Deutsche Geodätische Kommission
bei der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 674

Boris Kargoll

On the Theory and Application of Model Misspecification Tests in Geodesy

München 2012

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

ISSN 0065-5325 ISBN 978-3-7696-5086-0

Diese Arbeit ist gleichzeitig veröffentlicht in:
Schriftenreihe des Instituts für Geodäsie und Geoinformation
der Rheinischen Friedrich-Wilhelms Universität Bonn
ISSN 1864-1113, Nr. 8, Bonn 2008

DGK Deutsche Geodätische Kommission
bei der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 674

On the Theory and Application of Model Misspecification Tests in Geodesy

Inaugural-Dissertation zur
Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)
der Hohen Landwirtschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms Universität
zu Bonn

vorgelegt am 08.05.2007 von

Dipl.-Ing. Boris Kargoll

aus Karlsruhe

München 2012

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

ISSN 0065-5325 ISBN 978-3-7696-5086-0

Diese Arbeit ist gleichzeitig veröffentlicht in:
Schriftenreihe des Instituts für Geodäsie und Geoinformation
der Rheinischen Friedrich-Wilhelms Universität Bonn
ISSN 1864-1113, Nr. 8, Bonn 2008

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

Hauptberichterstatter: Prof. Dr.techn. W.-D. Schuh
Mitberichterstatter: Prof. Dr.rer.nat. H.-P. Helfrich

Tag der mündlichen Prüfung: 20.06.2007

ISSN 0065-5325 ISBN 978-3-7696-5086-0

On the Theory and Application of Model Misspecification Tests in Geodesy

Abstract

Many geodetic testing problems concerning parametric hypotheses may be formulated within the framework of testing linear constraints imposed on a linear Gauss-Markov model. Although geodetic standard tests for such problems are computationally convenient and intuitively sound, no rigorous attempt has yet been made to derive them from a unified theoretical foundation or to establish optimality of such procedures. Another shortcoming of current geodetic testing theory is that no standard approach exists for tackling analytically more complex testing problems, concerning for instance unknown parameters within the weight matrix.

To address these problems, it is proven that, under the assumption of normally distributed observation, various geodetic standard tests, such as Baarda's or Pope's test for outliers, multivariate significance tests, deformation tests, or tests concerning the specification of the a priori variance factor, are uniformly most powerful (UMP) within the class of invariant tests. UMP invariant tests are proven to be equivalent to likelihood ratio tests and Rao's score tests. It is also shown that the computation of many geodetic standard tests may be simplified by transforming them into Rao's score tests.

Finally, testing problems concerning unknown parameters within the weight matrix such as autoregressive correlation parameters or overlapping variance components are addressed. It is shown that, although strictly optimal tests do not exist in such cases, corresponding tests based on Rao's Score statistic are reasonable and computationally convenient diagnostic tools for deciding whether such parameters are significant or not. The thesis concludes with the derivation of a parametric test of normality as another application of Rao's Score test.

Zur Theorie und Anwendung von Modell-Misspezifikationstests in der Geodäsie

Zusammenfassung

Was das Testen von parametrischen Hypothesen betrifft, so lassen sich viele geodätische Testprobleme in Form eines Gauss-Markov-Modells mit linearen Restriktionen darstellen. Obwohl geodätische Standardtests rechnerisch einfach und intuitiv vernünftig sind, wurde bisher kein strenger Versuch unternommen, solche Tests ausgehend von einer einheitlichen theoretischen Basis herzuleiten oder die Optimalität solcher Tests zu begründen. Ein weiteres Defizit im gegenwärtigen Verständnis geodätischer Testtheorie besteht darin, dass kein Standardverfahren zum Lösen von analytisch komplexeren Testproblemen exisitiert, welche beispielsweise unbekannte Parameter in der Gewichtsmatrix betreffen.

Um diesen Problemen gerecht zu werden wird bewiesen, dass unter der Annahme normalverteilter Beobachtungen verschiedene geodätische Standardtests, wie z.B. Baardas oder Popes Ausreissertest, multivariate Signifikanztests, Deformationstests, oder Tests bzgl. der Angabe des a priori Varianzfaktors, allesamt gleichmäßig beste (engl.: uniformly most powerful - UMP) invariante Tests sind. Es wird ferner bewiesen dass UMP invariante Tests äquivalent zu Likelihood-Quotienten-Tests und Raos Score-Tests sind. Ausserdem wird gezeigt, dass sich die Berechnung vieler geodätischer Standardtests vereinfachen lässt indem diese als Raos Score-Tests formuliert werden.

Abschließend werden Testprobleme behandelt in Bezug auf unbekannte Parameter innerhalb der Gewichtsmatrix, beispielsweise in Bezug auf autoregressive Korrelationsparameter oder überlappende Varianzkomponenten. In solchen Fällen existieren keine im strengen Sinne besten Tests. Es wird aber gezeigt, dass entsprechende Tests, die auf Raos Score-Statistik beruhen, sinnvolle und vom Rechenaufwand her günstige Diagnose-Tools darstellen um festzustellen, ob Parameter wie die eingangs erwähnten signifikant sind oder nicht. Am Ende dieser Dissertation steht mit der Herleitung eines parametrischen Tests auf Normalverteilung eine weitere Anwendung von Raos Score-Test.

Contents

1 Introduction

1.1 Objective

Hypothesis testing is the foundation of all critical model analyses. Particularly relevant to geodesy is the practice of model misspecification testing which has the objective of determining whether a given observation model accurately describes the physical reality of the data. Examples of common testing problems include how to detect outliers, how to determine whether estimated parameter values or changes thereof are significant, or how to verify the measurement accuracy of a given instrument. Geodesists know how to handle such problems intuitively using standard parameter tests, but it often remains unclear in what mathematical sense these tests are optimal.

The first goal of this thesis is to develop a theoretical foundation which allows establishing optimality of such tests. The approach will be based on the theory of Neyman and Pearson (1928, 1933), whose celebrated fundamental lemma defines an optimal test as one which is most powerful among all tests with some particular significance level. As this concept is applicable only to very simple problems, tests must be considered that are most powerful in a wider sense. An intuitively appealing way to do so is based on the fact that complex testing problems may often be reduced to simple problems by exploiting symmetries. One mathematical description of symmetry is invariance, whose application to testing problems then leads to invariant tests. In this context, a uniformly most powerful invariant test defines a test which is optimal among all invariant tests available in the given testing problem. In this thesis, it will be demonstrated for the first time that many geodetic standard tests fit into this framework and share the property of being uniformly most powerful.

In order to be useful in practical situations, a testing procedure should not only be optimal, but it must also be computationally manageable. It is well known that hypothesis tests have different mathematical descriptions, which may vary considerably in computational complexity. Most geodetic standard tests are usually derived from likelihood ratio tests (see, for instance, Koch, 1999; Teunissen, 2000). An alternative, oftentimes much simpler representation is based on Rao's (1948) score test, which has not been acknowledged as such by geodesists although it has found its way into geodetic practice, for instance, via Baarda's outlier test. To shed light on this important topic, it is another major intent of this thesis to describe Rao's score method in a general and systematic way, and to demonstrate what types of geodetic testing problems are ideally handled by this technique.

1.2 Outline

The following Section 2 of this thesis begins with a review of classical testing theory. The focus is on *parametric testing problems*, that is, hypotheses to be tested are propositions concerning parameters of the data's probability distribution. We will then follow the classical approach of considering tests with fixed significance level and maximum power. In this context, the *Neyman-Pearson Lemma* and the resulting idea of a *most powerful test* will be explained, and the concept of a *uniformly most powerful test* will be introduced. The subsequent definition of *sufficiency* will play a central role in reducing the complexity of testing problems. Following this, we will examine more complex problems that require a simplification going beyond sufficiency. For this puropse, we will use the principle of *invariance*, which is the mathematical description of symmetry. We will see that invariant tests are tests with power distributed symmetrically over the space of parameters. This leads us to the notion of a *uniformly most powerful invariant (UMPI) test*, which is a designated optimal test among such invariant tests. Finally, we will explore the relationships of UMPI tests to *likelihood ratio tests* and *Rao's score tests*.

Section 3 extends the ideas developed in Section 2 to address the general problem of testing linear hypotheses in the Gauss-Markov model with normally distributed observations. Here we focus on the case in which the design matrix is of full rank and where the weight matrix is known. Then, the testing problem will be reduced by sufficiency and invariance, and UMPI tests derived for the two cases where the variance of unit weight is either known or unknown a priori. Emphasis will be placed on demonstrating further that these UMPI tests correspond to the tests already used in geodesy. Another key result of this section will be to show how all these tests are formulated as likelihood ratio and Rao's score tests. The section concludes with a discussion of various geodetic testing problems. It will be shown that many standard tests used so far, such as Baarda's and Pope's outlier test, multivariate parameter tests, deformation tests, or tests concerning the variance of unit weight, are optimal (UMPI) in a statistical sense, but that computational complexity can often be effectively reduced by using equivalent Rao's score tests instead.

Section 4 addresses a number of testing problems in generalized Gauss-Markov models for which no UMPI tests exist, because a reduction by sufficiency and invariance are not effective. The first problem considered will be testing for first-order autoregressive correlation. Rao's score test will be derived, and its power against several simple alternative hypotheses will be determined by carrying out a Monte Carlo simulation. The second application of this section will treat the case of testing for a single overlapping variance component,for which Rao's score test will be once again derived. The final problem consists of testing whether observations follow a normal distribution. It this situation, Rao's score test will be shown to lead to a test which measures the deviation of the sample's skewness and kurtosis from the theoretical values of a normal distribution.

Finally, Section 5 highlights the main conclusions of this work and gives an outlook on promising extensions to the theory and applications of the approach presented in this thesis.

2 Theory of Hypothesis Testing

2.1 The observation model

Let us assume that some data vector $\boldsymbol{y} = [y_1, \ldots, y_n]'$ is subject to a statistical analysis. As this thesis is concerned rather with exploring theoretical aspects of such analyses, it will be useful to see this data vector as one of many potential realizations of a vector $\boldsymbol{Y}$ of observables $Y_1, \ldots, Y_n$. This is reflected by the fact that measuring the same quantity multiple times does not result in identical data values, but rather in some frequency distribution of values according to some random mechanism. In geodesy, quantities that are subject to observation or measurement usually have a geometrical or physical meaning. In this sense, $\boldsymbol{Y}$, or its realization $\boldsymbol{y}$, will be viewed as being incorporated in some kind of model and thereby connected to some other quantities or parameters. Parametric observation models may be set up for multiple reasons. They are often used as a way to reduce great volumes of raw data to low-dimensional approximating functions. A model might also be used simply because the quantity of primary interest is not directly observable, but must be derived from other data. In reality, both aspects often go hand in hand.

To give these explanations a mathematical expression, let the random vector $\boldsymbol{Y}$ with values in $\mathbb{R}^n$ be part of a linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}, \tag{2.1-1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^m$ denotes a vector of unknown non-stochastic parameters and $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ a known matrix of non-stochastic coefficients reflecting the functional relationship. It will be assumed throughout that rank$\boldsymbol{A} = m$ and that $n > m$ so that (2.1-1) constitutes a genuine adjustment problem. $\boldsymbol{E}$ represents a real-valued random vector of unknown disturbances or errors, which are assumed to satisfy

$$\mathbf{E}\{\boldsymbol{E}\} = \mathbf{0} \quad \text{and} \quad \boldsymbol{\Sigma}\{\boldsymbol{E}\} = \sigma^2 \, \boldsymbol{P}_{\boldsymbol{\omega}}^{-1}. \tag{2.1-2}$$

We will occasionally refer to these two conditions as the *Markov conditions*. The weight matrix $\boldsymbol{P}_{\boldsymbol{\omega}}$ may be a function of unknown parameters $\boldsymbol{\omega}$, which allows for certain types of correlation and variance-change (or *heteroscedasticity*) models regarding the errors. Whenever such parameters do not appear, we will use $\boldsymbol{P}$ to denote the weight matrix.

To make the following testing procedures operable, these linear model specifications must be accompanied by certain assumptions regarding the type of probability distribution considered for $\boldsymbol{Y}$. For this purpose, it will be assumed that any such distribution P may be defined by a parametric density function

$$f(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\omega}, \boldsymbol{c}), \tag{2.1-3}$$

which possibly depends on additional unknown shape parameters $\boldsymbol{c}$ controlling, for instance, the skewness and kurtosis of the distribution. Now, let the vector $\boldsymbol{\theta} := [\boldsymbol{\beta}', \sigma^2, \boldsymbol{\omega}', \boldsymbol{c}']'$ comprise the totality of unknown parameters taking values in some u-dimensional space $\boldsymbol{\Theta}$. The **parameter space** $\boldsymbol{\Theta}$ then corresponds to a collection of densities

$$\mathcal{F} = \{f(\boldsymbol{y}; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}, \tag{2.1-4}$$

which in turn defines the contemplated collection of distributions

$$\mathcal{W} = \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}. \tag{2.1-5}$$

Example 2.1: An angle has been independently observed n times. Each observation $Y_1, \ldots, Y_n$ is assumed to follow a distribution that belongs to the class of normal distributions

$$\mathcal{W} = \left\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\right\} \tag{2.1-6}$$

with mean μ and variance σ^2, or in short notation $Y_i \sim N(\mu, \sigma^2)$. The relationship between $\boldsymbol{Y} = [Y_1, \ldots, Y_n]'$ and the mean parameter μ constitutes the simplest form of a linear model (2.1-1), where $\boldsymbol{X}$ is an $(n \times 1)$-vector of ones and $\boldsymbol{\beta}$ equals the single parameter μ. Furthermore, as the observations are independent with constant mean and variance, the joint normal density function $f(\boldsymbol{y}; \mu, \sigma^2)$ may be decomposed (i.e. *factorized*) into the product

$$f(\boldsymbol{y}; \mu, \sigma^2) = \prod_{i=1}^{n} f(y_i; \mu, \sigma^2) \tag{2.1-7}$$

of identical univariate normal density functions defined by

$$f(y_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{y_i - \mu}{\sigma}\right)^2\right\} \qquad (y_i \in \mathbb{R}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+, i = 1, \ldots, n). \tag{2.1-8}$$

Therefore, the class of densities $\mathcal{F}$ considered for $\boldsymbol{Y}$ may be written as

$$\mathcal{F} = \left\{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2\right\} : [\mu, \sigma^2]' \in \boldsymbol{\Theta}\right\} \tag{2.1-9}$$

with two-dimensional parameter space $\boldsymbol{\Theta} = \mathbb{R} \times \mathbb{R}^+$. □

2.2 The testing problem

The goal of any parametric statistical inference is to extract information from the given data $\boldsymbol{y}$ about the unknown true parameters $\bar{\boldsymbol{\theta}}$, which refer to the unknown true probability distribution $P_{\bar{\boldsymbol{\theta}}}$ and the true density function $f(\boldsymbol{y}; \bar{\boldsymbol{\theta}})$ with respect to the observables $\boldsymbol{Y}$. For this purpose, we will assume that $\bar{\boldsymbol{\theta}}$, $P_{\bar{\boldsymbol{\theta}}}$, and $f(\boldsymbol{y}; \bar{\boldsymbol{\theta}})$ are unique and identifiable elements of $\boldsymbol{\Theta}$, $\mathcal{W}$, and $\mathcal{F}$ respectively.

While **estimation** aims at determining the numerical values of $\bar{\boldsymbol{\theta}}$, that is, selecting one specific element from $\boldsymbol{\Theta}$, the goal of **testing** is somewhat simpler in that one only seeks to determine whether $\bar{\boldsymbol{\theta}}$ is an element of a subset $\boldsymbol{\Theta}_0$ of $\boldsymbol{\Theta}$ or not. Despite this seemingly great difference between the purpose of estimation and testing, which is reflected by a separate treatment of both topics in most statistical text books, certain concepts from estimation will turn out to be indispensable for the theory of testing. As this thesis is focussed on testing, the necessary estimation methodology will be introduced without a detailed analysis thereof.

In order to formulate the test problem, a non-empty and genuine subset $\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$ (corresponding to some $\mathcal{W}_0 \subset \mathcal{W}$ and $\mathcal{F}_0 \subset \mathcal{F}$) must be specified. Then, the **null hypothesis** is defined as the proposition

$$H_0 : \bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_0. \tag{2.2-10}$$

When the null hypothesis is such that $\boldsymbol{\Theta}_0$ represents one point $\boldsymbol{\theta}_0$ within the parameter space $\boldsymbol{\Theta}$, then the elements of $\boldsymbol{\theta}_0$ assign unique numerical values to all the elements in $\bar{\boldsymbol{\theta}}$, and (2.2-10) simplifies to the proposition

$$H_0 : \bar{\boldsymbol{\theta}} = \boldsymbol{\theta}_0. \tag{2.2-11}$$

In this case, H_0 is called a **simple null hypothesis**. On the other hand, if at least one element of $\bar{\boldsymbol{\theta}}$ is assigned a whole range of values, say $\mathbb{R}^+$, then H_0 is called a **composite null hypothesis**. In such a case, an equality relation as in (2.2-11) can clearly not be established for all the parameters in $\bar{\boldsymbol{\theta}}$. Unknown parameters whose true values are not uniquely fixed under H_0 are also called **nuisance parameters**.

Example 2.2 (Example 2.1 continued): On the basis of given observed numerical values $\boldsymbol{y} = [y_1, \ldots, y_n]'$, we want to test whether the observed angle is an exact right angle ($100\,gon$) or not. Let us investigate three different scenarios:

1. If σ^2 is known *a priori* to take the true value σ_0^2, then $\boldsymbol{\Theta} = \mathbb{R}$ is one-dimensional, and under the null hypothesis $H_0 : \bar{\mu} = 100$ the subset $\boldsymbol{\Theta}_0$ shrinks to the single point

   $$\boldsymbol{\Theta}_0 = \{100\}.$$

   Hence, H_0 is a simple null hypothesis by definition.

2. If μ and σ^2 are both unknown, then the null hypothesis, written as

   $$H_0 : \bar{\mu} = 100 \ \ (\sigma^2 \in \mathbb{R}^+),$$

   leaves the nuisance parameter σ^2 unspecified. Therefore, the subset

   $$\boldsymbol{\Theta}_0 = \left\{(100, \sigma^2) : \sigma^2 \in \mathbb{R}^+\right\}$$

   does not specify a single point, but an interval of values. Consequently, H_0 is composite under this scenario.

3. If the question is whether the observed angle is a $100\,gon$ <u>and</u> the standard deviation is really $3\,mgon$ (e.g. as promised by the producer of the instrument), then the null hypothesis

   $$H_0 : \bar{\mu} = 100, \bar{\sigma} = 0.003$$

   refers to the single point $\boldsymbol{\Theta}_0 = (100, 0.003^2)$ within $\boldsymbol{\Theta}$. In that case, H_0 is seen to be simple. □

2.3 The test decision

Imagine that the space S of all possible observations $\boldsymbol{y}$ consists of two complementary regions: a **region of acceptance** S_A, which consists of all values that support a certain null hypothesis H_0, and a **region of rejection** or **critical region** S_C, which comprises all the observations that contradict H_0 in some sense. A *test decision* could then be on based simply observing whether some given data values $\boldsymbol{y}$ are in S_A (which would imply acceptance of H_0), or whether $\boldsymbol{y} \in S_C$ (which would result in rejection of H_0).

It will be necessary to perceive any test decision as the realization of a random variable ϕ which, as a function of $\boldsymbol{Y}$, takes the value 1 in case of rejection and 0 in case of acceptance of H_0. This mapping, defined as

$$\phi(\boldsymbol{y}) = \begin{cases} 1, & \text{if } \boldsymbol{y} \in S_C, \\ 0, & \text{if } \boldsymbol{y} \in S_A, \end{cases} \tag{2.3-12}$$

is also called a **test** or **critical function**, for it indicates whether a given observation $\boldsymbol{y}$ falls into the *critical region* or not. (2.3-12) can be viewed as the mathematical implementation of a binary **decision rule**, which is typical for test problems. This notion now allows for the more formal definition of the regions S_A and S_C as

$$S_C = \phi^{-1}(1) = \{\boldsymbol{y} \in S \mid \phi(\boldsymbol{y}) = 1\}, \tag{2.3-13}$$

$$S_A = \phi^{-1}(0) = \{\boldsymbol{y} \in S \mid \phi(\boldsymbol{y}) = 0\}. \tag{2.3-14}$$

Example 2.3 (Ex. 2.2 continued): For simplicity, let Y $(n = 1)$ be the single observation of an angle, which is assumed to be normally distributed with unknown mean $\bar{\mu}$ and known standard deviation $\bar{\sigma} = \sigma_0 = 3\,mgon$. To test the hypothesis that the observed angle is a right angle ($H_0 : \bar{\mu} = 100$), an engineer suggests the following decision rule: Reject H_0, when the observed angle deviates from $100\,gon$ by at least five times the standard deviation. The critical function reads

$$\phi(\boldsymbol{y}) = \begin{cases} 1, & \text{if } y \leq 99.985 \text{ or } y \geq 100.015 \\ 0, & \text{if } 99.985 < y < 100.015. \end{cases} \tag{2.3-15}$$

The critical region is given by $S_C = (-\infty, 99.985] \cup [100.015, +\infty)$, and the region of acceptance by $S_A = (99.985, 100.015)$. □

Due to the random and binary nature of a test, two different types of error may occur. The **error of the first kind** or **Type I error** arises, when the data $\boldsymbol{y}$ truly stems from a distribution in $\mathcal{W}_0$ (specified by H_0), but happens to fall into the region of rejection S_C. Consequently, H_0 is falsely rejected. The **error of the second kind** or **Type II error** occurs, when the data $\boldsymbol{y}$ does not stem from a distribution in $\mathcal{W}_0$, but is an element of the region of acceptance S_A. Clearly, H_0 is then accepted by mistake.

From Example 2.3 it is not clear whether the suggested decision rule is in fact reasonable. The following subsection will demonstrate how the two above errors can be measured and how they can be used to find optimal decision rules.

2.4 The size and power of a test

As any test (2.3-12) is itself a random variable derived from the observations $\boldsymbol{Y}$, it is straightforward to ask for the probabilities with which these errors occur. Since tests with small error probabilities appear to be more desirable than tests with large errors, it is natural to use these probabilities in order to find optimal test procedures. For this purpose, let α denote the **probability of a Type I error**, and β (not to be confused with the parameter $\boldsymbol{\beta}$ of the linear model 2.1-1) the **probability of a Type II error**. Instead of β, it is more common to use the complementary quantity $\pi := 1 - \beta$, called the **power** of a test.

When H_0 is simple, i.e. when all the unknown parameter values are specified by H_0, then the numerical value for α may be computed from (2.3-12) by

$$\alpha = P_{\boldsymbol{\theta}_0}[\phi(\boldsymbol{Y}) = 1] = P_{\boldsymbol{\theta}_0}(\boldsymbol{Y} \in S_C) = \int_{S_C} f(\boldsymbol{y}; \boldsymbol{\theta}_0)\, d\boldsymbol{y}. \tag{2.4-16}$$

From (2.4-16) it becomes evident why α is also called the **size (of the critical region)**, because its value represents the area under the density function measured over S_C. Notice that for composite H_0, the value for α will generally depend on the values of the nuisance parameters. In that case, it is appropriate to define α as a function with

$$\alpha(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}[\phi(\boldsymbol{Y}) = 1] = P_{\boldsymbol{\theta}}(\boldsymbol{Y} \in S_C) = \int_{S_C} f(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y} \quad (\boldsymbol{\theta} \in \boldsymbol{\Theta}_0). \tag{2.4-17}$$

Example 2.4 (Example 2.3 continued): What is the *size of the critical region* or the *probability of the Type I error* for the test defined by (2.3-15)?

Recall that $\mu_0 = 100$ is the value assigned to $\bar{\mu}$ by H_0 and that $\sigma_0 = 0.003$ is the fixed value for $\bar{\sigma}$ assumed as known *a priori*. Then, after transforming Y into an $N(0,1)$-distributed random variable, the values of the standard normal distribution function Φ may be obtained from statistical tables (see, for instance, Kreyszig, 1998, p.423-424) to answer the above question.

$$\begin{aligned}\alpha &= P_{\theta_0}(Y \in S_C) = N_{\mu_0,\sigma_0^2}(Y \leq 99.985 \text{ or } Y \geq 100.015)\\ &= 1 - N_{\mu_0,\sigma_0^2}(99.985 < Y < 100.015)\\ &= 1 - N_{0,1}\left(\frac{99.985-\mu_0}{\sigma_0} < \frac{Y-\mu_0}{\sigma_0} < \frac{100.015-\mu_0}{\sigma_0}\right)\\ &= 1 - [\Phi(5) - \Phi(-5)]\\ &\approx 0.\end{aligned}$$

If $\bar{\sigma}$ was unknown, then the numerical value of α would depend on the value of σ. □

Let us finish the discussion of the size of a test by observing in Fig. 2.1 that different choices of the critical region may have the same total probability mass.

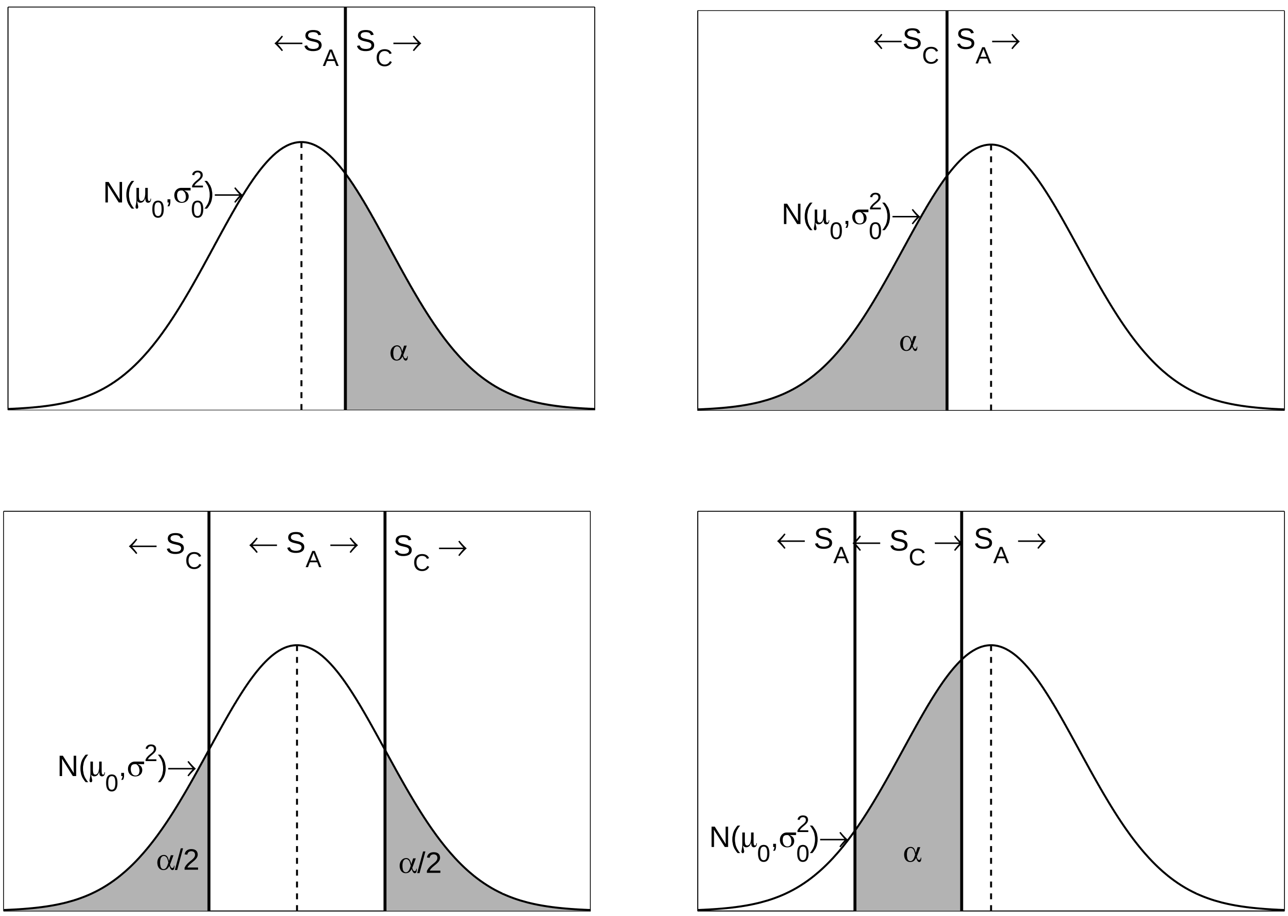

Fig. 2.1 Let $N(\mu_0, \sigma_0^2)$ denote the distribution of a single observation Y under a simple H_0 (with known and fixed variance σ_0^2). This figure presents four (out of infinitely many different ways) to specify a critical region S_C of fixed size α.

The computation of the **probability of a Type II error** is more intricate than that of α, because the premise of a false H_0 does not tell us anything about which distribution we should use to measure the event that $\boldsymbol{y}$ is in S_A. For this very reason, an alternative class of distributions $\mathcal{W}_1 \subset \mathcal{W}$ must be specified which contains the true distribution if H_0 is false. If we let $\mathcal{W}_1$ be represented by a corresponding non-empty parameter subset $\boldsymbol{\Theta}_1 \in \boldsymbol{\Theta}$, then we may define the **alternative hypothesis** as

$$H_1 : \bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_1 \quad (\emptyset \neq \boldsymbol{\Theta}_1 \subset \boldsymbol{\Theta}, \boldsymbol{\Theta}_1 \cap \boldsymbol{\Theta}_0 = \emptyset), \tag{2.4-18}$$

which may be *simple* or *composite* in analogy to H_0. The condition $\boldsymbol{\Theta}_1 \cap \boldsymbol{\Theta}_0 = \emptyset$ is necessary to avoid ambiguities due to overlapping hypotheses.

Example 2.5 (Example 2.2 continued): For testing the right angle hypothesis $H_0 : \bar{\mu} = 100$, we will assume that $\bar{\sigma}^2 = \sigma_0^2 = 0.003^2$ is fixed and known. Let us consider the following three situations.

1. Imagine that a map indicates that the observed angle is a right angle, while a second older map gives a value of say 100.018 *gon*. In this case, the data y could be used to test H_0 against the alternative $H_1 : \bar{\mu} = 100.018$. $\boldsymbol{\Theta}_1 = \{100.018\}$ represents one point in $\boldsymbol{\Theta}$, hence H_1 is *simple*.
2. If the right angle hypothesis is doubtful but there is evidence that the angle can definitely not be smaller than 100 *gon*, then the appropriate alternative reads $H_1 : \bar{\mu} > 100$, which is now *composite* due to $\boldsymbol{\Theta}_1 = \{\mu : \mu > 100\}$, and it is called **one-sided**, because the alternative values for μ are elements of a single interval.
3. When no prior information regarding potential alternative angle sizes is available, then $H_1 : \bar{\mu} \neq 100$ is a reasonable choice as we will see later. Since the alternative values for μ are split up into two intervals separated by the value under H_0, we speak of a **two-sided** (composite) H_1. □

With the specification of an alternative subspace $\boldsymbol{\Theta}_1 \subset \boldsymbol{\Theta}$, which the unknown true parameter $\bar{\boldsymbol{\theta}}$ is assumed to be an element of if H_0 is false, the **probability of a Type II error** follows to be either

$$\beta = P_{\boldsymbol{\theta}_1}[\phi(\boldsymbol{Y}) = 0] = P_{\boldsymbol{\theta}_1}(\boldsymbol{Y} \in S_A) = \int_{S_A} f(\boldsymbol{y}; \boldsymbol{\theta}_1)\, d\boldsymbol{y} \tag{2.4-19}$$

if H_1 is simple (i.e. if $\boldsymbol{\theta}_1$ is the unique element of $\boldsymbol{\Theta}_1$), or

$$\beta(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}[\phi(\boldsymbol{Y}) = 0] = P_{\boldsymbol{\theta}}(\boldsymbol{Y} \in S_A) = \int_{S_A} f(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y} \quad (\boldsymbol{\theta} \in \boldsymbol{\Theta}_1) \tag{2.4-20}$$

if $\boldsymbol{\Theta}_1$ is composed of multiple elements. As simple alternatives are rarely encountered in practical situations, the general notation of (2.4-20) will be maintained. As already mentioned, it is more common to use the **power of a test**, defined as

$$\Pi(\boldsymbol{\theta}) := 1 - P_{\boldsymbol{\theta}}(\boldsymbol{Y} \in S_A) = P_{\boldsymbol{\theta}}(\boldsymbol{Y} \in S_C) = P_{\boldsymbol{\theta}}[\phi(\boldsymbol{Y}) = 1] \quad (\boldsymbol{\theta} \in \boldsymbol{\Theta}_1). \tag{2.4-21}$$

The numerical values of Π may be interpreted as the **probabilities of avoiding a Type II error**.

When designing a test, it will be useful to determine the probability of rejecting H_0 as a function defined over the entire parameter space $\boldsymbol{\Theta}$. Such a function may be defined as

$$\mathrm{Pf}(\boldsymbol{\theta}) := P_{\boldsymbol{\theta}}[\phi(\boldsymbol{Y}) = 1] = P_{\boldsymbol{\theta}}(\boldsymbol{Y} \in S_C) \quad (\boldsymbol{\theta} \in \boldsymbol{\Theta}) \tag{2.4-22}$$

and will be called the **power function** of a test. Clearly, this function will in particular produce the sizes α for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ and the power values Π for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$. For all the other values of $\boldsymbol{\theta}$, this function will provide the *hypothetical power* of the test if the true parameter is neither assumed to be an element of $\boldsymbol{\Theta}_0$, nor of $\boldsymbol{\Theta}_1$.

Example 2.6 (Example 2.5 continued): Recall that the size of this test turned out to be approximately 0 as Ex. 2.4 demonstrated. Let us now ask, what the power of the test would be for testing $H_0 : \bar{\mu} = 100$ against $H_1 : \bar{\mu} = \mu_1 = 100.018$ with $\bar{\sigma}^2 = \sigma_0^2 = 0.003$ known *a priori*. Using the (2.4-21), we obtain

$$\begin{aligned}
\Pi &= 1 - P_{\mu_1,\sigma_0^2}(Y \in S_A) = 1 - N_{\mu_1,\sigma_0^2}(99.985 < Y < 100.015) \\
&= 1 - N_{0,1}\left(\frac{99.985 - \mu_1}{0.003} < \frac{Y - \mu_1}{0.003} < \frac{100.015 - \mu_1}{0.003}\right) \\
&= 1 - [\Phi(-1) - \Phi(-11)] \\
&\approx 0.8413.
\end{aligned}$$

Notice that ***the larger the difference between μ_1 and μ_0, the larger the power*** becomes. For instance, if H_1 had been specified as $\mu_1 = 100.021$, then the power would increase to $\Pi \approx 0.9772$, and for $\mu_1 = 100.024$ the power would already be very close to 1. This is intuitively understandable, because very similar hypotheses are expected to be harder to separate on the basis of some observed data than extremely different hypotheses. Figure 2.2 illustrates this point. □

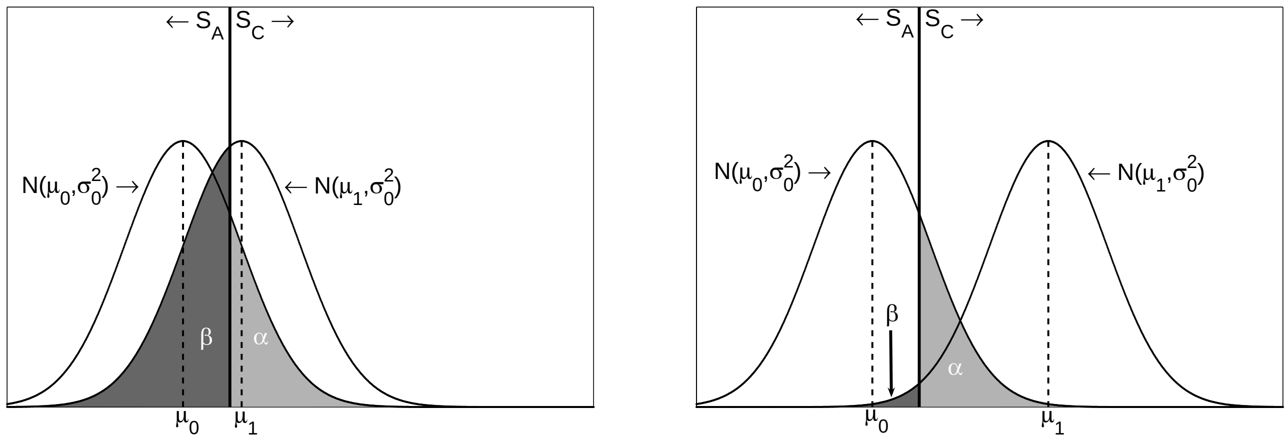

Fig. 2.2 The probability of a Type II error ($\beta = 1 - \Pi$) becomes smaller as the distance $\mu_1 - \mu_0$ (with identical variance σ_0^2) between the null hypothesis H_0 and the alternative H_1 increases.

Another important observation to make in this context is that, unfortunately, ***the errors of the first and second kind cannot be minimized independently***. For instance, when the critical region S_C is extended towards μ_0 (Fig. 2.3 left → right), then clearly its size becomes larger. In doing this, S_A shrinks, and the error of the second kind becomes smaller. This effect is explained by the fact that both errors are measured in complementary regions and thereby affect each other's size. Therefore, no critical function can exist that minimizes both error probabilities simultaneously. The purpose of the following subsection is to present a practical solution to resolve this conflict.

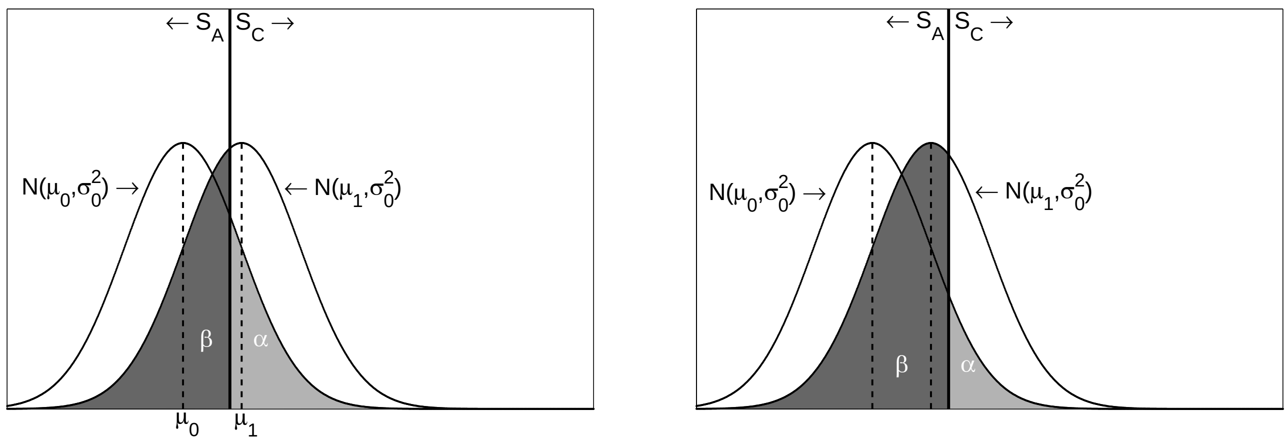

Fig. 2.3 Let $N(\mu_0, \sigma_0^2)$ and $N(\mu_1, \sigma_0^2)$ denote the distributions of a single observation Y under simple H_0 and H_1, respectively. Changing the S_C/S_A partitioning of the observation space (abscissa) necessarily causes an increase in probability of one error type and a decrease in probability of the other type.

2.5 Best critical regions

As pointed out in the previous section, shifting the critical region and making one error type more unlikely always causes the other error to become more probable. Therefore, the probabilities of Type I and Type II errors cannot be minimized simultaneously. One way to resolve this conflict is to keep the probability of a Type I error fixed at a relatively small value and to seek a critical region that minimizes the probability of a Type II error, or equivalently that maximizes the power of the test.

To make the mathematical concepts, necessary for this procedure, intuitively understandable, examples will be given mainly with respect to the class of observation models (2.1-6) introduced in Example 2.1. The remainder of this Section 2.5 is organized such that tests with best critical regions will be constructed for testing problems that are progressively complex within that class of models. The determination of optimal critical regions in the context of the general linear model (2.1-1) with general parametric densities as in (2.1-3) will be subject of detailed investigations in Sections 3 and 4.

2.5.1 Most powerful (MP) tests

The simplest kind of problem for which a critical region with optimal power may exist is that of testing a simple $H_0 : \bar{\theta} = \theta_0$ against a simple alternative hypothesis $H_1 : \bar{\theta} = \theta_1$ involving a single unknown parameter. Using definitions (2.4-16) and (2.4-21), the problem is to find a set S_C such that the restriction

$$\int_{S_C} f(\boldsymbol{y};\theta_0)\, d\boldsymbol{y} = \alpha \tag{2.5-23}$$

is satisfied, where α as a given size is also called the **(significance) level**, and

$$\int_{S_C} f(\boldsymbol{y};\theta_1) d\boldsymbol{y} \;\text{ is a maximum.} \tag{2.5-24}$$

Such a critical region will be called the **best critical region (BCR)**, and a test based on the BCR will be denoted as **most powerful (MP)** for testing H_0 against H_1 at level α. A solution to this problem may be found on the basis of the following lemma of Neyman and Pearson (see, for instance, Rao, 1973, p. 446).

Theorem 2.1 (Neyman-Pearson Lemma). *Suppose that $f(\boldsymbol{Y};\theta_0)$ and $f(\boldsymbol{Y};\theta_1)$ are two densities defined on a space S. Let $S_C \subset S$ be any critical region with*

$$\int_{S_C} f(\boldsymbol{y};\theta_0)\, d\boldsymbol{y} = \alpha, \tag{2.5-25}$$

where α has a given value. If there exists a constant k_α such that for the region $S_C^* \subset S$ with

$$\begin{cases} \dfrac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} > k_\alpha \;\text{ if }\; \boldsymbol{y} \in S_C^* \\ \dfrac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} < k_\alpha \;\text{ if }\; \boldsymbol{y} \notin S_C^*, \end{cases} \tag{2.5-26}$$

condition (2.5-25) is satisfied, then

$$\int_{S_C^*} f(\boldsymbol{y};\theta_1) d\boldsymbol{y} \geq \int_{S_C} f(\boldsymbol{y};\theta_1) d\boldsymbol{y}. \tag{2.5-27}$$

Notice if when $f(\boldsymbol{Y};\theta_0)$ and $f(\boldsymbol{Y};\theta_1)$ are densities under simple hypotheses H_0 and H_1, and if the conditions (2.5-25) and (2.5-26) hold for some k_α, then S_C^* denotes the BCR for testing H_0 versus H_1 at fixed level α, because (2.5-27) is equivalent to the desired maximum power condition (2.5-24). Also observe that (2.5-26) then defines the MP test, which may be written as

$$\phi(\boldsymbol{y}) = \begin{cases} 1 \;\text{ if }\; \dfrac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} > k_\alpha \\ 0 \;\text{ if }\; \dfrac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} < k_\alpha. \end{cases} \tag{2.5-28}$$

This condition (2.5-28) expresses that in order for a test to be most powerful, the critical region S_C must comprise all the observations $\boldsymbol{y}$, for which the so-called **density ratio** $f(\boldsymbol{y};\theta_1)/f(\boldsymbol{y};\theta_0)$ is larger than some

α-dependent number k_α. This can be explained by the following intuitions of Stuart et al. (1999, p. 176). Using definition (2.4-21), the power may be rewritten in terms of the density ratio as

$$\Pi = \int_{S_C} f(\boldsymbol{y};\theta_1)\,d\boldsymbol{y} = \int_{S_C} \frac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} f(\boldsymbol{y};\theta_0)\,d\boldsymbol{y}.$$

Since α has a fixed value, maximizing Π is equivalent to maximizing the quantity

$$\frac{\Pi}{\alpha} = \frac{\displaystyle\int_{S_C} \frac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} f(\boldsymbol{y};\theta_0)\,d\boldsymbol{y}}{\displaystyle\int_{S_C} f(\boldsymbol{y};\theta_0)\,d\boldsymbol{y}}.$$

In order for a test to have maximum power, its critical region S_C must clearly include all the observations $\boldsymbol{y}$,

1. for which the integral value in the denominator equals α, and
2. for which the density ratio in the nominator produces the largest possible values, whose lower bound may be defined as the number k_α (with the values of the additional factor $f(\boldsymbol{y};\theta_0)$ fixed by condition 1).

These are the very conditions given by the Neyman-Pearson Lemma. A more formal proof may be found, for instance, in Teunissen (2000, p. 30f.). The following example demonstrates how the BCR may be constructed for a simple test problem by applying the Neyman-Pearson Lemma.

Example 2.7: Test of the normal mean with known variance - Simple alternatives. Let $Y_1, \ldots, Y_n$ be independently and normally distributed observations with common unknown mean $\bar{\mu}$ and common known standard deviation $\bar{\sigma} = \sigma_0$. What is the BCR for a test of the simple null hypothesis $H_0 : \bar{\mu} = \mu_0$ against the simple alternative hypothesis $H_1 : \bar{\mu} = \mu_1$ at level α? (It is assumed that μ_0, μ_1, σ_0 and α have fixed numerical values.)

