Rückhalt vertrauen zu können. Meine Familie hat nie einen Zweifel daran gehabt, dass ich meine Arbeit erfolgreich abschließe und hat mich immer wieder bestärkt, wenn für mich das Licht am Ende des Tunnels nicht mehr zu erkennen war. Ich danke Rosi Filp, die es mir mit ihrer Liebe und Unterstützung überhaupt erst ermöglicht hat, mein Ziel zu erreichen und mir zu jeder Zeit den Rückhalt und die Kraft gegeben hat, weiterzugehen und so nicht nur im Sinne eines erfolgreichen Abschlusses der Arbeit endlich anzukommen.