In order to construct the BCR, we will first try to find a number k_α such that condition (2.5-26) about the density ratio $f(\boldsymbol{y};\theta_1)/f(\boldsymbol{y};\theta_0)$ holds. As the observations are independently distributed with common mean $\bar{\mu}$ and variance σ_0^2, the factorized form of the joint normal density function $f(\boldsymbol{y})$ according to Example 2.1 may be applied. This yields the expression

$$\frac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} = \frac{\displaystyle\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\mu_1}{\sigma_0}\right)^2\right\}}{\displaystyle\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\mu_0}{\sigma_0}\right)^2\right\}} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\displaystyle\sum_{i=1}^{n}(y_i-\mu_1)^2\right\}}{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\displaystyle\sum_{i=1}^{n}(y_i-\mu_0)^2\right\}} \quad (2.5\text{-}29)$$

for the density ratio. An application of the ordinary binomial formula allows us to split off a factor that does not depend on μ, that is

$$\frac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\displaystyle\sum_{i=1}^{n} y_i^2\right\} \exp\left\{-\frac{1}{2\sigma_0^2}\displaystyle\sum_{i=1}^{n}(-2y_i\mu_1+\mu_1^2)\right\}}{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\displaystyle\sum_{i=1}^{n} y_i^2\right\} \exp\left\{-\frac{1}{2\sigma_0^2}\displaystyle\sum_{i=1}^{n}(-2y_i\mu_0+\mu_0^2)\right\}}. \quad (2.5\text{-}30)$$

Now, the first two factors in the nominator and denominator cancel out due to their independence of μ. Rearranging the remaining terms leads to

$$\frac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} = \frac{\exp\left\{\frac{\mu_1}{\sigma_0^2}\displaystyle\sum_{i=1}^{n} y_i - \frac{n\mu_1^2}{2\sigma_0^2}\right\}}{\exp\left\{\frac{\mu_0}{\sigma_0^2}\displaystyle\sum_{i=1}^{n} y_i - \frac{n\mu_0^2}{2\sigma_0^2}\right\}} \quad (2.5\text{-}31)$$

$$= \exp\left\{\frac{\mu_1}{\sigma_0^2}\sum_{i=1}^{n} y_i - \frac{\mu_0}{\sigma_0^2}\sum_{i=1}^{n} y_i - \frac{n\mu_1^2}{2\sigma_0^2} + \frac{n\mu_0^2}{2\sigma_0^2}\right\} \quad (2.5\text{-}32)$$

$$= \exp\left\{\frac{1}{\sigma_0^2}(\mu_1-\mu_0)\sum_{i=1}^{n} y_i - \frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2)\right\}, \quad (2.5\text{-}33)$$

which reveals two remarkable facts: the simplified density ratio depends on the observations only through their sum $\sum_{i=1}^{n} y_i$, and the density ratio, as an exponential function, is a positive number. Therefore, we may choose another positive number k_α such that

$$\exp\left\{\frac{1}{\sigma_0^2}(\mu_1-\mu_0)\sum_{i=1}^{n} y_i - \frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2)\right\} > k_\alpha \tag{2.5-34}$$

always holds. Taking natural logarithms on both sides of this inequality yields

$$\frac{1}{\sigma_0^2}(\mu_1-\mu_0)\sum_{i=1}^{n} y_i - \frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2) > \ln k_\alpha$$

or, after multiplication with $2\sigma_0^2$ and expansion of the left side by $n\cdot\frac{1}{n}$,

$$2n(\mu_1-\mu_0)\frac{1}{n}\sum_{i=1}^{n} y_i > 2\sigma_0^2\ln k_\alpha + n(\mu_1^2-\mu_0^2).$$

Depending on whether $\mu_1 > \mu_0$ or $\mu_1 < \mu_0$, the sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ must satisfy

$$\bar{y} > \frac{2\sigma_0^2\ln k_\alpha + n(\mu_1^2-\mu_0^2)}{2n(\mu_1-\mu_0)} =: k'_\alpha \qquad (\text{if } \mu_1 > \mu_0)$$

or

$$\bar{y} < \frac{2\sigma_0^2\ln k_\alpha + n(\mu_1^2-\mu_0^2)}{2n(\mu_1-\mu_0)} =: k'_\alpha \qquad (\text{if } \mu_1 < \mu_0)$$

in order for the second condition (2.5-26) of the Neyman-Pearson Lemma to hold. Note that the quantities $\sigma_0^2, n, \mu_1, \mu_0$ are all constants fixed *a priori*, and k_α is a constant whose exact value is still to be determined. Thus, k'_α is itself an unknown constant.

Now, in order for the first condition (2.5-25) of the Neyman-Pearson Lemma to hold in addition, S_C must have size α under the null hypothesis. As mentioned above, the critical region S_C may be constructed solely by inspecting the value $\bar{y}$, which may be viewed as the outcome of the random variable $\bar{Y} := \frac{1}{n}\sum_{i=1}^{n} Y_i$. Under H_0, $\bar{Y}$ is normally distributed with expectation μ_0 (identical to the expectation of each of the original observations $Y_1,\ldots,Y_n$) and standard deviation $\sigma_0/\sqrt{n}$. Therefore, the size is determined by

$$\alpha = \begin{cases} N_{\mu_0,\sigma_0^2/n}(\bar{Y} > k'_\alpha) & \text{if } \mu_1 > \mu_0, \\ N_{\mu_0,\sigma_0^2/n}(\bar{Y} < k'_\alpha) & \text{if } \mu_1 < \mu_0. \end{cases}$$

It will be more convenient to standardize $\bar{Y}$ because this allows us to evaluate the size in terms of the standard normal distribution. The condition to be satisfied by k'_α then reads

$$\alpha = \begin{cases} N_{0,1}\left(\dfrac{\bar{Y}-\mu_0}{\sigma_0/\sqrt{n}} > \dfrac{k'_\alpha-\mu_0}{\sigma_0/\sqrt{n}}\right) & \text{if } \mu_1 > \mu_0, \\ N_{0,1}\left(\dfrac{\bar{Y}-\mu_0}{\sigma_0/\sqrt{n}} < \dfrac{k'_\alpha-\mu_0}{\sigma_0/\sqrt{n}}\right) & \text{if } \mu_1 < \mu_0, \end{cases}$$

or, using the standard normal distribution function Φ,

$$\alpha = \begin{cases} 1-\Phi\left(\dfrac{k'_\alpha-\mu_0}{\sigma_0/\sqrt{n}}\right) & \text{if } \mu_1 > \mu_0, \\ \Phi\left(\dfrac{k'_\alpha-\mu_0}{\sigma_0/\sqrt{n}}\right) & \text{if } \mu_1 < \mu_0. \end{cases}$$

Rewriting this as

$$\Phi\left(\frac{k'_\alpha-\mu_0}{\sigma_0/\sqrt{n}}\right) = \begin{cases} 1-\alpha & \text{if } \mu_1 > \mu_0, \\ \alpha & \text{if } \mu_1 < \mu_0 \end{cases}$$

allows us to determine the argument of Φ by applying the inverse standard normal distribution function Φ^{-1} to the previous equation, which yields

$$\frac{k'_\alpha - \mu_0}{\sigma_0/\sqrt{n}} = \begin{cases} \Phi^{-1}(1-\alpha) \text{ if } \mu_1 > \mu_0, \\ \Phi^{-1}(\alpha) \text{ if } \mu_1 < \mu_0, \end{cases}$$

from which the constant k'_α is obtained as

$$k'_\alpha = \begin{cases} \mu_0 + \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha) \text{ if } \mu_1 > \mu_0, \\ \mu_0 + \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(\alpha) \text{ if } \mu_1 < \mu_0, \end{cases}$$

or

$$k'_\alpha = \begin{cases} \mu_0 + \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha) \text{ if } \mu_1 > \mu_0, \\ \mu_0 - \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha) \text{ if } \mu_1 < \mu_0. \end{cases}$$

Consequently, depending on the sign of $\mu_1 - \mu_0$, there are two different values for k'_α that satisfy the first condition (2.5-25) of the Neyman-Pearson Lemma. When $\mu_1 > \mu_0$ the BCR is seen to consist of all the observations $\boldsymbol{y} \in S$, for which

$$\bar{y} > \mu_0 + \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha), \tag{2.5-35}$$

and when $\mu_1 < \mu_0$, the BCR reads

$$\bar{y} < \mu_0 - \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha). \tag{2.5-36}$$

In the first case ($\mu_1 > \mu_0$), the MP test is given by

$$\phi_u(\boldsymbol{y}) = \begin{cases} 1 \text{ if } \bar{y} > \mu_0 + \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha), \\ 0 \text{ if } \bar{y} < \mu_0 + \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha), \end{cases} \tag{2.5-37}$$

and in the second case ($\mu_1 < \mu_0$), the MP test is

$$\phi_l(\boldsymbol{y}) = \begin{cases} 1 \text{ if } \bar{y} < \mu_0 - \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha), \\ 0 \text{ if } \bar{y} > \mu_0 - \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha). \end{cases} \tag{2.5-38}$$

Observe that the critical regions depend solely on the value of the one-dimensional random variable $\bar{Y}$, which, as a function of the observations $\boldsymbol{Y}$, is also called a **statistic**. As this statistic appears in the specific context of hypothesis testing, we will speak of $\bar{Y}$ as a **test statistic**. We see from this that it is not necessary to actually specify an n-dimensional region S_C used as the BCR, but the BCR may be expressed conveniently in terms of one-dimensional intervals. For this purpose, let $(c_u, +\infty)$ and $(-\infty, c_l)$ denote the critical regions with respect to the sample mean $\bar{y}$ as defined by (2.5-35) and (2.5-36). The real constants

$$c_u := \mu_0 + \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha) \tag{2.5-39}$$

and

$$c_l := \mu_0 - \frac{\sigma_0}{\sqrt{n}}\,\Phi^{-1}(1-\alpha) \tag{2.5-40}$$

are called the **upper critical value** and the **lower critical value** corresponding to the BCR for testing H_0 versus H_1.

In a practical situation, it will be clear from the numerical specification of H_1 which of the tests (2.5-37) and (2.5-38) should be applied. Then, the test is carried out by computing the mean $\bar{y}$ of the given data $\boldsymbol{y}$ and by checking how large its value is in comparison to the critical value of (2.5-37) or (2.5-38), respectively. □

Example 2.8: A most powerful test about the Beta distribution. Let $Y_1, \ldots, Y_n$ be independently and $B(\alpha, \beta)$-distributed observations on $[0, 1]$ with common unknown parameter $\bar{\alpha}$ (which in this case is not to be confused with the size or level of the test) and common known parameter $\bar{\beta} = 1$ (not to be confused with the probability of a Type II error). What is the BCR for a test of the simple null hypothesis $H_0 : \bar{\alpha} = \alpha_0 = 1$ against the simple alternative hypothesis $H_1 : \bar{\alpha} = \alpha_1 = 2$ at level α^*?

The density function of the **univariate Beta distribution** in standard form is defined by

$$f(y; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \quad (0 < y < 1; \alpha, \beta > 0), \tag{2.5-41}$$

see Johnson and Kotz (1970b, p. 37) or Koch (1999, p. 115). Notice that (2.5-41) simplifies under H_0 to

$$f(y; \alpha_0) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} y^{1-1}(1-y)^{1-1} = 1 \quad (0 < y < 1), \tag{2.5-42}$$

and under H_1 to

$$f(y; \alpha_1) = \frac{\Gamma(3)}{\Gamma(2)\Gamma(1)} y^{2-1}(1-y)^{1-1} = 2y \quad (0 < y < 1) \tag{2.5-43}$$

where we used the facts that $\Gamma(1) = \Gamma(2) = 1$ and $\Gamma(3) = 2$. The density 2.5-42 defines the so-called **uniform distribution** with parameters $a = 0$ and $b = 1$, see Johnson and Kotz (1970b, p. 57) or Koch (2000, p. 21). We may now proceed as in Example 2.7 and determine the BCR by using the Neyman-Pearson Lemma (Theorem 2.1). For n independent observations, the joint density may be written as the product of the individual univariate densities, which results in the density ratio

$$\frac{f(\boldsymbol{y}; \alpha_1)}{f(\boldsymbol{y}; \alpha_0)} = \prod_{i=1}^{n} 2y_i / \prod_{i=1}^{n} 1 = 2^n \prod_{i=1}^{n} y_i, \tag{2.5-44}$$

where we assumed that each observation is strictly within the interval $(0, 1)$. As the density ratio is a positive number, we may choose a number k_{α^*} such that $2^n \prod_{i=1}^{n} y_i > k_{\alpha^*}$ holds. Division by 2^n and taking both sides to the power of $1/n$ yields the equivalent inequality

$$\left(\prod_{i=1}^{n} y_i\right)^{1/n} > \left(2^{-n} k_{\alpha^*}\right)^{1/n}.$$

Now we have found a seemingly convenient condition about the sample's **geometric mean** $\breve{Y} := \left(\prod_{i=1}^{n} Y_i\right)^{1/n}$ rather than about the entire sample $\boldsymbol{Y}$ itself. Then the second condition (2.5-26 or equivalently 2.5-28) of the Neyman-Pearson Lemma gives

$$\phi(\boldsymbol{y}) = \begin{cases} 1 & \text{if} \quad \breve{y} > \left(2^{-n} k_{\alpha^*}\right)^{1/n} =: k'_{\alpha^*} \\ 0 & \text{if} \quad \breve{y} < \left(2^{-n} k_{\alpha^*}\right)^{1/n} =: k'_{\alpha^*}. \end{cases}$$

To ensure that ϕ has some specified level α^*, the first condition (2.5-25) of the Neyman-Pearson Lemma requires that α^* equals the probability under H_0 that the geometric mean exceeds k'_{α^*}. Unfortunately, in contrast to the *arithmetic mean* $\bar{Y}$ of n independent normal variables, the *geometric mean* $\breve{Y}$ of n independent standard uniform variables does not have a standard distribution. However, as Stuart and Ord (2003, p. 393) demonstrate in their Example 11.15, the statistic

$$U := -\ln \breve{Y}^n = -\ln \prod_{i=1}^{n} Y_i = -\sum_{i=1}^{n} \ln Y_i$$

follows a Gamma distribution $G(b, p)$ with $b = 1$ and $p = n$, defined by Equation 2.107 in Koch (1999, p. 112). Thus the first Neyman-Pearson condition reads

$$\alpha^* = G_{1,n}(U > k''_{\alpha^*}) = 1 - F_{G_{1,n}}(k''_{\alpha^*}),$$

from which the critical value k''_{α^*} follows to be $k''_{\alpha^*} = F^{-1}_{G_{1,n}}(1 - \alpha^*)$, and which may be obtained in MATLAB by executing the command `CV = gaminv(1 − α*, n, 1)`. In summary, the MP test is given by

$$\phi(\boldsymbol{y}) = \begin{cases} 1 & \text{if} \quad u(\boldsymbol{y}) = -\sum_{i=1}^{n} \ln y_i > k''_{\alpha^*} = -\ln\left(2^{-n} k_{\alpha^*}\right) = F^{-1}_{G_{1,n}}(1 - \alpha^*), \\ 0 & \text{if} \quad u(\boldsymbol{y}) = -\sum_{i=1}^{n} \ln y_i < k''_{\alpha^*} = -\ln\left(2^{-n} k_{\alpha^*}\right) = F^{-1}_{G_{1,n}}(1 - \alpha^*). \end{cases}$$

□

2.5.2 Reduction to sufficient statistics

We saw in Example 2.7 that applying the conditions of the Neyman-Pearson Lemma to derive the BCR led to a condition about the sample mean $\bar{y}$ rather than about the original data $\boldsymbol{y}$. We might say that it was **sufficient** to use the mean value of the data for testing a hypothesis about the parameter μ of the normal distribution. This raises the important question of whether it is always possible to reduce the data in such a way.

To generalize this idea, let $\mathcal{F} = \{f(\boldsymbol{y};\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ be a collection of densities where the parameter $\boldsymbol{\theta}$ is unknown. Further, let each $f(\boldsymbol{y};\boldsymbol{\theta})$ depend on the value of a random function or statistic $T(\boldsymbol{Y})$ which is independent of $\boldsymbol{\theta}$. If any inference about $\boldsymbol{\theta}$, be it estimation or testing, depends on the observations $\boldsymbol{Y}$ only through the value of $T(\boldsymbol{Y})$, then this statistic will be called **sufficient for $\boldsymbol{\theta}$**.

This qualitative definition of **sufficiency** can be interpreted such that a sufficient statistic captures all the relevant information that the data contains about the unknown parameters. The point is that the data might have some additional information that does not contribute anything to solving the estimation or test problem. The following classical example highlights this distinction between information that is essential and information that is completely negligible for estimating an unknown parameter.

Example 2.9: Sufficient statistic in Bernoulli's random experiment. Let $Y_1, \ldots, Y_n$ denote independent binary observations within an idealized setting of Bernoulli's random experiment (see, for instance, Lehmann, 1959a, p. 17-18). The probability p of the elementary event *success* ($y_i = 1$) is assumed to be unknown, but valid for all observations. The probability of the second possible outcome *failure* ($y_i = 0$) is then $1 - p$.

Now, it is intuitively clear that in order to estimate the unknown success rate p, it is completely sufficient to know how many successes $T(\boldsymbol{y}) := \sum_{i=1}^{n} y_i$ occurred in total within n trials. The additional information regarding which specific observations were successes or failures does not contribute anything useful for determining the success rate p. In this sense the use of the statistic $T(\boldsymbol{Y})$ reduces the n data to a single value which carries all the essential information required to determine p. □

The concept of sufficiency provides a convenient tool to achieve a data reduction without any loss of information about the unknown parameters. The definition above, however, is not easily applicable when one has to deal with specific estimation or testing problems. As a remedy, **Neyman's Factorization Theorem** gives an easy-to-check condition for the existence of a sufficient statistic in any given parametric inference problem.

Theorem 2.2 (Neyman's Factorization Theorem)**.** *Let $\mathcal{F} = \{f(\boldsymbol{y};\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ be a collection of densities for a sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$. A vector of statistics $\boldsymbol{T}(\boldsymbol{Y})$ is sufficient for $\boldsymbol{\theta}$ if and only if there exist functions $g(\boldsymbol{T}(\boldsymbol{Y});\boldsymbol{\theta})$ and $h(\boldsymbol{Y})$ such that*

$$f(\boldsymbol{y};\boldsymbol{\theta}) = g(\boldsymbol{T}(\boldsymbol{y});\boldsymbol{\theta}) \cdot h(\boldsymbol{y}) \qquad (2.5\text{-}45)$$

holds for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{y} \in S$.

Proof. A deeper understanding of the sufficiency concept involves an investigation into conditional probabilities which is beyond the scope of this thesis. The reader familiar with conditional probabilities is referred to Lehmann and Romano (2005, p. 20) for a proof of this theorem. □

It is easy to see that the trivial choice $\boldsymbol{T}(\boldsymbol{y}) := \boldsymbol{y}$, $g(\boldsymbol{T}(\boldsymbol{y});\boldsymbol{\theta}) := f(\boldsymbol{y};\boldsymbol{\theta})$ and $h(\boldsymbol{y}) := 1$ is always possible, but achieves no data reduction. Far more useful is the fact that any reversible function of a sufficient statistic is also sufficient for $\boldsymbol{\theta}$ (cf. Casella and Berger, 2002, p. 280). In particular, multiplying a sufficient statistic with constants yields again a sufficient statistic. The following example will now establish sufficient statistics for the normal density with both parameters μ and σ^2 unknown.

Example 2.10: Suppose that observations $Y_1, \ldots, Y_n$ are independently and normally distributed with common unknown mean $\bar{\mu}$ and common unknown variance $\bar{\sigma}^2$. Let the sample mean and variance be defined as $\bar{Y} = \sum_{i=1}^{n} Y_i/n$ and $S^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)$, respectively. The joint normal density can then be written as

$$\begin{aligned} f(\boldsymbol{y};\mu,\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i-\mu)^2\right\} = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{n\mu^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} y_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2 - \frac{n-1}{2\sigma^2}s^2\right\} \cdot I_{\mathbb{R}^n}(\boldsymbol{y}) \end{aligned}$$

where $\boldsymbol{T}(\boldsymbol{Y}) := [\bar{Y}, S^2]'$ is **sufficient** for (μ, σ^2) and $h(\boldsymbol{y}) := I_{\mathbb{R}^n}(\boldsymbol{y}) = 1$ with I as the indicator function. □

The great practical value of Neyman's Factorization Theorem in connection to hypothesis testing lies in the simple fact that any density ratio will automatically simplify in the same way as in Example 2.7 from (2.5-30) to (2.5-31). What generally happens is that the factor $h(\boldsymbol{y})$ is the same for $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ due to its independence of any parameters, and thereby cancels out in the ratio, that is,

$$\frac{f(\boldsymbol{y};\boldsymbol{\theta}_1)}{f(\boldsymbol{y};\boldsymbol{\theta}_0)} = \frac{g(\boldsymbol{T}(\boldsymbol{y});\boldsymbol{\theta}_1)\cdot h(\boldsymbol{y})}{g(\boldsymbol{T}(\boldsymbol{y});\boldsymbol{\theta}_0)\cdot h(\boldsymbol{y})} = \frac{g(\boldsymbol{T}(\boldsymbol{y});\boldsymbol{\theta}_1)}{g(\boldsymbol{T}(\boldsymbol{y});\boldsymbol{\theta}_0)} \quad \text{(for all } \boldsymbol{y}\in S). \tag{2.5-46}$$

In addition, this ratio will now be a function of the observations $\boldsymbol{Y}$ through a statistic $\boldsymbol{T}(\boldsymbol{Y})$ which is usually low-dimensional, such as $[\,\bar{Y}, S^2\,]$ in Example 2.10. This usually reduces the complexity and dimensionality of the test problem greatly.

Example 2.7 revisited. Instead of starting the derivation of the BCR by setting up the density ratio $f(\boldsymbol{y};\theta_1)/f(\boldsymbol{y};\theta_0)$ of the raw data as in (2.5-29), we could save time by first reducing $\boldsymbol{Y}$ to the sufficient statistic $T(\boldsymbol{Y}) = \bar{Y}$ and by applying (2.5-46) in connection with the distribution $N(\mu,\sigma_0^2/n)$ of the sample mean. Then

$$\begin{aligned}\frac{g(\bar{y};\theta_1)}{g(\bar{y};\theta_0)} &= \frac{\dfrac{1}{\sqrt{2\pi}(\sigma_0/\sqrt{n})}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\bar{y}-\mu_1}{\sigma_0/\sqrt{n}}\right)^2\right\}}{\dfrac{1}{\sqrt{2\pi}(\sigma_0/\sqrt{n})}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\bar{y}-\mu_0}{\sigma_0/\sqrt{n}}\right)^2\right\}} = \exp\left\{-\frac{n}{2\sigma_0^2}(\bar{y}-\mu_1)^2 + \frac{n}{2\sigma_0^2}(\bar{y}-\mu_0)^2\right\}\\ &= \exp\left\{\frac{n}{\sigma_0^2}(\mu_1-\mu_0)\bar{y} - \frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2)\right\}\end{aligned}$$

leads to (2.5-33) more directly. □

We have seen so far that the sample mean is sufficient when μ is the only unknown parameter, and that the sample mean and variance are jointly sufficient when μ and σ^2 are unknown. Now, what is the maximal reduction generally possible for data that are generated by a more complex observation model, such as by (2.1-1)? Clearly, when a parametric estimation or testing problem comprises u unknown parameters that are not redundant, then a reduction from $n > u$ observations to u corresponding statistics appears to be maximal. It is difficult to give clear-cut conditions that would encompass all possible statistical models and that would also be easily comprehensible without going into too many mathematical details. Therefore, the problem will be addressed only by providing a working definition and a practical theorem, which will be applicable to most of the test problems in this thesis.

Now, to be more specific, we will call a sufficient statistic $\boldsymbol{T}(\boldsymbol{Y})$ **minimally sufficient** if, for any other sufficient statistic $\boldsymbol{T}'(\boldsymbol{Y})$, $\boldsymbol{T}(\boldsymbol{Y})$ is a function of $\boldsymbol{T}'(\boldsymbol{Y})$. As this definition is rather impractial, the following theorem of Lehmann and Scheffe will be a useful tool.

Theorem 2.3 (Lehmann-Scheffe)**.** *Let $f(\boldsymbol{y};\boldsymbol{\theta})$ denote the joint density function of observations $\boldsymbol{Y}$. Suppose there exists a statistic $T(\boldsymbol{Y})$ such that, for every two data points $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, the ratio $f(\boldsymbol{y}_1;\boldsymbol{\theta})/f(\boldsymbol{y}_2;\boldsymbol{\theta})$ is constant as a function of $\boldsymbol{\theta}$ if and only if $T(\boldsymbol{y}_1) = T(\boldsymbol{y}_2)$. Then $T(\boldsymbol{Y})$ is **minimally sufficient** for $\boldsymbol{\theta}$.*

Proof. See Casella and Berger (2002, p. 280-281). □

Example 2.11: Suppose that observations $Y_1,\ldots,Y_n$ are independently and normally distributed with common unknown mean $\bar{\mu}$ and common unknown variance $\bar{\sigma}^2$. Let $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ be two data points, and let $(\bar{y}_1, s_1^2)$ and $(\bar{y}_2, s_2^2)$ be the corresponding values of the sample mean $\bar{Y}$ and variance S^2. To prove that the sample mean and variance are minimally sufficient statistics, the ratio of densities is rewritten as

$$\begin{aligned}\frac{f(\boldsymbol{y}_1;\mu,\sigma^2)}{f(\boldsymbol{y}_2;\mu,\sigma^2)} &= \frac{\prod_{i=1}^{n}\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{1}{2\sigma^2}(y_{1,i}-\mu)^2\right\}}{\prod_{i=1}^{n}\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{1}{2\sigma^2}(y_{2,i}-\mu)^2\right\}} = \frac{(2\pi\sigma^2)^{-n/2}\exp\left\{[n(\bar{y}_1-\mu)^2+(n-1)s_1^2]/(-2\sigma^2)\right\}}{(2\pi\sigma^2)^{-n/2}\exp\left\{[n(\bar{y}_2-\mu)^2+(n-1)s_2^2]/(-2\sigma^2)\right\}}\\ &= \exp\left\{[-n(\bar{y}_1^2-\bar{y}_2^2)+2n\mu(\bar{y}_1-\bar{y}_2)-(n-1)(s_1^2-s_2^2)]/(2\sigma^2)\right\}.\end{aligned}$$

As this ratio is constant only if $\bar{y}_1 = \bar{y}_2$ and $s_1^2 = s_2^2$, the statistic $T(\boldsymbol{Y}) = (\bar{Y}, S^2)$ is indeed **minimally sufficient**. The observations $\boldsymbol{Y}$ cannot be reduced beyond $\boldsymbol{T}(\boldsymbol{Y})$ without losing relevant information. □

2.5.3 Uniformly most powerful (UMP) tests

The concept of the BCR for testing a simple H_0 against a simple H_1 about a single parameter, as defined by the Neyman-Pearson Lemma, is dissatifactory insofar that the great majority of test problems involves composite alternatives. The question to be addressed in this subsection is how a BCR may be defined for such problems.

Let us start with the basic premise that we seek an optimal critical function for testing the simple null

$$H_0 : \bar{\theta} = \theta_0 \tag{2.5-47}$$

versus a composite alternative hypothesis

$$H_1 : \bar{\theta} \in \Theta_1, \tag{2.5-48}$$

where the set of parameter values Θ_1 and θ_0 are disjoint subsets of a one-dimensional parameter space Θ. The most straightforward way to establish optimality under these conditions is to determine the BCR for testing H_0 against a fixed simple $H_1 : \bar{\theta} = \theta_1$ for an arbitrary $\theta_1 \in \Theta_1$ and to check whether the resulting BCR is independent of the specific value θ_1. If this is the case, then all the values $\theta_1 \in \Theta_1$ produce the same BCR, because θ_1 was selected arbitrarily. This critical region that all the simple alternatives H_1' in

$$H_1 = \{H_1' : \bar{\theta} = \theta_1 \text{ with } \theta_1 \in \Theta_1\} \tag{2.5-49}$$

have in common may then be defined as the **BCR for testing a simple $\mathbf{H_0}$ against a composite $\mathbf{H_1}$**. A test based on such a BCR is called **uniformly most powerful (UMP)** for testing H_0 versus H_1 at level α.

Now, it seems rather cumbersome to derive the BCR for a composite H_1 by applying the conditions of the Neyman-Pearson Lemma to each simple $H_1' \in H_1$. The following theorem replaces this infeasible procedure by conditions that can be verified more directly. These conditions say that in order for a UMP test to exist, (1) the test problem may have only **one unknown parameter** , (2) the alternative hypothesis must be **one-sided** , and (3) each distribution in $\mathcal{W}$ must have a so-called **monotone density ratio** . The third condition means that, for all $\theta_1 > \theta_0$ with $\theta_0, \theta_1 \in \Theta$, the ratio $f(\boldsymbol{y}; \theta_1)/f(\boldsymbol{y}; \theta_0)$ (or the ratio $g(t; \theta_1)/g(t; \theta_0)$ in terms of the sufficient statistic $T(\boldsymbol{Y})$) must be a strictly monotonical function of $T(\boldsymbol{Y})$. The following example will illuminate this issue.

Example 2.12: To show that the normal distribution $N(\mu, \sigma_0^2)$ with unknown μ and known σ_0^2 has a monotone density ratio, we may directly inspect the simplified density ratio (2.5-33) from Example 2.7. We see immediately that the ratio is an increasing function of $T(\boldsymbol{y}) := \sum_{i=1}^n y_i$ when $\mu_1 > \mu_0$. □

Theorem 2.4. *Let $\mathcal{W}$ be a class of distributions with a one-dimensional parameter space and monotone density ratio in some statistic $T(\boldsymbol{Y})$.*

1. *Suppose that $H_0 : \bar{\theta} = \theta_0$ is to be tested against the* ***upper one-sided alternative*** *$H_1 : \bar{\theta} > \theta_0$. Then, there exists a UMP test ϕ_u at level α and a constant C with*

$$\phi_u(T(\boldsymbol{y})) := \begin{cases} 1, \text{ if } T(\boldsymbol{y}) > C, \\ 0, \text{ if } T(\boldsymbol{y}) < C \end{cases} \tag{2.5-50}$$

and

$$P_{\theta_0}\{\phi_u(T(\boldsymbol{Y})) = 1\} = \alpha. \tag{2.5-51}$$

2. *For testing H_0 against the* ***lower one-sided alternative*** *$H_1 : \bar{\theta} < \theta_0$, there exists a UMP test ϕ_l at level α and a constant C with*

$$\phi_l(T(\boldsymbol{y})) := \begin{cases} 1, \text{ if } T(\boldsymbol{y}) < C \\ 0, \text{ if } T(\boldsymbol{y}) > C \end{cases} \tag{2.5-52}$$

and

$$P_{\theta_0}\{\phi_l(T(\boldsymbol{Y})) = 1\} = \alpha. \tag{2.5-53}$$

Proof. To prove (1), consider first the case of a simple alternative $H_1 : \bar{\theta} = \theta_1$ for some $\theta_1 > \theta_0$. With the values for θ_0 and θ_1 fixed, the density ratio can be written as

$$\frac{f(\boldsymbol{y};\theta_1)}{f(\boldsymbol{y};\theta_0)} = \frac{g(T(\boldsymbol{y});\theta_1)}{g(T(\boldsymbol{y});\theta_0)} = h(T(\boldsymbol{y})),$$

that is, as a function of the observations alone. According to the Neyman-Pearson Lemma 2.1, the ratio must be large enough, i.e. $h(T(\boldsymbol{y})) > k$ with k depending on α. Now, if $T(\boldsymbol{y}_1) < T(\boldsymbol{y}_2)$ holds for some $\boldsymbol{y}_1, \boldsymbol{y}_2 \in S$, then certainly also $h(T(\boldsymbol{y}_1)) \leq h(T(\boldsymbol{y}_2))$ due to the assumption that the density ratio is monotone in $T((\boldsymbol{Y}))$. In other words, the observation $\boldsymbol{y}_2$ is in both cases at least as suitable as $\boldsymbol{y}_1$ for making the ratio h sufficiently large. In this way, the BCR may be equally well constructed by all the data $\boldsymbol{y} \in S$ for which $T(\boldsymbol{y})$ is large enough, for instance $T(\boldsymbol{y}) > C$, where the constant C must be determined such that the size of this BCR equals the prescribed value α. As these implications are true regardless of the exact value θ_1, the BCR will be the same for all the simple alternatives with $\theta_1 > \theta_0$. Therefore, the test (2.5-50) is UMP. The proof of (2) follows the same sequence of arguments with all inequalities reversed. □

The next theorem is of great practical value as it ensures that most of the standard distributions used in hypothesis testing have a monotone density ratio even in their non-central forms.

Theorem 2.5. *The following* $1P$*-distributions (with possibly additional known parameters* μ_0*,* σ_0^2, p_0 *and known degrees of freedom* f_0*,* $f_{1,0}$*,* $f_{2,0}$*) have a density with monotone density ratio in some statistic* T*:*

1. *Multivariate independent normal distributions* $N(\mathbf{1}\mu, \sigma_0^2\boldsymbol{I})$ *and* $N(\mathbf{1}\mu_0, \sigma^2\boldsymbol{I})$*,*
2. *Gamma distribution* $G(b, p_0)$*,*
3. *Beta distribution* $B(\alpha, \beta_0)$*,*
4. *Non-central Student distribution* $t(f_0, \lambda)$*,*
5. *Non-central Chi-squared distribution* $\chi^2(f_0, \lambda)$*,*
6. *Non-central Fisher distribution* $F(f_{1,0}, f_{2,0}, \lambda)$*,*

Proof. The proofs of (1) and (2) may be elegantly based on the more general result that any density that is a member of the **one-parameter exponential family**, defined by

$$f(\boldsymbol{y};\theta) = h(\boldsymbol{y})c(\theta)\exp\{w(\theta)T(\boldsymbol{y})\}, \quad h(\boldsymbol{y}) \geq 0, c(\theta) \geq 0, \tag{2.5-54}$$

(cf. Olive, 2006, for more details) has a monotone density ratio (see Lehmann and Romano, 2005, p. 67), and that the normal and Gamma distributions can be written in this form (2.5-54).

1. The density function of $N(\mathbf{1}\mu, \sigma_0^2\boldsymbol{I})$ (2.5-29) can be rewritten as

$$f(\boldsymbol{y};\mu) = (2\pi\sigma_0^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n y_i^2\right\}\exp\left\{-\frac{n\mu^2}{2\sigma_0^2}\right\}\exp\left\{\frac{\mu}{\sigma_0^2}\sum_{i=1}^n y_i\right\},$$

where $h(\boldsymbol{y}) := (2\pi\sigma_0^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n y_i^2\right\} \geq 0$, $c(\theta) := \exp\{-\frac{n\mu^2}{2\sigma_0^2}\} \geq 0$, $w(\theta) := \frac{\mu}{\sigma_0^2}$, and $T(\boldsymbol{y}) := \sum_{i=1}^n y_i$ satisfy (2.5-54). Similarly, the density function of $N(\mathbf{1}\mu_0, \sigma^2\boldsymbol{I})$ reads, in terms of (2.5-54),

$$f(\boldsymbol{y};\sigma^2) = I_{\mathbb{R}^n}(\boldsymbol{y})(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\mu_0)^2\right\},$$

where $h(\boldsymbol{y})$ corresponds to the indicator function $I_{\mathbb{R}^n}(\boldsymbol{y})$ with definite value one, $c(\theta) := (2\pi\sigma^2)^{-n/2} \geq 0$, $w(\theta) := -\frac{1}{2\sigma^2}$, and $T(\boldsymbol{y}) := \sum_{i=1}^n (y_i - \mu_0)^2$.

2. The Gamma distribution, defined by Equation 2.107 in Koch (1999, p. 112), with known parameter p_0 may be directly written as

$$f(y;b) = \frac{y^{p_0-1}}{\Gamma(p_0)}b^{p_0}\exp\{-by\} \qquad (b > 0, p_0 > 0, y \in \mathbb{R}^+),$$

where $h(\boldsymbol{y}) := y^{p_0-1}/\Gamma(p_0) \geq 0$, $c(\theta) := b^{p_0} \geq 0$, $w(\theta) := b$, and $T(\boldsymbol{y}) := -y$ satisfy (2.5-54).

3. - 6. The proofs for these distributions are lengthy and may be obtained from Lehmann and Romano (2005, p. 224 and 307).

□

Example 2.13: Test of the normal mean with known variance (composite alternatives). We are now in a position to extend Example 2.7 and seek BCRs for composite alternative hypotheses. For demonstration purposes, both the raw definition of a UMP test and the more convenient Theorem 2.4 will applied if possible. Let us first look at the formal statement of the test problems.

Let $Y_1, \ldots, Y_n$ be independently and normally distributed observations with common unknown mean $\bar{\mu}$ and common known variance $\bar{\sigma}^2 = \sigma_0^2$. Do a UMP tests for testing the simple null hypothesis $H_0 : \bar{\mu} = \mu_0$ against the composite alternative hypothesis

1. $H_1 : \bar{\mu} > \mu_0$,
2. $H_1 : \bar{\mu} < \mu_0$,
3. $H_1 : \bar{\mu} \neq \mu_0$

exist at level α, and if so, what are the BCRs? (It is assumed that μ_0, σ_0 and α have fixed numerical values.)

1. Recall from Example 2.7 (2.5-35) that the BCR for the test of $H_0 : \bar{\mu} = \mu_0$ against the simple $H_1 : \bar{\mu} = \mu_1$ with $\mu_1 > \mu_0$ is given by all the observations satisfying

$$\bar{y} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} \Phi^{-1}(1-\alpha),$$

when $\mu_1 > \mu_0$. Evidently, the critical region is the same for all the simple alternatives

$$\{H_1' : \bar{\mu} = \mu_1 \text{ with } \mu_1 > \mu_0\},$$

because it is independent of μ_1. Therefore, the critical function (2.5-37) is UMP for testing H_0 against the composite alternative $H_1 : \bar{\mu} > \mu_0$. The following alternative proof makes direct use of Theorem 2.4.

In Example 2.12, the normal distribution $N(\mu, \sigma_0^2)$ with known variance was already demonstrated to have a monotone density ratio in the sufficient statistic $\sum_{i=1}^n Y_i$ or in $T(\boldsymbol{Y}) := \sum_{i=1}^n Y_i/n$ as a reversible function thereof. As the current testing problem is about a single parameter, a one-sided H_1, and a class of distribution with monotone density ratio, all the conditions of Theorem 2.4 are satisfied. It remains to find a constant C such that the critical region (2.5-50) has size α according to condition (2.5-51). It is found easily because we know already that $T(\boldsymbol{Y})$ is distributed as $N(\mu_0, \sigma_0^2)$ under H_0, so that

$$\begin{aligned}\alpha &= P_{\mu_0,\sigma_0^2/n}\{\phi(\boldsymbol{Y}) = 1\} = P_{\mu_0,\sigma_0^2/n}\{\boldsymbol{Y} \in S_C\} = N_{\mu_0,\sigma_0^2/n}\{T(\boldsymbol{Y}) > C\} = 1 - N_{0,1}\left(\frac{T(\boldsymbol{Y}) - \mu_0}{\sigma_0/\sqrt{n}} < \frac{C - \mu_0}{\sigma_0/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{C-\mu_0}{\sigma_0/\sqrt{n}}\right),\end{aligned}$$

from which C follows to be

$$C = \mu_0 + \frac{\sigma_0}{\sqrt{n}} \Phi^{-1}(1-\alpha).$$

Note that the number C would change to $C = n\mu_0 + \sqrt{n}\sigma_0\sqrt{n}\Phi^{-1}(1-\alpha)$ if $\sum_{i=1}^n Y_i$ was used as the sufficient statistic instead of $\sum_{i=1}^n Y_i/n$, because the mean and variance of the normal distribution are affected by the factor $1/n$.

2. The proof of existence and determination of the BCR of a UMP test for testing H_0 versus $H_1 : \bar{\mu} < \mu_0$ is analogous to the first case above. All the conditions required by Theorem 2.4 are satisfied, and the constant C appearing in the UMP test (2.5-52) and satisfying (2.5-53) is now found to be

$$C = \mu_0 - \frac{\sigma_0}{\sqrt{n}} \Phi^{-1}(1-\alpha).$$

3. In this case, there is no common BCR for testing H_0 against $H_1 : \bar{\mu} \neq \mu_0$. Although the BCRs (2.5-35) and (2.5-36) do not individually depend on the value of μ_1, they differ in signs through the location of μ_1 relative to μ_0. Consequently, there is no UMP test for the two-sided alternative. This fact is also reflected by Theorem 2.4, which requires the alternative to be one-sided. □

2.5.4 Reduction to invariant statistics

We will now tackle the problem of testing a generally multi-parameter and composite null hypothesis

$$H_0 : \bar{\theta} \in \Theta_0$$

against a possibly composite and two-sided alternative

$$H_1 : \bar{\theta} \in \Theta_1$$

with the usual assumption that Θ_0 and Θ_1 are non-empty and disjoint subsets of the parameter space Θ, which is connected to a parametric family of densities

$$\mathcal{F} = \{f(\boldsymbol{y}; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\},$$

or equivalently

$$\mathcal{F}_T = \{f_T(\boldsymbol{T}(\boldsymbol{y}); \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\},$$

when (minimally) sufficient statistics $\boldsymbol{T}(\boldsymbol{Y})$ are used as ersatz observations for Y. Recall that sufficiency only reduces the dimension of the observation space, whereas it always leaves the parameter space unchanged. The problem is now that no UMP test exists when the parameter space is multi-dimensional or when the alternative hypothesis is two-sided, because the conditions of Theorem 2.4 would be violated. To overcome this serious limitation, we will investigate a reduction technique that may be applied in addition to a reduction by sufficiency, and that will oftentimes produce a simplified test problem for which a UMP test then exists.

Since any reduction beyond minimal sufficiency is necessarily bound to cause a loss of relevant information, it is essential to understand what kind of information may be safely discarded in a given test problem, and what the equivalent mathematical transformation is. The following example gives a first demonstration about the nature of such transformations.

Example 2.14: Recall from Example 2.13 that there exists no UMP test for testing $H_0 : \bar{\mu} = \mu_0 = 0$ against the two-sided $H_1 : \bar{\mu} \neq \mu_0 = 0$ (with $\bar{\sigma}^2 = \sigma_0^2$ known), as the one-sidedness condition of Theorem 2.4 is violated. However, if we discard the sign of the sample mean, i.e. if we only measure the absolute deviation of the sample mean from μ_0, and if we use the sign-insensitive statistic $\bar{Y}^2$ instead of $\bar{Y}$, then the problem becomes one of testing $H_0 : \bar{\mu}^2 = 0$ against the one-sided $H_1 : \bar{\mu}^2 > 0$. This is so because $\bar{Y} \sim N(\mu, \sigma_0^2/n)$ implies that $\frac{n}{\sigma_0^2} \bar{Y}^2$ has a non-central chi-squared distribution $\chi^2(1, \lambda)$ with one degree of freedom and non-centrality parameter $\lambda = \frac{n}{\sigma_0^2} \mu^2$ (see Koch, 1999, p. 127). Then, $\bar{\mu} = 0$ is equivalent to $\bar{\lambda} = 0$ under H_0, and $\bar{\mu} \neq 0$ is equivalent to $\bar{\lambda} > 0$ under H_1. As the transformed test problem is about a one-sided alternative and a test statistic with a monotone density ratio (by virtue of Theorem 2.5-4), the UMP test according to (2.5-50) of Theorem 2.4 is given by

$$\phi(\frac{n}{\sigma_0^2} \bar{y}^2) := \begin{cases} 1, \text{ if } \frac{n}{\sigma_0^2} \bar{y}^2 > C, \\ 0, \text{ if } \frac{n}{\sigma_0^2} \bar{y}^2 < C, \end{cases} \tag{2.5-55}$$

where, according to condition (2.5-51), C is fixed such that the size of (2.5-55) equals the prescribed value α. Using definition (2.4-16) and the fact that $\frac{n}{\sigma_0^2} \bar{Y}^2$ has a central chi-squared distribution with one degree of freedom under H_0, this is

$$\alpha = 1 - \chi^2_{1,0} \left(\frac{n}{\sigma_0^2} \bar{Y}^2 < C \right) = 1 - F_{\chi^2_{1,0}}(C),$$

which yields as the critical value

$$C = F^{-1}_{\chi^2_{1,0}}(1 - \alpha).$$

We will call the transformed problem of testing $H_0 : \bar{\lambda} = 0$ against $H_1 : \bar{\lambda} > 0$ the **invariance-reduced testing problem**, and the corresponding test 2.5-55 (based on Theorem 2.4) the **UMP invariant test**. It will be interesting to compare the power function of this test with the power functions of the UMP tests for the one-sided alternatives from Example 2.13. Using (2.4-22), the power function of the invariant test (2.5-55) reads

$$\mathrm{Pf}(\mu) = 1 - \chi^2_{1,n\mu^2/\sigma_0^2}\left(\frac{n}{\sigma_0^2}\bar{Y}^2 < C\right) = 1 - F_{\chi^2_{1,n\mu^2/\sigma_0^2}}\left(F^{-1}_{\chi^2_{1,0}}(1-\alpha)\right).$$

The power functions of the upper and lower one-sided UMP tests derived in Example 2.13 (here with the specific value $\mu_0 = 0$) are found to be

$$\mathrm{Pf}^{(1)}(\mu) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}}{\sigma_0}\mu\right)$$

and

$$\mathrm{Pf}^{(2)}(\mu) = 1 - \Phi\left(-\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}}{\sigma_0}\mu\right),$$

respectively.

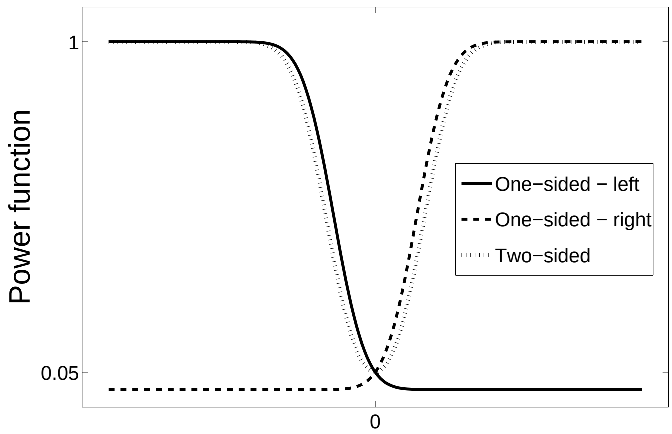

Fig. 2.4 Power functions for the two UMP tests at level $\alpha = 0.05$ about $H_1 : \bar{\mu} < 0$ ('lower one-sided') and $H_1 : \bar{\mu} > 0$ ('upper one-sided'), and a UMP invariant test about $H_1 : \bar{\mu} \neq 0$ reduced to $H_1 : \bar{\mu}^2 > 0$.

Figure 2.4 shows that each of the UMP tests has slightly higher power within their one-sided $\boldsymbol{\Theta}_1$-domains than the invariance-reduced test for the originally two-sided alternative. Observe that each of the UMP one-sided tests would have practically zero power when the value of the true parameter is unexpectedly on the other side of the parameter space. On the other hand, the invariance-reduced test guarantees reasonable power throughout the entire two-sided parameter space.

Clearly, the power function of the invariance-reduced test (2.5-55) is symmetrical with respect to $\mu = 0$, because the sign of the sample mean, and consequently that of the mean parameter, is not being considered. Therefore, we might say that this test has been designed to be **equally sensitive in both directions away from** $\mu_0 = 0$. In mathematical terminology, one would say that the test is **invariant under sign changes** $\bar{Y} \to \pm\bar{Y}$, and $\bar{Y}^2$ is a **sign-invariant statistic**, i.e. a random variable whose value remains unchanged when the sign of $\bar{Y}$ changes. Notice that the one-sidedness condition of the Theorem 2.4 is restored by virtue of the fact the parameter $\lambda = \frac{n}{\sigma_0^2}\mu^2$ of the new test statistic $\bar{Y}^2$ is now non-negative, thereby resulting in a one-sided H_1. The crucial point is, however, that the hypotheses of the reduced testing problem remain equivalent to the original hypotheses. □

Reduction by invariance is not only suitable for transforming a test problem about a two-sided H_1 into one about a one-sided H_1. In fact, we will see that the concept of invariance may also be applied to transform a testing problem involving multiple unknown parameters into a test problem with a single unknown parameter, as required by Theorem 2.4. To make this approach operable within the framework of general linear models, a number of definitions and theorems will be introduced now.

To begin with, it will be assumed throughout the remainder of this section that the original observations $\boldsymbol{Y}$ with sample space S have been reduced to minimally sufficient ersatz observations $\boldsymbol{T}(\boldsymbol{Y})$ with values in S_T

and with a collection of densities $\mathcal{F}_T$. In fact, Arnold (1985) showed that any inference based on the following invariance principles is exactly the same for $\boldsymbol{Y}$ and $T(\boldsymbol{Y})$. Then, let us consider an invertible transformation g of the ersatz observations from S_T to S_T (such as the sign change of the sample mean $T(\boldsymbol{Y}) = \bar{Y}$ in Example 2.14). Typically, such a statistic $g(\boldsymbol{T})$ will induce a corresponding transformation $\bar{g}(\boldsymbol{\theta})$ of the parameters from $\boldsymbol{\Theta}$ to $\boldsymbol{\Theta}$ (such as $\bar{Y} \rightarrow \pm\bar{Y}$ induces $\mu \rightarrow \pm\mu$ in Example 2.14).

What kind of transformation g is suitable for reducing a test problem in a meaningful way? According to Arnold (1981, p. 11), the first desideratum is that any transformation g with induced $\bar{g}$ leaves the hypotheses of the given test problem unchanged. In other words, we require that

$$(1)\ \bar{\boldsymbol{g}}(\boldsymbol{\Theta}) := \{\bar{g}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\} = \boldsymbol{\Theta} \tag{2.5-56}$$

$$(2)\ \bar{\boldsymbol{g}}(\boldsymbol{\Theta}_0) := \{\bar{g}(\boldsymbol{\theta}_0) : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\} = \boldsymbol{\Theta}_0 \tag{2.5-57}$$

$$(3)\ \bar{\boldsymbol{g}}(\boldsymbol{\Theta}_1) := \{\bar{g}(\boldsymbol{\theta}_1) : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1\} = \boldsymbol{\Theta}_1 \tag{2.5-58}$$

$$(4)\ g(\boldsymbol{T}) \text{ has a density function in } \{f_T(g(\boldsymbol{T}); \bar{g}(\boldsymbol{\theta})) : \bar{g}(\boldsymbol{\theta}) \in \boldsymbol{\Theta}\} \tag{2.5-59}$$

holds (see also Cox and Hinkley, 1974, p. 157). If such a transformation of the testing problem exists, we will say that **the testing problem is invariant under g (with induced $\bar{g}$)**. In Example 2.14 we have seen that the hypotheses in terms of the parameter μ is equivalent to the hypotheses in terms of the new parameter $\lambda = g(\mu)$ when the reversal of the sign is used as transformation g.

The second desideratum (cf. Arnold, 1981, p. 11) is that any transformation g with induced $\bar{g}$ leaves the test decision, that is, the critical function ϕ unchanged. Mathematically, this is expressed by the condition

$$\phi(g(\boldsymbol{t})) = \phi(\boldsymbol{t}) \qquad (\text{for all } \boldsymbol{t} \in S_T). \tag{2.5-60}$$

If this is the case for some transformation g, we will say that **the critical function** or **test is invariant under g (with induced $\bar{g}$)**.

The first desideratum, which defines an invariant test problem, may also be interpreted such that if we observe $g(\boldsymbol{t})$ (with some density function $f_T(g(\boldsymbol{t}); \bar{g}(\boldsymbol{\theta})))$ rather than $\boldsymbol{t}$ (with density function $f_T(\boldsymbol{t}; \boldsymbol{\theta})$), and if the hypotheses are equivalent in the sense that $H_0 : \bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_0 \Leftrightarrow \bar{g}(\bar{\boldsymbol{\theta}}) \in \boldsymbol{\Theta}_0$ and $H_1 : \bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_1 \Leftrightarrow \bar{g}(\bar{\boldsymbol{\theta}}) \in \boldsymbol{\Theta}_1$, then the test problem about the transformed data $g(\boldsymbol{T})$ is clearly the same as that in terms of the original data $\boldsymbol{T}$. Then it seems logical to apply a decision rule ϕ which yields the same result no matter if $g(\boldsymbol{t})$ or $\boldsymbol{t}$ has been observed. But this is the very proposition of the second desideratum, which says that $\phi(g(\boldsymbol{t}))$ should equal $\phi(\boldsymbol{t})$.

Example 2.14 constitutes the rare case that a test problem is invariant under a single transformation g. Usually, test problems are invariant under a certain collection $\mathcal{G}$ of (invertible) transformations g within the data domain with a corresponding collection $\bar{\mathcal{G}}$ of (invertible) transformations $\bar{g}$ within the parameter domain. The following proposition reflects a very useful fact about such collections of transformations (see Arnold, 1981, p. 12).

Proposition 2.1. *If a test problem is invariant under some invertible transformations $g \in \mathcal{G}$, $g_1 \in \mathcal{G}$, and $g_2 \in \mathcal{G}$ (from a space S_T to S_T) with induced transformations $\bar{g} \in \bar{\mathcal{G}}$, $\bar{g}_1 \in \bar{\mathcal{G}}$, and $\bar{g}_2 \in \bar{\mathcal{G}}$ (from a space Θ to Θ), then it is also invariant under the inverse transformation g^{-1} and the composition $g_1 \circ g_2$ of two transformations, and the induced transformations are $\overline{g^{-1}} = \bar{g}^{-1}$ and $\overline{g_1 \circ g_2} = \bar{g}_1 \circ \bar{g}_2$, respectively.*

If a test problem remains invariant under each $g \in \mathcal{G}$ with induced $\bar{g} \in \bar{\mathcal{G}}$, then this proposition says that both $\mathcal{G}$ and $\bar{\mathcal{G}}$ are **closed under compositions and inverses** (which will again be elements of S_T and Θ, respectively). In that case, $\mathcal{G}$ and $\bar{\mathcal{G}}$ are said to be **groups**. Let us now investigate how invariant tests may be generally constructed.

We have seen in Example 2.14 that a reasonable test may be based on an **invariant statistic**, which remains unchanged by transformations in $\mathcal{G}$ (such as $M(\bar{Y}) := \bar{Y}^2$ under $g(\bar{Y}) = \pm\bar{Y}$). Clearly, any statistic $\boldsymbol{M}(\boldsymbol{T})$ that is to be invariant under a collection G of transformations on S_T must satisfy

$$\boldsymbol{M}(\boldsymbol{T}) = \boldsymbol{M}(\boldsymbol{g}(\boldsymbol{T})) \tag{2.5-61}$$

for all $\boldsymbol{g} \in \mathcal{G}$. However, the invariance condition (2.5-61) alone does not necessarily guarantee that a test which is based on such a statistic $\boldsymbol{M}(\boldsymbol{T})$ is itself invariant. In fact, whenever two data points $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ from S_T produce the same value $\boldsymbol{M}(\boldsymbol{t}_1) = \boldsymbol{M}(\boldsymbol{t}_2)$ for the invariant statistic, the additional condition

$$\boldsymbol{t}_1 = \boldsymbol{g}(\boldsymbol{t}_2) \tag{2.5-62}$$

is required to hold for some $\boldsymbol{g} \in \mathcal{G}$. An invariant statistic which satisfies also (2.5-62) is called a **maximal invariant**. Condition (2.5-62) ensures that $\mathcal{G}$ is the largest collection under which the testing problem is invariant.

Example 2.15: As in Example 2.10, let $\boldsymbol{T}(\boldsymbol{Y}) := [\bar{Y},\ S^2]'$ be the vector of jointly sufficient statistics for independently and normally distributed observations $Y_1, \ldots, Y_n$ with common unknown mean $\bar{\mu}$ and common unknown variance $\bar{\sigma}^2$. The problem of testing $H_0 : \bar{\mu} = 0$ versus $H_1 : \bar{\mu} \neq 0$ is invariant under the transformation

$$\boldsymbol{g}\left(\begin{bmatrix} \bar{Y} \\ S^2 \end{bmatrix}\right) = \begin{bmatrix} (-1) \cdot \bar{Y} \\ S^2 \end{bmatrix},$$

which we will write in the form

$$\boldsymbol{g}\left(\bar{Y}, S^2\right) = \left((-1) \cdot \bar{Y}, S^2\right)$$

for convenience. To see this, we first notice that $\boldsymbol{g}$ induces the transformation

$$\bar{\boldsymbol{g}}\left(\mu, \sigma^2\right) = \left((-1) \cdot \mu, \sigma^2\right),$$

because $(-1) \cdot \bar{Y} \sim N(-\mu, \sigma^2)$, while S^2 (and thus its distribution) remains unchanged. With $\boldsymbol{\Theta} = \mathbb{R} \times \mathbb{R}^+$, $\boldsymbol{\Theta}_0 = \{0\} \times \mathbb{R}^+$ and $\boldsymbol{\Theta}_1 = \mathbb{R} - \{0\} \times \mathbb{R}^+$, we obtain

$$\begin{aligned} \bar{\boldsymbol{g}}(\boldsymbol{\Theta}_0) &= \{\bar{\boldsymbol{g}}\left(\mu, \sigma^2\right) : \mu = 0, \sigma^2 \in \mathbb{R}^+\} = \{\left(0, \sigma^2\right) : \sigma^2 \in \mathbb{R}^+\} = \boldsymbol{\Theta}_0, \\ \bar{\boldsymbol{g}}(\boldsymbol{\Theta}_1) &= \{\bar{\boldsymbol{g}}\left(\mu, \sigma^2\right) : \mu \neq 0, \sigma^2 \in \mathbb{R}^+\} = \{\left(-\mu, \sigma^2\right) : \mu \neq 0, \sigma^2 \in \mathbb{R}^+\} = \boldsymbol{\Theta}_1. \end{aligned}$$

Due to $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$, $\bar{g}(\boldsymbol{\Theta}) = \boldsymbol{\Theta}$ also holds. Thus, the above testing problem is invariant under the transformation $\boldsymbol{g}$. Consider now the statistic

$$M(\boldsymbol{T}) = M\left(\bar{Y}, S^2\right) := \frac{\bar{Y}^2}{S^2}.$$

This statistic is invariant because of

$$M\left(\boldsymbol{g}\left(\bar{Y}, S^2\right)\right) = \left((-1) \cdot \bar{Y}, S^2\right) = \frac{(-1)^2 \cdot \bar{Y}^2}{S^2} = \frac{\bar{Y}^2}{S^2} = M\left(\bar{Y}, S^2\right).$$

Let us now investigate the question whether M is also maximally invariant. Suppose that $\boldsymbol{t}_1 = [\bar{y}_1,\ s_1^2]'$ and $\boldsymbol{t}_2 = [\bar{y}_2,\ s_2^2]'$ are two realizations of $\boldsymbol{T}(\boldsymbol{Y})$. Then $M(\boldsymbol{t}_1) = M(\boldsymbol{t}_2)$ is seen to hold e.g. for $\bar{y}_2 = 2\bar{y}_1$ and $s_2^2 = 4s_1^2$ because of

$$M\left(\boldsymbol{t}_2\right) = M\left(\bar{y}_2, \sigma_2^2\right) = M\left(2\bar{y}_1, 4\sigma_1^2\right) = \frac{4\bar{y}_1^2}{4s_1^2} = \frac{\bar{y}_1^2}{s_1^2} = M\left(\bar{y}_1, \sigma_1^2\right) = M\left(\boldsymbol{t}_1\right).$$

However, the necessary condition $\boldsymbol{t}_1 = \boldsymbol{g}(\boldsymbol{t}_2)$ is not satisfied, since

$$\boldsymbol{g}\left(\bar{y}_2, \sigma_2^2\right) = \left((-1) \cdot \bar{y}_2, \sigma_2^2\right) = \left(-2\bar{y}_1, 4\sigma_1^2\right) \neq \boldsymbol{t}_1.$$

Consequently, M must be invariant under a larger group of transformations than $\boldsymbol{g}$. Indeed, M can be shown to be maximally invariant under the group of transformations defined by

$$\boldsymbol{g}_c\left(\bar{Y}, S^2\right) = \left(c\bar{Y}, c^2 S^2\right) \qquad (c \neq 0),$$

which includes the above transformation with $c = -1$. Arnold (1981, Section 1.5) demonstrates a technique for proving maximality, which shall be outlined here as well. First, we assume that $\boldsymbol{t}_1 = [\bar{y}_1,\ \sigma_1^2]'$ and $\boldsymbol{t}_2 = [\bar{y}_2,\ \sigma_2^2]'$ are two realizations of $\boldsymbol{T}(\boldsymbol{Y})$ for which $M(\boldsymbol{t}_1) = M(\boldsymbol{t}_2)$ holds. If we find some $c \neq 0$ for which $\boldsymbol{t}_1 = \boldsymbol{g}_c(\boldsymbol{t}_2)$ is satisfied, then M follows to be maximally invariant. Observe next that, using the above definition of M, the assumption $M(\bar{y}_1, \sigma_1^2) = M(\bar{y}_2, \sigma_2^2)$ is equivalent to $\bar{y}_1^2/s_1^2 = \bar{y}_2^2/s_2^2$ or $\left(\bar{y}_1/\bar{y}_2\right)^2 = s_1^2/s_2^2$. Then, if we define $c := \bar{y}_1/\bar{y}_2$, we see immediately that $\bar{y}_1 = c\bar{y}_2$ and $s_1^2 = c^2 s_2^2$, and we have

$$\boldsymbol{t}_1 = \begin{bmatrix} \bar{y}_1 \\ s_1^2 \end{bmatrix} = \begin{bmatrix} c\bar{y}_2 \\ c^2 s_2^2 \end{bmatrix} = \boldsymbol{g}_c\left(\begin{bmatrix} \bar{y}_2 \\ s_2^2 \end{bmatrix}\right) = \boldsymbol{g}_c(\boldsymbol{t}_2)$$

as desired. □

The following proposition from Arnold (1981, p. 13) ensures that maximal invariants exist generally.

Proposition 2.2. *For any group $\mathcal{G}$ of invertible transformations g from an arbitrary space S_T to S_T there exists a maximal invariant.*

The next theorem provides the maximal invariants for some groups of transformations that will be particularly useful for reducing testing problems.

Theorem 2.6. *Let $\boldsymbol{T}$ be a random vector, S and T random variables, and c a positive real number. Then,*

1. *$M(T, S^2) = T^2/S^2$ is a maximal invariant statistic under the group $\mathcal{G}$ of* ***scale changes***

$$\boldsymbol{g}(T, S^2) = (cT, c^2S^2) \quad (c > 0). \tag{2.5-63}$$

2. *$M(T^2, S^2) = T^2/S^2$ is a maximal invariant statistic under the group $\mathcal{G}$ of* ***scale changes***

$$\boldsymbol{g}(T^2, S^2) = (c^2T^2, c^2S^2) \quad (c > 0). \tag{2.5-64}$$

3. *$M(T) = T^2$ is a maximal invariant statistic under the* ***sign change***

$$g(T) = (-1) \cdot T. \tag{2.5-65}$$

4. *$\boldsymbol{M}(T, S^2) = (T^2, S^2)$ is a maximal invariant under the* ***sign change***

$$\boldsymbol{g}(T, S^2) = ((-1) \cdot T, S^2). \tag{2.5-66}$$

5. *$M(\boldsymbol{T}) = \boldsymbol{T}'\boldsymbol{T}$ is a maximal invariant statistic under the group $\mathcal{G}$ of* ***orthogonal transformations***

$$\boldsymbol{g}(\boldsymbol{T}) = \boldsymbol{\Gamma}\boldsymbol{T}, \tag{2.5-67}$$

where $\boldsymbol{\Gamma}$ is an arbitrary orthogonal matrix.

6. *$\boldsymbol{M}(\boldsymbol{T}, S^2) = (\boldsymbol{T}'\boldsymbol{T}, S^2)$ is a maximal invariant statistic under the group $\mathcal{G}$ of* ***orthogonal transformations***

$$\boldsymbol{g}(\boldsymbol{T}, S^2) = (\boldsymbol{\Gamma}\boldsymbol{T}, S^2), \tag{2.5-68}$$

where $\boldsymbol{\Gamma}$ is an arbitrary orthogonal matrix.

Proof.

1. See Example 2.15.

2. $M(T^2, S^2)$ is an invariant statistic because of

$$M(\boldsymbol{g}(T^2, S^2)) = M(c^2T^2, c^2S^2) = \frac{c^2T^2}{c^2S^2} = \frac{T^2}{S^2} = M(T^2, S^2).$$

To prove maximality, suppose that $M(t_1^2, s_1^2) = M(t_2^2, s_2^2)$ holds. From this, the equivalent conditions $t_1^2/s_1^2 = t_2^2/s_2^2$ and $t_1^2/t_2^2 = s_1^2/s_2^2$ follow. Defining $c^2 := t_1^2/t_2^2$ results in $t_1^2 = c^2t_2^2$ and $s_1^2 = c^2s_2^2$, that is

$$\begin{bmatrix} t_1^2 \\ s_1^2 \end{bmatrix} = \begin{bmatrix} c^2t_2^2 \\ c^2s_2^2 \end{bmatrix} = \boldsymbol{g}\left(\begin{bmatrix} t_2^2 \\ s_2^2 \end{bmatrix}\right)$$

as required.

3. Invariance of $M(T)$ follows from

$$M(g(T)) = M((-1) \cdot T) = (-1)^2 \cdot T^2 = T^2 = M(T).$$

Then, let t_1 and t_2 be two realizations of T for which $M(t_1) = M(t_2)$ holds. This equation is equivalent to $t_1^2 = t_2^2$, which is satisfied by $t_1 = -t_2$. Hence, $t_1 = g(t_2)$, which proves that $M(T)$ is a maximally invariant statistic under g.

4. The proof of this fact follows from the same line of reasoning as 3.

5. As any orthogonal matrix satisfies $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}$, we obtain

$$M(\boldsymbol{g}(\boldsymbol{T})) = M(\boldsymbol{\Gamma}\boldsymbol{T}) = (\boldsymbol{\Gamma}\boldsymbol{T})'(\boldsymbol{\Gamma}\boldsymbol{T}) = \boldsymbol{T}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{T} = \boldsymbol{T}'\boldsymbol{T} = M(\boldsymbol{T}),$$

which shows that $M(\boldsymbol{T})$ is an invariant statistic. To prove maximality, let $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ be two non-zero realizations of $\boldsymbol{T}$, for which $M(\boldsymbol{t}_1) = M(\boldsymbol{t}_2)$, or equivalently, $\boldsymbol{t}_1'\boldsymbol{t}_1 = \boldsymbol{t}_2'\boldsymbol{t}_2$ holds. This condition expresses that the vectors $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ must have equal length. Then, there always exists an orthogonal transformation $\boldsymbol{\Gamma}\boldsymbol{t}_1$ which does not change the length of $\boldsymbol{t}_1$ (see Meyer, 2000, Characterization #4 regarding the matrix $\boldsymbol{P}$, p. 321), that is, which satisfies $\boldsymbol{t}_1'\boldsymbol{t}_1 = \boldsymbol{t}_2'\boldsymbol{t}_2$ for some vector $\boldsymbol{t}_2$.

6. The proof of this fact follows from the same line of reasoning as 5.

□

2.5.5 Uniformly most powerful invariant (UMPI) tests

Let us begin the current section with the following definition. Since any group $\mathcal{G}$ of transformations acting on the observation space S_T induces a corresponding group $\bar{\mathcal{G}}$ of transformations acting on the parameter space $\boldsymbol{\Theta}$, there will exist a maximal invariant $\overline{\boldsymbol{M}}(\boldsymbol{\theta})$ under $\bar{\mathcal{G}}$, which will be called the **parameter maximal invariant**.

Theorem 2.7. *$\phi(\boldsymbol{T})$ is an invariant critical function if and only if there exists $\phi^*(\boldsymbol{M}(\boldsymbol{T}))$ such that $\phi(\boldsymbol{t}) = \phi^*(\boldsymbol{M}(\boldsymbol{t}))$ holds for every $\boldsymbol{t} \in S_T$. Then, the distribution of $\boldsymbol{M}(\boldsymbol{T})$ depends only on $\overline{\boldsymbol{M}}(\boldsymbol{\theta})$, the induced maximal invariant under $\bar{\mathcal{G}}$.*

Proof. See Arnold (1981, p. 13). □

From Theorem 2.7 it becomes evident that we may restrict attention to the invariance-reduced problem of testing

$$H_0 : \overline{\boldsymbol{M}}(\bar{\boldsymbol{\theta}}) \in \overline{\boldsymbol{M}}(\Theta_0)$$

against

$$H_1 : \overline{\boldsymbol{M}}(\bar{\boldsymbol{\theta}}) \in \overline{\boldsymbol{M}}(\Theta_1)$$

based on maximally invariant statistics $\boldsymbol{M}(\boldsymbol{T})$ with distribution depending on parameters $\overline{\boldsymbol{M}}(\boldsymbol{\theta})$. If a complete reduction by invariance is possible, then $\boldsymbol{M}(\boldsymbol{T})$ will be a scalar test statistic depending on a single parameter $\overline{\boldsymbol{M}}(\boldsymbol{\theta})$, and the transformed spaces $\overline{\boldsymbol{M}}(\Theta_0)$ and $\overline{\boldsymbol{M}}(\Theta_1)$ will represent a single point (simple H_0) and a one-sided interval (one-sided H_1), respectively. Given that such a one-dimensional test statistic $M(\boldsymbol{T})$ has a monotone density ratio, all the requirements of Theorem 2.4 are satisfied by the fully invariance-reduced test problem. The UMP critical function for the invariant test problem then reads

$$\phi(M(\boldsymbol{t})) := \begin{cases} 1, \text{ if } M(\boldsymbol{t}) > C, \\ 0, \text{ if } M(\boldsymbol{t}) < C \end{cases} \tag{2.5-69}$$

if H_1 is an upper one-sided alternative, and

$$\phi(M(\boldsymbol{t})) := \begin{cases} 1, \text{ if } M(\boldsymbol{t}) < C, \\ 0, \text{ if } M(\boldsymbol{t}) > C \end{cases} \tag{2.5-70}$$

if H_1 is a lower one-sided alternative hypothesis. In both cases the critical value must satisfy the condition

$$P_{\theta_0} \{\phi(M(\boldsymbol{T})) = 1\} = \alpha, \tag{2.5-71}$$

which guarantees that the test ϕ has fixed level α. Recall also that $\boldsymbol{t} = \boldsymbol{T}(\boldsymbol{y})$ contains the values of the sufficient statistics $\boldsymbol{T}$ at the observed data $\boldsymbol{y}$. Since such a test (if it exists as presumed here) is UMP (at level α) for the invariance-reduced test problem (using group $\mathcal{G}$), it is UMP among all tests that are invariant under $\mathcal{G}$. Therefore, ϕ will be called the **UMP invariant (UMPI) test (at level α)** for testing the original hypotheses $H_0 : \bar{\theta} \in \Theta_0$ against $H_1 : \bar{\theta} \in \Theta_1$.

Parenthesis: Let us return for a moment to Example 2.14 and the problem of testing $H_0 : \bar{\mu} = 0$ versus $H_1 : \bar{\mu} \neq 0$. If we inspect the power function of the UMPI test in Figure 2.4, we see that it does not fall below the level α. This generally desirable property of a test is called **unbiasedness**. Arnold (1981, Theorem 1.13) states that any UMPI test is also unbiased. Without going into details, it should be mentioned that there exist testing problems for which no UMPI tests exist, but for which a test can be found which is UMP within the class of all unbiased tests. However, we will not be concerned with such testing problems in this thesis. Instead, the reader interested in the concept of such optimally unbiased (UMPU) tests is referred to Koch (1999, p. 277), where conditions for the existence of UMPU tests are given, or to Lehmann and Romano (2005, Chapters 4 and 5) for a detailed discussion of that topic.

Example 2.16 (Example 2.14 restated): Test of the normal mean with known variance - Two-sided alternative. Let $Y_1, \ldots, Y_n$ be independently and normally distributed observations with common unknown mean $\bar{\mu}$ and common known variance $\bar{\sigma}^2 = \sigma_0^2$. What is the best critical region for a test of the simple null hypothesis $H_0 : \bar{\mu} = \mu_0$ against the two-sided alternative hypothesis $H_1 : \bar{\mu} \neq \mu_0$ at level α?

This example is slightly more general than Example 2.14, because the hypotheses are not centered around 0. However, the simple transformation $Y_i' = Y_i - \mu_0$ of the original random variables Y_i into variables Y_i' solves this technical problem. Such a procedure is justified in light of the fact that the distribution of $\boldsymbol{Y} \sim N(\mathbf{1}\mu, \sigma_0^2 \boldsymbol{I})$ is transformed into $N(\mathbf{1}(\mu - \mu_0), \sigma_0^2 \boldsymbol{I})$ for $\boldsymbol{Y}'$ without changing the second moment. Thus the true mean of the Y_i' is now $\bar{\mu}' := \bar{\mu} - \mu_0$, and the hypotheses become $H_0 : \bar{\mu}' = \bar{\mu} - \mu_0 = 0$ and $H_1 : \bar{\mu}' = \bar{\mu} - \mu_0 \neq 0$, respectively. Therefore, we may restrict ourselves to the simple case $\mu_0 = 0$ knowing that a test problem about $\mu_0 \neq 0$ may always be centered by transforming the observations. It should be mentioned here that the cases of correlated and/or heteroscedastic observations will be discussed in the context of the normal Gauss-Markov model in Section 3.

Now, the test problem

$$\begin{aligned} &\boldsymbol{Y} \sim N(\mathbf{1}\mu, \sigma_0^2 \boldsymbol{I}) \\ &H_0 : \bar{\mu} = 0 \ \text{ against } \ H_1 : \bar{\mu} \neq 0 \end{aligned} \tag{2.5-72}$$

does not allow for a UMP test since H_1 is two-sided. This fact does not change after reducing the problem about $\boldsymbol{Y}$ to the equivalent test problem

$$\begin{aligned} &T(\boldsymbol{Y}) = \bar{Y} \sim N(\mu, \sigma_0^2/n) \\ &H_0 : \bar{\mu} = 0 \ \text{ against } \ H_1 : \bar{\mu} \neq 0 \end{aligned} \tag{2.5-73}$$

about the sample mean $\bar{Y}$ used as a sufficient statistic $T(\boldsymbol{Y})$ for μ. However, by using the invariance principle, this test problem may be transformed into a problem about a one-sided H_1. To be more specific, the test problem is invariant under sign changes

$$g(\bar{Y}) = (-1) \cdot \bar{Y} \sim N(-\mu, \sigma_0^2/n),$$

and the induced transformation acting on the parameter space is obviously

$$\bar{g}(\mu) = (-1) \cdot \mu.$$

Due to $\bar{g}(\Theta_0) = \Theta_0$ (with Θ_0 degenerating to the single point $\mu_0 = 0$), $\bar{g}(\Theta_1) = \Theta_1$ (with $\Theta_1 = \mathbb{R} - 0$) and $\bar{g}(\Theta) = \Theta$, the problem is indeed invariant under g. From Theorem 2.6-3 it follows that $M(\bar{Y}) = \bar{Y}^2$ is a maximal invariant under sign changes. To obtain a test statistic with a standard distribution, it is more convenient to use the standardized sample mean $\frac{\sqrt{n}}{\sigma_0} \bar{Y} \sim N(\frac{\sqrt{n}}{\sigma_0} \mu, 1)$ as a sufficient statistic, which is possible, because any reversible function of a sufficient statistic is itself sufficient. Then, the maximally invariant test statistic

$$M(\boldsymbol{Y}) = (\frac{\sqrt{n}}{\sigma_0} \bar{Y})^2$$

has a non-central chi-squared distribution $\chi^2(1, \lambda)$ with one degree of freedom and non-negative non-centrality parameter $\lambda = \frac{n}{\sigma_0^2} \mu^2$. Now, it is easily seen that the new testing problem

$$\begin{aligned} &M(\boldsymbol{Y}) = \frac{n}{\sigma_0^2} \bar{Y}^2 \sim \chi^2(1, \lambda) \quad \text{with } \lambda = \frac{n}{\sigma_0^2} \mu^2 \\ &H_0 : \bar{\lambda} = 0 \ \text{ against } \ H_1 : \bar{\lambda} > 0 \end{aligned} \tag{2.5-74}$$

is equivalent to the original problem of testing $H_0 : \bar{\mu} = 0$ against $H_1 : \bar{\mu} \neq 0$, because $\bar{\mu} = 0$ is equivalent $\bar{\lambda} = 0$ (when H_0 is true), and $\bar{\mu} \neq 0$ is equivalent to $\bar{\lambda} > 0$ (when H_1 is true). As this reduced testing problem involves only one unknown parameter λ (which corresponds to $\overline{\boldsymbol{M}(\boldsymbol{\theta})}$ in Theorem 2.7) and a one-sided H_1, and since the distribution of $M(\boldsymbol{Y})$ has a monotone density ratio by virtue of Theorem 2.5-5, a UMP test ϕ exists as a consequence of Theorem 2.4 with

$$\phi(\boldsymbol{y}) := \begin{cases} 1, \text{ if } \ M(\boldsymbol{y}) = n \, \frac{\bar{Y}^2}{\sigma_0^2} > k_{1-\alpha}^{\chi^2(1)}, \\ 0, \text{ if } \ M(\boldsymbol{y}) = n \, \frac{\bar{Y}^2}{\sigma_0^2} < k_{1-\alpha}^{\chi^2(1)}, \end{cases} \tag{2.5-75}$$

and critical value $k_{1-\alpha}^{\chi^2(1,\lambda=0)}$. Recall that the critical value is always computed under the assumption of a true H_0, which is why $\lambda = 0$. Then, ϕ is also the UMPI test (at level α) for the original test problem. This test may be written equivalently in terms of the $N(0,1)$-distributed test statistic $\sqrt{M(\boldsymbol{Y})}$, that is,

$$\phi(\boldsymbol{y}) := \begin{cases} 1, \text{ if } \sqrt{M(\boldsymbol{y})} = \sqrt{n}\,\frac{|\bar{Y}|}{\sigma_0} > k_{1-\alpha/2}^{N(0,1)}, \\ 0, \text{ if } \sqrt{M(\boldsymbol{y})} = \sqrt{n}\,\frac{|\bar{Y}|}{\sigma_0} < k_{1-\alpha/2}^{N(0,1)}, \end{cases} \tag{2.5-76}$$

with critical value $k_{1-\alpha/2}^{N(0,1)} = \Phi^{-1}(1-\alpha/2)$. □

Example 2.17: Test of the normal mean with unknown variance - Two-sided alternative. Let $Y_1, \ldots, Y_n$ be independently and normally distributed observations with common unknown mean $\bar{\mu}$ and common unknown variance $\bar{\sigma}^2$. What is the best critical region for a test of the composite null hypothesis $H_0 : \bar{\mu} = \mu_0$ $(\bar{\sigma}^2 > 0)$ against the two-sided alternative hypothesis $H_1 : \bar{\mu} \neq \mu_0$ $(\bar{\sigma}^2 > 0)$ at level α?

As demonstrated in Example 2.16, it will be sufficient to consider $\mu_0 = 0$ without loss of generality. We have seen in Example 2.10 that the observations $\boldsymbol{Y}$ may be reduced without loss of information to the jointly sufficient statistic $\boldsymbol{T}(\boldsymbol{Y}) = [\bar{Y}, S^2]'$ where $S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$ denotes the sample variance. Therefore, the given test problem

$$\begin{aligned} &\boldsymbol{Y} \sim N(\mathbf{1}\mu, \sigma^2\boldsymbol{I}) \\ &H_0 : \bar{\mu} = 0 \;\text{ against }\; H_1 : \bar{\mu} \neq 0 \end{aligned}$$

may be written as

$$\begin{aligned} &T_1(\boldsymbol{Y}) = \bar{Y} \sim N(\mu, \sigma^2/n) \\ &T_2(\boldsymbol{Y}) = (n-1)S^2/\sigma^2 \sim \chi^2(n-1) \\ &H_0 : \bar{\mu} = 0 \;\text{ against }\; H_1 : \bar{\mu} \neq 0 \end{aligned} \tag{2.5-77}$$

In the present case we are not only faced with the problem of a two-sided H_1 (which we already learnt to handle in Example 2.16), but with the additional challenge of a two-dimensional parameter space. Let us investigate both problems separately by finding suitable groups of transformations first. We will then combine the results later on to obtain the final solution to the test problem.

To begin with, it is easily verified that the test problem (2.5-77) in terms of $T_1(\boldsymbol{Y})$ and $T_2(\boldsymbol{Y})$ is invariant under the group $\mathcal{G}_1$ of sign changes acting on the sample mean. With

$$\boldsymbol{g}_1(\bar{Y}, S^2) = ((-1)\cdot\bar{Y}, S^2)$$

the induced transformation is identified as

$$\bar{\boldsymbol{g}}_1(\mu, \sigma^2) = ((-1)\cdot\mu, \sigma^2)$$

due to the change in distribution

$$\bar{Y} \sim N(\mu, \sigma^2/n) \longrightarrow (-1)\cdot\bar{Y} \sim N(-\mu, \sigma^2/n).$$

As already explained in Example 2.16, sign changes $\bar{\boldsymbol{g}}_1(\mu, \sigma^2)$ do not affect the hypotheses as they are symmetrical about 0. According to Theorem 2.6-4, maximal invariants under $\boldsymbol{g}_1$ are $[\bar{Y}^2, S^2]$, or $\boldsymbol{M}_1(\bar{Y}, S^2) := [\frac{n}{\sigma^2}\bar{Y}^2, (n-1)S^2]$ after rescaling, which leads to the reduced problem

$$\begin{aligned} &M_{1,1}(\boldsymbol{Y}) = \frac{n}{\sigma^2}\,\bar{Y}^2 \sim \chi^2(1,\lambda) \quad \text{with } \lambda = \frac{n}{\sigma^2}\,\mu^2 \\ &M_{1,2}(\boldsymbol{Y}) = (n-1)S^2/\sigma^2 \sim \chi^2(n-1) \\ &H_0 : \bar{\lambda} = 0 \;\text{ against }\; H_1 : \bar{\lambda} > 0 \end{aligned} \tag{2.5-78}$$

Although the alternative hypothesis is now one-sided, there are still two statistics for the two unknown parameters λ and σ^2. Therefore, we conclude that reduction by sign invariance alone does not go far enough in this case.

In addition to being sign-invariant, the test problem (2.5-77) can also be shown to be scale-invariant, that is, invariant under the group $\mathcal{G}_2$ of scale changes

$$\boldsymbol{g}_2(\bar{Y}, S^2) = (c\bar{Y}, c^2S^2).$$

This transformation arises when the scale of the original observations $\boldsymbol{Y}$ is changed by multiplying them with a positive constant c, because the distribution of $c\boldsymbol{Y}$ changes to $N(\mathbf{1}c\mu, c^2\sigma^2\boldsymbol{I})$, and the sufficient statistics to $c\bar{Y}$ and c^2S^2, respectively. Evidently, $\mathcal{G}_2$ induces $\bar{\mathcal{G}}_2$ with

$$\bar{\boldsymbol{g}}_2(\mu, \sigma^2) = (c\mu, c^2\sigma^2),$$

because of the transitions in distribution

$$\begin{aligned} \bar{Y} \sim N(\mu, \sigma^2/n) &\rightarrow c \cdot \bar{Y} \sim N(c\mu, c^2\sigma^2/n), \\ S^2 \sim G((n-1)/2, 2\sigma^2)/(n-1) &\rightarrow c^2S^2 \sim G((n-1)/2, 2c^2\sigma^2)/(n-1). \end{aligned}$$

That the test problem is indeed scale-invariant is seen from the fact that $\bar{\mathcal{G}}_2$ (with $c > 0$) does not change the hypotheses due to

$$\begin{aligned} \bar{\boldsymbol{g}}_2(\Theta_0) &= \bar{\boldsymbol{g}}_2(\{(\mu, \sigma^2) : \mu = 0, \sigma^2 \in \mathbb{R}^+\}) \\ &= \{(c \cdot 0, c^2\sigma^2) : \sigma^2 \in \mathbb{R}^+\} \\ &= \{(0, \sigma_c^2) : \sigma_c^2 \in \mathbb{R}^+\} \\ &= \Theta_0 \end{aligned}$$

and

$$\begin{aligned} \bar{\boldsymbol{g}}_2(\Theta_1) &= \bar{\boldsymbol{g}}_2(\{(\mu, \sigma^2) : \mu \in \mathbb{R} - \{0\}, \sigma^2 \in \mathbb{R}^+\}) \\ &= \{(c\mu, c^2\sigma^2) : \mu \in \mathbb{R} - \{0\}, \sigma^2 \in \mathbb{R}^+\} \\ &= \{(\mu_c, \sigma_c^2) : \mu_c \in \mathbb{R} - \{0\}, \sigma_c^2 \in \mathbb{R}^+\} \\ &= \Theta_1. \end{aligned}$$

By virtue of Theorem 2.6-1 a maximal invariant (rescaled by $\sqrt{n}$) under $\mathcal{G}_2$ is given by

$$M_2(\boldsymbol{Y}) = \frac{\sqrt{n}\,\bar{Y}}{\sqrt{S^2}} = \sqrt{n}\,\frac{\bar{Y}}{S} \tag{2.5-79}$$

which has a $t(n-1, \lambda)$-distribution with $n-1$ degrees of freedom and non-centrality parameter $\lambda = \sqrt{n}\,\frac{\mu}{\sigma}$. The resulting test problem

$$\begin{aligned} &M_2(\boldsymbol{Y}) = \sqrt{n}\,\frac{\bar{Y}}{S} \sim t(n-1, \lambda) \\ &H_0 : \bar{\lambda} = 0 \;\text{ against }\; H_1 : \bar{\lambda} \neq 0 \end{aligned} \tag{2.5-80}$$

now has a reduced parameter space in light of the single parameter λ. However, due to $\sigma > 0$, the originally two-sided alternative $H_1 : \bar{\mu} \neq 0$ is only equivalent to $\bar{\lambda} \neq 0$ because a negative $\bar{\mu}$ will cause a negative $\bar{\lambda}$, and a positive $\bar{\mu}$ leads to a positive $\bar{\lambda}$. In summary, reduction by scale invariance could successfully produce an equivalent test problem about one single parameter, but the problem concerning the two-sidedness of H_1 could not be resolved.

Parenthesis: Since the present test problem is invariant under two different groups of transformations, and since either group does not simplify the problem far enough, it is logical to seek a maximal invariant as a test statistic that corresponds to a test problem which is invariant under both groups. The following theorem is of great practical value as it allows us determine a maximal invariant step by step.

Theorem 2.8. *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two groups of transformations from S_T to S_T and let $\mathcal{G}$ be the smallest group containing $\mathcal{G}_1$ and $\mathcal{G}_2$. Suppose that $M_1(\boldsymbol{T})$ is a maximal invariant under $\mathcal{G}_1$ and that $M_1(\boldsymbol{T})$ satisfies $M_1(g_2(\boldsymbol{T})) = \widehat{g}_2(M_1(\boldsymbol{T}))$. Further, let $M_2(\boldsymbol{T})$ be a maximal invariant under the group $\widehat{\mathcal{G}}_2$ of transformations $\widehat{g}_2$. Then $M(\boldsymbol{T}) = M_2(M_1(\boldsymbol{T}))$ is a maximal invariant under $\mathcal{G}$.*

Proof. See Stuart et al. (1999, p. 297). □

Example 2.17 (continued): Let us now combine these complementary results. Theorem 2.8 allows us to determine the maximal invariant under the union $\mathcal{G}$ of the two sub-groups $\mathcal{G}_1$ and $\mathcal{G}_2$ sequentially. $\boldsymbol{M}_1$ as the maximal invariant under the group $\mathcal{G}_1$ of sign changes with $\boldsymbol{g_1}(\bar{Y}, S^2) = ((-1)\cdot\bar{Y}, S^2)$ can be shown to satisfy $\boldsymbol{M}_1(\boldsymbol{g_2}(\boldsymbol{T})) = \widehat{\boldsymbol{g}}_2(\boldsymbol{M}_1(\boldsymbol{T}))$, because there exists a transformation $\widehat{g}_2$ such that

$$\boldsymbol{M}_1(\boldsymbol{g_2}(\bar{Y}^2, S^2)) = (c^2\bar{Y}^2, c^2S^2) = \widehat{\boldsymbol{g}}_2(\bar{Y}^2, S^2),$$

where $\widehat{\mathcal{G}}_2$ is the group of transformations $\widehat{\boldsymbol{g}}_2$ from Theorem 2.6-2, which gives $M_2(\bar{Y}, S^2) = \bar{Y}^2/S^2$ as its maximal invariant. It follows from Theorem 2.8 that

$$M(\bar{Y}, S^2) = M_2(\boldsymbol{M}_1(\bar{Y}, S^2)) = \bar{Y}^2/S^2 \tag{2.5-81}$$

is the total maximal invariant under the union of $\mathcal{G}_1$ and $\mathcal{G}_2$. Now recall that we may always use rescaled versions of maximal invariants. Then, due to $\frac{n}{\sigma^2}\,\bar{Y}^2 \sim \chi^2(1,\lambda)$ with $\lambda = \frac{n}{\sigma^2}\,\mu^2$ and $(n-1)\,\frac{S^2}{\sigma^2} \sim \chi^2(n-1)$, the ratio

$$M(\boldsymbol{Y}) := \frac{\frac{n}{\sigma^2}\,\bar{Y}^2}{(n-1)\,\frac{S^2}{\sigma^2}/(n-1)} = \frac{n\bar{Y}^2}{S^2} \tag{2.5-82}$$

follows an $F(1, n-1, \lambda)$-distribution. Since $\mu = 0$ is equivalent to $\lambda = 0$ and $\mu \neq 0$ to $\lambda > 0$ (with arbitrary $\sigma^2 > 0$), the hypotheses of the original test problem (2.5-77) are equivalent to $H_0 : \bar{\lambda} = 0$ versus $H_1 : \bar{\lambda} > 0$. In summary, the fully reduced, both sign- and scale-invariant test problem reads

$$M(\boldsymbol{Y}) = \frac{n\bar{Y}^2}{S^2} \sim F(1, n-1, \lambda) \quad \text{with } \lambda = \frac{n}{\sigma^2}\,\mu^2 \tag{2.5-83}$$

$$H_0 : \bar{\lambda} = 0 \ \text{ against } \ H_1 : \bar{\lambda} > 0 \tag{2.5-84}$$

As Theorem 2.5-6 shows that the non-central F-distribution with known degrees of freedom and unknown non-centrality parameter λ has a monotone density ratio, all three conditions (one unknown parameter, one-sided H_1, and a test statistic with monotone density ratio) for the existence of UMP test are satisfied, and Theorem 2.4 gives the best test

$$\phi(\boldsymbol{y}) := \begin{cases} 1, \text{ if } M(\boldsymbol{y}) = n\,\frac{\bar{Y}^2}{S^2} > k_{1-\alpha}^{F(1,n-1)}, \\ 0, \text{ if } M(\boldsymbol{y}) = n\,\frac{\bar{Y}^2}{S^2} < k_{1-\alpha}^{F(1,n-1)}, \end{cases} \tag{2.5-85}$$

with critical value is given by $k_{1-\alpha}^{F(1,n-1)}$. By definition it follows that ϕ is the UMPI test for the original test problem (2.5-77). This test is usually given in terms of $\sqrt{M(\boldsymbol{y})}$ which has Student's distribution $t(n-1)$, that is,

$$\phi(\boldsymbol{y}) := \begin{cases} 1, \text{ if } \sqrt{M(\boldsymbol{y})} = \sqrt{n}\,\frac{|\bar{Y}|}{S} > k_{1-\alpha/2}^{t(n-1)}, \\ 0, \text{ if } \sqrt{M(\boldsymbol{y})} = \sqrt{n}\,\frac{|\bar{Y}|}{S} < k_{1-\alpha/2}^{t(n-1)}, \end{cases} \tag{2.5-86}$$

with critical value $k_{1-\alpha/2}^{t(n-1)}$. □

The purpose of Examples 2.16 and 2.17 was to demonstrate that the standard tests concerning the mean with the variance either known (2.5-76) or unknown (2.5-86) are optimal within the class of invariant tests.

Equipped with these tools for reducing the space of observations and the space of parameters to one-dimensional intervals, we could now proceed and investigate more complex test problems, which occur very often in the context of linear models. Linear models essentially constitute generalizations of the observation model $N(\boldsymbol{1}\mu, \sigma^2\boldsymbol{I})$ to $N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{P}^{-1})$, where $\boldsymbol{X\beta}$ represents a possibly multi-dimensional and non-constant mean, and where $\sigma^2\boldsymbol{P}^{-1}$ indicates that the observations might be correlated and of non-constant variance. However, to keep the theoretical explanations short, the reader interested in tests within the context of linear models is referred to Section 3. At this point we will continue the current section by presenting convenient one-step reduction techniques that are oftentimes equivalent to a sequential reduction by sufficiency and invariance.

2.5.6 Reduction to the Likelihood Ratio and Rao's Score statistic

At the beginning of the current Section 2.5, where we discussed the case of testing against a simple H_1, we have seen that the density ratio $f(\boldsymbol{y};\theta_1)/f(\boldsymbol{y};\theta_0)$, used as a test statistic for the MP test as defined in (2.5-26) by the Neyman-Pearson Lemma, may be simplified such that a single sufficient statistic can be used as an equivalent test statistic. Whenever a test problem has a one-sided H_1 (but still only one single unknown parameter), then that sufficient statistic must have a distribution with a monotone density ratio in order for a UMP test to exist. If multiple parameters are unknown, then our approach was to shrink the dimension of the parameter space to 1 in order to have a one-dimensional test statistic at one's disposal. Such a test statistic was derived as the maximal invariant under a group of transformations, which lead us to tests that are UMP among all tests invariant under these transformations.

As a test problem may be invariant under numerous sub-groups of transformations (such as the one in Example 2.17), a manual step-wise reduction of a test problem by invariance can become quite cumbersome. Therefore, we will investigate ways for obtaining a UMPI test in a more direct manner. We will see that there are in fact two equivalent methods for reducing a test problem about n observations (or equally well about m minimal sufficient statistics) and m unknown parameters to a one-parameter problem with a one-sided H_1.

Reduction to the Likelihood Ratio statistic. Let us consider the problem of testing $H_0 : \bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_0$ against $H_1 : \bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_1$ on the basis of observations $\boldsymbol{Y}$ with true density function in

$$\mathcal{F} = \{f(\boldsymbol{y};\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}\,.$$

Inspection of the density ratio $f(\boldsymbol{y};\theta_1)/f(\boldsymbol{y};\theta_0)$, used for testing a simple H_0 versus a simple H_1 by the Neyman-Pearson Lemma, reveals that this quantity is not unique anymore if the hypotheses are composite, i.e. if θ_0 and θ_1 are elements of intervals $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$. In that case it would not be clear at which values θ_0 and θ_1 the density ratio should be evaluated. This situation is of course not improved if the densities comprise multiple unknown parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. One approach to removing the ambiguity of the density ratio consists in taking the maximum value of the density function over $\boldsymbol{\Theta}_0$ and over $\boldsymbol{\Theta}_1$, that is, to determine the value of

$$\frac{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_1} f(\boldsymbol{y};\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} f(\boldsymbol{y};\boldsymbol{\theta})}. \tag{2.5-87}$$

Since the densities in (2.5-87) are now treated as functions of $\boldsymbol{\theta}$ rather than of $\boldsymbol{y}$, it is necessary to switch the arguments, or formally to introduce a new function from $\boldsymbol{\Theta}$ to $\mathbb{R}$, which is defined as

$$L(\boldsymbol{\theta};\boldsymbol{y}) := f(\boldsymbol{y};\boldsymbol{\theta}), \tag{2.5-88}$$

and which treats $\boldsymbol{y}$ as given constants. L is called the **likelihood function (for $\boldsymbol{y}$)**, and the fraction

$$\frac{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} L(\boldsymbol{\theta};\boldsymbol{y})}{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_1} L(\boldsymbol{\theta};\boldsymbol{y})} \tag{2.5-89}$$

denotes the **generalized likelihood ratio**. Notice that in this definition the nominator and denominator have been switched with respect to the generalized density ratio (2.5-87). To take this change into account when comparing this ratio with the critical value, we only need to switch the </>-relation accordingly. If the hypotheses are such that $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$, then we may modify (2.5-89) slightly into

$$GLR := \frac{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} L(\boldsymbol{\theta};\boldsymbol{y})}{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta};\boldsymbol{y})}. \tag{2.5-90}$$

The only difference between (2.5-90) and (2.5-89) is that (2.5-90) may take the value 1, because $\boldsymbol{\Theta}_0$ is a subset of $\boldsymbol{\Theta}$ (see also Koch, 1999, p. 279, for a discussion of the properties of the generalized likelihood ratio).

All the examples discussed so far and all the applications to be investigated in Sections 3 and 4 allow us to rewrite the hypotheses $H_0 : \bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_0$ and $H_1 : \bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_1$ in the form of **linear constraints (restrictions)**

$$H_0 : \boldsymbol{H}\bar{\boldsymbol{\theta}} = \boldsymbol{w} \quad \text{versus} \quad H_1 : \boldsymbol{H}\bar{\boldsymbol{\theta}} \left\{ \begin{matrix} < \\ > \\ \neq \end{matrix} \right\} \boldsymbol{w}, \tag{2.5-91}$$

where $\boldsymbol{H}$ is an $(r \times u)$-matrix with known constants and rank r, and where $\boldsymbol{w}$ is an $(r \times 1)$-vector of known constants.

Example 2.18: In the Examples 2.13, 2.14, and 2.16 we considered the problems of testing the specification $H_0 : \bar{\mu} = \mu_0$ of the mean parameter against the alternative specifications $H_1 : \bar{\mu} < \mu_0$, $H_1 : \bar{\mu} > \mu_0$, and $H_1 : \bar{\mu} \neq \mu_0$ on the basis of normally distributed observations with known variance. These hypotheses may be rewritten in the form (2.5-91) by using the vectors/matrices $\bar{\theta} := [\,\bar{\mu}\,]$, $H := [\,1\,]$, and $w := [\,\mu_0\,]$, which are all scalars in this case.

In Example 2.17, we investigated the problem of testing $H_0 : \bar{\mu} = \mu_0$ versus $H_1 : \bar{\mu} \neq \mu_0$ in the same class of normal distributions, but with unknown variance. These hypotheses are expressed as in (2.5-91) by defining

$$\bar{\boldsymbol{\theta}} := \begin{bmatrix} \bar{\mu} \\ \bar{\sigma}^2 \end{bmatrix}, \; \boldsymbol{H} := [\,1,\; 0\,], \text{ and } \boldsymbol{w} := [\,\mu_0\,]. \qquad \square$$

When the hypotheses are given in terms of linear restrictions (2.5-91), then the maxima in (2.5-90) may be interpreted in the following way. The value $\widetilde{\boldsymbol{\theta}}$ for which the likelihood function in the nominator of (2.5-90) attains its maximum over $\boldsymbol{\Theta}_0$, or equivalently for which the constraint $\boldsymbol{H}\bar{\boldsymbol{\theta}} = \boldsymbol{w}$ holds, is called the **restricted maximum likelihood (ML) estimate for $\boldsymbol{\theta}$**. On the other hand, the value $\widehat{\boldsymbol{\theta}}$ for which the likelihood function in the denominator of (2.5-90) attains its maximum over the entire parameter space $\boldsymbol{\Theta}$, denotes then the **unrestricted maximum likelihood (ML) estimate for $\boldsymbol{\theta}$**.

If we assume that the likelihood function is at least twice differentiable with positive definite *Hessian*, then the restricted ML estimate $\widetilde{\boldsymbol{\theta}}$ is obtained as the solution of

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left(L(\boldsymbol{\theta}; \boldsymbol{y}) - \boldsymbol{k}'(\boldsymbol{H}\boldsymbol{\theta} - \boldsymbol{w}) \right) = \boldsymbol{0}, \tag{2.5-92}$$

where $\boldsymbol{k}$ denotes an $(r \times 1)$-vector of unknown Lagrange multipliers. The unrestricted ML estimate $\widehat{\boldsymbol{\theta}}$ follows as the solution of the likelihood equation

$$\frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}; \boldsymbol{y}) = \boldsymbol{0}. \tag{2.5-93}$$

Then, rewriting (2.5-90) in terms of the ML estimators and the random vector $\boldsymbol{Y}$ yields

$$GLR(\boldsymbol{Y}) = \frac{L(\widetilde{\boldsymbol{\theta}}; \boldsymbol{Y})}{L(\widehat{\boldsymbol{\theta}}; \boldsymbol{Y})}. \tag{2.5-94}$$

This is the reciprocal of the statistic that Koch (1999, Chap. 4.2) and Teunissen (2000, Chap. 3) use to derive the test of the general hypothesis in the normal Gauss-Markov model. In that case, which will also be addressed in detail in Section 3 of this thesis, the restricted und unrestricted ML estimates are equivalent to the restricted and unrestricted least squares estimates. However, it shall already be mentioned here that there are important cases where the Gauss-Markov model is not restricted to the class of normal distributions, but where the likelihood function may depend on additional distribution parameters (see Application 7). For this reason, we will maintain the more general notation in terms of the likelihood function and the restricted/unrestricted ML estimates, and we will speak of restricted/unrestricted least squares estimates only if we apply a class of normal distributions.

In certain cases, it will sometimes be more convenient to use a logarithmic version of the GLR, that is,

$$-2 \ln GLR = -2 \ln \frac{L(\widetilde{\boldsymbol{\theta}}; \boldsymbol{Y})}{L(\widehat{\boldsymbol{\theta}}; \boldsymbol{Y})} = -2 \left(\ln L(\widetilde{\boldsymbol{\theta}}; \boldsymbol{Y}) - \ln L(\widehat{\boldsymbol{\theta}}; \boldsymbol{Y}) \right). \tag{2.5-95}$$

Due to the strictly increasing monotonicity of the *logarithmus naturalis*, the estimates $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ remain unchanged if, as in (2.5-95), the so-called **log-likelihood function**

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) := \ln L(\boldsymbol{\theta}; \boldsymbol{y}) \tag{2.5-96}$$

is maximized instead of the likelihood function (2.5-88). This property guarantees that the restricted ML estimate $\widetilde{\boldsymbol{\theta}}$ is also the solution of

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left(\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) - \boldsymbol{k}'(\boldsymbol{H}\boldsymbol{\theta} - \boldsymbol{w}) \right) = \boldsymbol{0}, \tag{2.5-97}$$

and that the unrestricted ML estimate $\widehat{\boldsymbol{\theta}}$ is the solution of the log-likelihood equation

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \boldsymbol{0}. \tag{2.5-98}$$

One advantage of this approach is that the logarithm of a Gaussian density will cancel out with the exponential operator, which results in a function that is easier to handle (see Example 2.19). We call the statistic T_{LR}, defined by

$$T_{LR}(\boldsymbol{Y}) := -2\ln GLR = -2\left(\mathcal{L}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) - \mathcal{L}(\widehat{\boldsymbol{\theta}};\boldsymbol{Y})\right), \tag{2.5-99}$$

the **likelihood ratio (LR) statistic**.

A test which uses (2.5-94) as test statistic is called the **generalized likelihood ratio (GLR) test**, given by

$$\phi_{GLR}(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \ L(\widetilde{\boldsymbol{\theta}};\boldsymbol{y})/L(\widehat{\boldsymbol{\theta}};\boldsymbol{y}) < k^*_\alpha, \\ 0 & \text{if } \ L(\widetilde{\boldsymbol{\theta}};\boldsymbol{y})/L(\widehat{\boldsymbol{\theta}};\boldsymbol{y}) > k^*_\alpha, \end{cases} \tag{2.5-100}$$

where the critical value k^*_α is such that ϕ has level α. Alternatively, the test

$$\phi_{LR}(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \ -2\left(\mathcal{L}(\widetilde{\boldsymbol{\theta}};\boldsymbol{y}) - \mathcal{L}(\widehat{\boldsymbol{\theta}};\boldsymbol{y})\right) > k_\alpha, \\ 0 & \text{if } \ -2\left(\mathcal{L}(\widetilde{\boldsymbol{\theta}};\boldsymbol{y}) - \mathcal{L}(\widehat{\boldsymbol{\theta}};\boldsymbol{y})\right) < k_\alpha, \end{cases} \tag{2.5-101}$$

based on the statistic (2.5-99) is called the **likelihood ratio (LR) test**. Both tests are truly equivalent because both the corresponding statistics and critical values are strictly monotonic functions of each other.

It is easily verified that the test (2.5-100) is equivalent to the MP test (2.5-26) of Neyman and Pearson if both H_0 and H_1 are simple hypotheses and if $\boldsymbol{\theta}$ is a single parameter, because the maxima then equal the point values of the densities at $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, respectively. Furthermore, if H_1 is a one-sided hypothesis (with $\boldsymbol{\theta}$ still being a single parameter) and if $\boldsymbol{Y}$ has a density with monotone density ratio, then the GLR/LR test is also equal to the UMP test in Theorem 2.4 (see Lemma 2 in Birkes, 1990). Even more importantly, if a test problem involves multiple parameters in a *normal Gauss-Markov model*, then the GLR/LR test is also identical to the UMPI test obtained from a step-wise reduction by invariance. We will demonstrate this fact, which has been proven by Lehmann (1959b), in the following simple example and in greater detail in Section 3.

Example 2.19 (Example 2.17 revisited): The LR test of the normal mean with unknown variance. Let $Y_1, \ldots, Y_n$ be independently and normally distributed observations with common unknown mean $\bar{\mu}$ and common unknown variance $\bar{\sigma}^2$. What is the LR test for testing $H_0 : \bar{\mu} = \mu_0$ $(\bar{\sigma}^2 > 0)$ versus $H_1 : \bar{\mu} \neq \mu_0$ $(\bar{\sigma}^2 > 0)$ at level α?

Using the fact that the joint density of independently distributed observations is the product of the univariate densities, we obtain for the log-likelihood function

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) &:= \ln f(\boldsymbol{y};\mu,\sigma^2) \\
&= \ln \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right\} \\
&= \sum_{i=1}^n \ln\left[(2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right\}\right] \\
&= \sum_{i=1}^n \left[0 - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y_i-\mu)^2\right] \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2
\end{aligned}$$

Let us first determine the unrestricted ML estimates for μ and σ^2 by applying (2.5-98). From the first order conditions

$$\frac{\partial}{\partial\mu}\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) = \frac{1}{\sigma^2}\sum_{i=1}^n (y_i-\mu) = 0, \tag{2.5-102}$$

$$\frac{\partial}{\partial\sigma^2}\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (y_i-\mu)^2 = 0, \tag{2.5-103}$$

we obtain the solutions

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{2.5-104}$$

and

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\mu})^2. \tag{2.5-105}$$

Notice that $\widehat{\sigma}^2$ differs from the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \widehat{\mu})^2$. The restricted estimates result from solving (2.5-97), that is,

$$\begin{aligned}
\frac{\partial}{\partial \mu}\left(\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) - k(\mu - \mu_0)\right) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) - k = 0,\\
\frac{\partial}{\partial \sigma^2}\left(\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) - k(\mu - \mu_0)\right) &= -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2 = 0,\\
\frac{\partial}{\partial k}\left(\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) - k(\mu - \mu_0)\right) &= -(\mu - \mu_0) = 0.
\end{aligned}$$

The third equation reproduces the restriction, that is

$$\widetilde{\mu} = \mu_0. \tag{2.5-106}$$

Using this result, the second equation gives

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_0)^2. \tag{2.5-107}$$

Substituting $\widetilde{\mu} = \mu_0$ and $\widetilde{\sigma}^2$ into the first equation results in the estimate

$$\widetilde{k} = \frac{1}{\widetilde{\sigma}^2}\sum_{i=1}^{n}(y_i - \mu_0) \tag{2.5-108}$$

for the Lagrange multiplier. To evaluate the test statistics based on the generalized likelihood ratio, we need to compute the likelihood function both at the unrestricted and the restricted estimates, which leads to

$$L(\widehat{\mu}, \widehat{\sigma}^2; \boldsymbol{y}) = (2\pi\widehat{\sigma}^2)^{-n/2} \exp\left\{-\frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}(y_i - \widehat{\mu})^2\right\} = (2\pi)^{-n/2}(\widehat{\sigma}^2)^{-n/2} \exp\left\{-\frac{n}{2}\right\}$$

and

$$L(\widetilde{\mu}, \widetilde{\sigma}^2; \boldsymbol{y}) = (2\pi\widetilde{\sigma}^2)^{-n/2} \exp\left\{-\frac{1}{2\widetilde{\sigma}^2}\sum_{i=1}^{n}(y_i - \mu_0)^2\right\} = (2\pi)^{-n/2}(\widetilde{\sigma}^2)^{-n/2} \exp\left\{-\frac{n}{2}\right\}.$$

With this, the GLR in (2.5-94) tales the value

$$GLR(\boldsymbol{y}) = \frac{L(\widetilde{\mu}, \widetilde{\sigma}^2; \boldsymbol{y})}{L(\widehat{\mu}, \widehat{\sigma}^2; \boldsymbol{y})} = \left(\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2}\right)^{-n/2},$$

and the value of the LR statistic in (2.5-99) becomes

$$T_{LR}(\boldsymbol{y}) = -2\ln GLR = n\ln\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2}.$$

We will now show that $T_{LR}(\boldsymbol{Y})$ (thus also the GLR statistic) is equivalent to the statistic $M(\boldsymbol{Y})$ in (2.5-83) of the UMPI test (2.5-85) for testing $H_0 : \bar{\mu} = \mu_0 = 0$ ($\bar{\sigma}^2 > 0$) versus $H_1 : \bar{\mu} \neq \mu_0 = 0$ ($\bar{\sigma}^2 > 0$) at level α. Recall that $M(\boldsymbol{Y}) = n\frac{\bar{Y}^2}{S^2}$ with sample mean $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$. Then, due to $\bar{Y} = \widehat{\mu}$ and $(n-1)S^2 = n\widehat{\sigma}^2$, we have

$$1 + \frac{M(\boldsymbol{Y})}{n-1} = 1 + \frac{n\bar{Y}^2}{(n-1)S^2} = 1 + \frac{\widehat{\mu}^2}{\widehat{\sigma}^2} = \frac{\widehat{\sigma}^2 + \widehat{\mu}^2}{\widehat{\sigma}^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\mu})^2 + \widehat{\mu}^2}{\widehat{\sigma}^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} y_i^2}{\widehat{\sigma}^2} = \frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2}.$$

Thus, $T_{LR}(\boldsymbol{Y}) = n\ln(1+\frac{M(\boldsymbol{Y})}{n-1})$ is a strictly monotonically increasing function of $M(\boldsymbol{Y})$. If we transform the critical value $C = k_{1-\alpha}^{F(1,n-1)}$ of the UMPI test (2.5-85) accordingly into $C^* = n\ln(1+\frac{C}{n-1})$, then the LR test

$$\phi_{LR}(\boldsymbol{y}) := \begin{cases} 1, \text{ if } T_{LR}(\boldsymbol{y}) = n\ln\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} > C^*, \\ 0, \text{ if } T_{LR}(\boldsymbol{y}) = n\ln\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} < C^*, \end{cases} \tag{2.5-109}$$

will produce the same result as the UMPI test (2.5-85). □

Reduction to Rao's Score statistic. Another way to formulate the Likelihood Ratio statistic (2.5-99) results from applying a two-term Taylor series to the log-likelihood function (2.5-96). For this purpose, we will assume throughout this thesis that the first two derivatives of the log-likelihood function exist. If the unrestricted ML estimate is used as Taylor point, then we obtain

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) = \mathcal{L}(\widehat{\boldsymbol{\theta}};\boldsymbol{y}) + \left.\frac{\partial\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}}\right|_{\widehat{\boldsymbol{\theta}}} + \frac{1}{2}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})' \left.\frac{\partial^2\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}). \tag{2.5-110}$$

The vector of first partial derivatives

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\theta};\boldsymbol{y}) := \frac{\partial\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}} \tag{2.5-111}$$

is called the **(log-likelihood or efficient) score**. The *Hessian* matrix of second partial derivatives will be denoted by

$$\boldsymbol{\mathcal{H}}(\boldsymbol{\theta};\boldsymbol{y}) := \frac{\partial^2\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}. \tag{2.5-112}$$

This matrix, which appears in the exact residual term of the Taylor series (2.5-110), is evaluated at possibly different points between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$. Now, it follows from the log-likelihood equations (2.5-98) that the score vector vanishes at $\widehat{\boldsymbol{\theta}}$, that is $\boldsymbol{\mathcal{S}}(\widehat{\boldsymbol{\theta}};\boldsymbol{y}) = \boldsymbol{0}$. Then, evaluation of (2.5-110) at the restricted ML estimate $\widetilde{\boldsymbol{\theta}}$ yields

$$\mathcal{L}(\widetilde{\boldsymbol{\theta}};\boldsymbol{y}) = \mathcal{L}(\widehat{\boldsymbol{\theta}};\boldsymbol{y}) + \frac{1}{2}(\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}})'\boldsymbol{\mathcal{H}}(\boldsymbol{\theta}^*;\boldsymbol{y})(\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}),$$

or

$$-2\left(\mathcal{L}(\widetilde{\boldsymbol{\theta}};\boldsymbol{y}) - \mathcal{L}(\widehat{\boldsymbol{\theta}};\boldsymbol{y})\right) = -(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}})'\boldsymbol{\mathcal{H}}(\boldsymbol{\theta}^*;\boldsymbol{y})(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}).$$

We will now use an argument by Stuart et al. (1999, p. 57) stating that, in terms of random variables, $\boldsymbol{\mathcal{H}}(\boldsymbol{\theta}^*;\boldsymbol{Y}) \approx \boldsymbol{E}\{\boldsymbol{\mathcal{H}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})\}$ for large n. Note that if the log-likelihood function is naturally given as a quadratic function of $\boldsymbol{\theta}$, then the Hessian will be a matrix of constants. In that case, we will write the Hessian as $\boldsymbol{\mathcal{H}}_{\boldsymbol{Y}}$, which is then identical to $\boldsymbol{E}\{\boldsymbol{\mathcal{H}}_{\boldsymbol{Y}}\}$. The expectation of the negative Hessian of the log-likelihood function, that is

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta};\boldsymbol{Y}) := \boldsymbol{E}\{-\boldsymbol{\mathcal{H}}(\boldsymbol{\theta};\boldsymbol{Y})\} \tag{2.5-113}$$

is called the **information matrix**. With this, we obtain for the test statistic

$$\begin{aligned} T_{LR}(\boldsymbol{Y}) &= -2\left(\mathcal{L}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) - \mathcal{L}(\widehat{\boldsymbol{\theta}};\boldsymbol{Y})\right) \\ &\approx (\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}})'\boldsymbol{E}\{-\boldsymbol{\mathcal{H}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}) \\ &= (\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}})'\boldsymbol{\mathcal{I}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}) \end{aligned} \tag{2.5-114}$$

In a second step we apply a one-term Taylor series to the score with Taylor point $\widetilde{\boldsymbol{\theta}}$, that is

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\theta};\boldsymbol{y}) = \boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{y}) + \left.\frac{\partial\boldsymbol{\mathcal{S}}(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}^{**}}(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}).$$

Then we evaluate the score at the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$, which gives

$$\boldsymbol{\mathcal{S}}(\widehat{\boldsymbol{\theta}};\boldsymbol{y}) = \boldsymbol{0} = \boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{y}) + \boldsymbol{\mathcal{H}}(\boldsymbol{\theta}^{**};\boldsymbol{y})(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}).$$

Now, the same argument applies as above, i.e. $\boldsymbol{\mathcal{H}}(\boldsymbol{\theta}^{**};\boldsymbol{Y}) \approx \boldsymbol{E}\{\boldsymbol{\mathcal{H}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})\}$ for large n, which gives

$$\boldsymbol{\mathcal{I}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}) \approx \boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}),$$

or

$$\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}} \approx \boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})\boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})$$

Again, if the log-likelihood function is quadratic in $\boldsymbol{\theta}$, then the above approximations become exact.

Substituting the last equation for $\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}$ into (2.5-114) finally yields

$$T_{LR}(\boldsymbol{Y}) \approx \boldsymbol{\mathcal{S}}'(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})\boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})\boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) =: T_{RS}(\boldsymbol{Y}). \tag{2.5-115}$$

The statistic $T_{RS}(\boldsymbol{Y})$ is called (the) **Rao's Score (RS) statistic** (see Equation 6e.3.6 in Rao, 1973, p. 418), which was originally proposed in Rao (1948) for the problem of testing $H_0 : \bar{\theta} = \theta_0$ versus $H_1 : \bar{\theta} > \theta_0$ with a single unknown parameter θ. In this one-dimensional case, Rao's Score statistic takes the simple form

$$T_{RS}(\boldsymbol{Y}) = \mathcal{S}^2(\theta_0;\boldsymbol{Y})/\mathcal{I}(\theta_0;\boldsymbol{Y}). \tag{2.5-116}$$

Example 2.20 (Example 2.17 revisited): The RS test of the normal mean with unknown variance. Let $Y_1,\ldots,Y_n$ be independently and normally distributed observations with common unknown mean $\bar{\mu}$ and common unknown variance $\bar{\sigma}^2$. What is the RS test for testing $H_0 : \bar{\mu} = \mu_0$ ($\bar{\sigma}^2 > 0$) versus $H_1 : \bar{\mu} \neq \mu_0$ ($\bar{\sigma}^2 > 0$) at level α?

To determine the value of Rao's Score statistic (2.5-115), we need to determine the log-likelihood score and the inverse of the information matrix, and then evaluate these quantities at the restricted ML estimates. The first partial derivatives of the log-likelihood function with respect to μ and σ^2 have already been determined as (2.5-102) and (2.5-103) in Example 2.19. Thus, the log-likelihood score vector follows to be

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\theta};\boldsymbol{y}) = \begin{bmatrix} \frac{\partial\mathcal{L}(\mu,\sigma^2;\boldsymbol{y})}{\partial\mu} \\ \frac{\partial\mathcal{L}(\mu,\sigma^2;\boldsymbol{y})}{\partial\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2}\sum\limits_{i=1}^n (y_i-\mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum\limits_{i=1}^n (y_i-\mu)^2 \end{bmatrix}.$$

The Hessian of the log-likelihood function comprises the second partial derivatives with respect to all unknown parameters. For the current example, these are

$$\begin{aligned}
\frac{\partial^2\mathcal{L}(\mu,\sigma^2;\boldsymbol{y})}{\partial\mu\partial\mu} &= -\frac{n}{\sigma^2}, \\
\frac{\partial^2\mathcal{L}(\mu,\sigma^2;\boldsymbol{y})}{\partial\mu\partial\sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^n (y_i-\mu), \\
\frac{\partial^2\mathcal{L}(\mu,\sigma^2;\boldsymbol{y})}{\partial\sigma^2\partial\mu} &= -\frac{1}{\sigma^4}\sum_{i=1}^n (y_i-\mu), \\
\frac{\partial^2\mathcal{L}(\mu,\sigma^2;\boldsymbol{y})}{\partial\sigma^2\partial\sigma^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (y_i-\mu)^2.
\end{aligned}$$

Then, the information matrix follows to be

$$\begin{aligned}
\boldsymbol{\mathcal{I}}(\boldsymbol{\theta};\boldsymbol{Y}) &= \boldsymbol{E}\{-\boldsymbol{\mathcal{H}}(\boldsymbol{\theta};\boldsymbol{Y})\} \\
&= \boldsymbol{E}\left\{-\begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4}\sum\limits_{i=1}^n (Y_i-\mu) \\ -\frac{1}{\sigma^4}\sum\limits_{i=1}^n (Y_i-\mu)^2 & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum\limits_{i=1}^n (Y_i-\mu)^2 \end{bmatrix}\right\} \\
&= \begin{bmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^4}\sum\limits_{i=1}^n (E\{Y_i\}-\mu) \\ \frac{1}{\sigma^4}\sum\limits_{i=1}^n (E\{Y_i\}-\mu) & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum\limits_{i=1}^n E\{(Y_i-\mu)^2\} \end{bmatrix}.
\end{aligned}$$

Due to the definitions $E\{Y_i\} = \mu$ and $E\{(Y_i - \mu)^2\} = \sigma^2$ of the first moment and the second central moment, respectively, the off-diagonal components of the information matrix vanish, and we obtain

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}; \boldsymbol{Y}) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}.$$

Then, using the fact that the restricted ML estimate for μ is $\widetilde{\mu} = \mu_0$, Rao's Score statistic in (2.5-115) becomes

$$\begin{aligned} T_{RS}(\boldsymbol{Y}) &= \boldsymbol{\mathcal{S}}'(\widetilde{\mu}, \widetilde{\sigma}^2; \boldsymbol{Y})\boldsymbol{\mathcal{I}}^{-1}(\widetilde{\mu}, \widetilde{\sigma}^2; \boldsymbol{Y})\boldsymbol{\mathcal{S}}(\widetilde{\mu}, \widetilde{\sigma}^2; \boldsymbol{Y}) \\ &= \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2}\sum\limits_{i=1}^{n}(Y_i - \mu_0) \\ -\frac{n}{2\widetilde{\sigma}^2} + \frac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(Y_i - \mu_0)^2 \end{bmatrix}' \begin{bmatrix} \frac{n}{\widetilde{\sigma}^2} & 0 \\ 0 & \frac{n}{2\widetilde{\sigma}^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2}\sum\limits_{i=1}^{n}(Y_i - \mu_0) \\ -\frac{n}{2\widetilde{\sigma}^2} + \frac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(Y_i - \mu_0)^2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2}\sum\limits_{i=1}^{n}(Y_i - \mu_0) \\ 0 \end{bmatrix}' \begin{bmatrix} \frac{\widetilde{\sigma}^2}{n} & 0 \\ 0 & \frac{2\widetilde{\sigma}^2}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2}\sum\limits_{i=1}^{n}(Y_i - \mu_0) \\ 0 \end{bmatrix} \\ &= \frac{1}{n\widetilde{\sigma}^2}\left(\sum_{i=1}^{n}(Y_i - \mu_0)\right)^2. \end{aligned}$$

Two aspects are typical for Rao's Score statistic. Firstly, the log-likelihood score vanishes in the direction of σ^2. This happens necessarily because σ^2 is not restricted by H_0. Therefore, the unrestricted ML estimate of such a free parameter will certainly maximize the log-likelihood function in that direction. Secondly, the information matrix is diagonal, reflecting the fact that both parameters are determined independently. These two properties, which are true also for more complex testing problems such as for the applications in Sections 3 and 4, simplify the determination of Rao's Score statistic considerably.

If we recall from (2.5-108) in Example 2.19 that $\frac{1}{\widetilde{\sigma}^2}\sum_{i=1}^{n}(Y_i - \mu_0)$ is the estimator for the Lagrange multiplier, we may rewrite Rao's Score statistic in the form

$$T_{RS}(\boldsymbol{Y}) = \frac{\widetilde{\sigma}^2}{n}\left(\frac{1}{\widetilde{\sigma}^2}\sum_{i=1}^{n}(Y_i - \mu_0)\right)^2 = \frac{\widetilde{\sigma}^2}{n}\widetilde{k}^2.$$

For this reason, Rao's Score statistic is often called the **Lagrange Multiplier (LM) statistic**, a term which was probably first used by Silvey (1959) and which is used typically in the field of econometrics.

Let us assume that the observations have been centered such that the hypotheses are about $\mu_0 = 0$, as demonstrated in Example 2.16. As for the relation between T_{LR} and the UMPI test statistic derived in Example 2.19, we can show that T_{RS} is not identical with the UMPI test statistic, but a strictly monotonic function thereof. Recall from Example 2.19 that $M(\boldsymbol{Y})/(n-1) = \widehat{\mu}^2/\widehat{\sigma}^2$ and $1 + M(\boldsymbol{Y})/(n-1) = (\widehat{\sigma}^2 + \widehat{\mu}^2)/\widehat{\sigma}^2$. With this, we obtain

$$n\frac{\frac{1}{n-1}M(\boldsymbol{Y})}{1 + \frac{1}{n-1}M(\boldsymbol{Y})} = n\frac{\widehat{\mu}^2}{\widehat{\sigma}^2}\Big/\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} = n\frac{\widehat{\mu}^2}{\widetilde{\sigma}^2} = \frac{1}{n\widetilde{\sigma}^2}\left(\sum_{i=1}^{n}Y_i\right)^2 = T_{RS}.$$

Therefore, Rao's Score test

$$\phi_{RS}(\boldsymbol{y}) := \begin{cases} 1, & \text{if} \quad T_{RS}(\boldsymbol{y}) = \frac{\widetilde{\sigma}^2}{n}\widetilde{k}^2 > C^*, \\ 0, & \text{if} \quad T_{RS}(\boldsymbol{y}) = \frac{\widetilde{\sigma}^2}{n}\widetilde{k}^2 < C^*, \end{cases} \tag{2.5-117}$$

with critical value $C^* = n\frac{C/n-1}{1+C/(n-1)}$ will be exactly the same as the UMPI test (2.5-85) with critical value $C = k_{1-\alpha}^{F(1,n-1)}$. □

3 Theory and Applications of Misspecification Tests in the Normal Gauss-Markov Model

3.1 Introduction

In this section we will consider a number of very common test problems within the context the **normal Gauss-Markov model (GMM)**

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}, \quad (3.1\text{-}118)$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\{\boldsymbol{E}\} = \sigma^2 \boldsymbol{P}^{-1} \quad (3.1\text{-}119)$$

with normally distributed zero-mean errors $\boldsymbol{E}$, known design matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ of full rank, known positive definite weight matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$, and parameters $\boldsymbol{\beta} \in \mathbb{R}^{m\times 1}$ and $\sigma^2 \in \mathbb{R}^+$, respectively. Thus, we may write the resulting class of distributions with respect to the observables $\boldsymbol{Y}$ as

$$\mathcal{W} = \{N\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{P}^{-1}\right) : \boldsymbol{\beta} \in \mathbb{R}^{m\times 1}, \sigma^2 \in \mathbb{R}^+\}, \quad (3.1\text{-}120)$$

which corresponds to the space $\boldsymbol{\Theta} = \mathbb{R}^{m\times 1} \times \mathbb{R}^+$ of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ and to the class of multivariate normal density functions

$$\mathcal{F} = \{f\left(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2\right) : \boldsymbol{\beta} \in \mathbb{R}^{m\times 1}, \sigma^2 \in \mathbb{R}^+\},$$

of multivariate normal density functions, defined by

$$f(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2) = (2\pi)^{-n/2}(\det \sigma^2 \boldsymbol{P}^{-1})^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\} \quad (3.1\text{-}121)$$

(see Equation 2.125 in Koch, 1999, p. 117). We will further assume that the unknown true parameter vector $\bar{\boldsymbol{\theta}}$ is one element of $\boldsymbol{\Theta}$.

Frequently, the numerical value σ_0^2 for σ^2 is known *a priori*. In this case we will rewrite the class of distributions as

$$\mathcal{W} = \{N\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma_0^2\boldsymbol{P}^{-1}\right) : \boldsymbol{\beta} \in \mathbb{R}^{m\times 1}\}, \quad (3.1\text{-}122)$$

the space of parameters $\boldsymbol{\theta} = \boldsymbol{\beta}$ as $\boldsymbol{\Theta} = \mathbb{R}^{m\times 1}$, and the corresponding class of density functions as

$$\mathcal{F} = \{f\left(\boldsymbol{y}; \boldsymbol{\beta}\right) : \boldsymbol{\beta} \in \mathbb{R}^{m\times 1}\} \quad (3.1\text{-}123)$$

with

$$f(\boldsymbol{y}; \boldsymbol{\beta}) = (2\pi)^{-n/2}(\det \sigma_0^2 \boldsymbol{P}^{-1})^{-1/2} \exp\left\{-\frac{1}{2\sigma_0^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\}$$

Notice that, by setting $\boldsymbol{X} := \boldsymbol{1}$ and $\boldsymbol{P} := \boldsymbol{I}$, we obtain the observation model used in Examples 2.16 and 2.17 (depending on whether σ^2 is known or unknown *a priori*).

As the parameter space comprises two types of parameters, we will naturally consider two categories of test problems. The first one is about testing the parameters $\boldsymbol{\beta}$ appearing in the functional model (3.1-118), and the second one is about testing the variance factor σ^2, which is part of the stochastic model. Solutions to these test problems are well known (see, for instance, Koch, 1999; Teunissen, 2000) and belong to the standard procedures of geodetic adjustment theory. Therefore, rather than to repeat common knowledge, the purpose of this section is to reconceptualize these tests, in particular the test statistics, by deriving them as *optimal* procedures. For this purpose, we will exploit symmetry assumptions, that is, invariance principles with respect to the power function in the same way as demonstrated in Section 2 for some simpler test problems. It will turn out that the standard outlier tests, significance tests, tests of linear hypotheses, and the test of the variance factor owe much of their uselfulness to the fact that they may all be derived as *uniformly most powerful invariant tests*.

3.2 Derivation of optimal tests concerning parameters of the functional model

So far we have confined ourselves to problems, where the hypothesis that parameters take particular values was to be tested against some simple or composite alternative. However, limitation to such problems is unnecessarily restrictive. There are situations where we would rather want to test whether a set of linear functions of the parameters takes particular values. As a common example, it is desired in deformation analysis to test whether differences of coordinates are zero, or whether they differ significantly (see Application 3). We have already seen in (2.5-91) of Section 2.5.6 that hypotheses may take the form of linear constraints (restrictions) concerning the parameters to be tested. This model also fits conveniently into the framework of the normal GMM (3.1-118 + 3.1-119). In the current section, we shall restrict attention to hypotheses concerning parameters $\boldsymbol{\beta}$ within the functional model (3.1-118), which may then be written as constraints

$$H_0 : \boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{w} \quad \text{versus} \quad H_1 : \boldsymbol{H}\bar{\boldsymbol{\beta}} \neq \boldsymbol{w}, \tag{3.2-124}$$

where $\boldsymbol{H} \in \mathbb{R}^{m_2 \times m}$ (with $m_2 \leq m$) denotes a matrix of full rank.

This general model setup may be simplified in various ways before addressing the fundamental question of optimality procedures. The first step will be to reparameterize the GMM and the constraints such that the hypotheses become direct propositions about the values of the unknown parameters rather than about the values of functions thereof. Furthermore, to exploit symmetries within the parameter space effectively, it will also be convenient to center these transformed hypotheses about zero. Finally, we shall simplify the stochastic model (3.1-119) by transforming the observations into uncorrelated variables with constant variance.

After carrying out these preprocessing steps, we will reduce the testing problem by sufficiency and invariance in a similar manner to the approach presented in Section 2. Then, after reversing the preprocessing steps, we will obtain, as the main result of this section, the UMPI test for testing the hypotheses in (3.2-124).

The individual steps of this preprocessing and reduction procedure will now be carried out within the following subsections:

1. Reparameterization of the test problem.
2. Centering of the hypotheses.
3. Full decorrelation/homogenization of the observations.
4. Reduction to independent sufficient statistics with elimination of additional functional parameters.
5. Reduction to a maximal invariant statistic.
6. Back-substitution (reversal of steps 1.-4).

3.2.1 Reparameterization of the test problem

Following Meissl (1982, Section C.2.2), we expand the $m_2 \times m$-matrix $\boldsymbol{H}$ by some $(m - m_2) \times m$-matrix $\boldsymbol{M}$ into an invertible $m \times m$ block matrix and introduce new parameters $\boldsymbol{\beta}_1^{(r)} \in \mathbb{R}^{m_1 \times 1}$ (where $m_1 := m - m_2$) and $\boldsymbol{\beta}_2^{(r)} \in \mathbb{R}^{m_2 \times 1}$ with

$$\begin{bmatrix} \boldsymbol{\beta}_1^{(r)} \\ \boldsymbol{\beta}_2^{(r)} \end{bmatrix} := \begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{H} \end{bmatrix} \boldsymbol{\beta}. \tag{3.2-125}$$

Using the invertibility assumption we obtain the equivalent relation

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{H} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\beta}_1^{(r)} \\ \boldsymbol{\beta}_2^{(r)} \end{bmatrix}.$$

Then, multiplying this equation with $\boldsymbol{X}$ from the left yields

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X} \begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{H} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\beta}_1^{(r)} \\ \boldsymbol{\beta}_2^{(r)} \end{bmatrix} =: \begin{bmatrix} \boldsymbol{X}_1^{(r)} \ \boldsymbol{X}_2^{(r)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1^{(r)} \\ \boldsymbol{\beta}_2^{(r)} \end{bmatrix}. \tag{3.2-126}$$

Notice that this definition allows us to derive the following expression for the original design matrix $\boldsymbol{X}$ from the implication

$$\boldsymbol{X}\begin{bmatrix}\boldsymbol{M}\\ \boldsymbol{H}\end{bmatrix}^{-1} =: \left[\boldsymbol{X}_1^{(r)}\ \boldsymbol{X}_2^{(r)}\right] \Rightarrow \boldsymbol{X} = \left[\boldsymbol{X}_1^{(r)}\ \boldsymbol{X}_2^{(r)}\right]\begin{bmatrix}\boldsymbol{M}\\ \boldsymbol{H}\end{bmatrix}. \tag{3.2-127}$$

Using (3.2-126) we may substitute the original functional model (3.1-118) by

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(r)}\boldsymbol{\beta}_2^{(r)} \tag{3.2-128}$$

and, in light of (3.2-125), the linear restriction (3.2-124) by

$$\boldsymbol{H}\bar{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}}_2^{(r)}. \tag{3.2-129}$$

This reparameterization leads to an equivalent testing problem which involves the transformed version $\boldsymbol{X}^{(r)} = [\,\boldsymbol{X}_1^{(r)}\ \boldsymbol{X}_2^{(r)}\,]$ of the original design matrix $\boldsymbol{X} = [\,\boldsymbol{X}_1\ \boldsymbol{X}_2\,]$ in partitioned form. We will describe this simplified class of test problems by the new observation model

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(r)}\boldsymbol{\beta}_2^{(r)}, \sigma^2\boldsymbol{P}^{-1}\right) \tag{3.2-130}$$

(where the true value of σ^2 may be known or unknown *a priori*), and by the new hypotheses

$$H_0 : \bar{\boldsymbol{\beta}}_2^{(r)} = \boldsymbol{w} \quad versus \quad H_1 : \bar{\boldsymbol{\beta}}_2^{(r)} \neq \boldsymbol{w}. \tag{3.2-131}$$

3.2.2 Centering of the hypotheses

Similarly to the data transformation in Example 2.16 we may subtract the generalized constant mean $\boldsymbol{X}_2^{(r)}\boldsymbol{w}$ from the data, that is,

$$\boldsymbol{Y}^{(c)} := \boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\boldsymbol{w}. \tag{3.2-132}$$

While this transformation leaves the covariance matrix as the second central moment unchanged, it changes the expectation to

$$\boldsymbol{E}\left\{\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\boldsymbol{w}\right\} = \boldsymbol{E}\left\{\boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(r)}\boldsymbol{\beta}_2^{(r)} - \boldsymbol{X}_2^{(r)}\boldsymbol{w}\right\} = \boldsymbol{E}\left\{\boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(r)}\left(\boldsymbol{\beta}_2^{(r)} - \boldsymbol{w}\right)\right\}. \tag{3.2-133}$$

Setting $\boldsymbol{\beta}_2^{(rc)} := \boldsymbol{\beta}_2^{(r)} - \boldsymbol{w}$ leads to the new observation model

$$\boldsymbol{Y}^{(c)} \sim N\left(\boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(r)}\boldsymbol{\beta}_2^{(rc)}, \sigma^2\boldsymbol{P}^{-1}\right), \tag{3.2-134}$$

where the true value of σ^2 may be known or unknown *a priori*. The hypotheses in terms of ersatz parameters $\boldsymbol{\beta}_2^{(rc)}$ are then evidently given by

$$H_0 : \bar{\boldsymbol{\beta}}_2^{(rc)} = \boldsymbol{0} \quad versus \quad H_1 : \bar{\boldsymbol{\beta}}_2^{(rc)} \neq \boldsymbol{0}. \tag{3.2-135}$$

3.2.3 Full decorrelation/homogenization of the observations

A Cholesky decomposition of the weight matrix into

$$\boldsymbol{P} = \boldsymbol{G}\boldsymbol{G}', \tag{3.2-136}$$

where $\boldsymbol{G}$ stands for an invertible lower triangular matrix, allows for a full decorrelation or homogenization of the observations by virtue of the one-to-one transformation

$$\boldsymbol{Y}^{(ch)} := \boldsymbol{G}'\boldsymbol{Y}^{(c)} \tag{3.2-137}$$

(cf. Koch, 1999, p. 154). The expectation of the transformed observables becomes

$$\boldsymbol{E}\left\{\boldsymbol{G}'\boldsymbol{Y}^{(c)}\right\} = \boldsymbol{G}'\left(\boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(r)}\boldsymbol{\beta}_2^{(rc)}\right) = \boldsymbol{G}'\boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{G}'\boldsymbol{X}_2^{(r)}\boldsymbol{\beta}_2^{(rc)}.$$

The fact that $\boldsymbol{Y}^{(ch)}$ has covariance matrix $\sigma^2\boldsymbol{I}$ follows directly from an application of error propagation to the linear function $\boldsymbol{G}'\boldsymbol{Y}^{(c)}$, which yields with (3.2-136)

$$\begin{aligned}\boldsymbol{\Sigma}\left\{\boldsymbol{G}'\boldsymbol{Y}^{(c)}\right\} &= \boldsymbol{G}'\boldsymbol{\Sigma}\left\{\boldsymbol{Y}^{(c)}\right\}\boldsymbol{G} = \boldsymbol{G}'\sigma^2\boldsymbol{P}^{-1}\boldsymbol{G} = \sigma^2\boldsymbol{G}'\left(\boldsymbol{G}\boldsymbol{G}'\right)^{-1}\boldsymbol{G}\\ &= \sigma^2\boldsymbol{G}'(\boldsymbol{G}')^{-1}\boldsymbol{G}^{-1}\boldsymbol{G} = \sigma^2\boldsymbol{I}.\end{aligned}$$

After introducing the transformed block design matrices $\boldsymbol{X}_1^{(rh)} := \boldsymbol{G}'\boldsymbol{X}_1^{(r)}$ and $\boldsymbol{X}_2^{(rh)} := \boldsymbol{G}'\boldsymbol{X}_2^{(r)}$, the transformed observation model then reads

$$\boldsymbol{Y}^{(ch)} \sim N\left(\boldsymbol{X}_1^{(rh)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(rh)}\boldsymbol{\beta}_2^{(rc)}, \sigma^2\boldsymbol{I}\right), \tag{3.2-138}$$

where the true value of σ^2 may be known or unknown *a priori*. As homogenization does not transform parameters, the hypotheses may still be written as in (3.2-135), that is,

$$H_0 : \bar{\boldsymbol{\beta}}_2^{(rc)} = \boldsymbol{0} \quad \text{versus} \quad H_1 : \bar{\boldsymbol{\beta}}_2^{(rc)} \neq \boldsymbol{0}. \tag{3.2-139}$$

Whenever a test problem with structure (3.1-120/3.1-122, 3.2-124) or (3.2-130, 3.2-131) or (3.2-134, 3.2-135) is given, it may be transformed directly into (3.2-138, 3.2-139), which will turn out to be the most suitable structure for subsequent reductions by sufficiency and invariance.

3.2.4 Reduction to minimal sufficient statistics with elimination of nuisance parameters

To reduce the observations $\boldsymbol{Y}^{(ch)}$ by sufficiency, we need to generalize the result of Examples 2.10 and 2.11 to the case of the linear model.

Proposition 3.1. *In the normal Gauss-Markov model $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, the least squares estimators*

$$\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{Y} \tag{3.2-140}$$

and

$$(n-m)\widehat{\sigma}^2 = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{3.2-141}$$

constitute independently distributed and minimally sufficient statistics for $\boldsymbol{\beta}$ and σ^2, respectively.

Proof. Using the estimates defined by (3.2-140) and (3.2-141), the multivariate normal density (3.1-121) may be rewritten as

$$\begin{aligned}f(\boldsymbol{y};\boldsymbol{\beta},\sigma^2) &= (2\pi)^{-n/2}(\det\sigma^2\boldsymbol{I})^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\right\}\\ &= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left((n-m)\widehat{\sigma}^2 + (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right)\right\}I_{\mathbb{R}^n}(\boldsymbol{y}).\end{aligned}$$

Thus, it follows from Neyman's Factorization Theorem 2.2 that $\widehat{\boldsymbol{\beta}}$ and $(n-m)\widehat{\sigma}^2$ are jointly sufficient statistics for $\boldsymbol{\beta}$ and σ^2. Then, Arnold (1981, p. 65) shows that these statistics are *complete*, which implies minimality (see Arnold, 1990, p. 346), and independently distributed with

$$\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}) \tag{3.2-142}$$

and

$$(n-m)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(n-m). \tag{3.2-143}$$

□

Next, we will rewrite this fundamental result in terms of partitioned parameters as demanded by the observation model (3.2-138).

Proposition 3.2. *In the normal linear Gauss-Markov model* $\boldsymbol{Y} \sim N(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2, \sigma^2\boldsymbol{I})$ *with partitioned parameters* $\boldsymbol{\beta}_1 \in \mathbb{R}^{m_1\times 1}$ *and* $\boldsymbol{\beta}_2 \in \mathbb{R}^{m_2\times 1}$*, the least squares estimators*

$$\begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{Y} \\ \boldsymbol{X}_2'\boldsymbol{Y} \end{bmatrix}, \text{ short} : \begin{bmatrix} \boldsymbol{N}_{11} & \boldsymbol{N}_{12} \\ \boldsymbol{N}_{21} & \boldsymbol{N}_{22} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{n}_1 \\ \boldsymbol{n}_2 \end{bmatrix} \quad (3.2\text{-}144)$$

and

$$(n - m_1 - m_2)\widehat{\sigma}^2 = (\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2) \quad (3.2\text{-}145)$$

constitute minimally sufficient statistics for $\boldsymbol{\beta}_1$*,* $\boldsymbol{\beta}_2$*, and* σ^2*, respectively. Furthermore, the statistics* $[\widehat{\boldsymbol{\beta}}_1' \ \widehat{\boldsymbol{\beta}}_2']$ *and* $\widehat{\sigma}^2$ *are independently distributed with*

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \sigma^2 \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix}^{-1} \right) \quad (3.2\text{-}146)$$

and

$$(n - m_1 - m_2)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(n - m_1 - m_2). \quad (3.2\text{-}147)$$

Before considering a reduction of $\boldsymbol{Y}^{(ch)}$ to sufficient statistics we may notice that the observation model (3.2-138) comprises functional parameters $\boldsymbol{\beta}_1^{(r)}$ not subject to hypotheses. Therefore, we may eliminate them from the normal equations (3.2-144) (cf. Schuh, 2006a, Section 1.2.1) without changing the test problem itself. First we rewrite the partitioned normal equations (3.2-144) in terms of reparameterized, centered, and homogenized quantities, that is,

$$\boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_1^{(rh)}\widehat{\boldsymbol{\beta}}_1^{(r)} + \boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)} = \boldsymbol{X}_1'^{(rh)}\boldsymbol{Y}^{(ch)} \quad (3.2\text{-}148)$$

$$\boldsymbol{X}_2'^{(rh)}\boldsymbol{X}_1^{(rh)}\widehat{\boldsymbol{\beta}}_1^{(r)} + \boldsymbol{X}_2'^{(rh)}\boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)} = \boldsymbol{X}_2'^{(rh)}\boldsymbol{Y}^{(ch)} \quad (3.2\text{-}149)$$

Isolation of $\widehat{\boldsymbol{\beta}}_1^{(r)}$ in (3.2-148) yields

$$\widehat{\boldsymbol{\beta}}_1^{(r)} = \left(\boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_1^{(rh)}\right)^{-1}\left(\boldsymbol{X}_1'^{(rh)}\boldsymbol{Y}^{(ch)} - \boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)}\right), \quad (3.2\text{-}150)$$

and after substitution into (3.2-149)

$$\widehat{\boldsymbol{\beta}}_2^{(rc)} = \boldsymbol{N}_{22}^{(-1)}\left(\boldsymbol{X}_2'^{(rh)} - \boldsymbol{X}_2'^{(rh)}\boldsymbol{X}_1^{(rh)}\left(\boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_1^{(rh)}\right)^{-1}\boldsymbol{X}_1'^{(rh)}\right)\boldsymbol{Y}^{(ch)} \quad (3.2\text{-}151)$$

with Schur complement

$$\boldsymbol{N}_{22}^{(-1)} := \left(\boldsymbol{X}_2'^{(rh)}\boldsymbol{X}_2^{(rh)} - \boldsymbol{X}_2'^{(rh)}\boldsymbol{X}_1^{(rh)}\left(\boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_1^{(rh)}\right)^{-1}\boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_2^{(rh)}\right)^{-1} \quad (3.2\text{-}152)$$

as abbreviation. We will not give $\boldsymbol{N}_{22}^{(-1)}$ the index $.^{(rh)}$ because (1) this matrix naturally refers to the model with two groups of parameters $\boldsymbol{\beta}_1^{(r)}$ and $\boldsymbol{\beta}_1^{(r)}$, and (2) it may be written directly in terms of non-homogeneous quantities, that is,

$$\boldsymbol{N}_{22}^{(-1)} = \left(\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)} - \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\right)^{-1}.$$

The residuals in the observation model (3.2-138) are defined as

$$\widehat{\boldsymbol{E}}^{(rch)} = \boldsymbol{Y}^{(ch)} - \boldsymbol{X}_1^{(rh)}\widehat{\boldsymbol{\beta}}_1^{(r)} - \boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)}, \quad (3.2\text{-}153)$$

and may, after substitution of (3.2-150), be written as

$$\widehat{\boldsymbol{E}}^{(rch)} = \left(\boldsymbol{I} - \boldsymbol{X}_1^{(rh)}\left(\boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_1^{(rh)}\right)^{-1}\boldsymbol{X}_1'^{(rh)}\right)\left(\boldsymbol{Y}^{(ch)} - \boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)}\right). \quad (3.2\text{-}154)$$

If the true value of the variance of unit weight must be estimated, then these two alternative formulations for the residuals correspond to the following expressions for the estimator of σ^2:

$$\begin{aligned}(n-m_1-m_2)\widehat{\sigma}^2_{(rch)} &= \widehat{\boldsymbol{E}}'^{(rch)}\widehat{\boldsymbol{E}}^{(rch)}\\ &= \left(\boldsymbol{Y}^{(ch)}-\boldsymbol{X}_1^{(rh)}\widehat{\boldsymbol{\beta}}_1^{(r)}-\boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)}\right)'\left(\boldsymbol{Y}^{(ch)}-\boldsymbol{X}_1^{(rh)}\widehat{\boldsymbol{\beta}}_1^{(r)}-\boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)}\right)\\ &= \left(\boldsymbol{Y}^{(ch)}-\boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)}\right)'\left(\boldsymbol{I}-\boldsymbol{X}_1^{(rh)}\left(\boldsymbol{X}_1'^{(rh)}\boldsymbol{X}_1^{(rh)}\right)^{-1}\boldsymbol{X}_1'^{(rh)}\right)\left(\boldsymbol{Y}^{(ch)}-\boldsymbol{X}_2^{(rh)}\widehat{\boldsymbol{\beta}}_2^{(rc)}\right).\end{aligned}$$

Instead of using the vector $\widehat{\boldsymbol{\beta}}_2^{(rc)}$ with possibly fully populated weight matrix, it will be much more convenient to operate with the fully decorrelated and homogenized vector

$$\widehat{\boldsymbol{\beta}}_2^{(rch)} := \boldsymbol{G}'_{22}\widehat{\boldsymbol{\beta}}_2^{(rc)} \tag{3.2-155}$$

which is also sufficient for $\boldsymbol{\beta}_2^{(rc)}$ as a one-to-one function of $\widehat{\boldsymbol{\beta}}_2^{(rc)}$. Here, $\boldsymbol{G}_{22}$ represents the (invertible) lower triangular matrix obtained from the Cholesky factorization $\boldsymbol{P}_{22} := \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} = \boldsymbol{G}_{22}\boldsymbol{G}'_{22}$. After reducing the observation model (3.2-138) by sufficiency, we now have a test problem about ersatz observations $[\,\widehat{\boldsymbol{\beta}}_2'^{(rch)},\ \widehat{\sigma}^2\,]$ with reduced dimension (m_2+1). The hypotheses

$$H_0 : \bar{\boldsymbol{\beta}}_2^{(rch)} = \boldsymbol{0} \quad \text{versus} \quad H_1 : \bar{\boldsymbol{\beta}}_2^{(rch)} \neq \boldsymbol{0} \tag{3.2-156}$$

follow from (3.2-156) by observing that $\bar{\boldsymbol{\beta}}_2^{(rch)} = \boldsymbol{0}$ if and only if $\bar{\boldsymbol{\beta}}_2^{(rc)} = \boldsymbol{0}$ and $\bar{\boldsymbol{\beta}}_2^{(rch)} \neq \boldsymbol{0}$ if and only if $\bar{\boldsymbol{\beta}}_2^{(rc)} \neq \boldsymbol{0}$.

3.2.5 Reduction to a maximal invariant statistic

In this step we seek to reduce the test problem in terms of independent sufficient statistics by invariance in the same way as we did in Example 2.17. The only difference will be that in the present case we cannot apply sign invariance as the expectation is now given by a non-constant mean vector. Instead we will verify that the test problem is invariant under the group of orthogonal transformations acting on the mean vector $\boldsymbol{\beta}_2^{(rch)}$.

Case 1: $\bar{\sigma}^2 = \sigma_0^2$ known. Let us begin with the simpler case that the true value of the variance factor is known *a priori*. First we note that each orthogonal transformation

$$\boldsymbol{g}_1(\widehat{\boldsymbol{\beta}}_2^{(rch)}) = \boldsymbol{\Gamma}\widehat{\boldsymbol{\beta}}_2^{(rch)}$$

in $\mathcal{G}_1$ results in a change of distribution from $\widehat{\boldsymbol{\beta}}_2^{(rch)} \sim N(\boldsymbol{\beta}_2^{(rch)}, \sigma_0^2\boldsymbol{I})$ to $\boldsymbol{\Gamma}\widehat{\boldsymbol{\beta}}_2^{(rch)} \sim N(\boldsymbol{\Gamma}\boldsymbol{\beta}_2^{(rch)}, \sigma_0^2\boldsymbol{I})$, where the covariance matrix of $\boldsymbol{\Gamma}\widehat{\boldsymbol{\beta}}_2^{(rch)}$ remains unchanged due to the property of any orthogonal matrix $\boldsymbol{\Gamma}$ that $\boldsymbol{\Gamma}\boldsymbol{\Gamma}' = \boldsymbol{I}$. From this the induced transformation within the parameter domain is seen to be

$$\bar{\boldsymbol{g}}_1(\boldsymbol{\beta}_2^{(rch)}) = \boldsymbol{\Gamma}\boldsymbol{\beta}_2^{(rch)}.$$

Then, Theorem 2.6-5 gives $\widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}$ as the maximal invariant with respect to the transformation $\boldsymbol{\Gamma}\widehat{\boldsymbol{\beta}}_2^{(rch)}$. We must still prove that the original test problem remains itself invariant under $\mathcal{G}_1$ with induced group of transformations $\bar{\mathcal{G}}_1$. This is truly the case because

$$\bar{\boldsymbol{g}}_1(\boldsymbol{\Theta}_0) = \bar{\boldsymbol{g}}_1(\{\boldsymbol{0}\}) = \{\boldsymbol{\Gamma}\boldsymbol{0}\} = \{\boldsymbol{0}\} = \boldsymbol{\Theta}_0$$

and

$$\begin{aligned}\bar{\boldsymbol{g}}_1(\boldsymbol{\Theta}_1) &= \bar{\boldsymbol{g}}_1(\{\boldsymbol{\beta}_2^{(rch)} : \boldsymbol{\beta}_2^{(rch)} \in \mathbb{R}^{m_2} - \{\boldsymbol{0}\}) = \{\boldsymbol{\Gamma}\boldsymbol{\beta}_2^{(rch)} : \boldsymbol{\beta}_2^{(rch)} \in \mathbb{R}^{m_2} - \{\boldsymbol{0}\}\}\\ &= \{\boldsymbol{\beta}_2^{(rch\Gamma)} : \boldsymbol{\beta}_2^{(rch\Gamma)} \in \mathbb{R}^{m_2} - \{\boldsymbol{0}\}\} = \boldsymbol{\Theta}_1\end{aligned}$$

leaves the hypotheses unchanged. To formulate the invariant test problem we need to find the distribution of the maximal invariant. From $\widehat{\boldsymbol{\beta}}_2^{(rch)} \sim N(\boldsymbol{\beta}_2^{(rch)}, \sigma_0^2\boldsymbol{I})$ it follows that $\widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2'^{(rch)}/\sigma_0^2 \sim \chi^2(m_2, \lambda)$ with non-centrality parameter $\lambda = \boldsymbol{\beta}_2'^{(rch)}\boldsymbol{\beta}_2^{(rch)}/\sigma_0^2$ (see Koch, 1999, p. 127). Notice now that $\lambda = 0$ if and only if $\boldsymbol{\beta}_2^{(rch)} = \boldsymbol{0}$, and that $\lambda > 0$ if and only if $\boldsymbol{\beta}_2^{(rch)} \neq \boldsymbol{0}$. Therefore, we may write the two-sided hypothesis testing problem in terms of the maximal parameter invariant λ, which will take a positive value if H_1 is true. Consequently, the invariant test problem

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2^{\prime(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}/\sigma_0^2 \sim \chi^2(m_2,\lambda)$$
$$H_0 : \bar{\lambda} = 0 \text{ against } H_1 : \bar{\lambda} > 0 \tag{3.2-157}$$

with single parameter $\lambda = \boldsymbol{\beta}_2^{\prime(rch)}\boldsymbol{\beta}_2^{(rch)}/\sigma_0^2$ has a one-sided alternative hypothesis. Furthermore, the non-central χ^2-distribution with fixed degree of freedom has a monotone density ratio according to Theorem 2.5-5. Therefore, Theorem 2.4 is applicable, which gives the UMP test

$$\phi(\boldsymbol{y}) = \begin{cases} 1, & \text{if} \quad M(\boldsymbol{y}) > k_{1-\alpha}^{\chi^2(m_2)}, \\ 0, & \text{if} \quad M(\boldsymbol{y}) < k_{1-\alpha}^{\chi^2(m_2)}, \end{cases} \tag{3.2-158}$$

(at level α). It follows that ϕ is the UMPI test for testing $H_0 : \boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{w}$ versus $H_1 : \boldsymbol{H}\bar{\boldsymbol{\beta}} \neq \boldsymbol{w}$ in the original observation model $\boldsymbol{Y} \sim N\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma_0^2\boldsymbol{P}^{-1}\right)$, in which the variance factor has been assumed to be known *a priori*.

Case 2: $\bar{\sigma}^2$ unknown. In this case, the statistic $\widehat{\sigma}^2_{(rch)}$ acts as an additional ersatz observation. The group of orthogonal transformations, acting on the generally multi-dimensional statistic $\widehat{\boldsymbol{\beta}}_2^{(rch)}$, is defined by

$$\boldsymbol{g}_1(\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2_{(rch)}) = (\boldsymbol{\Gamma}\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2),$$

and causes the distribution to change from $\widehat{\boldsymbol{\beta}}_2^{(rch)} \sim N(\boldsymbol{\beta}_2^{(rch)}, \sigma^2\boldsymbol{I})$ to $\boldsymbol{\Gamma}\widehat{\boldsymbol{\beta}}_2^{(rch)} \sim N(\boldsymbol{\Gamma}\boldsymbol{\beta}_2^{(rch)}, \sigma^2\boldsymbol{I})$. As the statistic $\widehat{\sigma}^2_{(rch)}$ is not changed by any transformation $\boldsymbol{g}_1 \in \mathcal{G}_1$, its distribution also remains unchanged. The induced group of transformations follows to be

$$\bar{\boldsymbol{g}}_1(\boldsymbol{\beta}_2^{(rch)}, \sigma^2) = (\boldsymbol{\Gamma}\boldsymbol{\beta}_2^{(rch)}, \sigma^2).$$

Theorem 2.6-6 gives $\boldsymbol{M}_1(\boldsymbol{Y}) = [\widehat{\boldsymbol{\beta}}_2^{\prime(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2_{(rch)}]'$ as the maximal invariants with respect to the transformation $\boldsymbol{\Gamma}\widehat{\boldsymbol{\beta}}_2^{(rch)}$. The test problem is invariant because of

$$\bar{\boldsymbol{g}}_1(\boldsymbol{\Theta}_0) = \bar{g}_1(\{(\boldsymbol{\beta}_2^{(rch)}, \sigma^2) : \boldsymbol{\beta}_2^{(rch)} = \boldsymbol{0}, \sigma^2 \in \mathbb{R}^+\}) = \{(\boldsymbol{\Gamma}\boldsymbol{0}, \sigma^2) : \sigma^2 \in \mathbb{R}^+\} = \{(\boldsymbol{0}, \sigma^2) : \sigma^2 \in \mathbb{R}^+\} = \boldsymbol{\Theta}_0$$

and

$$\begin{aligned} \bar{\boldsymbol{g}}_1(\boldsymbol{\Theta}_1) &= \bar{g}_1(\{(\boldsymbol{\beta}_2^{(rch)}, \sigma^2) : \boldsymbol{\beta}_2^{(rch)} \in \mathbb{R}^{m_2} - \{\boldsymbol{0}\}, \sigma^2 \in \mathbb{R}^+\}) \\ &= \{(\boldsymbol{\Gamma}\boldsymbol{\beta}_2^{(rch)}, \sigma^2) : \boldsymbol{\beta}_2^{(rch)} \in \mathbb{R}^{m_2} - \{\boldsymbol{0}\}, \sigma^2 \in \mathbb{R}^+\} \\ &= \{(\boldsymbol{\beta}_2^{(rch\Gamma)}, \sigma^2) : \boldsymbol{\beta}_2^{(rch\Gamma)} \in \mathbb{R}^{m_2} - \{\boldsymbol{0}\}, \sigma^2 \in \mathbb{R}^+\} = \boldsymbol{\Theta}_1. \end{aligned}$$

To further reduce $\boldsymbol{M}_1(\boldsymbol{Y})$, observe that the test problem is also invariant under the group $\mathcal{G}_2$ of scale changes

$$\boldsymbol{g}_2(\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2_{(rch)}) = (c\widehat{\boldsymbol{\beta}}_2^{(rch)}, c^2\widehat{\sigma}^2_{(rch)}),$$

which induces the group $\bar{\mathcal{G}}_2$ of parameter transformations

$$\bar{\boldsymbol{g}}_2(\boldsymbol{\beta}_2^{(rch)}, \sigma^2) = (c\boldsymbol{\beta}_2^{(rch)}, c^2\sigma^2),$$

because of the transitions in distribution

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_2^{(rch)} \sim N(\boldsymbol{\beta}_2^{(rch)}, \sigma^2\boldsymbol{I}) \quad &\rightarrow \quad c \cdot \widehat{\boldsymbol{\beta}}_2^{(rch)} \sim N(c\boldsymbol{\beta}_2^{(rch)}, c^2\sigma^2\boldsymbol{I}), \\ \widehat{\sigma}^2_{(rch)} \sim G((n-m)/2, 2\sigma^2)/(n-m) \quad &\rightarrow \quad c^2\widehat{\sigma}^2 \sim G((n-m)/2, 2c^2\sigma^2)/(n-m). \end{aligned}$$

Next, we observe that

$$\boldsymbol{M}_1(\boldsymbol{g}_2(\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2_{(rch)})) = \boldsymbol{M}_1(c\widehat{\boldsymbol{\beta}}_2^{(rch)}, c^2\widehat{\sigma}^2_{(rch)}) = (c^2\widehat{\boldsymbol{\beta}}_2^{\prime(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}, c^2\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{g}}_2(\boldsymbol{M}_1(\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2_{(rch)}))$$

holds if

$$\widehat{\boldsymbol{g}}_2(\widehat{\boldsymbol{\beta}}_2^{\prime(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2_{(rch)}) = (c^2\widehat{\boldsymbol{\beta}}_2^{\prime(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}, c^2\widehat{\sigma}^2_{(rch)})$$

defines the group $\widehat{\mathcal{G}}_2$ of scale changes. It follows from Theorem 2.6-2 that

$$M_2(\widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2_{(rch)}) = \frac{\widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}}{\widehat{\sigma}^2_{(rch)}}$$

is a maximal invariant under $\widehat{\mathcal{G}}_2$. Then, Theorem 2.8 implies that

$$M_2(\boldsymbol{M}_1(\widehat{\boldsymbol{\beta}}_2^{(rch)}, \widehat{\sigma}^2_{(rch)})) = \frac{\widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}}{\widehat{\sigma}^2_{(rch)}}$$

is the statistic maximally invariant under orthogonal transformations and scale changes. Since $\widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}/\sigma$ has a non-central χ^2-distribution with m_2 degrees of freedom and $(n-m)\widehat{\sigma}^2_{(rch)}/\sigma^2$ a central χ^2-distribution with $n-m$ degrees of freedom (both statistics being independently distributed), the maximal invariant $M(\boldsymbol{Y}) := (n-m)\widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}/m_2\widehat{\sigma}^2_{(rch)}$ is distributed as $F(m_2, n-m, \lambda)$ (see Koch, 1999, p. 130). The invariant test problem is finally given by

$$M(\boldsymbol{Y}) = \frac{\widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}}{m_2\widehat{\sigma}^2_{(rch)}} \sim F(m_2, n-m, \lambda) \quad \text{with } \lambda = \boldsymbol{\beta}_2'^{(rch)}\boldsymbol{\beta}_2^{(rch)}/\sigma^2$$
$$H_0 : \bar{\lambda} = 0 \text{ against } H_1 : \bar{\lambda} > 0,$$

which is about one single unknown parameter (λ), a two-sided alternative hypothesis, and a test statistic whose distribution has a monotone density ratio (see Theorem 2.5-6). Therefore, there exists a UMP test for the invariance-reduced test problem (see Theorem 2.4) at level α, which is given by

$$\phi(\boldsymbol{y}) = \begin{cases} 1, & \text{if} \quad M(\boldsymbol{y}) > k_{1-\alpha}^{F(m_2,n-m)}, \\ 0, & \text{if} \quad M(\boldsymbol{y}) < k_{1-\alpha}^{F(m_2,n-m)}. \end{cases} \qquad (3.2\text{-}159)$$

It follows that ϕ is the UMPI test for testing $H_0 : \boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{w}$ versus $H_1 : \boldsymbol{H}\bar{\boldsymbol{\beta}} \neq \boldsymbol{w}$ in the original observation model $\boldsymbol{Y} \sim N\left(\boldsymbol{X\beta}, \sigma^2\boldsymbol{P}^{-1}\right)$, in which the variance factor has been assumed to be unknown *a priori*.

3.2.6 Back-substitution

The test statistic $M(\boldsymbol{Y})$ is inconvenient to compute as it comprises quantities transformed in multiple ways. Therefore, we will express $M(\boldsymbol{Y})$ in terms of the original quantities of models (3.1-118) and (3.1-119). This will be achieved in three steps reversing the transformations in 3.2.1-3.2.4, where each step covers a particular equivalent form of the test problem often encountered in practice.

Case 1: $\bar{\sigma}^2 = \sigma_0^2$ known.

Proposition 3.3. *The invariant test statistic*

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2^{\prime(rch)} \widehat{\boldsymbol{\beta}}_2^{(rch)} / \sigma_0^2 \tag{3.2-160}$$

for the UMP test (3.2-158) (at level α) regarding the hypotheses

$$H_0 : \bar{\lambda} = 0 \quad \text{versus} \quad H_1 : \bar{\lambda} > 0, \tag{3.2-161}$$

i.e. for the UMPI test (3.2-158) (at level α) regarding the original hypotheses $H_0 : \boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{w}$ versus $H_1 : \boldsymbol{H}\bar{\boldsymbol{\beta}} \neq \boldsymbol{w}$, is identical to:

1. *the test statistic*

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2^{\prime(rc)} \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \widehat{\boldsymbol{\beta}}_2^{(rc)} / \sigma_0^2 \tag{3.2-162}$$

for the equivalent test problem

$$\boldsymbol{Y}^{(ch)} \sim N\left(\boldsymbol{X}_1^{(rh)} \boldsymbol{\beta}_1^{(rc)} + \boldsymbol{X}_2^{(rh)} \boldsymbol{\beta}_2^{(rc)}, \sigma_0^2 \boldsymbol{I}\right) \tag{3.2-163}$$

$$H_0 : \boldsymbol{\beta}_2^{(rc)} = \boldsymbol{0} \quad \textit{versus} \quad H_1 : \boldsymbol{\beta}_2^{(rc)} \neq \boldsymbol{0}, \tag{3.2-164}$$

which we will call ***the problem of testing the significance of additional parameters $\boldsymbol{\beta}_2^{(rc)}$ with known variance factor*** *σ_0^2, if the least squares estimator*

$$\widehat{\boldsymbol{\beta}}_2^{(rc)} = \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2^{\prime(rh)} \boldsymbol{Y}^{(ch)} - \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2^{\prime(rh)} \boldsymbol{X}_1^{(rh)} \left(\boldsymbol{X}_1^{\prime(rh)} \boldsymbol{X}_1^{(rh)}\right)^{-1} \boldsymbol{X}_1^{\prime(rh)} \boldsymbol{Y}^{(ch)} \tag{3.2-165}$$

with residuals

$$\widehat{\boldsymbol{E}}^{(rch)} = \left(\boldsymbol{I} - \boldsymbol{X}_1^{(rh)} \left(\boldsymbol{X}_1^{\prime(rh)} \boldsymbol{X}_1^{(rh)}\right)^{-1} \boldsymbol{X}_1^{\prime(rh)}\right) \left(\boldsymbol{Y}^{(ch)} - \boldsymbol{X}_2^{(rh)} \widehat{\boldsymbol{\beta}}_2^{(r)}\right) \tag{3.2-166}$$

is used. Whenever a test problem is naturally given in the form (3.2-163) and (3.2-164), i.e. by

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2, \sigma_0^2 \boldsymbol{I}\right) \tag{3.2-167}$$

$$H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0} \quad \textit{versus} \quad H_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{0} \tag{3.2-168}$$

which we will call ***the natural problem of testing the significance of additional parameters $\boldsymbol{\beta}_2$ with known variance factor*** *σ_0^2, then all the indices are omitted, in which case the test statistic of the UMPI test reads*

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \widehat{\boldsymbol{\beta}}_2 / \sigma_0^2 \tag{3.2-169}$$

with least squares estimator

$$\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2' \boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2' \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1' \boldsymbol{Y} \tag{3.2-170}$$

and residuals

$$\widehat{\boldsymbol{E}} = \left(\boldsymbol{I} - \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1'\right) \left(\boldsymbol{Y} - \boldsymbol{X}_2 \widehat{\boldsymbol{\beta}}_2\right). \tag{3.2-171}$$

2. *identical to the test statistic*

$$M(\boldsymbol{Y}) = \left(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}\right)' \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \left(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}\right) / \sigma_0^2 \tag{3.2-172}$$

for the equivalent test problem

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(r)}\boldsymbol{\beta}_2^{(r)}, \sigma_0^2 \boldsymbol{P}^{-1}\right) \tag{3.2-173}$$

$$H_0 : \boldsymbol{\beta}_2^{(r)} = \boldsymbol{w} \quad versus \quad H_1 : \boldsymbol{\beta}_2^{(r)} \neq \boldsymbol{w} \tag{3.2-174}$$

if the least squares estimator

$$\widehat{\boldsymbol{\beta}}_2^{(r)} = \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2'^{(r)} \boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2'^{(r)} \boldsymbol{P}\boldsymbol{X}_1^{(r)} \left(\boldsymbol{X}_1'^{(r)} \boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1} \boldsymbol{X}_1'^{(r)} \boldsymbol{P}\boldsymbol{Y} \tag{3.2-175}$$

with residuals

$$\widehat{\boldsymbol{E}}^{(r)} = \left(\boldsymbol{I} - \boldsymbol{X}_1^{(r)} \left(\boldsymbol{X}_1'^{(r)} \boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1} \boldsymbol{X}_1'^{(r)} \boldsymbol{P}\right) \left(\boldsymbol{Y} - \boldsymbol{X}_2^{(r)} \widehat{\boldsymbol{\beta}}_2^{(r)}\right) \tag{3.2-176}$$

is used. Whenever a test problem is naturally given in the form (3.2-173) and (3.2-174), i.e. by

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2, \sigma_0^2 \boldsymbol{P}^{-1}\right) \tag{3.2-177}$$

$$H_0 : \boldsymbol{\beta}_2 = \boldsymbol{w} \quad versus \quad H_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{w} \tag{3.2-178}$$

then all the indices are omitted, in which case the test statistic of the UMPI test reads

$$M(\boldsymbol{Y}) = \left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{w}\right)' \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{w}\right) / \sigma_0^2 \tag{3.2-179}$$

with least squares estimator

$$\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2' \boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2' \boldsymbol{P}\boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{P}\boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1' \boldsymbol{P}\boldsymbol{Y} \tag{3.2-180}$$

and residuals

$$\widehat{\boldsymbol{E}} = \left(\boldsymbol{I} - \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{P}\boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1' \boldsymbol{P}\right) \left(\boldsymbol{Y} - \boldsymbol{X}_2 \widehat{\boldsymbol{\beta}}_2\right). \tag{3.2-181}$$

3. *identical to the test statistic*

$$M(\boldsymbol{Y}) = (\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{w})' \left(\boldsymbol{H}(\boldsymbol{A}'\boldsymbol{P}\boldsymbol{A})^{-1}\boldsymbol{H}'\right)^{-1} (\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{w})/\sigma_0^2, \tag{3.2-182}$$

for the original test problem

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma_0^2 \boldsymbol{P}^{-1}\right) \tag{3.2-183}$$

$$H_0 : \boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{w} \quad versus \quad H_1 : \boldsymbol{H}\bar{\boldsymbol{\beta}} \neq \boldsymbol{w} \tag{3.2-184}$$

if the least squares estimator

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{Y} \tag{3.2-185}$$

with residuals

$$\widehat{\boldsymbol{E}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{3.2-186}$$

is used.

Proof. Part 1: Reversing the parameter homogenization by using (3.2-155) yields

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2'^{(rch)} \widehat{\boldsymbol{\beta}}_2^{(rch)} / \sigma_0^2 = \widehat{\boldsymbol{\beta}}_2'^{(rc)} \boldsymbol{G}_{22} \boldsymbol{G}_{22}' \widehat{\boldsymbol{\beta}}_2^{(rc)} / \sigma_0^2 = \widehat{\boldsymbol{\beta}}_2'^{(rc)} \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \widehat{\boldsymbol{\beta}}_2^{(rc)} / \sigma_0^2,$$

which proves equality of (3.2-160) and (3.2-162). The hypotheses (3.2-161) and (3.2-164) have already been shown to be equivalent by virtue of invariance of the hypotheses (Section 3.2.5, Case 1). Furthermore, (3.2-185) is the sufficient statistic in the observation model (3.2-163) introduced in Section 3.2.4.

Part 2: Reversing the data homogenization (3.2-137) to (3.2-151) results in

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_2^{(rc)} &= \boldsymbol{N}_{22}^{(-1)}\left(\boldsymbol{X}_2'^{(r)}\boldsymbol{G} - \boldsymbol{X}_2'^{(r)}\boldsymbol{G}\boldsymbol{G}'\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{G}\boldsymbol{G}'\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{G}\right)\boldsymbol{G}'\boldsymbol{Y}^{(c)}\\ &= \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y}^{(c)} - \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y}^{(c)}.\end{aligned}$$

Using the non-centered observations (3.2-132) and the definition (3.2-152) of the Schur complement, we obtain

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_2^{(rc)} &= \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\left(\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\boldsymbol{w}\right) - \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\left(\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\boldsymbol{w}\right)\\ &= \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y}\\ &\quad -\boldsymbol{N}_{22}^{(-1)}\left(\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)} - \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\right)\boldsymbol{w}\\ &= \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)}\left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1}\boldsymbol{w}\\ &= \widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}.\end{aligned}$$

From Proposition 3.2 it follows that $\widehat{\boldsymbol{\beta}}_2^{(r)}$ is indeed the least squares estimator for $\boldsymbol{\beta}_2^{(r)}$ in the partitioned observation model (3.2-173). Using this result, the test statistics 3.2-162) and (3.2-172) are identical if the least squares estimators within the corresponding observation models are applied.

Part 3: As a first step towards proving the third part of the propostion we will now prove the identity

$$\left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} = \left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1}.$$

Using the expression (3.2-127) for $\boldsymbol{X}$ we obtain

$$\begin{aligned}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X} &= [\,\boldsymbol{M}' \;\; \boldsymbol{H}'\,]\begin{bmatrix}\boldsymbol{X}_1'^{(r)}\\ \boldsymbol{X}_2'^{(r)}\end{bmatrix}\boldsymbol{P}\left[\boldsymbol{X}_1^{(r)} \;\; \boldsymbol{X}_2^{(r)}\right]\begin{bmatrix}\boldsymbol{M}\\ \boldsymbol{H}\end{bmatrix}\\ &= [\,\boldsymbol{M}' \;\; \boldsymbol{H}'\,]\begin{bmatrix}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)} & \boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\\ \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)} & \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\end{bmatrix}\begin{bmatrix}\boldsymbol{M}\\ \boldsymbol{H}\end{bmatrix}.\end{aligned}$$

Inverting both sided yields

$$(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1} = \begin{bmatrix}\boldsymbol{M}\\ \boldsymbol{H}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)} & \boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\\ \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)} & \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\end{bmatrix}^{-1}[\,\boldsymbol{M}' \;\; \boldsymbol{H}'\,]^{-1}. \tag{3.2-187}$$

It follows that

$$\begin{bmatrix}\boldsymbol{M}\\ \boldsymbol{H}\end{bmatrix}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}[\,\boldsymbol{M}' \;\; \boldsymbol{H}'\,] = \begin{bmatrix}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)} & \boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\\ \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)} & \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\end{bmatrix}^{-1}.$$

After expanding the left side and introducing blocks of the total inverse, we have

$$\begin{bmatrix}\boldsymbol{M}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{M}' & \boldsymbol{M}(\boldsymbol{A}'\boldsymbol{P}\boldsymbol{A})^{-1}\boldsymbol{H}'\\ \boldsymbol{H}(\boldsymbol{A}'\boldsymbol{P}\boldsymbol{A})^{-1}\boldsymbol{M}' & \boldsymbol{H}(\boldsymbol{A}'\boldsymbol{P}\boldsymbol{A})^{-1}\boldsymbol{H}'\end{bmatrix} = \begin{bmatrix}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} & (\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)}\\ (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} & (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)}\end{bmatrix}.$$

Using the definition (3.2-152) of the Schur complement, the identity

$$\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{H}' = \left(\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\right)^{(-1)} = \boldsymbol{N}_{22}^{(-1)}$$

is seen to hold. Inversion of this equation provides us with the desired result

$$\left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1} = \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \tag{3.2-188}$$

In a second step we will prove the equality $\widehat{\boldsymbol{\beta}}_2^{(r)} = \boldsymbol{H}\widehat{\boldsymbol{\beta}}$. Using (3.2-127) and (3.2-187), the least squares estimator (3.2-185) may be expressed as

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{H} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)} & \boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)} \\ \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)} & \boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)} \end{bmatrix}^{-1} [\,\boldsymbol{M}' \;\; \boldsymbol{H}'\,]^{-1} [\,\boldsymbol{M}' \;\; \boldsymbol{H}'\,] \begin{bmatrix} \boldsymbol{X}_1'^{(r)} \\ \boldsymbol{X}_2'^{(r)} \end{bmatrix} \boldsymbol{P}\boldsymbol{Y} \\ &= \begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{H} \end{bmatrix}^{-1} \begin{bmatrix} (\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} & (\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)} \\ (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} & (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1'^{(r)} \\ \boldsymbol{X}_2'^{(r)} \end{bmatrix} \boldsymbol{P}\boldsymbol{Y}. \end{aligned} \tag{3.2-189}$$

Now, the pre-multiplied version of this equation, that is,

$$\begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{H} \end{bmatrix} \widehat{\boldsymbol{\beta}} = \begin{bmatrix} (\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} & (\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)} \\ (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} & (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1'^{(r)} \\ \boldsymbol{X}_2'^{(r)} \end{bmatrix} \boldsymbol{P}\boldsymbol{Y}$$

clearly implies that

$$\boldsymbol{H}\widehat{\boldsymbol{\beta}} = \left[(\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} \;\; (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)} \right] \begin{bmatrix} \boldsymbol{X}_1'^{(r)} \\ \boldsymbol{X}_2'^{(r)} \end{bmatrix} \boldsymbol{P}\boldsymbol{Y}.$$

Observing that $(\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)} = N_{22}^{(-1)}$ and using Equation 1.111 in Koch (1999, p. 33) to obtain the expression

$$(\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} = -N_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}$$

for the other block of the inverse, we obtain

$$\begin{aligned}\boldsymbol{H}\widehat{\boldsymbol{\beta}} &= \left[-N_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1} \;\; N_{22}^{(-1)} \right] \begin{bmatrix} \boldsymbol{X}_1'^{(r)} \\ \boldsymbol{X}_2'^{(r)} \end{bmatrix} \boldsymbol{P}\boldsymbol{Y} \\ &= -N_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y} + N_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y} = \widehat{\boldsymbol{\beta}}_2^{(r)}\end{aligned}$$

according to (3.2-175). This proves that

$$\left(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}\right)' \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \left(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}\right) = \left(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{w}\right)' \left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1} \left(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{w}\right),$$

i.e. the statistics (3.2-182) and (3.2-172) are identical. □

Case 2: $\bar{\sigma}^2$ unknown.

Proposition 3.4. *The invariant test statistic*

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2'^{(rch)} \widehat{\boldsymbol{\beta}}_2^{(rch)} / (m_2 \widehat{\sigma}^2_{(rch)}) \tag{3.2-190}$$

for the UMP test (3.2-159) (at level α) regarding the hypotheses

$$H_0 : \bar{\lambda} = 0 \quad \text{versus} \quad H_1 : \bar{\lambda} > 0, \tag{3.2-191}$$

i.e. for the UMPI test (3.2-159) (at level α) regarding the original hypotheses $H_0 : \boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{w}$ versus $H_1 : \boldsymbol{H}\bar{\boldsymbol{\beta}} \neq \boldsymbol{w}$, is identical to:

1. *the test statistic*

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2'^{(rc)} \left(\boldsymbol{N}_{22}^{(-1)} \right)^{-1} \widehat{\boldsymbol{\beta}}_2^{(rc)} / (m_2 \widehat{\sigma}^2_{(rch)}) \tag{3.2-192}$$

for the equivalent test problem

$$\boldsymbol{Y}^{(ch)} \sim N\left(\boldsymbol{X}_1^{(rh)} \boldsymbol{\beta}_1^{(rc)} + \boldsymbol{X}_2^{(rh)} \boldsymbol{\beta}_2^{(rc)}, \sigma^2 \boldsymbol{I} \right) \tag{3.2-193}$$

$$H_0 : \boldsymbol{\beta}_2^{(rc)} = \boldsymbol{0} \quad \textit{versus} \quad H_1 : \boldsymbol{\beta}_2^{(rc)} \neq \boldsymbol{0}, \tag{3.2-194}$$

which we will call ***the problem of testing the significance of additional parameters $\boldsymbol{\beta}_2^{(rc)}$ with unknown variance factor****, if the least squares estimators*

$$\widehat{\boldsymbol{\beta}}_2^{(rc)} = \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2'^{(rh)} \boldsymbol{Y}^{(ch)} - \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2'^{(rh)} \boldsymbol{X}_1^{(rh)} \left(\boldsymbol{X}_1'^{(rh)} \boldsymbol{X}_1^{(rh)} \right)^{-1} \boldsymbol{X}_1'^{(rh)} \boldsymbol{Y}^{(ch)} \tag{3.2-195}$$

$$\widehat{\sigma}^2_{(rch)} = \widehat{\boldsymbol{E}}'^{(rch)} \widehat{\boldsymbol{E}}^{(rch)} / (n-m) \tag{3.2-196}$$

with residuals

$$\widehat{\boldsymbol{E}}^{(rch)} = \left(\boldsymbol{I} - \boldsymbol{X}_1^{(rh)} \left(\boldsymbol{X}_1'^{(rh)} \boldsymbol{X}_1^{(rh)} \right)^{-1} \boldsymbol{X}_1'^{(rh)} \right) \left(\boldsymbol{Y}^{(ch)} - \boldsymbol{X}_2^{(rh)} \widehat{\boldsymbol{\beta}}_2^{(r)} \right) \tag{3.2-197}$$

are used. Whenever a test problem is naturally given in the form (3.2-193) and (3.2-194), i.e. by

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2, \sigma^2 \boldsymbol{I} \right) \tag{3.2-198}$$

$$H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0} \quad \textit{versus} \quad H_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{0}, \tag{3.2-199}$$

which we will call ***the natural problem of testing the significance of additional parameters $\boldsymbol{\beta}_2$ with unknown variance factor****, then all the indices are omitted, in which case the test statistic of the UMPI test reads*

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{N}_{22}^{(-1)} \right)^{-1} \widehat{\boldsymbol{\beta}}_2 / (m_2 \widehat{\sigma}^2) \tag{3.2-200}$$

with least squares estimators

$$\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2' \boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)} \boldsymbol{X}_2' \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1' \boldsymbol{Y} \tag{3.2-201}$$

$$\widehat{\sigma}^2 = \widehat{\boldsymbol{E}}' \widehat{\boldsymbol{E}} / (n-m) \tag{3.2-202}$$

and residuals

$$\widehat{\boldsymbol{E}} = \left(\boldsymbol{I} - \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1' \right) \left(\boldsymbol{Y} - \boldsymbol{X}_2 \widehat{\boldsymbol{\beta}}_2 \right). \tag{3.2-203}$$

2. *identical to the test statistic*

$$M(\boldsymbol{Y}) = \left(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w} \right)' \left(\boldsymbol{N}_{22}^{(-1)} \right)^{-1} \left(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w} \right) / (m_2 \widehat{\sigma}^2_{(r)}) \tag{3.2-204}$$

for the equivalent test problem

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1^{(r)} + \boldsymbol{X}_2^{(r)}\boldsymbol{\beta}_2^{(r)}, \sigma^2\boldsymbol{P}^{-1}\right) \tag{3.2-205}$$

$$H_0 : \boldsymbol{\beta}_2^{(r)} = \boldsymbol{w} \quad versus \quad H_1 : \boldsymbol{\beta}_2^{(r)} \neq \boldsymbol{w} \tag{3.2-206}$$

if the least squares estimators

$$\widehat{\boldsymbol{\beta}}_2^{(r)} = \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y} \tag{3.2-207}$$

$$\widehat{\sigma}^2_{(r)} = \widehat{\boldsymbol{E}}'^{(r)}\boldsymbol{P}\widehat{\boldsymbol{E}}^{(r)}/(n-m) \tag{3.2-208}$$

with residuals

$$\widehat{\boldsymbol{E}}^{(r)} = \left(\boldsymbol{I} - \boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\right)\left(\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\widehat{\boldsymbol{\beta}}_2^{(r)}\right) \tag{3.2-209}$$

are used. Whenever a test problem is naturally given in the form (3.2-205) and (3.2-206), i.e. by

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2, \sigma^2\boldsymbol{P}^{-1}\right) \tag{3.2-210}$$

$$H_0 : \boldsymbol{\beta}_2 = \boldsymbol{w} \quad versus \quad H_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{w} \tag{3.2-211}$$

then all the indices are omitted, in which case the test statistic of the UMPI test reads

$$M(\boldsymbol{Y}) = \left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{w}\right)'\left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1}\left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{w}\right)/(m_2\widehat{\sigma}^2) \tag{3.2-212}$$

with least squares estimators

$$\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'\boldsymbol{P}\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{P}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{P}\boldsymbol{Y} \tag{3.2-213}$$

$$\widehat{\sigma}^2 = \widehat{\boldsymbol{E}}'\boldsymbol{P}\widehat{\boldsymbol{E}}/(n-m) \tag{3.2-214}$$

and residuals

$$\widehat{\boldsymbol{E}} = \left(\boldsymbol{I} - \boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{P}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{P}\right)\left(\boldsymbol{Y} - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right). \tag{3.2-215}$$

3. *identical to the test statistic*

$$M(\boldsymbol{Y}) = (\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{w})'\left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1}(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{w})/(m_2\widehat{\sigma}^2), \tag{3.2-216}$$

for the original test problem

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{P}^{-1}\right) \tag{3.2-217}$$

$$H_0 : \boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{w} \quad versus \quad H_1 : \boldsymbol{H}\bar{\boldsymbol{\beta}} \neq \boldsymbol{w} \tag{3.2-218}$$

if the least squares estimators

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{Y} \tag{3.2-219}$$

$$\widehat{\sigma}^2 = \widehat{\boldsymbol{E}}'\boldsymbol{P}\widehat{\boldsymbol{E}}/(n-m) \tag{3.2-220}$$

with residuals

$$\widehat{\boldsymbol{E}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{3.2-221}$$

are used.

Proof. Part 1: Reversing the parameter homogenization by using (3.2-155) yields

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2'^{(rch)}\widehat{\boldsymbol{\beta}}_2^{(rch)}/(m_2\widehat{\sigma}^2_{(rch)}) = \widehat{\boldsymbol{\beta}}_2'^{(rc)}\boldsymbol{G}_{22}\boldsymbol{G}_{22}'\widehat{\boldsymbol{\beta}}_2^{(rc)}/(m_2\widehat{\sigma}^2_{(rch)}) = \widehat{\boldsymbol{\beta}}_2'^{(rc)}\left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1}\widehat{\boldsymbol{\beta}}_2^{(rc)}/(m_2\widehat{\sigma}^2_{(rch)}),$$

which proves equality of (3.2-190) and (3.2-192). The hypotheses (3.2-191) and (3.2-194) have already been shown to be equivalent by virtue of invariance of the hypotheses (Section 3.2.5, Case 2). In addition, (3.2-219) and (3.2-220) are the sufficient statistics in the observation model (3.2-193) introduced in Section 3.2.4.

Part 2: In Part 2 of the proof of Proposition 3.3, it was already shown that $\widehat{\boldsymbol{\beta}}_2^{(rc)} = \widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}$. Now we show in addition that $\widehat{\sigma}^2_{(ch)} = \widehat{\sigma}^2$. Reversing the homogenization (3.2-137), centering (3.2-132), and the result $\widehat{\boldsymbol{\beta}}_2^{(rc)} = \widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}$, the residuals (3.2-154) may be rewritten as

$$\begin{aligned}
\widehat{\boldsymbol{E}}^{(rch)} &= \left(\boldsymbol{I} - \boldsymbol{P}^{\frac{1}{2}}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}'^{\frac{1}{2}}\right)\left(\boldsymbol{P}^{\frac{1}{2}}(\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\boldsymbol{w}) - \boldsymbol{P}^{\frac{1}{2}}\boldsymbol{X}_2^{(r)}(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w})\right)\\
&= \left(\boldsymbol{I} - \boldsymbol{P}^{\frac{1}{2}}\boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}'^{\frac{1}{2}}\right)\boldsymbol{P}^{\frac{1}{2}}\left(\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\boldsymbol{w} - \boldsymbol{X}_2^{(r)}\widehat{\boldsymbol{\beta}}_2^{(r)} + \boldsymbol{X}_2^{(r)}\boldsymbol{w})\right)\\
&= \boldsymbol{P}^{\frac{1}{2}}\left(\boldsymbol{I} - \boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\right)\left(\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\widehat{\boldsymbol{\beta}}_2^{(r)}\right)\\
&= \boldsymbol{P}^{\frac{1}{2}}\widehat{\boldsymbol{E}}^{(r)}.
\end{aligned}$$

Therefore, $\widehat{\sigma}^2_{(rch)} = \widehat{\boldsymbol{E}}'^{(rch)}\widehat{\boldsymbol{E}}^{(rch)}/(n-m) = \widehat{\boldsymbol{E}}'^{(r)}\boldsymbol{P}\widehat{\boldsymbol{E}}^{(r)}/(n-m) = \widehat{\sigma}^2$. In other words, the estimator for the variance of unit weight is not affected by transformations that involve centering or homogenization of the observation model. Proposition 3.2 shows that $\widehat{\sigma}^2_{(r)}$ is the least squares estimator of σ^2. Using the above equalities, the test statistics 3.2-192) and (3.2-204) are identical if the least squares estimators within the corresponding observation models are applied.

Part 3: In addition to the result given by Part 3 of the proof of Proposition 3.3, it remains to prove that also $\widehat{\sigma}^2 = \widehat{\sigma}^2_{(r)}$. From the equivalent expressions (3.2-189) and (3.2-127) for the least squares estimator (3.2-219) and $\boldsymbol{X}$, respectively, it follows that

$$\begin{aligned}
\widehat{\boldsymbol{E}} &= \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\\
&= \boldsymbol{Y} - \left[\boldsymbol{X}_1^{(r)}\ \boldsymbol{X}_2^{(r)}\right]\begin{bmatrix}\boldsymbol{M}\\ \boldsymbol{H}\end{bmatrix}\begin{bmatrix}\boldsymbol{M}\\ \boldsymbol{H}\end{bmatrix}^{-1}\begin{bmatrix}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} & (\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)}\\ (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)} & (\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_1'^{(r)}\\ \boldsymbol{X}_2'^{(r)}\end{bmatrix}\boldsymbol{P}\boldsymbol{Y}\\
&= \boldsymbol{Y} - \Big[\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)}\boldsymbol{X}_1'^{(r)} + \boldsymbol{X}_2^{(r)}(\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{(-1)}\boldsymbol{X}_1'^{(r)} + \boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)})^{(-1)}\boldsymbol{X}_2'^{(r)}\\
&\quad + \boldsymbol{X}_2^{(r)}\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\Big]\boldsymbol{P}\boldsymbol{Y}
\end{aligned}$$

Using the identity relations (Equation 1.111 in Koch, 1999, p. 33) for submatrices of the inverse of a 2×2-block matrix, we obtain

$$\begin{aligned}
\widehat{\boldsymbol{E}} &= \boldsymbol{Y} - \boldsymbol{X}_1^{(r)}\left[(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1} + (\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\right]\\
&\quad \times\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y} + \boldsymbol{X}_2^{(r)}\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y}\\
&\quad +\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y}\\
&= \boldsymbol{Y} - \boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y}\\
&\quad -\boldsymbol{X}_2^{(r)}\left(\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y}\right)\\
&\quad +\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_2^{(r)}\left(\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{Y}\right)\\
&= \left(\boldsymbol{I} - \boldsymbol{X}_1^{(r)}\left(\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\boldsymbol{X}_1^{(r)}\right)^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{P}\right)\left(\boldsymbol{Y} - \boldsymbol{X}_2^{(r)}\widehat{\boldsymbol{\beta}}_2^{(r)}\right) = \widehat{\boldsymbol{E}}^{(r)}
\end{aligned}$$

according to (3.2-153). This proves that the reparameterization does not affect the estimator of the residuals. Due to $\widehat{\sigma}^2 = \widehat{\boldsymbol{E}}'\boldsymbol{P}\widehat{\boldsymbol{E}}/(n-m) = \widehat{\boldsymbol{E}}'^{(r)}\boldsymbol{P}\widehat{\boldsymbol{E}}^{(r)}/(n-m) = \widehat{\sigma}^2_{(r)}$, the reparameterization does not change the estimator of the variance factor either. This completes the proof of

$$\left(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}\right)'\left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1}\left(\widehat{\boldsymbol{\beta}}_2^{(r)} - \boldsymbol{w}\right)/(m\widehat{\sigma}^2_{(r)}) = \left(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{w}\right)'\left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1}\left(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{w}\right)(m\widehat{\sigma}^2).$$

□

3.2.7 Equivalent forms of the UMPI test concerning parameters of the functional model

We will now prove that the Generalized Likelihood Ratio test is equivalent to the UMPI test if the set of linear restrictions is tested. We will restrict attention to *the problem of testing the significance of additional parameters* (Case 1 of Proposition 3.3), knowing that if the test problem naturally involves a set of linear restrictions (3.2-124) (Case 3 of Proposition 3.3), it may be transformed into the first form.

Case 1: $\bar{\sigma}^2 = \sigma_0^2$ known.

Proposition 3.5. *For testing $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$ in the (possibly reparameterized, centered and homogenized) linear model $\boldsymbol{Y} \sim N(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2, \sigma_0^2\boldsymbol{I})$, the statistic (3.2-162) with*

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1 \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1'\boldsymbol{X}_2\right) \widehat{\boldsymbol{\beta}}_2/\sigma_0^2 \tag{3.2-222}$$

of the UMPI test (3.2-158) is identical to:

1. *the Likelihood Ratio statistic (2.5-99) with*

$$T_{LR}(\boldsymbol{Y}) = -2\left(\mathcal{L}(\widetilde{\boldsymbol{\beta}}_1, \widetilde{\boldsymbol{\beta}}_2; \boldsymbol{Y}) - \mathcal{L}(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2; \boldsymbol{Y})\right), \tag{3.2-223}$$

2. *Rao's Score statistic (2.5-115) with*

$$T_{RS}(\boldsymbol{Y}) = \widetilde{\boldsymbol{U}}'\boldsymbol{X}_2 \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'\widetilde{\boldsymbol{U}}/\sigma_0^2, \tag{3.2-224}$$

if the least squares estimators $\widetilde{\boldsymbol{\beta}}_1$ (under the restrictions of H_0) and $\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$ (unrestricted) are used.

Proof. Part 1: The GLR/LR test compares, on the basis of given data, the likelihood of the observation model under H_0 to that of the model under H_1. If H_0 is true, then the above observation model is a Gauss-Markov model with restrictions $\boldsymbol{\beta}_2 = \mathbf{0}$. As this restricted model may also be written as $\boldsymbol{Y} \sim N(\boldsymbol{X}_1\boldsymbol{\beta}_1, \sigma_0^2\boldsymbol{I})$, the restricted least squares estimators are given by

$$\boldsymbol{X}_1'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 = \boldsymbol{X}_1'\boldsymbol{Y}, \tag{3.2-225}$$

$$\widetilde{\boldsymbol{\beta}}_2 = \mathbf{0}. \tag{3.2-226}$$

If, on the other hand, H_1 is true, then the observation model constitutes an unrestricted Gauss-Markov model, and the unrestricted least squares estimators read

$$\boldsymbol{X}_1'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_1'\boldsymbol{Y}, \tag{3.2-227}$$

$$\boldsymbol{X}_2'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_2'\boldsymbol{Y}. \tag{3.2-228}$$

Koch (1999, p. 161) proved that, due to the normal distribution of the observations, these least squares estimators are identical to the maximum likelihood estimators. Therefore, the GLR becomes, according to (2.5-94),

$$\begin{aligned}
GLR(\boldsymbol{Y}) &= \frac{L(\widetilde{\boldsymbol{\beta}}_1, \mathbf{0}; \boldsymbol{Y})}{L(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2; \boldsymbol{Y})} = \frac{\left(2\pi\sigma_0^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2}(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widetilde{\boldsymbol{\beta}}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widetilde{\boldsymbol{\beta}}_2)\right\}}{\left(2\pi\sigma_0^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2}(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)\right\}} \\
&= \frac{\exp\left\{-\frac{1}{2\sigma_0^2}\left(\boldsymbol{Y}'\boldsymbol{Y} - 2\widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{Y} - 2\widetilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Y} + \widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 + 2\widetilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 + \widetilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widetilde{\boldsymbol{\beta}}_2\right)\right\}}{\exp\left\{-\frac{1}{2\sigma_0^2}\left(\boldsymbol{Y}'\boldsymbol{Y} - 2\widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{Y} - 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Y} + \widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right)\right\}} \\
&= \exp\left\{-\frac{1}{2\sigma_0^2}\left(-2\widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{Y} - 2\widetilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Y} + \widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 + 2\widetilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 + \widetilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widetilde{\boldsymbol{\beta}}_2\right.\right. \\
&\quad \left.\left. + 2\widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{Y} + 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right)\right\}.
\end{aligned}$$

Notice now that $2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Y}$ equals $2\widehat{\boldsymbol{\beta}}_2'$ times (3.2-228). Substitution of this and of (3.2-226) yields

$$\begin{aligned}
GLR(\boldsymbol{Y}) &= \exp\left\{-\frac{1}{2\sigma_0^2}\left(-2\widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{Y} + \widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 + 2\widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{Y} + 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right.\right. \\
&\quad \left.\left. - \widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma_0^2}\left(-2\widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{Y} + \widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 + 2\widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right)\right\}
\end{aligned}$$

Then, substitution of $\widetilde{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}$ from (3.2-225) and $\widehat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{Y} - \boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)$ from (3.2-227) leads to

$$\begin{aligned}
GLR(\boldsymbol{Y}) &= \exp\left\{-\frac{1}{2\sigma_0^2}\left(-2\boldsymbol{Y}'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}\right.\right.\\
&\quad +2(\boldsymbol{X}_1'\boldsymbol{Y} - \boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y} - (\boldsymbol{X}_1'\boldsymbol{Y} - \boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\\
&\quad \left.\left.\times(\boldsymbol{X}_1'\boldsymbol{Y} - \boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2) + \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right)\right\}\\
&= \exp\left\{-\frac{1}{2\sigma_0^2}\left(\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 - \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right)\right\}\\
&= \exp\left\{-\frac{1}{2\sigma_0^2}\widehat{\boldsymbol{\beta}}_2'\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2\right\}.
\end{aligned}$$

Using definitions (2.5-95) and (2.5-99), the LR statistic

$$T_{LR}(\boldsymbol{Y}) = -2\ln GLR(\boldsymbol{Y}) = \frac{1}{\sigma_0^2}\widehat{\boldsymbol{\beta}}_2'\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2.$$

is indeed equal to $M(\boldsymbol{Y})$.

Part 2: The RS test determines whether the estimates for the Gauss-Markov model with restrictions (valid under H_0) satisfy the likelihood equations for the unrestricted Gauss-Markov model (valid under H_1). Therefore, we must first determine log-likelihood function of the unrestricted model, which is given by

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y}) &= \ln\left[\left(2\pi\sigma_0^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma_0^2}(\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{\beta}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{\beta}_2)\right\}\right]\\
&= -\frac{n}{2}\ln(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}(\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{\beta}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{\beta}_2).
\end{aligned}$$

From this, the log-likelihood score (2.5-111) follows to be

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y}) = \begin{bmatrix} \dfrac{\partial\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_1; \boldsymbol{Y})}{\partial\boldsymbol{\beta}_1} \\ \dfrac{\partial\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_1; \boldsymbol{Y})}{\partial\boldsymbol{\beta}_2} \end{bmatrix} = \frac{1}{\sigma_0^2}\begin{bmatrix} \boldsymbol{X}_1'(\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{\beta}_2) \\ \boldsymbol{X}_2'(\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{\beta}_2) \end{bmatrix}$$

Evaluating the score at the restricted least squares estimators $\widetilde{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}$ and $\widetilde{\boldsymbol{\beta}}_2 = \boldsymbol{0}$ as in (3.2-225) and (3.2-226), and using the corresponding residuals $\widetilde{\boldsymbol{U}} = \boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1$, yields

$$\boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\beta}}_1, \boldsymbol{0}; \boldsymbol{Y}) = \frac{1}{\sigma_0^2}\begin{bmatrix} \boldsymbol{X}_1'(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1) \\ \boldsymbol{X}_2'(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1) \end{bmatrix} = \frac{1}{\sigma_0^2}\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}_2'\widetilde{\boldsymbol{u}} \end{bmatrix}$$

The Hessian and the information of $\boldsymbol{Y}$ are then

$$\boldsymbol{\mathcal{H}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y}) = \begin{bmatrix} \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1'} & \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2'} \\ \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y})}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_1'} & \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y})}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_2'} \end{bmatrix} = -\frac{1}{\sigma_0^2}\begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix}$$

and

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y}) = E_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left\{-\boldsymbol{\mathcal{H}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{Y})\right\} = \frac{1}{\sigma_0^2}\begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix}.$$

Now, using the definition of the RS statistic (2.5-115), we obtain

$$\begin{aligned}
T_{RS} &= \frac{1}{\sigma_0^2}\begin{bmatrix} \mathbf{0} \\ \boldsymbol{X}_2'\widetilde{\boldsymbol{U}} \end{bmatrix}'\begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0} \\ \boldsymbol{X}_2'\widetilde{\boldsymbol{U}} \end{bmatrix} \\
&= \frac{1}{\sigma_0^2}\begin{bmatrix} \mathbf{0}' & \widetilde{\boldsymbol{U}}'\boldsymbol{X}_2 \end{bmatrix}\begin{bmatrix} (\boldsymbol{X}_1'\boldsymbol{X}_1)^{(-1)} & (\boldsymbol{X}_1'\boldsymbol{X}_2)^{(-1)} \\ (\boldsymbol{X}_2'\boldsymbol{X}_1)^{(-1)} & (\boldsymbol{X}_2'\boldsymbol{X}_2)^{(-1)} \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \boldsymbol{X}_2'\widetilde{\boldsymbol{U}} \end{bmatrix} \\
&= \frac{1}{\sigma_0^2}\widetilde{\boldsymbol{U}}'\boldsymbol{X}_2\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}} \\
&= \frac{1}{\sigma_0^2}\widetilde{\boldsymbol{U}}'\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}},
\end{aligned}$$

where we used the Schur complement as defined in (3.2-152). The statistic T_{RS} is defined in terms of the restricted estimates $\widetilde{\boldsymbol{\beta}}_1$ through the residuals $\widetilde{\boldsymbol{u}}$. On the other hand, M is a function of unrestricted estimates $\widehat{\boldsymbol{\beta}}_2$. To show that both statistics are identical, we will use (3.2-225) - (3.2-228) to express M as a function of $\widetilde{\boldsymbol{\beta}}_1$. The first step is to combine (3.2-225) and (3.2-227), which yields

$$\boldsymbol{X}_1'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_1'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1,$$

which in turn implies that

$$\widehat{\boldsymbol{\beta}}_1 = \widetilde{\boldsymbol{\beta}}_1 - (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2.$$

Substitution of this result into (3.2-228) gives

$$\boldsymbol{X}_2'\boldsymbol{X}_1\left(\widetilde{\boldsymbol{\beta}}_1 - (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right) + \boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_2'\boldsymbol{Y},$$

or, after rearranging terms,

$$\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_2'\boldsymbol{Y} - \boldsymbol{X}_2'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1,$$

and finally

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_2 &= \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1) \\
&= \boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}}.
\end{aligned}$$

Now we can rewrite the statistic M as

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2'\left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1}\widehat{\boldsymbol{\beta}}_2/\sigma_0^2 = \widetilde{\boldsymbol{U}}'\boldsymbol{X}_2\boldsymbol{N}_{22}^{(-1)}\left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1}\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{U}}'\boldsymbol{X}_2\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}},$$

which completes the proof that $M = T_{RS}$. □

Case 2: $\bar{\sigma}^2$ unknown.

Proposition 3.6. *For testing $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$ in the (possibly reparameterized, centered and homogenized) linear model $\boldsymbol{Y} \sim N(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2, \sigma^2\boldsymbol{I})$, the UMPI test (3.2-159), based on the statistic (3.2-192) with*

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1 \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1'\boldsymbol{X}_2\right) \widehat{\boldsymbol{\beta}}_2/(m_2\widehat{\sigma}^2) \tag{3.2-229}$$

is equivalent to:

1. *the Likelihood Ratio test (2.5-100), based on the statistic (2.5-94) with*

$$T_{LR}(\boldsymbol{Y}) = n \ln\left(1 + \frac{m_2}{n-m} M(\boldsymbol{Y})\right), \tag{3.2-230}$$

2. *Rao's Score test, based on the statistic (2.5-115) with*

$$\begin{aligned} T_{RS}(\boldsymbol{Y}) &= \widetilde{\boldsymbol{U}}'\boldsymbol{X}_2 \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'\widetilde{\boldsymbol{U}}/\widetilde{\sigma}^2_{ML} & (3.2\text{-}231) \\ &= n\frac{\frac{m_2}{n-m}M(\boldsymbol{Y})}{1+\frac{m_2}{n-m}M(\boldsymbol{Y})} & (3.2\text{-}232) \end{aligned}$$

if the maximum likelihood estimators $\widetilde{\boldsymbol{\beta}}_1$, $\widetilde{\sigma}^2$ (under the restriction of H_0) and $\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$, $\widehat{\sigma}^2$ (unrestricted) are used.

Proof. Part 1: The only difference to Part 1 of Proposition 3.5 is that the variance factor must be estimated here in addition to the functional paramaters. Now, if H_0 is true, then the observation model is a Gauss-Markov model with restrictions, which may be written in the simple form $\boldsymbol{Y} \sim N(\boldsymbol{X}_1\boldsymbol{\beta}_1, \sigma^2\boldsymbol{I})$. The restricted least squares estimators are given by

$$\begin{aligned} \boldsymbol{X}_1'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 &= \boldsymbol{X}_1'\boldsymbol{Y}, & (3.2\text{-}233) \\ \widetilde{\boldsymbol{\beta}}_2 &= \mathbf{0}, & (3.2\text{-}234) \\ \widetilde{\sigma}^2 &= (\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1)'(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1)/(n-m_1). & (3.2\text{-}235) \end{aligned}$$

Notice that the estimators (3.2-233) and (3.2-234) (when the variance factor is unknown) are exactly the same as the estimators (3.2-225) and (3.2-226) (when the variance factor is known). The alternative observation model (under H_1) reads $\boldsymbol{Y} \sim N(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2, \sigma^2\boldsymbol{I})$, for which the least squares estimators are given by

$$\begin{aligned} \boldsymbol{X}_1'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 &= \boldsymbol{X}_1'\boldsymbol{Y}, & (3.2\text{-}236) \\ \boldsymbol{X}_2'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 &= \boldsymbol{X}_2'\boldsymbol{Y}, & (3.2\text{-}237) \\ \widehat{\sigma}^2 &= (\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)/(n-m). & (3.2\text{-}238) \end{aligned}$$

Again, the fact that σ^2 must be estimated does not affect the structure of the estimators for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Consequently, the estimators (3.2-236) and (3.2-237) of the current model are identical to (3.2-227) and (3.2-228) of the model with known variance factor. It will also be useful to note that the variance estimators (3.2-235) and (3.2-238) are divided by different scaling factors or degrees of freedom (m_1 and m, respectively) in light of the fact that the functional models under H_0 and H_1 have different numbers of functional parameters (m_1 parameters $\boldsymbol{\beta}_1$ and $m = m_1 + m_2$ parameters $[\,\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2'\,]$, respectively).

As in Part 1 of Proposition 3.5, the least squares estimators (3.2-233), (3.2-234), (3.2-236) and (3.2-237) for the parameters of the functional model are exactly the same as the maximum likelihood estimators. However, as far as the estimation of σ^2 is concerned, Koch (1999, p. 162) shows that the (unbiased) least squares estimators (3.2-235) and (3.2-238) differ from the corresponding (biased) maximum likelihood estimators, given by

$$\widetilde{\sigma}^2_{ML} = (\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1)'(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1)/n \tag{3.2-239}$$

for the observation model under H_0, and

$$\widehat{\sigma}^2_{ML} = (\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)/n \tag{3.2-240}$$

for the alternative observation model under H_1. However, if these maximum likelihood estimators are adjusted such that the scaling factor n is replaced in each case by the correct degree of freedom, then we obtain the least squares estimators (3.2-235) and (3.2-238), which we might call the *bias-corrected maximum likelihood estimators of* σ^2.

After these preliminary remarks, we may evaluate the Generalized Likelihood Ratio, that is,

$$\begin{aligned} GLR(\boldsymbol{Y}) &= \frac{L(\widetilde{\boldsymbol{\beta}}_1, \boldsymbol{0}, \widetilde{\sigma}^2; \boldsymbol{Y})}{L(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\sigma}^2; \boldsymbol{Y})} = \frac{\left(2\pi\widetilde{\sigma}^2_{ML}\right)^{-n/2} \exp\left\{-\frac{1}{2\widetilde{\sigma}^2_{ML}}(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\boldsymbol{0})'(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\boldsymbol{0})\right\}}{\left(2\pi\widehat{\sigma}^2_{ML}\right)^{-n/2} \exp\left\{-\frac{1}{2\widehat{\sigma}^2_{ML}}(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)\right\}} \\ &= \left(\frac{\widetilde{\sigma}^2_{ML}}{\widehat{\sigma}^2_{ML}}\right)^{-n/2} \exp\left\{-\frac{1}{2\widetilde{\sigma}^2_{ML}} n\widetilde{\sigma}^2_{ML} + \frac{1}{2\widehat{\sigma}^2_{ML}} n\widehat{\sigma}^2_{ML}\right\} = \left(\frac{\widetilde{\sigma}^2_{ML}}{\widehat{\sigma}^2_{ML}}\right)^{-n/2}. \end{aligned}$$

By virtue of the definitions (2.5-95) and (2.5-99) regarding the Likelihood Ratio statistic, we obtain

$$T_{LR} = -2\ln GLR(\boldsymbol{Y}) = -2\ln\left(\frac{\widetilde{\sigma}^2_{ML}}{\widehat{\sigma}^2_{ML}}\right)^{-n/2} = n\ln\left(\frac{\widetilde{\sigma}^2_{ML}}{\widehat{\sigma}^2_{ML}}\right). \tag{3.2-241}$$

We will now prove that $n\ln(1 + \frac{m_2}{n-m}M) = n\ln\left(\frac{\widetilde{\sigma}^2_{ML}}{\widehat{\sigma}^2_{ML}}\right)$ or, equivalently, that $1 + \frac{m_2}{n-m}M = \frac{\widetilde{\sigma}^2_{ML}}{\widehat{\sigma}^2_{ML}}$.

$$\begin{aligned} 1 + \frac{m_2}{n-m}M &= 1 + \frac{m_2}{n-m}\widehat{\boldsymbol{\beta}}_2'\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2/(m_2\widehat{\sigma}^2) \\ &= 1 + \widehat{\boldsymbol{\beta}}_2'\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2/((n-m)\widehat{\sigma}^2). \end{aligned}$$

Using the equality $(n-m)\widehat{\sigma}^2 = n\widehat{\sigma}^2_{ML}$, we obtain

$$\begin{aligned} 1 + \frac{m_2}{n-m}M &= \frac{1}{n\widehat{\sigma}^2_{ML}}\left\{n\widehat{\sigma}^2_{ML} + \widehat{\boldsymbol{\beta}}_2'\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2\right\} \\ &= \frac{1}{n\widehat{\sigma}^2_{ML}}\left\{(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)\right. \\ &\quad \left.+\widehat{\boldsymbol{\beta}}_2'\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2\right\}. \end{aligned}$$

Observe now that the equality $2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Y} = 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + 2\widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2$ when (3.2-237) is multiplied by $2\widehat{\boldsymbol{\beta}}_2'$, so that some of terms cancel out, leading to

$$1 + \frac{m_2}{n-m}M = \frac{1}{n\widehat{\sigma}^2_{ML}}\left\{\boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{Y}'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \widehat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right\}.$$

Notice that, by using (3.2-233), (3.2-236) may be rewritten as $\widehat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{Y} - \boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2) = \widetilde{\boldsymbol{\beta}}_1 - (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2$. Substitution of this expression allows for the simplification

$$\begin{aligned} 1 + \frac{m_2}{n-m}M &= \frac{1}{n\widehat{\sigma}^2_{ML}}\left\{\boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{Y}'\boldsymbol{X}_1\left[\widetilde{\boldsymbol{\beta}}_1 - \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right] + \left[\widetilde{\boldsymbol{\beta}}_1' - \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right]\right. \\ &\quad \left.\times\boldsymbol{X}_1'\boldsymbol{X}_1\left[\widetilde{\boldsymbol{\beta}}_1 - \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right] - \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2\right\} \\ &= \frac{1}{n\widehat{\sigma}^2_{ML}}\left\{\boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{Y}'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1 + \widetilde{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1\right\} = \frac{1}{n\widehat{\sigma}^2_{ML}}\left(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1\right)'\left(\boldsymbol{Y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1\right) \\ &= \frac{n\widetilde{\sigma}^2_{ML}}{n\widehat{\sigma}^2_{ML}}. \end{aligned}$$

From this follows directly the desired equality $n\ln(1 + \frac{m_2}{n-m}M) = n\ln\left(\frac{\widetilde{\sigma}^2_{ML}}{\widehat{\sigma}^2_{ML}}\right) = T_{LR}$. The LR test is truly equivalent to the UMPI test because T_{LR} is seen to be a strictly monotonically increasing function of M. Therefore, if the UMPI test is based on the statistic $T_{LR}(\boldsymbol{Y})$ instead of $M(\boldsymbol{Y})$, then the critical value may be transformed accordingly by this function, and the transformed region of rejection is equivalent to the original region.

The difference between Case 1 (σ^2 known *a priori*) and Case 2 (σ^2 unknown) is that, in the first case, the test statistics $T_{LR}(\boldsymbol{Y})$ and $M(\boldsymbol{Y})$, hence their distributions, therefore also the region of rejection remains unchanged, while in the second case, all of these quantities do change, but remain equivalent.

Part 2: From the log-likelihood function

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y}) &= \ln\left[\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2)'(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2)\right\}\right]\\ &= -\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2)'(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2).\end{aligned}$$

of the observation model (including both H_0 and H_1), the log-likelihood score is derived as

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y}) = \begin{bmatrix}\dfrac{\partial\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1,\sigma^2;\boldsymbol{Y})}{\partial\boldsymbol{\beta}_1}\\ \dfrac{\partial\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1,\sigma^2;\boldsymbol{Y})}{\partial\boldsymbol{\beta}_2}\\ \dfrac{\partial\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1,\sigma^2;\boldsymbol{Y})}{\partial\sigma^2}\end{bmatrix} = \begin{bmatrix}\frac{1}{\sigma^2}\boldsymbol{X}_1'\left(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2\right)\\ \frac{1}{\sigma^2}\boldsymbol{X}_2'\left(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2\right)\\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\left(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2\right)'\left(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2\right)\end{bmatrix}.$$

If $\boldsymbol{U}=\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1-\boldsymbol{X}_2\boldsymbol{\beta}_2$ denote the residuals, the Hessian and the information matrix read

$$\begin{aligned}\boldsymbol{\mathcal{H}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y}) &= \begin{bmatrix}\dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1'} & \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2'} & \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\boldsymbol{\beta}_1\partial\sigma^2}\\ \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_1'} & \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_2'} & \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\boldsymbol{\beta}_2\partial\sigma^2}\\ \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\sigma^2\partial\boldsymbol{\beta}_1} & \frac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\sigma^2\partial\boldsymbol{\beta}_2} & \dfrac{\partial^2\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})}{\partial\sigma^2\partial\sigma^2}\end{bmatrix}\\ &= \begin{bmatrix}-\frac{1}{\sigma^2}\boldsymbol{X}_1'\boldsymbol{X}_1 & -\frac{1}{\sigma^2}\boldsymbol{X}_1'\boldsymbol{X}_2 & -\frac{1}{\sigma^4}\boldsymbol{X}_1'\boldsymbol{U}\\ -\frac{1}{\sigma^2}\boldsymbol{X}_2'\boldsymbol{X}_1 & -\frac{1}{\sigma^2}\boldsymbol{X}_2'\boldsymbol{X}_2 & -\frac{1}{\sigma^4}\boldsymbol{X}_2'\boldsymbol{U}\\ -\frac{1}{\sigma^4}\boldsymbol{U}'\boldsymbol{X}_1 & -\frac{1}{\sigma^4}\boldsymbol{U}'\boldsymbol{X}_2 & \frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\boldsymbol{U}'\boldsymbol{U}\end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y}) &= E\left\{-\boldsymbol{\mathcal{H}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma^2;\boldsymbol{Y})\right\}\\ &= \begin{bmatrix}\frac{1}{\sigma^2}\boldsymbol{X}_1'\boldsymbol{X}_1 & \frac{1}{\sigma^2}\boldsymbol{X}_1'\boldsymbol{X}_2 & \frac{1}{\sigma^4}\boldsymbol{X}_1'E\{\boldsymbol{U}\}\\ \frac{1}{\sigma^2}\boldsymbol{X}_2'\boldsymbol{X}_1 & \frac{1}{\sigma^2}\boldsymbol{X}_2'\boldsymbol{X}_2 & \frac{1}{\sigma^4}\boldsymbol{X}_2'E\{\boldsymbol{U}\}\\ \frac{1}{\sigma^4}E\{\boldsymbol{U}'\}\boldsymbol{X}_1 & \frac{1}{\sigma^4}E\{\boldsymbol{U}'\}\boldsymbol{X}_2 & -\frac{n}{2\sigma^4}+\frac{1}{\sigma^6}E\{\boldsymbol{U}'\boldsymbol{U}\}\end{bmatrix} = \begin{bmatrix}\frac{1}{\sigma^2}\boldsymbol{X}_1'\boldsymbol{X}_1 & \frac{1}{\sigma^2}\boldsymbol{X}_1'\boldsymbol{X}_2 & \boldsymbol{0}\\ \frac{1}{\sigma^2}\boldsymbol{X}_2'\boldsymbol{X}_1 & \frac{1}{\sigma^2}\boldsymbol{X}_2'\boldsymbol{X}_2 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \frac{n}{2\sigma^4}\end{bmatrix},\end{aligned}$$

where $E\{\boldsymbol{U}\}=\boldsymbol{0}$ and $E\{\boldsymbol{U}'\boldsymbol{U}\}=n\sigma^2$ by virtue of the Markov conditions. From the definition of Rao's Score statistic (), we obtain

$$\begin{aligned}T_{RS}(\boldsymbol{Y}) &= \boldsymbol{\mathcal{S}}'(\widetilde{\boldsymbol{\beta}}_1,\widetilde{\boldsymbol{\beta}}_2,\widetilde{\sigma}^2;\boldsymbol{Y})\,\boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\beta}}_1,\widetilde{\boldsymbol{\beta}}_2,\widetilde{\sigma}^2;\boldsymbol{Y})\,\boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\beta}}_1,\widetilde{\boldsymbol{\beta}}_2,\widetilde{\sigma}^2;\boldsymbol{Y})\\ &= \begin{bmatrix}\boldsymbol{0}\\ \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}}\\ \boldsymbol{0}\end{bmatrix}'\begin{bmatrix}\frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}_1'\boldsymbol{X}_1 & \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}_1'\boldsymbol{X}_2 & \boldsymbol{0}\\ \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}_2'\boldsymbol{X}_1 & \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}_2'\boldsymbol{X}_2 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \frac{n}{2\widetilde{\sigma}^4_{ML}}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{0}\\ \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}}\\ \boldsymbol{0}\end{bmatrix}\\ &= \frac{1}{\widetilde{\sigma}^2_{ML}}\widetilde{\boldsymbol{U}}'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{(-1)}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}}\\ &= \widetilde{\boldsymbol{U}}'\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2-\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}}/\widetilde{\sigma}^2_{ML}.\end{aligned}$$

as proposed in (3.2-231). To prove (3.2-232), rewrite T_{RS} as

$$\begin{aligned}T_{RS}(\boldsymbol{Y}) &= \frac{1}{\widetilde{\sigma}^2_{ML}}\left(\boldsymbol{Y}-\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1\right)'\boldsymbol{X}_2\boldsymbol{N}_{22}^{(-1)}\boldsymbol{X}_2'\left(\boldsymbol{Y}-\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1\right)\\ &= \frac{1}{\widetilde{\sigma}^2_{ML}}\left(\boldsymbol{X}_2'\boldsymbol{Y}-\boldsymbol{X}_2'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1\right)'\boldsymbol{N}_{22}^{(-1)}\left(\boldsymbol{X}_2'\boldsymbol{Y}-\boldsymbol{X}_2'\boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1\right),\end{aligned}$$

substitute (3.2-233) for $\widetilde{\boldsymbol{\beta}}_1$, that is,

$$T_{RS}(\boldsymbol{Y}) = \frac{1}{\widetilde{\sigma}^2_{ML}}\left(\boldsymbol{X}_2'\boldsymbol{Y}-\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}\right)'\boldsymbol{N}_{22}^{(-1)}\left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1}\boldsymbol{N}_{22}^{(-1)}\left(\boldsymbol{X}_2'\boldsymbol{Y}-\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}\right).$$

Substitution of $\widehat{\boldsymbol{\beta}}_2$ as given in (3.2-201) yields

$$T_{RS}(\boldsymbol{Y}) = \frac{1}{\widetilde{\sigma}^2_{ML}} \widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \widehat{\boldsymbol{\beta}}_2.$$

From Part 1 of the current proof, we already know that $1 + \frac{m_2}{n-m} M(\boldsymbol{Y}) = \widetilde{\sigma}^2_{ML}/\widehat{\sigma}^2_{ML}$. Isolation of $\widetilde{\sigma}^2_{ML}$ and substitution into T_{RS} then results in

$$T_{RS}(\boldsymbol{Y}) = \frac{\widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \widehat{\boldsymbol{\beta}}_2}{\widehat{\sigma}^2_{ML} \left(1 + \frac{m_2}{n-m} M\right)}.$$

Using the relationship $(n-m)\widehat{\sigma}^2 = n\widehat{\sigma}^2_{ML}$ between the least squares and the maximum likelihood estimator for σ^2, we obtain

$$T_{RS}(\boldsymbol{Y}) = \frac{\widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \widehat{\boldsymbol{\beta}}_2}{\frac{n-m}{n}\widehat{\sigma}^2 \left(1 + \frac{m_2}{n-m} M(\boldsymbol{Y})\right)} = \frac{\frac{n}{n-m}\widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{N}_{22}^{(-1)}\right)^{-1} \widehat{\boldsymbol{\beta}}_2/\widehat{\sigma}^2}{1 + \frac{m_2}{n-m} M(\boldsymbol{Y})} = n\,\frac{\frac{m_2}{n-m} M(\boldsymbol{Y})}{1 + \frac{m_2}{n-m} M(\boldsymbol{Y})}.$$

As with the relationship between the statistics T_{LR} and M, established in Part 1 of this proof, we see that the statistic T_{RS} is a strictly monotonically increasing function of M. Therefore, the UMPI test (3.2-159) may be based upon T_{RS} instead of M if the critical value is transformed according to (3.2-232). In this sense, we say that Rao's Score test is equivalent to the UMPI test. If the true value of σ^2 is known *a priori*, then Proposition 3.5 states that the statistics T_{RS} and M, hence their distributions, and therefore the corresponding critical regions are identical. □

3.3 Application 1: Testing for outliers

Consider the Gauss-Markov model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\nabla} + \boldsymbol{U}, \tag{3.3-242}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2 \boldsymbol{I}, \tag{3.3-243}$$

where $\boldsymbol{X}\boldsymbol{\beta}$ represents the deterministic trend model underlying observations $\boldsymbol{Y}$, and where $\boldsymbol{Z}\boldsymbol{\nabla}$ denotes an additional **mean shift model** . Both design matrices $\boldsymbol{X} \in \mathbb{R}^{n \times m_1}$ and $\boldsymbol{Z} \in \mathbb{R}^{n \times m_2}$ are assumed to be known and of full rank. The mean shift model takes its simplest form if $\boldsymbol{Z}$ is a vector with zeros in the components $1, \ldots, i-1, i+1, \ldots, n$ and a one in the i-th component, that is,

$$Z = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]'. \tag{3.3-244}$$

Then, the mean shift parameter ∇ may be viewed as a single **additive outlier** or gross error affecting observation Y_i. If a model for multiple outliers is desired, then $\boldsymbol{Z}$ is simply expanded by additional columns. A test for a single (or multiple) outlier(s) may then be based on the hypotheses

$$H_0 : \bar{\boldsymbol{\nabla}} = \boldsymbol{0} \text{ versus } H_1 : \bar{\boldsymbol{\nabla}} \neq \boldsymbol{0}. \tag{3.3-245}$$

Clearly, if H_0 is true, then no outliers are present in the data, and if H_1 is true, then at least one outlier is present.

If the errors $\boldsymbol{U}$ follow a normal distribution with expectation $\boldsymbol{E}\{\boldsymbol{U}\} = \boldsymbol{0}$, then we may rewrite the Gauss-Markov model (3.3-242)+(3.3-243) as

$$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\nabla}, \sigma^2 \boldsymbol{I}). \tag{3.3-246}$$

This observation model, together with the hypotheses (3.3-245), is seen to constitute a *natural problem of testing the significance of additional parameters* $\boldsymbol{\nabla}$ according to (3.2-167)+(3.2-168) or (3.2-198)+(3.2-199), depending on whether the variance factor σ^2 is known or unknown *a priori*. We will investigate both case separately in subsequent sections 3.3.1 and 3.3.2.

Example 3.1: Linear regression with a group of adjacent outliers. The following dataset (Fig. 3.1) has been analyzed by Rousseeuw and Leroy (2003, p. 26) in the context of outlier testing and robust parameter estimation. As the observations between 1964 and 1969 are seen to clearly mismatch the rest of the data, which may be approximated reasonably well by a straight line, we could take this fact into consideration by introducing additional mean shift parameters $\nabla_1, \ldots, \nabla_6$.

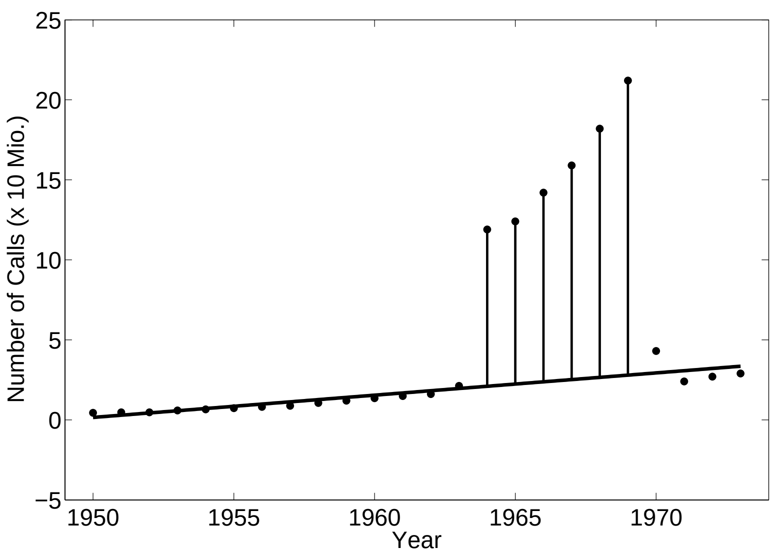

Figure 3.1. Linear regression model with six additional adjacent mean shift parameters.

The functional model may be written according to (3.3-242), that is,

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{14} \\ Y_{15} \\ Y_{16} \\ Y_{17} \\ Y_{18} \\ Y_{19} \\ Y_{20} \\ Y_{21} \\ \vdots \\ Y_{24} \end{bmatrix} = \begin{bmatrix} 1 & 1950 \\ \vdots & \vdots \\ 1 & 1963 \\ 1 & 1964 \\ 1 & 1965 \\ 1 & 1966 \\ 1 & 1967 \\ 1 & 1968 \\ 1 & 1969 \\ 1 & 1970 \\ \vdots & \vdots \\ 1 & 1973 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \nabla_1 \\ \nabla_2 \\ \nabla_3 \\ \nabla_4 \\ \nabla_5 \\ \nabla_6 \end{bmatrix} + \begin{bmatrix} U_1 \\ \vdots \\ U_{14} \\ U_{15} \\ U_{16} \\ U_{17} \\ U_{18} \\ U_{19} \\ U_{20} \\ U_{21} \\ \vdots \\ U_{24} \end{bmatrix}$$

□

3.3.1 Baarda's test

If the true value of the variance factor in (3.3-243) is known to take the *a priori* value $\bar{\sigma}^2 = \sigma_0^2$, then Proposition 3.3 guarantees that there exists a UMPI test for the outlier test problem

$$\boldsymbol{Y} \sim N(\boldsymbol{X\beta} + \boldsymbol{Z\nabla}, \sigma_0^2 \boldsymbol{I}). \tag{3.3-247}$$

and hypotheses (3.3-245). According to (3.2-158), the UMPI test is given by

$$\phi(\boldsymbol{y}) = \begin{cases} 1, & \text{if} \quad M(\boldsymbol{y}) > k_{1-\alpha}^{\chi^2(m_2)}, \\ 0, & \text{if} \quad M(\boldsymbol{y}) < k_{1-\alpha}^{\chi^2(m_2)}, \end{cases}$$

(where m_2 denotes the number of modeled outliers) with statistic

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\nabla}}' \left(\boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{Z}'\boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Z} \right) \widehat{\boldsymbol{\nabla}} / \sigma_0^2 \tag{3.3-248}$$

following from (3.2-169), and least squares estimator

$$\widehat{\boldsymbol{\nabla}} = \left(\boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{Z}'\boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Z} \right)^{-1} \boldsymbol{Z}'\boldsymbol{Y} - \left(\boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{Z}'\boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Z} \right)^{-1} \boldsymbol{Z}'\boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Y} \tag{3.3-249}$$

for the outliers $\boldsymbol{\nabla}$ corresponding to (3.2-180). This test is called **Baarda's test** (Baarda, 1967, 1968). We may rewrite (3.3-249) in the more common form

$$\widehat{\boldsymbol{\nabla}} = \left(\boldsymbol{Z}' \left(\boldsymbol{I} - \boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}' \right) \boldsymbol{Z} \right)^{-1} \boldsymbol{Z}' \left(\boldsymbol{I} - \boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}' \right) \boldsymbol{Y} \tag{3.3-250}$$

$$= \left(\boldsymbol{Z}' \boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\} \boldsymbol{Z} \right)^{-1} \boldsymbol{Z}' \boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\} \boldsymbol{Y} \tag{3.3-251}$$

where $\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}$ evidently denotes the cofactor matrix of residuals $\widetilde{\boldsymbol{U}} = \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ in the outlier-free Gauss-Markov model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{U}, \tag{3.3-252}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2 \boldsymbol{I}. \tag{3.3-253}$$

As $\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}$ is also the projector into the orthogonal space with respect to the column space of $\boldsymbol{X}$, $\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}\boldsymbol{Y} = \widetilde{\boldsymbol{U}}$ is easily verified, and (3.3-251) becomes

$$\widehat{\boldsymbol{\nabla}} = \left(\boldsymbol{Z}'\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\widetilde{\boldsymbol{U}}. \tag{3.3-254}$$

Let us now look at the case where $\boldsymbol{Z}$ contains only a single column as in (3.3-244), that is where the observation model (3.3-247) comprises a single outlier parameter ∇. Then, observing that $\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}_{ii} = \boldsymbol{Z}'\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}\boldsymbol{Z}$ and $\widetilde{U}_i = \boldsymbol{Z}'\widetilde{\boldsymbol{U}}$, (3.3-254) simplifies to

$$\widehat{\nabla} = \left(\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}_{ii}\right)^{-1}\widetilde{U}_i = \frac{\widetilde{U}_i}{r_i}, \tag{3.3-255}$$

where r_i, as the value of the i-th main diagonal element of $\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}$, denotes the partial redundancy of Y_i. Using the definition of $\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}$ and of the partial redundancy, as well as the scalar outlier estimator (3.3-255), the statistic $M(\boldsymbol{Y})$ in (3.3-248) takes the considerably simpler form

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\nabla}}'\left(\boldsymbol{Z}'\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}\boldsymbol{Z}\right)\widehat{\boldsymbol{\nabla}}/\sigma_0^2 = \widehat{\boldsymbol{\nabla}}'\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}_{ii}\widehat{\boldsymbol{\nabla}}/\sigma_0^2 = \frac{r_i\widehat{\nabla}^2}{\sigma_0^2}. \tag{3.3-256}$$

Notice that this test statistic requires that the additional parameter ∇ is estimated. However, this is not necessary as we may substitute the second part of (3.3-255) for $\widehat{\nabla}$ in (3.3-256), which yields the alternative test statistic

$$M(\boldsymbol{Y}) = \frac{\widetilde{U}_i^2}{r_i\sigma_0^2}. \tag{3.3-257}$$

It is instructive to see that this expression is nothing else than Rao's Score statistic given in Part 2 of the equivalence Proposition 3.5. To see this, rewrite (3.2-224) first in terms of matrix $\boldsymbol{X}$ and vector $\boldsymbol{Z}$ as

$$T_{RS}(\boldsymbol{Y}) = \widetilde{\boldsymbol{U}}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\widetilde{\boldsymbol{U}}/\sigma_0^2, \tag{3.3-258}$$

then substitute $\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}$, which yields

$$T_{RS}(\boldsymbol{Y}) = \widetilde{\boldsymbol{U}}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\widetilde{\boldsymbol{U}}/\sigma_0^2 = \widetilde{U}_i\left(\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}_{ii}\right)^{-1}\widetilde{U}_i/\sigma_0^2 = \frac{\widetilde{U}_i^2}{r_i\sigma_0^2}. \tag{3.3-259}$$

Therefore, both statistics $M(\boldsymbol{Y})$ in (3.3-256) and $T_{RS}(\boldsymbol{Y})$ in (3.3-259) are identical. However, Rao's Score statistic $T_{RS}(\boldsymbol{Y})$, being naturally based on the residuals of the outlier-free Gauss-Makov model, is more convenient to compute and simpler to implement than the invariance-reduced statistic $M(\boldsymbol{Y})$ based on the estimator of the outlier. We may directly apply the UMPI test (3.2-158) in terms of Rao's Score statistic to the current outlier test problem and write

$$\phi_{Baarda}(\boldsymbol{y}) = \begin{cases} 1, & \text{if} \quad \dfrac{\widetilde{u}_i^2}{r_i\sigma_0^2} > k_{1-\alpha}^{\chi^2(1)}, \\ 0, & \text{if} \quad \dfrac{\widetilde{u}_i^2}{r_i\sigma_0^2} < k_{1-\alpha}^{\chi^2(1)}, \end{cases} \tag{3.3-260}$$

which is the test proposed by Baarda (1967, p. 23). An alternative expression of Baarda's test for a single outlier (3.3-260) may be written as

$$\phi_{Baarda}(\boldsymbol{y}) = \begin{cases} 1, & \text{if} \quad \dfrac{\widetilde{u}_i}{\sqrt{r_i}\sigma_0} > k_{1-\alpha}^{N(0,1)}, \\ 0, & \text{if} \quad \dfrac{\widetilde{u}_i}{\sqrt{r_i}\sigma_0} < k_{1-\alpha}^{N(0,1)}. \end{cases} \tag{3.3-261}$$

Example 3.2: Testing for multiple outliers. The Gravity Dataset given in Appendix 6.2, which has been kindly communicated by Dr. Diethard Ruess, consists of $n = 91$ gravity differences between the old and the new Austrian gravity network (Österreichisches Schweregrundnetz, ÖSGN). To approximate this data, we use the polynomial

$$y_i = \beta_1 + \beta_2\phi_i + \beta_3\lambda_i \quad (i = 1, \ldots, 91) \tag{3.3-262}$$

as functional model, where ϕ_i denote the latitudes and λ_i the longitudes (in decimal degrees). The data is assumed to be uncorrelated and of constant standard deviation $\sigma_0 = 0.08\,mgal$. Schuh (2006b) suggested that the observations $y_3, y_6, y_{10}, y_{42}, y_{45}, y_{78}, y_{87}, y_{89}$ are outliers. To test the data against this hypothesis, we add eight additional shift parameters $\nabla_1, \ldots, \nabla_8$ to the functional model (3.3-262). The observation equations then read

$$\begin{array}{rcccccccccl}
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=1,2)\\
y_3 &=& \beta_1 &+& \beta_2\phi_3 &+& \beta_3\lambda_3 &+& \nabla_1 &+& u_1\\
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=4,5)\\
y_6 &=& \beta_1 &+& \beta_2\phi_6 &+& \beta_3\lambda_6 &+& \nabla_2 &+& u_6\\
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=7,8,9)\\
y_{10} &=& \beta_1 &+& \beta_2\phi_{10} &+& \beta_3\lambda_{10} &+& \nabla_3 &+& u_{10}\\
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=11,\ldots,41)\\
y_{42} &=& \beta_1 &+& \beta_2\phi_{42} &+& \beta_3\lambda_{42} &+& \nabla_4 &+& u_{42}\\
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=43,\ldots,44)\\
y_{45} &=& \beta_1 &+& \beta_2\phi_{45} &+& \beta_3\lambda_{45} &+& \nabla_5 &+& u_{45}\\
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=46,\ldots,77)\\
y_{78} &=& \beta_1 &+& \beta_2\phi_{78} &+& \beta_3\lambda_{78} &+& \nabla_6 &+& u_{78}\\
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=79,\ldots,86)\\
y_{87} &=& \beta_1 &+& \beta_2\phi_{87} &+& \beta_3\lambda_{87} &+& \nabla_7 &+& u_{87}\\
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=88)\\
y_{89} &=& \beta_1 &+& \beta_2\phi_{89} &+& \beta_3\lambda_{89} &+& \nabla_8 &+& u_{89}\\
y_i &=& \beta_1 &+& \beta_2\phi_i &+& \beta_3\lambda_i & & &+& u_i \quad (i=90,91)
\end{array}$$

Hence, the observation model is given as in (3.3-247) with (rounded) design matrices

$$\boldsymbol{X} = \begin{bmatrix}
1 & 48.1710 & 16.3168\\
1 & 48.2328 & 16.3373\\
1 & 48.3120 & 16.4308\\
1 & 48.7192 & 16.3033\\
1 & 48.5125 & 16.6217\\
1 & 48.6927 & 16.8707\\
1 & 48.3620 & 15.4038\\
1 & 48.1980 & 14.5280\\
1 & 48.1257 & 14.8788\\
1 & 48.1593 & 15.1010\\
1 & 48.2242 & 15.3580\\
\vdots & \vdots & \vdots\\
1 & 47.4573 & 09.6407\\
1 & 47.4307 & 09.7563\\
1 & 47.1325 & 10.1218
\end{bmatrix}, \quad
\boldsymbol{Z} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}.$$

Estimation of the mean shift parameters according to (3.3-249) yields

$$\widehat{\boldsymbol{\nabla}} = [-0.3024,\ 0.1331,\ -0.1608,\ -0.1944,\ -0.2100,\ 0.1703,\ -0.2756,\ 0.1899]',$$

and the test statistic (3.3-248) takes the value $M(\boldsymbol{y}) = 54.46$, which is larger than $k_{0.95}^{\chi^2(8)} = 15.51$. Thus, Baarda's test rejects H_0, i.e. vector of outlier parameters is significant. To carry out this test, we could also compute Rao's Score statistic (3.3-258) via the estimated residuals $\widetilde{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ based on the restricted parameter estimates

$$\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = [13.8830,\ 0.0186,\ 0.0188]'.$$

This gives the identical result $T_{RS} = 54.46$. □

3.3.2 Pope's test

If, in contrast to the situation in Section 3.3.1, the variance factor is unknown *a priori*, then we must apply Part 1 of Proposition 3.4, which states that there exists a UMPI test for the outlier test problem

$$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\nabla}, \sigma^2 \boldsymbol{I}). \tag{3.3-263}$$

with hypotheses (3.3-245). According to (3.2-159), the UMPI test is given by

$$\phi(\boldsymbol{y}) = \begin{cases} 1, \text{ if } M(\boldsymbol{y}) > k_{1-\alpha}^{F(m_2,n-m)}, \\ 0, \text{ if } M(\boldsymbol{y}) < k_{1-\alpha}^{F(m_2,n-m)}, \end{cases}$$

with statistic

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\nabla}}' \left(\boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{Z}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\right)\widehat{\boldsymbol{\nabla}}/(m_2\widehat{\sigma}^2) \tag{3.3-264}$$

following from (3.2-200). The least squares estimator for the outliers $\boldsymbol{\nabla}$, rearranged as in (3.3-251) by using the residuals $\widetilde{\boldsymbol{U}} = \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ of the non-extended model,

$$\widehat{\boldsymbol{\nabla}} = \left(\boldsymbol{Z}'\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\{\widetilde{\boldsymbol{U}}\}\boldsymbol{Y} \tag{3.3-265}$$

follow from (3.2-213) and the least squares estimator

$$\widehat{\sigma}^2 = \widehat{\boldsymbol{U}}'\widehat{\boldsymbol{U}}/(n-m) \tag{3.3-266}$$

for σ^2 from (3.2-214), where the residuals

$$\widehat{\boldsymbol{U}} = \left(\boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)\left(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\nabla}}\right) \tag{3.3-267}$$

of the extended model are given by (3.2-203). If only a single outlier is modeled, in other words, if $\boldsymbol{Z}$ has only $m_2 = 1$ columns, then (3.3-265) becomes

$$\widehat{\nabla} = \frac{\widetilde{U}_i}{r_i}, \tag{3.3-268}$$

which is identical to (3.3-255) for Baarda's test. The test statistic (3.3-264) then simplifies to

$$M(\boldsymbol{Y}) = \frac{r_i\widehat{\nabla}^2}{\widehat{\sigma}^2}. \tag{3.3-269}$$

The only difference between the statistics (3.3-257) and (3.3-269) is that the former is based on the *a priori* variance factor σ_0^2 and the latter on the estimator of the variance factor in the extended model. As the non-extended model is easier to adjust, it is more convenient to apply Rao's Score statistic whose general form (3.2-231) is simplified for the current test problem as follows:

$$T_{RS}(\boldsymbol{Y}) = \widetilde{\boldsymbol{U}}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\widetilde{\boldsymbol{U}}/\widetilde{\sigma}_{ML}^2 = \frac{\widetilde{U}_i^2}{r_i\widetilde{\sigma}_{ML}^2} = \frac{n\widetilde{U}_i^2}{(n-m)r_i\widetilde{\sigma}^2}. \tag{3.3-270}$$

The UMPI test (3.2-159) in terms of Rao's Score statistic takes the expression

$$\phi_{Pope}(\boldsymbol{y}) = \begin{cases} 1, \text{ if } \dfrac{n\widetilde{u}_i^2}{(n-m)r_i\widetilde{\sigma}^2} > k_{1-\alpha}^{F(1,n-m)}, \\ 0, \text{ if } \dfrac{n\widetilde{u}_i^2}{(n-m)r_i\widetilde{\sigma}^2} < k_{1-\alpha}^{F(1,n-m)}, \end{cases} \tag{3.3-271}$$

which was proposed in Pope (1976, p. 17). **Pope's test** is sometimes written in a square-root version of (3.3-271) as

$$\phi_{Pope}(\boldsymbol{y}) = \begin{cases} 1, \text{ if } \dfrac{\widetilde{u}_i}{\sqrt{r_i}\widetilde{\sigma}} > k_{1-\alpha}^{\tau(1,n-m)}, \\ 0, \text{ if } \dfrac{\widetilde{u}_i}{\sqrt{r_i}\widetilde{\sigma}} < k_{1-\alpha}^{\tau(1,n-m)}, \end{cases} \tag{3.3-272}$$

where $k_{1-\alpha}^{\tau(1,n-m)}$ denotes the critical value of the Tau distribution. Koch (1999, p. 303) shows that the Tau distribution is a function of the F distribution, and that both forms (3.3-271) and (3.3-272) of Pope's test are identical.

3.4 Application 2: Testing for extensions of the functional model

Suppose that we want to approximate given observables $\boldsymbol{Y}$ by a Gauss-Markov model

$$Y_i = \sum_{j=0}^{p} B_j(x_i)a_j + U_i, \quad \boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2\boldsymbol{I}, \tag{3.4-273}$$

where $B_j(x_i)$ denote base functions evaluated at known, not necessarily equidistant, nodes or locations x_i $(i = 1, \ldots, n)$ and $\boldsymbol{\beta} = [\,a_0, \ldots, a_p\,]$ the unknown parameters. Frequently used base functions B_j $(j = 0, \ldots, p)$ are, for instance, polynomials, trigonometric functions, or spherical harmonics. Let us further assume that the errors U_i $(i = 1, \ldots, n)$ are uncorrelated, homoscedastic, and normally distributed variables. If they were correlated and/or heteroscedastic with weight matrix $\boldsymbol{P}$, then we would preprocess the observation equations by decorrelation and/or homogenization as in Sect. 3.2.3.

In a practical situation with insufficient knowledge about the physical or geometrical relationship between the data and the nodes, it might not be clear how high the degree m of, for example, a polynomial expansion should be. Let us say we believe that the degree of the expansion should be specified by p_1, but we would like to check whether the base function B_{p_2} with $p_2 = p_1 + 1$ should be added to the model. One approach would be to estimate the parameters of the model (3.4-273) up to degree p_2 and to perform a significance test of the parameter a_{p_2}. If a_{p_2} turns out to be insignificant, then we carry out a new adjustment of the model (3.4-273) with degree p_1.

If we define $\boldsymbol{\beta}_1 := [\,a_0, \ldots, a_{p_1}\,]$ and $\beta_2 := a_{p_2}$, and if we let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ contain the values of the base functions evaluated at the locations, then we may rewrite (3.4-273) as

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{u}, \quad \boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2\boldsymbol{I}, \tag{3.4-274}$$

and the significance test of a_{p_2} is about the hypotheses

$$H_0 : \beta_2 = 0 \quad \text{versus} \quad H_1 : \beta_2 \neq 0.$$

As this problem is one of testing the significance of $m_2 = 1$ additional parameter β_2 with either known or unknown variance factor, the UMPI test (3.2-158) or (3.2-159) is based on either the statistic

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2/\sigma_0^2$$

or

$$M(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1\boldsymbol{X}_2\right)\widehat{\boldsymbol{\beta}}_2/(m_2\widehat{\sigma}^2)$$

by virtue of (3.2-169) and (3.2-200) in Propositions 3.4/3.5.

If however, on the grounds of prior information, we favor the model up to degree p_1 over the model up to degree p_2, then it seems more reasonable to adjust the smaller model (with p_1) first, and to estimate the additional parameters only after they have been verified to be significant. To imlement such a significance test, which is a reversed version of the significance test based on $M(\boldsymbol{Y})$, we may use Rao's Score statistic

$$T_{RS}(\boldsymbol{Y}) = \widetilde{\boldsymbol{U}}'\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}}/\sigma_0^2$$

(if the variance factor is known) or

$$T_{RS}(\boldsymbol{Y}) = \widetilde{\boldsymbol{U}}'\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\widetilde{\boldsymbol{U}}/\widetilde{\sigma}_{ML}^2$$

(if the variance factor must be estimated) as defined in (3.2-224) and (3.2-231). These statistics are based on the estimated residuals of the Gauss-Markov model (3.4-274) with the restriction $\beta_2 = 0$, which is equivalent to the ordinary Gauss-Markov model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{u}, \quad \boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2\boldsymbol{I}. \tag{3.4-275}$$

Rao's Score statistics for testing the significance of additional base functions are merely a generalization of Baarda's and Pope's statistic for testing the significance of outliers. This fact has already been pointed out by Jaeger et al. (2005, Section 5.4.5.5) in the context of testing the significance of additional transformation parameters.

Example 3.3: Testing for extension of a two-dimensional polynomial. We will demonstrate the two different approaches to significance testing of additional model parameters by further analyzing the Gravity Dataset from Example 3.2. For simplicity, the outlying observations $y_3, y_6, y_{10}, y_{42}, y_{45}, y_{78}, y_{87}, y_{89}$ will not be used, i.e. the corresponding rows are eliminated from $\boldsymbol{y}$ and $\boldsymbol{X}$. Now, we could consider

$$Y_i = a_0 + \phi_i a_1 + \lambda_i a_2 \qquad (i = 1, \ldots, 83) \tag{3.4-276}$$

as the model we favor under H_0, and the extension to degree 2

$$Y_i = a_0 + \phi_i a_1 + \lambda_i a_2 + \phi_i^2 a_3 + \phi_i \lambda_i a_4 + \lambda_i^2 a_5 \qquad (i = 1, \ldots, 83) \tag{3.4-277}$$

as an alternative model specification. This model is simply a two-dimensional polynomial version of the model in (3.4-273). We see that the null model (3.4-276) is obtained from the extended model (3.4-277) if the additional parameters a_3, a_4, a_5 are restricted to zero. To test whether the null model is misspecified, we will rewrite the extended model in the form (3.4-274) with $\sigma_0 = 0.08$ known *a priori*. In the present example, we define $\boldsymbol{\beta}_1 := [a_0,\ a_1,\ a_2]'$ and $\boldsymbol{\beta}_2 := [a_3,\ a_4,\ a_5]'$, and use the hypotheses $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0}$ versus $H_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{0}$. The design matrix $\boldsymbol{X}_1$ is obtained from $\boldsymbol{X}$ in Example 3.2 by deleting rows as described above.

$$\boldsymbol{X}_1 = \begin{bmatrix} 1 & 48.1710 & 16.3168 \\ 1 & 48.2328 & 16.3373 \\ 1 & 48.7192 & 16.3033 \\ 1 & 48.5125 & 16.6217 \\ 1 & 48.3620 & 15.4038 \\ 1 & 48.1980 & 14.5280 \\ 1 & 48.1257 & 14.8788 \\ 1 & 48.2242 & 15.3580 \\ \vdots & \vdots & \vdots \\ 1 & 47.4307 & 09.7563 \\ 1 & 47.1325 & 10.1218 \end{bmatrix}, \quad \boldsymbol{X}_2 = \begin{bmatrix} 2320.4452 & 785.9982 & 266.2391 \\ 2326.4062 & 787.9959 & 266.9085 \\ 2373.5572 & 794.2848 & 265.7987 \\ 2353.4627 & 806.3586 & 276.2798 \\ 2338.8830 & 744.9602 & 237.2781 \\ 2323.0472 & 700.2205 & 211.0628 \\ 2316.0798 & 716.0538 & 221.3797 \\ 2325.5703 & 740.6268 & 235.8682 \\ \vdots & \vdots & \vdots \\ 2249.6681 & 462.7494 & 95.1860 \\ 2221.4726 & 477.0673 & 102.4515 \end{bmatrix}$$

Notice the close similarity of this testing problem with the problem of testing for outliers in Example 3.2. The only difference is that the additional parameters $\boldsymbol{\beta}_2$ appear in every single observation equations while each mean shift parameter in $\boldsymbol{\nabla}$ affects only a single observation. Here we could analogously estimate $\widehat{\boldsymbol{\beta}}_2$ as in (3.2-170) and then compute the value of the test statistic (3.2-169). This gives

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_2 &= \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1\boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2'(\boldsymbol{I} - \boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1')\boldsymbol{y} \\ &= [-0.0287,\ -0.0170,\ -0.0044]' \end{aligned}$$

and

$$M(\boldsymbol{y}) = \widehat{\boldsymbol{\beta}}_2' \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1\boldsymbol{X}_2\right) \widehat{\boldsymbol{\beta}}_2/\sigma_0^2 = 7.74,$$

which is smaller than the critical value $k_{0.95}^{\chi^2(3)} = 7.81$. This test again yields the same result if only the parameters $\boldsymbol{\beta}_1$ of the null model are estimated and if then the corresponding residuals $\widetilde{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}_1\widetilde{\boldsymbol{\beta}}_1$ are used to determine the value of Rao's Score statistic (3.2-224). This would result in

$$\widetilde{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y} = [14.1459,\ 0.0122,\ 0.0224]$$

and

$$T_{RS}(\boldsymbol{y}) = \widetilde{\boldsymbol{u}}'\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\widetilde{\boldsymbol{u}}/\sigma_0^2 = 7.74.$$

The values in $\widetilde{\boldsymbol{\beta}}_1$ differ slightly from those in $\widetilde{\boldsymbol{\beta}}$ from Example 3.2 as a consequence of deleting observations. Hence, we could not reject the null model, i.e. the joint set of additional parameters could not be proven to be significant.

3.5 Application 3: Testing for point displacements

In Meissl (1982, Sect. 5.4) the following test problem is discussed. A leveling network has been measured twice (see Fig. 3.2), and the question is whether three of the points (7,8, and 9), which are located on a dam, changed in heights between both measurement campaigns. Point 1 has a fixed height, and the heights of the points 2, 3, 4, 5, and 6 are assumed to be unknown, but constant over time.

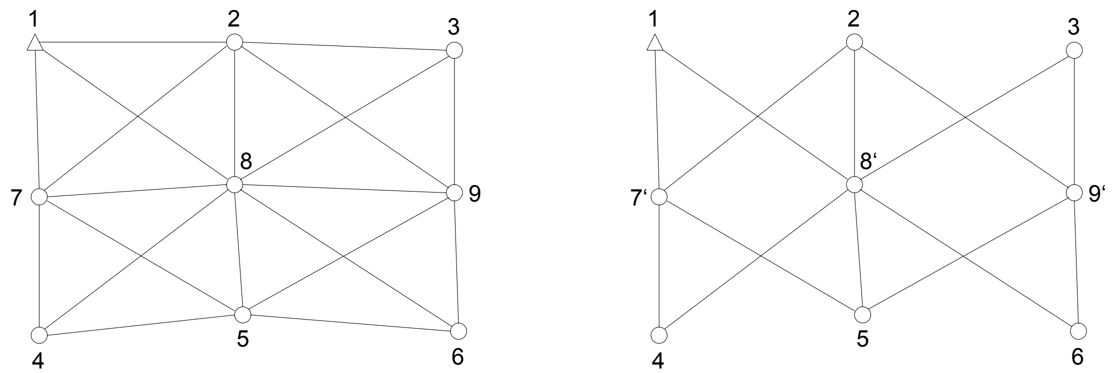

Fig. 3.2 A leveling network observed in a first campaign (left) and a second campaign (right) later in time.

The general structure of the observation model is specified as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2 \boldsymbol{I}, \tag{3.5-278}$$

where the functional model comprises $n = 34$ leveling observations $\boldsymbol{y}$ (i.e. observed height differences) with unknown accuracy σ^2 (see Appendix 6.1 for the numerical values). To allow for height displacements, the parameter vector

$$\boldsymbol{\beta} = [\,H_2,\ H_3,\ H_4,\ H_5,\ H_6,\ H_7,\ H_8,\ H_9,\ H_{7'},\ H_{8'},\ H_{9'}\,]'$$

contains the set of dam heights (H_7, H_8, H_9) regarding the first campaign and the set $(H_{7'}, H_{8'}, H_{9'})$ for the possibly different points modeled with respect to the second campaign. Let, for example, y_5 represent the observed height difference between points 2 and 8 (made in the first campaign) and y_{24} the observed height difference between points 2 and 8' (made in the second campaign). Then, the corresponding observation equations read

$$\begin{aligned} y_5 &= H_8 - H_2 + u_5, \\ y_{24} &= H_{8'} - H_2 + u_{24}, \end{aligned}$$

and the corresponding rows of the design matrix are given by

$$\begin{aligned} \boldsymbol{X}_5 &= [\,-1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0\,], \\ \boldsymbol{X}_{24} &= [\,-1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0\,]. \end{aligned}$$

The mathematical formulation of the above question, whether the points 7, 8, and 9 shifted significantly, is given by the null hypothesis $H_{7'} = H_7, H_{8'} = H_8, H_{8'} = H_8$ versus the alternative hypothesis $H_{7'} \neq H_7, H_{8'} \neq H_8, H_{8'} \neq H_8$, or in matrix notation by

$$H_0 : \boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{0} \quad \text{versus} \quad H_1 : \boldsymbol{H}\bar{\boldsymbol{\beta}} \neq \boldsymbol{0}$$

with

$$\boldsymbol{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & +1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & +1 \end{bmatrix}.$$

For this testing problem, Meissl (1982) gives the statistic

$$M(\boldsymbol{Y}) = (\boldsymbol{H}\widehat{\boldsymbol{\beta}})' \left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1} (\boldsymbol{H}\widehat{\boldsymbol{\beta}})/(3\widehat{\sigma}^2), \tag{3.5-279}$$

which is distributed as $F(3, n-m)$, because the model has three restrictions.

As an additional result however, we may conclude from Proposition 3.4(3) that this statistic leads to an optimal invariant test, because the testing problem is of the form (3.2-217) + (3.2-218) with $\boldsymbol{P} = \boldsymbol{I}$ and $\boldsymbol{w} = \boldsymbol{0}$. Consequently, using the test statistic (3.2-216) with $m_2 = 3$ and least squares estimates

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \tag{3.5-280}$$

$$\widehat{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}, \tag{3.5-281}$$

$$\widehat{\sigma}^2 = \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}/(n-m) \tag{3.5-282}$$

according to (3.2-219) - (3.2-221) leads to the UMPI test as given in (3.2-159). With the given numerical values in Appendix 6.1, the test statistic takes the value $M(\boldsymbol{y}) = 53.61$, which exceeds for instance the critical value $k_{0.95}^{F(3,23)} = 3.03$. Therefore, we conclude that the data shows significant evidence for a shift in height.

We will now demonstrate how this testing problem is reparameterized as in 3.2.1. Notice first that the functional model may be partitioned into $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$ with

$$\begin{aligned}\boldsymbol{\beta}_1 &= [H_2,\ H_3,\ H_4,\ H_5,\ H_6,\ H_7,\ H_8,\ H_9]',\\ \boldsymbol{\beta}_2 &= [H_{7'},\ H_{8'},\ H_{9'}]'\end{aligned}$$

and, regarding observations y_5 and y_{24},

$$\begin{aligned}\boldsymbol{X}_{5,1} &= [-1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0], \qquad \boldsymbol{X}_{5,2} = [0,\ 0,\ 0],\\ \boldsymbol{X}_{24,1} &= [-1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0], \qquad \boldsymbol{X}_{24,2} = [0,\ 1,\ 0].\end{aligned}$$

Now, if we use

$$\begin{aligned}\Delta H_7 &:= H_{7'} - H_7\\ \Delta H_8 &:= H_{8'} - H_8\\ \Delta H_9 &:= H_{9'} - H_9,\end{aligned}$$

i.e. the transformed quantities $\boldsymbol{H\beta}$, as parameters instead of $H_{7'}$, $H_{8'}$, and $H_{9'}$, then the new parameter vector reads

$$\boldsymbol{\beta}^{(r)} = [\boldsymbol{\beta}_1'^{(r)},\ \boldsymbol{\beta}_2'^{(r)}]' = [H_2,\ H_3,\ H_4,\ H_5,\ H_6,\ H_7,\ H_8,\ H_9,\ \Delta H_7,\ \Delta H_8,\ \Delta H_9]'$$

with components

$$\begin{aligned}\boldsymbol{\beta}_1^{(r)} &= [H_2,\ H_3,\ H_4,\ H_5,\ H_6,\ H_7,\ H_8,\ H_9]',\\ \boldsymbol{\beta}_2^{(r)} &= [\Delta H_7,\ \Delta H_8,\ \Delta H_9]'.\end{aligned}$$

Clearly, the hypotheses

$$H_0 : \bar{\boldsymbol{\beta}}_2^{(r)} = \boldsymbol{0} \quad \text{versus} \quad H_1 : \bar{\boldsymbol{\beta}}_2^{(r)} \neq \boldsymbol{0}$$

in terms of the new parameters are identical to the original hypotheses. To see how the design matrix changes, we first rewrite the observation equations, for instance, with respect to y_5 and y_{24}

$$\begin{aligned}y_5 &= H_8 - H_2 + u_5,\\ y_{24} &= H_{8'} - H_8 + H_8 - H_2 + u_{24} = H_8 + \Delta H_8 - H_2 + u_{24}.\end{aligned}$$

The corresponding rows of the design matrix with respect to the new parameters $\boldsymbol{\beta}^{(r)}$ read

$$\begin{aligned}\boldsymbol{X}_{5,1}^{(r)} &= [-1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0], \qquad \boldsymbol{X}_{5,2}^{(r)} = [0,\ 0,\ 0],\\ \boldsymbol{X}_{24,1}^{(r)} &= [-1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0], \qquad \boldsymbol{X}_{24,2}^{(r)} = [0,\ 1,\ 0].\end{aligned}$$

We see that $\boldsymbol{\beta}_1^{(r)} = \boldsymbol{\beta}_1$, $\boldsymbol{X}_{5,1}^{(r)} = \boldsymbol{X}_{5,1}$, $\boldsymbol{X}_{5,2}^{(r)} = \boldsymbol{X}_{5,2}$, and $\boldsymbol{X}_{24,2}^{(r)} = \boldsymbol{X}_{24,2}$. For the present example these equations in fact hold for all rows of the design matrix, so that the reparameterized observation model is given by

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2^{(r)} + \boldsymbol{u},\\ \boldsymbol{\Sigma} &= \sigma^2\boldsymbol{I}.\end{aligned}$$

According to Proposition 3.4-1, the UMPI test for this reparameterized testing problem is based on the value of the statistic

$$M(\boldsymbol{y}) = \widehat{\boldsymbol{\beta}}_2'^{(r)} \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_2 \right) \widehat{\boldsymbol{\beta}}_2^{(r)} / (3\widehat{\sigma}^2) \tag{3.5-283}$$

with estimates

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_2^{(r)} &= \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_2 \right)^{-1} \boldsymbol{X}_2' \left(\boldsymbol{I} - \boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)} \right) \boldsymbol{y}, && (3.5\text{-}284) \\
\widehat{\boldsymbol{u}} &= \left(\boldsymbol{I} - \boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)} \right) \left(\boldsymbol{y} - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2^{(r)} \right), && (3.5\text{-}285) \\
\widehat{\sigma}^2 &= \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}/(n-m). && (3.5\text{-}286)
\end{aligned}$$

With the given data, the displacement parameters take the values $\widehat{\boldsymbol{\beta}}_2^{(r)} = [\,0.0044, \;\; 0.0047, \;\; 0.0056\,]'$, and the test statistic becomes $M(\boldsymbol{y}) = 53.61$, which is of course the same value as determined above for 3.5-279. Notice that, as demonstrated in the proof of Proposition 3.4, the quantities $M(\boldsymbol{y})$, $\widehat{\boldsymbol{u}}$, and $\widehat{\sigma}^2$ remain unchanged by the reparameterization of the observation equations. This transformation allows us to apply Proposition 3.6 and compute the value of Rao's Score statistic by

$$T_{RS}(\boldsymbol{y}) = \widetilde{\boldsymbol{u}}'\boldsymbol{X}_2 \left(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1^{(r)}(\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_2 \right)^{-1} \boldsymbol{X}_2'\widetilde{\boldsymbol{u}}/\widetilde{\sigma}_{ML}^2, \tag{3.5-287}$$

which requires the estimates

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1'^{(r)}\boldsymbol{X}_1^{(r)})^{-1}\boldsymbol{X}_1'^{(r)}\boldsymbol{y}, && (3.5\text{-}288) \\
\widetilde{\boldsymbol{u}} &= \boldsymbol{y} - \boldsymbol{X}_1^{(r)}\widetilde{\boldsymbol{\beta}}_1, && (3.5\text{-}289) \\
\widetilde{\sigma}_{ML}^2 &= \widetilde{\boldsymbol{u}}'\widetilde{\boldsymbol{u}}/n. && (3.5\text{-}290)
\end{aligned}$$

With the given data, we obtain $T_{RS}(\boldsymbol{y}) = 29.74$. As the relationship between the statistics M and T_{RS} has been shown to be (3.2-232), we may apply this formula to check the validity of the results and to compute the critical value valid for T_{RS}. We find that

$$T_{RS}(\boldsymbol{y}) = 34 \cdot \frac{\frac{3}{23}M(\boldsymbol{y})}{1 + \frac{3}{23}M(\boldsymbol{y})} \quad \text{and} \quad k_{0.95}^{T_{RS}} = 34 \cdot \frac{\frac{3}{23}k_{0.95}^{F(3,23)}}{1 + \frac{3}{23}k_{0.95}^{F(3,23)}} = 9.63.$$

As Rao's Score statistic is determined under the assumption that H_0 is true, the null hypothesis $H_0 : \boldsymbol{\beta}_2^{(r)} = \boldsymbol{0}$ acts as a restriction on the Gauss-Markov model, and thus eliminates the parameters $\boldsymbol{\beta}_2^{(r)}$ from the reparameterized observation equations. In other words, if H_0 is true, then the Gauss-Markov model with restrictions

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2^{(r)} + \boldsymbol{u}, \\
\boldsymbol{\beta}_2^{(r)} &= \boldsymbol{0} \\
\boldsymbol{\Sigma} &= \sigma^2\boldsymbol{I}
\end{aligned}$$

is equivalent to the Gauss-Markov model

$$\boldsymbol{y} = \boldsymbol{X}_1^{(r)}\boldsymbol{\beta}_1 + \boldsymbol{u}, \quad \boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}, \tag{3.5-291}$$

for which the estimates are given by (3.5-288) - (3.5-290). The main advantage of Rao's Score statistic T_{RS} in (3.5-287) over M in (3.5-283) is that the restricted estimates $\widetilde{\boldsymbol{\beta}}_1$ are clearly less complex to compute than the unrestricted estimates $\widehat{\boldsymbol{\beta}}_2^{(r)}$. Furthermore, T_{RS} has an advantage over M in (3.5-279), because the restricted reparameterized Gauss-Markov model (3.5-291) has less unknown parameters to be estimated than the original model (3.5-278).

3.6 Derivation of an optimal test concerning the variance factor

So far we have only discussed testing problems concerning the parameters $\boldsymbol{\beta}$ of the functional model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}.$$

In the current section, we will derive an optimal procedure for testing hypotheses concerning the variance factor σ^2 in the stochastic model

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\{\boldsymbol{E}\} = \sigma^2 \boldsymbol{P}^{-1}.$$

As usual we will assume that the design matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ is known and of full rank, and that the weight matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ is positive definite. If we decompose the weight matrix into $\boldsymbol{P} = \boldsymbol{G}\boldsymbol{G}'$ as in Section 3.2.3, then the observations and the design matrix may be transformed into $\boldsymbol{Y}^{(h)} := \boldsymbol{G}'\boldsymbol{Y}$ and $\boldsymbol{X}^{(h)} := \boldsymbol{G}'\boldsymbol{X}$, where $\boldsymbol{Y}^{(h)}$ has the covariance matrix $\sigma^2\boldsymbol{I}$.

Let us now consider the problem of testing

$$H_0 : \bar{\sigma}^2 = \sigma_0^2 \quad \text{versus} \quad H_1 : \bar{\sigma}^2 > \sigma_0^2$$

in the observation model

$$\boldsymbol{Y}^{(h)} \sim N(\boldsymbol{X}^{(h)}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}).$$

Such a testing problem arises if we suspect that the given measurement accuracy σ_0^2 is too optimistic.

Proposition 3.1 allows us now to reduce $\boldsymbol{Y}^{(h)}$ to the set of minimally sufficient statistics $\boldsymbol{T}(\boldsymbol{Y}^{(h)}) := [\widehat{\boldsymbol{\beta}}', \widehat{\sigma}^2]'$ with $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'^{(h)}\boldsymbol{X}^{(h)})^{-1}\boldsymbol{X}'^{(h)}\boldsymbol{Y}^{(h)}$ and $(n-m)\widehat{\sigma}^2 = (\boldsymbol{Y}^{(h)} - \boldsymbol{X}^{(h)}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y}^{(h)} - \boldsymbol{X}^{(h)}\widehat{\boldsymbol{\beta}})$. The reduced observation model then reads

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &\sim N\left(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'^{(h)}\boldsymbol{X}^{(h)})^{-1}\right), \\ (n-m)\widehat{\sigma}^2 &\sim \chi^2(n-m)\cdot\sigma^2, \end{aligned}$$

and the hypotheses are still given by

$$H_0 : \bar{\sigma}^2 = \sigma_0^2 \quad \text{versus} \quad H_1 : \bar{\sigma}^2 > \sigma_0^2$$

This testing problem is invariant under the group G of translations

$$g\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\sigma}^2 \end{bmatrix}\right) = \begin{bmatrix} \widehat{\boldsymbol{\beta}} + \boldsymbol{a} \\ \widehat{\sigma}^2 \end{bmatrix} \tag{3.6-292}$$

with $\boldsymbol{a} \in \mathbb{R}^{m\times 1}$ acting on $\widehat{\boldsymbol{\beta}}$. Each of these tranformations will cause a change of distribution from $\boldsymbol{\beta} \sim N\left(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'^{(h)}\boldsymbol{X}^{(h)})^{-1}\right)$ to $\widehat{\boldsymbol{\beta}} + \boldsymbol{a} \sim N\left(\boldsymbol{\beta} + \boldsymbol{a}, \sigma^2(\boldsymbol{X}'^{(h)}\boldsymbol{X}^{(h)})^{-1}\right)$, while the second central moment of $\widehat{\boldsymbol{\beta}}$ and the distribution of $\widehat{\sigma}^2$ remain unaffected by these translations. Thus, the induced group $\bar{G}$ of transformations within the parameter domain is given by

$$\bar{g}\left(\begin{bmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{\beta} + \boldsymbol{a} \\ \sigma^2 \end{bmatrix}. \tag{3.6-293}$$

Evidently, neither the parameter space nor the hypotheses are changed under $\bar{G}$. Moreover, we see that $\widehat{\sigma}^2$, or more conveniently $M(\boldsymbol{Y}^{(h)}) := (n-m)\widehat{\sigma}^2$ is a maximal invariant under G. From the fact that $(n-m)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(n-m) = G(\frac{n-m}{2}, 2)$ (where G stands here for the Gamma distribution) it follows that $(n-m)\widehat{\sigma}^2 \sim G(\frac{n-m}{2}, 2\sigma^2)$. Thus, the invariant test problem

$$\begin{aligned} &M(\boldsymbol{Y}) = (n-m)\widehat{\sigma}^2 \sim G((n-m)/2, 2\sigma^2) \\ &H_0 : \bar{\sigma}^2 = \sigma_0^2 \;\text{ versus }\; H_1 : \bar{\sigma}^2 > \sigma_0^2, \end{aligned}$$

has one unknown parameter, a one-sided alternative hypothesis, and a test distribution with a monotone density ratio by virtue of Theorem 2.5-2. For this reduced testing problem, Theorem 2.4 gives the UMP test

$$\phi(\boldsymbol{y}) = \begin{cases} 1, \text{ if } M(\boldsymbol{y}) > k_{1-\alpha}^{G((n-m)/2, 2\sigma_0^2)}, \\ 0, \text{ if } M(\boldsymbol{y}) < k_{1-\alpha}^{G((n-m)/2, 2\sigma_0^2)}, \end{cases} = \begin{cases} 1, \text{ if } (n-m)\widehat{\sigma}^2/\sigma_0^2 > k_{1-\alpha}^{\chi^2(n-m)}, \\ 0, \text{ if } (n-m)\widehat{\sigma}^2/\sigma_0^2 < k_{1-\alpha}^{\chi^2(n-m)}, \end{cases} \tag{3.6-294}$$

which is UMPI for the original problem of testing $H_0 : \bar{\sigma}^2 = \sigma_0^2$ against $H_1 : \bar{\sigma}^2 > \sigma_0^2$ in the observation model $\boldsymbol{Y}^{(h)} \sim N(\boldsymbol{X}^{(h)}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$. This test is the same as given in Koch (1999, Section 4.2.4), but it was shown here in addition that (3.6-294) is optimal within the class of all tests with equal power in each direction of $\boldsymbol{\beta}$.

4 Applications of Misspecification Tests in Generalized Gauss-Markov models

4.1 Introduction

In this section, we will look at testing problems, where parameters of the distribution or stochastic model are hypothesized. In each of these problems, the null hypothesis states that the errors are distributed as $\boldsymbol{E} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$, whereas the alternative hypothesis represents one of the following types of model errors:

1. the errors are not uncorrelated, but correlated through an autoregressive process (Section 4.2);
2. the errors are not homoscedastic, but the variance changes according to an additive variance component (Section 4.3);
3. the errors are not normally distributed (Section 4.4).

In contrast to the testing problems in Section 3, where only the parameters of the functional model were subject to hypothesis, there will be no UMPI tests under the above scenarios, as these problems cannot be reduced to a single maximal invariant statistic. However, they can be given a suitable mathematical expression such that, at least, reductions to Likelihood Ratio or Rao's Score statistics are feasible. Although these statistics will then not be optimal in a strictly mathematical sense, one may hope that they will remain *reasonable tools* for detecting the above model errors. In this sense, Engle (1984), used the term **diagnostic** in his insightful review article.

The first step towards deriving such diagnostics for each of the above cases is to extend the mathematical model by estimable parameters that allow the data to be correlated, heteroscedastic, or non-normally distributed. H_0 then restricts these additional parameters to zero and thereby reduces the extended model to an ordinary normal Gauss-Markov model, while under H_1, these parameters remain unrestricted. Then, the Likelihood Ratio test compares the value of the likelihood function evaluated at the unrestricted ML estimate with the value of the likelihood function obtained at the restricted estimate. Therefore, if the restriction under H_0 reduces the likelihood significantly, the test statistic will take a large value and thus indicate that H_0 should be rejected. On the other hand, if the restriction is reflected by the given data, then the restricted likelihood will be close to the unrestricted likelihood, which will probably cause the statistic to take an insignificant value. Rao's Score test, on the other hand, does not require computing the unrestricted ML estimates (which may be computationally expensive if, for instance, variance components are present), because it measures the extent to which the scores (i.e. the first partial derivatives of the log-likelihood function) differ from zero if the restricted estimates are used.

Although the testing procedures based on Rao's Score statistic will be computationally feasible and statistically powerful, we will have to deal with one inconvenience: in contrast to the testing problems in Section 3, where the distribution of Rao's Score statistic was always exact as a strictly monotonically increasing function of a χ^2- or F-distribution, there will be no exact test distributions available for the problems above. Instead we will have to confine ourselves to using approximative test distributions, that is to critical values which are valid *asymptotically*. Therefore, the testing problems stated in the current section should be applied only when a large number (i.e. at least 100) of observations is given. It is beyond the scope of this thesis to explain in detail the definition of *asymptotic distribution* as this would require a rather lengthy discussion of various types of *convergence of random variables*. The interested reader shall therefore be referred to Godfrey (1988, p. 13-15), who gives a proof and more technical explanation of the following proposition.

Proposition 4.1. *Suppose that $Y_1, \ldots, Y_n$ are identically distributed observations with true density function in*

$$\mathcal{F} = \{f(\boldsymbol{y}; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$$

and consider the problem of testing $H_0 : \boldsymbol{H}\bar{\boldsymbol{\theta}} = \boldsymbol{w}$ versus $H_1 : \boldsymbol{H}\bar{\boldsymbol{\theta}} \neq \boldsymbol{w}$, where $\boldsymbol{H}$ is a known $(p \times u)$-matrix with $p < u$ and full row rank p. Then, under H_0, the asymptotic distribution of the LR and the equivalent RS statistic is given by:

$$-2\left(\mathcal{L}(\widetilde{\boldsymbol{\theta}}; \boldsymbol{Y}) - \mathcal{L}(\widehat{\boldsymbol{\theta}}; \boldsymbol{Y})\right) \approx \boldsymbol{\mathcal{S}}'(\widetilde{\boldsymbol{\theta}}; \boldsymbol{Y})\boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\theta}}; \boldsymbol{Y})\boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\theta}}; \boldsymbol{Y}) \overset{a}{\sim} \chi^2(p), \qquad (4.1\text{-}295)$$

where $\widetilde{\boldsymbol{\theta}}$ is the ML estimator for $\boldsymbol{\theta}$ restricted by H_0 and $\widehat{\boldsymbol{\theta}}$ the unrestricted ML estimator.

4.2 Application 5: Testing for autoregressive correlation

We will now discuss a linear model

$$Y_i = \boldsymbol{X}_i\boldsymbol{\beta} + E_i \quad (i = 1, \ldots, n), \tag{4.2-296}$$

where $\boldsymbol{X}_i$ represents the i-th row of the design matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ (with rank$\boldsymbol{X} = m$) and where $\boldsymbol{\beta} \in \mathbb{R}^{m\times 1}$ denotes a vector of unknown functional parameters. We will assume that each error E_i follows a **first-order autoregressive error model/process**, or **AR(1) model/process**, defined by

$$E_i = \alpha E_{i-1} + U_i \quad (i = 1, \ldots, n), \tag{4.2-297}$$

in which α is an unknown parameter. Let the stochastic model of the errors U_i be given by

$$\boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2\boldsymbol{I}. \tag{4.2-298}$$

If normally distributed error variables E_i are to have expectations, variances, and covariances that are independent of the index i (that are, for instance, independent of the absolute time or location), then we must require the AR(1) model (4.2-297) to be weakly stationary up to second order. This requirement imposes certain restrictions on the specification of the numerical value for α.

Let us explore the nature of these restrictions by investigating stationarity with repect to the first moment. We may rewrite (4.2-297) as

$$E_i \;=\; \alpha(\alpha E_{i-2} + U_{i-1}) + U_i = \alpha(\cdots(\alpha(\alpha E_0 + U_1) + U_2) + \ldots) + U_i = \alpha^{i-1}U_1 + \alpha^{i-2}U_2 + \ldots + U_i.$$

Taking the expected value of both sides of this equation yields

$$E\{E_i\} = \alpha^{i-1}E\{U_1\} + \alpha^{i-2}E\{U_2\} + \ldots + E\{U_i\}.$$

Under the condition of constant mean $\mu = E\{U_1\} = E\{U_2\} = \ldots = E\{U_n\}$, we obtain

$$E\{E_i\} = \mu\left(\alpha^{i-1} + \alpha^{i-2} + \ldots + 1\right) = \begin{cases} \mu\left(\dfrac{1-\alpha^i}{1-\alpha}\right), & \text{for } \alpha \neq 1 \\ \mu \cdot i, & \text{for } \alpha = 1. \end{cases}.$$

We see from this result that the error variables E_i have constant mean $\mu_E = E\{E_1\} = E\{E_2\} = \ldots = E\{E_n\}$ only if $\mu = 0$ holds, because only this condition eliminates the dependence of $E\{E_i\}$ on the index i. In other words, an AR(1) model is **weakly stationary up to order one** if the mean μ of the independent errors U_i $(i = 1, \ldots, n)$ is zero.

Similarly, we obtain for the covariance of two variables separated by distance h

$$\begin{aligned} E\{E_iE_{i+h}\} \;&=\; E\{\left(\alpha^{i-1}U_1 + \alpha^{i-2}U_2 + \ldots + U_i\right)\left(\alpha^{i+h-1}U_1 + \alpha^{i+h-2}U_2 + \ldots + U_{i+h}\right)\} \\ &=\; E\{\alpha^{2i+h-2}U_1^2 + \alpha^{2i+h-4}U_2^2 + \ldots + \alpha^h U_i^2 + 2\alpha^{2i+h-3}U_1U_2 + \ldots\}. \end{aligned} \tag{4.2-299}$$

Notice that, due to the stochastic model (4.2-298), all the expected values of the mixed terms $2\alpha^{2i+h-3}U_1U_2, \ldots$, i.e. all the covariances between any two distinct error variables U_i, U_j $(i \neq j)$, are zero. Furthermore, (4.2-298) expresses that the variances of all the U_i are constant with $\sigma^2 = E\{U_1^2\} = E\{U_2^2\} = \ldots = E\{U_n^2\}$. With this, we may rewrite (4.2-299) as

$$E\{E_iE_{i+h}\} = \sigma^2\left(\alpha^{2i+h-2} + \alpha^{2i+h-4} + \ldots + \alpha^h\right) = \begin{cases} \sigma^2\alpha^h\left(\dfrac{1-\alpha^{2i}}{1-\alpha^2}\right), & \text{for } \alpha \neq 1 \\ \sigma^2 \cdot i, & \text{for } \alpha = 1. \end{cases}. \tag{4.2-300}$$

It is seen that $E\{E_iE_{i+h}\}$ is independent of i only if $\sigma^2 = 0$. However, in that case all the errors E_i would be exactly zero, which is a nonsensical requirement if the E_i represent measurement errors. On the other hand, if $\sigma^2 > 0$, then we could resort to the following type of asymptotic stationarity. If $|\alpha| < 1$, then $\lim_{i\to\infty}\sigma^2\alpha^h\left(\frac{1-\alpha^{2i}}{1-\alpha^2}\right) = \frac{\sigma^2\alpha^h}{1-\alpha^2}$. It follows that the variance $(h = 0)$ and the covariance $(h > 0)$ of the errors E_i are asymptotically independent of the index i.

In summary, the conditions for the AR(1) model (4.2-298) to be **asymptotically weakly stationary up to order two** are given by (1) $E\{U_i\} = 0$ for all $i = 1, \ldots, n$, and (2) $|\alpha| < 1$. If these conditions are presumed, the covariance matrix of the error variables $\boldsymbol{E}$ is easily deduced from the limit $\frac{\sigma^2\alpha^h}{1-\alpha^2}$ of (4.2-300). If we let h run from $0, \ldots, n-1$, we obtain

$$\boldsymbol{\Sigma}\{\boldsymbol{E}\} := \sigma^2 \boldsymbol{Q}_\alpha = \sigma^2 \cdot \frac{1}{1-\alpha^2} \cdot \begin{bmatrix} 1 & \alpha & \cdots & \alpha^{n-1} \\ \alpha & 1 & \cdots & \alpha^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{n-1} & \alpha^{n-2} & \cdots & 1 \end{bmatrix}, \tag{4.2-301}$$

where we take into account that the cofactor matrix $\boldsymbol{Q}_\alpha$ of the autoregressive errors $\boldsymbol{E}$ depends on the unknown parameter α. According to Peracchi (2001, p. 277), the weight matrix is given by

$$\boldsymbol{P}_\alpha = \boldsymbol{Q}_\alpha^{-1} = \begin{bmatrix} 1 & -\alpha & & & \\ -\alpha & 1+\alpha^2 & \ddots & & \\ & \ddots & \ddots & & \\ & & & 1+\alpha^2 & -\alpha \\ & & & -\alpha & 1 \end{bmatrix}. \tag{4.2-302}$$

The weight matrix is tridiagonal and positive definite. We see that the errors $\boldsymbol{E}$ are uncorrelated only if $\alpha = 0$. If $\alpha \neq 0$, then we say that the errors have **serial correlations**. This term was coined due to the fact that autoregressive error models have traditionally been applied to time series.

To give a geodetic example, Schuh (1996, Chap. 3) used an *autoregressive moving average* (ARMA) model, which is a generalization of the AR(1) model, to obtain the covariance matrix of satellite data that have a band-limited error spectrum. Such data may be treated as a time series recorded along the satellite's orbit. Typically, such time series comprise very large numbers of observations, which justifies the use of asymptotic covariance matrices as in (4.2-301).

Now, the observation model (4.2-296) + (4.2-301) under the additional assumption of normally distributed errors may be summarized as

$$\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{P}_\alpha^{-1}). \tag{4.2-303}$$

To keep the observation model as simple as possible, we might want to check the data whether serial correlation is significant or not. Such a test may be based on the hypotheses

$$H_0 : \bar{\alpha} = 0 \quad \text{versus} \quad H_1 : \bar{\alpha} \neq 0. \tag{4.2-304}$$

For this purpose, we should clearly apply Rao's Score statistic, because it avoids estimation of the parameter α. To see this, recall that Rao's Score statistic is based on the residuals of the Gauss-Markov model with restriction H_0. As this restriction reduces the observation model (4.2-303) to the simpler model $\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$, we will only have to estimate $\boldsymbol{\beta}$ and σ^2. To derive Rao's Score statistic, we need to determine the log-likelihood score and the information with respect to the extended parameterization in (4.2-303), and then to evaluate these quantities at the estimates of the restricted model.

As the observations are now correlated and heteroscedastic, the joint density and consequently the log-likelihood do not factorize into the product of identical univariate densities. However, such a factorization is made possible quite easily through a transformation of the functional model (4.2-296) as in Section 3.2.3. If we decompose the weight matrix into $\boldsymbol{P}_\alpha = \boldsymbol{G}_\alpha' \boldsymbol{G}_\alpha$ with

$$\boldsymbol{G}_\alpha = \begin{bmatrix} \sqrt{1-\alpha^2} & 0 & & & \\ -\alpha & 1 & \ddots & & \\ & \ddots & \ddots & & \\ & & & & 0 \\ & & & -\alpha & 1 \end{bmatrix}, \tag{4.2-305}$$

which may be verified directly by multiplication, and transform the observations and the design matrix by $\boldsymbol{Y}^{(h)} = \boldsymbol{G}_\alpha \boldsymbol{Y}$ and $\boldsymbol{X}^{(h)} = \boldsymbol{G}_\alpha \boldsymbol{X}$, then the homogenized observations $\boldsymbol{Y}^{(h)}$ will have unity weight matrix. This transformation, which may also be written component-by-component as

$$Y_i^{(h)} = \begin{cases} \sqrt{1-\alpha^2} Y_i, & \text{for } i=1, \\ Y_i - \alpha Y_{i-1}, & \text{for } i=2,\dots,n \end{cases}, \quad \boldsymbol{X}_i^{(h)} = \begin{cases} \sqrt{1-\alpha^2} \boldsymbol{X}_i, & \text{for } i=1, \\ \boldsymbol{X}_i - \alpha \boldsymbol{X}_{i-1}, & \text{for } i=2,\dots,n \end{cases}, \tag{4.2-306}$$

is also known as the **Prais-Winsten transformation** (see, for instance, Peracchi, 2001, p. 278).

Now we may use the factorized form of the log-likelihood in (2.5-96), that is

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y}) = \ln \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{Y_i^{(h)} - \boldsymbol{X}_i^{(h)}\boldsymbol{\beta}}{\sigma}\right)^2 \right\} = \ln(2\pi\sigma^2)^{-n/2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_1^{(h)} - \boldsymbol{X}_1^{(h)}\boldsymbol{\beta}\right)^2.$$

Reversing the transformation (4.2-306), we obtain

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y}) &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left(\sqrt{1-\alpha^2}Y_1 - \sqrt{1-\alpha^2}\boldsymbol{X}_1\boldsymbol{\beta}\right)^2 \\ &\quad - \frac{1}{2\sigma^2}\sum_{i=2}^{n}(Y_i - \alpha Y_{i-1} - (\boldsymbol{X}_i - \alpha\boldsymbol{X}_{i-1})\boldsymbol{\beta})^2 \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1-\alpha^2}{2\sigma^2}(Y_1 - \boldsymbol{X}_1\boldsymbol{\beta})^2 - \frac{1}{2\sigma^2}\sum_{i=2}^{n}\left((Y_i - \boldsymbol{X}_i\boldsymbol{\beta}) - \alpha(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta})\right)^2. \end{aligned}$$

Before determining the first partial derivatives, it will be convenient to expand the second sum and to move the parameter α outside the summuation, which gives.

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y}) &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2 + \frac{\alpha^2}{2\sigma^2}(Y_1 - \boldsymbol{X}_1\boldsymbol{\beta})^2 \\ &\quad + \frac{\alpha}{\sigma^2}\sum_{i=2}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta}) - \frac{\alpha^2}{2\sigma^2}\sum_{i=2}^{n}(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta})^2. \end{aligned}$$

The first partial derivatives of the log-likelihood function give the log-likelihood scores (2.5-111), that is

$$\begin{aligned} \mathcal{S}_{\beta_j}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\beta_j} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})X_{i,j} - \frac{\alpha^2}{\sigma^2}(Y_1 - \boldsymbol{X}_1\boldsymbol{\beta})X_{1,j} - \frac{\alpha}{\sigma^2}\sum_{i=2}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})X_{i-1,j} \\ &\quad - \frac{\alpha}{\sigma^2}\sum_{i=2}^{n}(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta})X_{i,j} + \frac{\alpha^2}{\sigma^2}\sum_{i=2}^{n}(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta})X_{i-1,j}, \\ \mathcal{S}_{\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2 - \frac{\alpha^2}{2\sigma^4}(Y_1 - \boldsymbol{X}_1\boldsymbol{\beta})^2 \\ &\quad - \frac{\alpha}{\sigma^4}\sum_{i=2}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta}) + \frac{\alpha^2}{2\sigma^4}\sum_{i=2}^{n}(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta})^2, \\ \mathcal{S}_{\alpha}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\alpha} = \frac{\alpha}{\sigma^2}(Y_1 - \boldsymbol{X}_1\boldsymbol{\beta})^2 + \frac{1}{\sigma^2}\sum_{i=2}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta}) \\ &\quad - \frac{\alpha}{\sigma^2}\sum_{i=2}^{n}(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta})^2. \end{aligned}$$

The Hessian (2.5-112) which comprises the second partial derivatives follows to be

$$\begin{aligned} \mathcal{H}_{\beta_j\beta_k}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial^2\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\beta_j\partial\beta_k} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}X_{i,k}X_{i,j} + \frac{\alpha^2}{\sigma^2}X_{1,k}X_{1,j} + \frac{\alpha}{\sigma^2}\sum_{i=2}^{n}X_{i,k}X_{i-1,j} \\ &\quad + \frac{\alpha}{\sigma^2}\sum_{i=2}^{n}X_{i-1,k}X_{i,j} - \frac{\alpha^2}{\sigma^2}\sum_{i=2}^{n}X_{i-1,k}X_{i-1,j}, \\ \mathcal{H}_{\beta_j\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial^2\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\beta_j\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})X_{i,j} + \frac{\alpha^2}{\sigma^4}(Y_1 - \boldsymbol{X}_1\boldsymbol{\beta})X_{1,j} + \frac{\alpha}{\sigma^4}\sum_{i=2}^{n}(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})X_{i-1,j} \\ &\quad + \frac{\alpha}{\sigma^4}\sum_{i=2}^{n}(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta})X_{i,j} - \frac{\alpha^2}{\sigma^4}\sum_{i=2}^{n}(Y_{i-1} - \boldsymbol{X}_{i-1}\boldsymbol{\beta})X_{i-1,j}, \end{aligned}$$

$$\begin{aligned}
\mathcal{H}_{\beta_j\alpha}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial^2\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\beta_j\partial\alpha} = -\frac{2\alpha}{\sigma^2}(Y_1-\boldsymbol{X}_1\boldsymbol{\beta})X_{1,j} - \frac{1}{\sigma^2}\sum_{i=2}^{n}(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})X_{i-1,j}\\
&\quad -\frac{1}{\sigma^2}\sum_{i=2}^{n}(Y_{i-1}-\boldsymbol{X}_{i-1}\boldsymbol{\beta})X_{i,j} + \frac{2\alpha}{\sigma^2}\sum_{i=2}^{n}(Y_{i-1}-\boldsymbol{X}_{i-1}\boldsymbol{\beta})X_{i-1,j},\\
\mathcal{H}_{\sigma^2\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial^2\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\sigma^2\partial\sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})^2 + \frac{\alpha^2}{\sigma^6}(Y_1-\boldsymbol{X}_1\boldsymbol{\beta})^2\\
&\quad +\frac{2\alpha}{\sigma^6}\sum_{i=2}^{n}(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})(Y_{i-1}-\boldsymbol{X}_{i-1}\boldsymbol{\beta}) - \frac{\alpha^2}{\sigma^6}\sum_{i=2}^{n}(Y_{i-1}-\boldsymbol{X}_{i-1}\boldsymbol{\beta})^2,\\
\mathcal{H}_{\sigma^2\alpha}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial^2\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\sigma^2\partial\alpha} = -\frac{\alpha}{\sigma^4}(Y_1-\boldsymbol{X}_1\boldsymbol{\beta})^2 - \frac{1}{\sigma^4}\sum_{i=2}^{n}(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})(Y_{i-1}-\boldsymbol{X}_{i-1}\boldsymbol{\beta})\\
&\quad +\frac{\alpha}{\sigma^4}\sum_{i=2}^{n}(Y_{i-1}-\boldsymbol{X}_{i-1}\boldsymbol{\beta})^2,\\
\mathcal{H}_{\alpha\alpha}(\boldsymbol{\theta};\boldsymbol{Y}) &:= \frac{\partial^2\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial\alpha\partial\alpha} = \frac{1}{\sigma^2}(Y_1-\boldsymbol{X}_1\boldsymbol{\beta})^2 - \frac{1}{\sigma^2}\sum_{i=2}^{n}(Y_{i-1}-\boldsymbol{X}_{i-1}\boldsymbol{\beta})^2.
\end{aligned}$$

This gives for the information matrix (2.5-113) in terms of the errors $E_i = Y_i - \boldsymbol{X}_i\boldsymbol{\beta}$

$$\begin{aligned}
\mathcal{I}_{\beta_j\beta_k}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\beta_j\beta_k}(\boldsymbol{\theta};\boldsymbol{Y})\} = \frac{1}{\sigma^2}\sum_{i=1}^{n}X_{i,k}X_{i,j} - \frac{\alpha^2}{\sigma^2}X_{1,k}X_{1,j} - \frac{\alpha}{\sigma^2}\sum_{i=2}^{n}X_{i,k}X_{i-1,j}\\
&\quad -\frac{\alpha}{\sigma^2}\sum_{i=2}^{n}X_{i-1,k}X_{i,j} + \frac{\alpha^2}{\sigma^2}\sum_{i=2}^{n}X_{i-1,k}X_{i-1,j},\\
\mathcal{I}_{\beta_j\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\beta_j\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y})\} = \frac{1}{\sigma^4}\sum_{i=1}^{n}E\{E_i\}X_{i,j} - \frac{\alpha^2}{\sigma^4}E\{E_i\}X_{1,j} - \frac{\alpha}{\sigma^4}\sum_{i=2}^{n}E\{E_i\}X_{i-1,j}\\
&\quad -\frac{\alpha}{\sigma^4}\sum_{i=2}^{n}E\{E_{i-1}\}X_{i,j} + \frac{\alpha^2}{\sigma^4}\sum_{i=2}^{n}E\{E_{i-1}\}X_{i-1,j},\\
\mathcal{I}_{\beta_j\alpha}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\beta_j\alpha}(\boldsymbol{\theta};\boldsymbol{Y})\} = \frac{2\alpha}{\sigma^2}E\{E_1\}X_{1,j} + \frac{1}{\sigma^2}\sum_{i=2}^{n}E\{E_i\}X_{i-1,j} + \frac{1}{\sigma^2}\sum_{i=2}^{n}E\{E_{i-1}\}X_{i,j}\\
&\quad -\frac{2\alpha}{\sigma^2}\sum_{i=2}^{n}E\{E_{i-1}\}X_{i-1,j},\\
\mathcal{I}_{\sigma^2\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\sigma^2\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y})\} = -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n}E\{E_i^2\} - \frac{\alpha^2}{\sigma^6}E\{E_i^2\} - \frac{2\alpha}{\sigma^6}\sum_{i=2}^{n}E\{E_iE_{i-1}\}\\
&\quad +\frac{\alpha^2}{\sigma^6}\sum_{i=2}^{n}E\{E_{i-1}^2\},\\
\mathcal{I}_{\sigma^2\alpha}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\sigma^2\alpha}(\boldsymbol{\theta};\boldsymbol{Y})\} = \frac{\alpha}{\sigma^4}E\{E_1^2\} + \frac{1}{\sigma^4}\sum_{i=2}^{n}E\{E_iE_{i-1}\} - \frac{\alpha}{\sigma^4}\sum_{i=2}^{n}E\{E_{i-1}^2\},\\
\mathcal{I}_{\alpha\alpha}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\alpha\alpha}(\boldsymbol{\theta};\boldsymbol{Y})\} = -\frac{1}{\sigma^2}E\{E_1^2\} + \frac{1}{\sigma^2}\sum_{i=2}^{n}E\{E_{i-1}^2\}.
\end{aligned}$$

Evaluation of the scores at the restricted maximum likelihood estimates $\widetilde{\boldsymbol{\theta}} = [\widetilde{\boldsymbol{\beta}}' \;\; \widetilde{\sigma}^2_{ML} \;\; \widetilde{\alpha}]'$ yields

$$\begin{aligned}
\mathcal{S}_{\beta_j}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \frac{1}{\widetilde{\sigma}^2_{ML}}\sum_{i=1}^{n}(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})X_{i,j} = 0,\\
\mathcal{S}_{\sigma^2}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= -\frac{n}{2\widetilde{\sigma}^2_{ML}} + \frac{1}{2\widetilde{\sigma}^4_{ML}}\sum_{i=1}^{n}(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})^2 = 0,\\
\mathcal{S}_{\alpha}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \frac{1}{\widetilde{\sigma}^2_{ML}}\sum_{i=2}^{n}(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})(Y_{i-1}-\boldsymbol{X}_{i-1}\widetilde{\boldsymbol{\beta}}) = n\widetilde{\rho}_1
\end{aligned}$$

where we used the orthogonality relation $(\boldsymbol{Y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})\boldsymbol{X}^{(j)}=0$ (where $\boldsymbol{X}^{(j)}$ denotes the j-th column of the design matrix $\boldsymbol{X}$), the maximum likelihood variance estimator $\widetilde{\sigma}^2_{ML}=\sum_{i=1}^n(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})^2/n$, and the autocorrelation estimator $\widetilde{\rho}_1=\sum_{i=2}^n(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})(Y_{i-1}-\boldsymbol{X}_{i-1}\widetilde{\boldsymbol{\beta}})/\sum_{i=1}^n(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})^2$ for lag $h=1$.

Under H_0 the observations are uncorrelated. Hence, the information matrix at $\widetilde{\boldsymbol{\theta}}$ is given by

$$\begin{aligned}
\mathcal{I}_{\beta_j\beta_k}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \frac{1}{\widetilde{\sigma}^2_{ML}}\sum_{i=1}^n X_{i,k}X_{i,j},\\
\mathcal{I}_{\beta_j\sigma^2}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \mathcal{I}_{\beta_j\alpha}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})=\mathcal{I}_{\sigma^2\alpha}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})=0,\\
\mathcal{I}_{\sigma^2\sigma^2}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \frac{n}{2\widetilde{\sigma}^4_{ML}},\\
\mathcal{I}_{\alpha\alpha}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= n-2.
\end{aligned}$$

To obtain the total score vector $\boldsymbol{\mathcal{S}}_\theta(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})$ and the entire information matrix $\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\theta}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})$, we must express all the individual components in matrix notation. In particular, we set up the $(m\times 1)$-subvector $\boldsymbol{\mathcal{S}}_\beta(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})=\boldsymbol{0}$ consisting of all the entries $\mathcal{S}_{\beta_j}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})$ $(j=1,\ldots,m)$, and the $(m\times m)$-submatrix $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})=\frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}'\boldsymbol{X}$ containing all the elements $\mathcal{I}_{\beta_j\beta_k}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})$. Then, we obtain for Rao's Score statistic

$$\begin{aligned}
T_{RS} &= \boldsymbol{\mathcal{S}}'(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2_{ML},\widetilde{\alpha};\boldsymbol{Y})\boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2_{ML},\widetilde{\alpha};\boldsymbol{Y})\boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2_{ML},\widetilde{\alpha};\boldsymbol{Y})\\
&= \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ n\widetilde{\rho}_1\end{bmatrix}'\begin{bmatrix}\frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{n}{2\widetilde{\sigma}^4_{ML}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & n-2\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ n\widetilde{\rho}_1\end{bmatrix}.
\end{aligned}$$

The information matrix is block-diagonal, hence the three parameter groups are independent. We see immediately that

$$T_{RS}=\frac{n^2}{n-2}\widetilde{\rho}_1^{\,2}\approx n\widetilde{\rho}_1^{\,2}, \tag{4.2-307}$$

where the approximation will certainly be sufficient for large n. This statistic is called the **Durbin-Watson statistic** (see, for example, Krämer and Sonnberger, 1986, p. 17) and asymptotically follows a $\chi^2(1)$-distribution under H_0 (see Proposition 4.1). This statistic measures the size of the absolute value of the empirical autocorrelation for lag $h=1$, which is a reasonable procedure in light of the fact that the cofactor matrix $\boldsymbol{Q}_\alpha$ is dominated by the cofactors α on the secondary diagonal. In fact, if $|\alpha|$ is much smaller than 1, then the cofactors for the higher lags decay quite rapidly according to α^h.

The procedure explained here to obtain a significance test of the parameter α of an AR(1) model may be generalized quite easily to an AR(p) model, which is defined as

$$E_i=\alpha_1E_{i-1}+\ldots+\alpha_pE_{i-p}+U_i\quad(i=1,\ldots,n). \tag{4.2-308}$$

If a joint significant test is desired with respect to the parameters $\alpha_1,\ldots,\alpha_p$ of this model, then the log-likelihood function, the log-likelihood score, and the information matrix may be obtained according to the derivations above. The main difference is that the empirical autocorrelations up to lag p will appear in the score vector. In that case, Rao's Score statistic can be shown to take the form

$$T_{RS}=n\sum_{j=1}^p\widetilde{\rho}_j^{\,2}, \tag{4.2-309}$$

which is also known as the **Portmanteau** or **Box-Pierce statistic**, which is asymptotically distributed as $\chi^2(p)$ (see, for instance, Peracchi, 2001, p. 367).

Example 4.1: Linear regression model with AR(1) errors. Let us inspect the power of the exact Durbin-Watson test (4.2-307) for detecting correlations following an AR(1) model by performing a Monte Carlo simulation. For this purpose, suppose that the functional model

$$Y_i = \boldsymbol{X}_i\boldsymbol{\beta} + E_i$$

is represented by a straight line through the origin with $\boldsymbol{X}_i = [1,\ i]$ and $\boldsymbol{\beta} = [0,\ 0.01]'$ $(i = 1, \ldots, 1000)$. The error variables are assumed to follow the autoregressive model

$$E_i = \alpha E_{i-1} + U_i$$

with $\sigma^2 = 1$. The observations are uncorrelated under the null hypothesis, that is $H_0 : \bar{\alpha} = 0$, whereas the alternative hypothesis allows for correlations according to a non-zero value of α, that is $H_1 : \bar{\alpha} \neq 0$. Depending on the given problem, the deterministic model could be far more complex than the one we adopted here, but the main purpose of this example is to explain how the power of a test is examined empirically.

Now, we would intuitively expect the power of a reasonable diagnostic to increase as the value of α becomes larger. To verify our intuition, we generate $M = 1000$ vectors with dimension (1000×1) of standard-normally distributed random numbers. These vectors $\boldsymbol{u}^{(j)}$ $(j = 1, \ldots, M)$ represent random realizations of the uncorrelated error variables U_i. Then we transform these errors into possibly correlated errors e_i by using the set of parameter values $\alpha = \{0,\ -0.05,\ -0.1,\ -0.15\}$. The value $\alpha = 0$ represents the case of no correlations, that is of a true H_0. The values -0.05 and -0.1 reflect moderate negative correlation, while the value -0.15 produces a strong negative correlation (see Fig. 4.1). Now, adding the above linear trend to the error vectors $\boldsymbol{e}^{(j)}$ yields the data vectors $\boldsymbol{y}^{(j)}$ $(j = 1, \ldots, M)$, which will be tested in the following.

The first step in the testing procedure described in the current section consists in estimating the M sets of line parameters by $\widetilde{\boldsymbol{\beta}}^{(j)} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, where potential correlations are neglected. From these estimates the residual vectors follow to be $\widetilde{\boldsymbol{u}}^j = \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(j)}$. Then, we need to compute the autocorrelation for lag $h = 1$ with respect to each of these residual vectors, that is $\widetilde{\rho}_1^{(j)} = \sum_{i=2}^{n} \widetilde{u}_i^j \widetilde{u}_{i-1}^j / \sum_{i=1}^{n} (\widetilde{u}_i^j)^2$. This quantity is simply a standardized empirical version of the theoretical covariance $\sigma^2\alpha^h/(1-\alpha^2)$ of the AR(1) error model (see Fig. 4.1, right). Finally, we determine the values of the Durbin-Watson statistic $T_{RS}^{(j)} = n(\widetilde{\rho}_1^{(j)})^2$ and compare these to the critical value of the $\chi^2(1)$-distribution for instance at level 0.05. To obtain empirical values of the power function evaluated at $\alpha = \{0,\ -0.05,\ -0.1,\ -0.15\}$, we only need to count how many times the test rejects H_0, i.e. determine $N_R = \#\left(T_{RS}^{(j)} > k_{0.95}^{\chi^2(1)}\right)$ $(j = 1, \ldots, M)$ and divide this number by the number M of trials. Then, the ratio N_R/M is an estimate for the probability $\Pi(\alpha)$ that H_0 is rejected, which we expected to depend on the value α of the autoregressive parameter. The results of this simualtion are summarized in the following table:

α	0	-0.05	-0.1	-0.15
N_R/M	0.058	0.414	0.898	0.998

We see that for $\alpha = 0$, the level of the test is reproduced reasonably well, and that the power of the test is almost 1 for $\alpha = -0.15$. To obtain the finer details of the empirical power function, we would only have to extend this simulation to a finer grid of α-values. Naturally, we could also improve on the accuracy of the power estimates by generating a higher number M of random samples $\boldsymbol{U}^{(j)}$, say $M = 10000$.

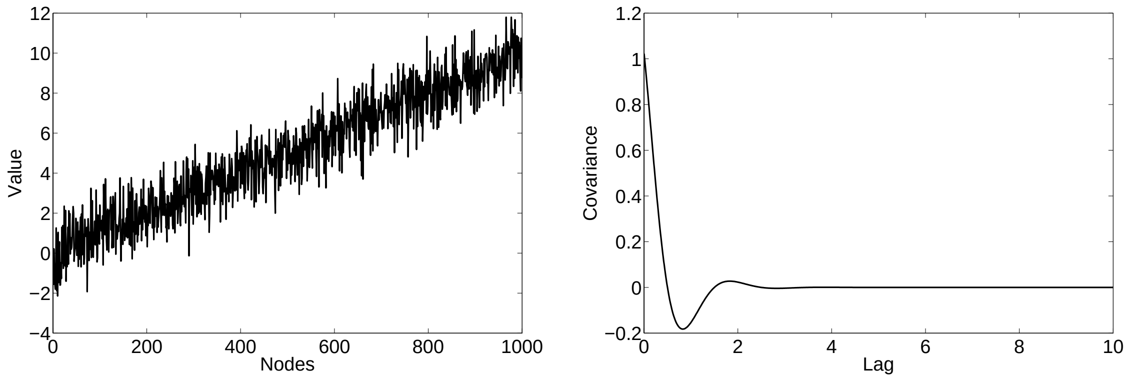

Fig. 4.1 A single realization of an AR(1) error process with parameter $\alpha = -0.15$ superimposed on a linear trend (left); theoretical covariance function for the same process.

4.3 Application 6: Testing for overlapping variance components

Suppose that observations $\boldsymbol{Y}$ are approximated by a linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E} \tag{4.3-310}$$

with normally distributed zero-mean errors $\boldsymbol{E}$, known design matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ of full rank, and parameters $\boldsymbol{\beta} \in \mathbb{R}^{m\times 1}$. Regarding the stochastic model, we assume that the covariance matrix is written as

$$\boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2 \boldsymbol{I} + \gamma \boldsymbol{V}, \tag{4.3-311}$$

where $\boldsymbol{V}$ is a positive known diagonal matrix. From these specifications it follows that the errors are uncorrelated, and that they possibly have unequal variances

$$\sigma_i^2 := \sigma_{U_i}^2 = \sigma^2 + \gamma V_{ii}. \tag{4.3-312}$$

The model (4.3-310) + (4.3-311) represents a **Gauss-Markov model with two overlapping variance components $\boldsymbol{\sigma^2}$ and $\boldsymbol{\gamma}$**. This model is a particular version of the general Gauss-Markov model with k unknown variance and covariance components, defined in Koch (1999, Equation 3.268). A test about the hypotheses

$$H_0 : \bar{\gamma} = 0 \ \text{ versus } \ H_1 : \bar{\gamma} > 0. \tag{4.3-313}$$

is most conveniently based on Rao's Score statistic, which avoids computation of the additional parameter γ. Recall that Rao's Score statistic requires that the first and second partial derivatives with respect to all unknown parameters, i.e. the score vector and the information matrix, are evaluated at the maximum likelihood estimates under the restriction H_0. Therefore, we must first determine the log-likelihood function for $\boldsymbol{Y}$. Using the definition of the density for the univariate normal distribution with parameters $\boldsymbol{\theta} = [\boldsymbol{\beta}' \ \ \sigma^2 \ \ \gamma]$, we find

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y}) &= \ln \prod_{i=1}^n f(Y_i;\boldsymbol{\theta}) = \sum_{i=1}^n \ln f(Y_i;\boldsymbol{\theta}) \\
&= \sum_{i=1}^n \ln(2\pi\sigma_i^2)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{Y_i - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma_i}\right)^2\right\} \\
&= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^n \ln(\sigma^2 + \gamma V_{ii}) - \frac{1}{2}\sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2}{\sigma^2 + \gamma V_{ii}}.
\end{aligned}$$

The first partial derivatives of the log-likelihood function give the scores

$$\mathcal{S}_{\beta_j}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \beta_j} = \sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})X_{i,j}}{\sigma^2 + \gamma V_{ii}}, \tag{4.3-314}$$

$$\mathcal{S}_{\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \sigma^2} = -\frac{1}{2}\sum_{i=1}^n \frac{1}{\sigma^2 + \gamma V_{ii}} + \frac{1}{2}\sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2}{(\sigma^2 + \gamma V_{ii})^2}, \tag{4.3-315}$$

$$\mathcal{S}_{\gamma}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \gamma} = -\frac{1}{2}\sum_{i=1}^n \frac{V_{ii}}{\sigma^2 + \gamma V_{ii}} + \frac{1}{2}\sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2}{(\sigma^2 + \gamma V_{ii})^2} V_{ii}. \tag{4.3-316}$$

The second partial derivatives follow to be

$$\mathcal{H}_{\beta_j\beta_k}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^n \frac{-X_{i,j}X_{i,k}}{\sigma^2 + \gamma V_{ii}}, \tag{4.3-317}$$

$$\mathcal{H}_{\beta_j\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \beta_j \partial \sigma^2} = -\sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})X_{i,j}}{(\sigma^2 + \gamma V_{ii})^2}, \tag{4.3-318}$$

$$\mathcal{H}_{\beta_j\gamma}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \beta_j \partial \gamma} = \frac{1}{2}\sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})X_{i,j}}{(\sigma^2 + \gamma V_{ii})^2} V_{ii}, \tag{4.3-319}$$

$$\mathcal{H}_{\sigma^2\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \sigma^2 \partial \sigma^2} = \frac{1}{2}\sum_{i=1}^n \frac{1}{(\sigma^2 + \gamma V_{ii})^2} - \sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2}{(\sigma^2 + \gamma V_{ii})^3}, \tag{4.3-320}$$

$$\mathcal{H}_{\sigma^2\gamma}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \sigma^2 \partial \gamma} = \frac{1}{2}\sum_{i=1}^n \frac{V_{ii}}{(\sigma^2 + \gamma V_{ii})^2} - \sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2}{(\sigma^2 + \gamma V_{ii})^3} V_{ii}, \tag{4.3-321}$$

$$\mathcal{H}_{\gamma\gamma}(\boldsymbol{\theta};\boldsymbol{Y}) := \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \gamma \partial \gamma} = \frac{1}{2}\sum_{i=1}^n \frac{V_{ii}^2}{(\sigma^2 + \gamma V_{ii})^2} - \sum_{i=1}^n \frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2}{(\sigma^2 + \gamma V_{ii})^3} V_{ii}^2. \tag{4.3-322}$$

Using the Markov conditions $E\{E_i\} = 0$ and $E\{E_i^2\} = \sigma_i^2$ about the errors $E_i = Y_i - \boldsymbol{X}_i\boldsymbol{\beta}$, we obtain for the components of the information matrix:

$$\begin{aligned}
\mathcal{I}_{\beta_j\beta_k}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\beta_j\beta_k}(\boldsymbol{\theta};\boldsymbol{Y})\} = \sum_{i=1}^{n} \frac{X_{i,j}X_{i,k}}{\sigma^2+\gamma V_{ii}},\\
\mathcal{I}_{\beta_j\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\beta_j\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y})\} = \sum_{i=1}^{n} \frac{E\{Y_i-\boldsymbol{X}_i\boldsymbol{\beta}\}X_{i,j}}{(\sigma^2+\gamma V_{ii})^2} = 0,\\
\mathcal{I}_{\beta_j\gamma}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\beta_j\gamma}(\boldsymbol{\theta};\boldsymbol{Y})\} = -\frac{1}{2}\sum_{i=1}^{n} \frac{E\{Y_i-\boldsymbol{X}_i\boldsymbol{\beta}\}X_{i,j}}{(\sigma^2+\gamma V_{ii})^2}V_{ii} = 0,\\
\mathcal{I}_{\sigma^2\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\sigma^2\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y})\} = -\frac{1}{2}\sum_{i=1}^{n} \frac{1}{(\sigma^2+\gamma V_{ii})^2} + \sum_{i=1}^{n} \frac{E\{(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})^2\}}{(\sigma^2+\gamma V_{ii})^3} = \sum_{i=1}^{n} \frac{1}{2(\sigma^2+\gamma V_{ii})^2},\\
\mathcal{I}_{\sigma^2\gamma}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\sigma^2\gamma}(\boldsymbol{\theta};\boldsymbol{Y})\} = -\frac{1}{2}\sum_{i=1}^{n} \frac{V_{ii}}{(\sigma^2+\gamma V_{ii})^2} + \sum_{i=1}^{n} \frac{E\{(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})^2\}}{(\sigma^2+\gamma V_{ii})^3}V_{ii} = \sum_{i=1}^{n} \frac{V_{ii}}{2(\sigma^2+\gamma V_{ii})^2},\\
\mathcal{I}_{\gamma\gamma}(\boldsymbol{\theta};\boldsymbol{Y}) &:= -E\{\mathcal{H}_{\gamma\gamma}(\boldsymbol{\theta};\boldsymbol{Y})\} = -\frac{1}{2}\sum_{i=1}^{n} \frac{V_{ii}^2}{(\sigma^2+\gamma V_{ii})^2} + \sum_{i=1}^{n} \frac{E\{(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})^2\}}{(\sigma^2+\gamma V_{ii})^3}V_{ii}^2 = \sum_{i=1}^{n} \frac{V_{ii}^2}{2(\sigma^2+\gamma V_{ii})^2}.
\end{aligned}$$

Evaluation of the scores at the restricted maximum likelihood estimates $\widetilde{\boldsymbol{\theta}} = [\,\widetilde{\boldsymbol{\beta}}' \;\; \widetilde{\sigma}^2_{ML} \;\; \widetilde{\gamma}\,]'$ under H_0 yields

$$\mathcal{S}_{\beta_j}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) = \sum_{i=1}^{n} \frac{(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})X_{i,j}}{\widetilde{\sigma}^2_{ML}} = \frac{\sum_{i=1}^{n}\widetilde{U}_iX_{i,j}}{\widetilde{\sigma}^2_{ML}} = 0, \tag{4.3-323}$$

$$\mathcal{S}_{\sigma^2}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) = -\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\widetilde{\sigma}^2_{ML}} + \frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})^2}{\widetilde{\sigma}^4_{ML}} = -\frac{n}{2\widetilde{\sigma}^2_{ML}} + \frac{\sum_{i=1}^{n}\widetilde{U}_i^2}{2\widetilde{\sigma}^4_{ML}} = 0, \tag{4.3-324}$$

$$\begin{aligned}
\mathcal{S}_{\gamma}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= -\frac{1}{2}\sum_{i=1}^{n}\frac{V_{ii}}{\widetilde{\sigma}^2_{ML}} + \frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i-\boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})^2}{\widetilde{\sigma}^4_{ML}}V_{ii} = -\frac{1}{2}\sum_{i=1}^{n}\frac{V_{ii}}{\widetilde{\sigma}^2_{ML}} + \frac{1}{2}\sum_{i=1}^{n}\frac{\widetilde{U}_i^2V_{ii}}{\widetilde{\sigma}^4_{ML}}\\
&= \frac{1}{2\widetilde{\sigma}^2_{ML}}\sum_{i=1}^{n}\left(\frac{\widetilde{U}_i^2}{\widetilde{\sigma}^2_{ML}}-1\right)V_{ii} = \frac{1}{2\widetilde{\sigma}^2_{ML}}\sum_{i=1}^{n}\bar{U}_iV_{ii} = \frac{1}{2\widetilde{\sigma}^2_{ML}}\boldsymbol{1}'\boldsymbol{V}\bar{\boldsymbol{U}},
\end{aligned} \tag{4.3-325}$$

where we may use the standardized residuals

$$\bar{U}_i := \frac{\widetilde{U}_i^2}{\widetilde{\sigma}^2_{ML}} - 1 = \frac{\widetilde{U}_i^2}{\widetilde{\boldsymbol{U}}'\widetilde{\boldsymbol{U}}/n} - 1 \tag{4.3-326}$$

and the $n \times 1$-vector $\boldsymbol{1}$ of ones to allow for a more compact notation. Evaluation of the information at $\widetilde{\boldsymbol{\theta}}$ yields

$$\begin{aligned}
\mathcal{I}_{\beta_j\beta_k}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \frac{\sum_{i=1}^{n}X_{i,j}X_{i,k}}{\widetilde{\sigma}^2_{ML}} = \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}'\boldsymbol{X},\\
\mathcal{I}_{\beta_j\sigma^2}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \mathcal{I}_{\beta_j\gamma}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) = 0,\\
\mathcal{I}_{\sigma^2\sigma^2}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \frac{n}{2\widetilde{\sigma}^4_{ML}},\\
\mathcal{I}_{\sigma^2\gamma}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \frac{\sum_{i=1}^{n}V_{ii}}{2\widetilde{\sigma}^4_{ML}} = \frac{\mathrm{tr}\boldsymbol{V}}{2\widetilde{\sigma}^4_{ML}} = \frac{1}{2\widetilde{\sigma}^4_{ML}}\boldsymbol{1}'\boldsymbol{V}\boldsymbol{1},\\
\mathcal{I}_{\gamma\gamma}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) &= \frac{\sum_{i=1}^{n}V_{ii}^2}{2\widetilde{\sigma}^4_{ML}} = \frac{\mathrm{tr}\boldsymbol{V}\boldsymbol{V}}{2\widetilde{\sigma}^4_{ML}} = \frac{1}{2\widetilde{\sigma}^4_{ML}}\boldsymbol{1}'\boldsymbol{V}\boldsymbol{V}\boldsymbol{1}.
\end{aligned}$$

For computational purposes, the expressions for $\mathcal{I}_{\sigma^2\gamma}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})$ and $\mathcal{I}_{\gamma\gamma}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})$ in terms of the trace of $\boldsymbol{V}$ or $\boldsymbol{V}\boldsymbol{V}$ are more convenient to use. However, to construct the test statistic itself, we will use these equations in matrix notation.

Now we obtain for Rao's Score statistic

$$
\begin{aligned}
T_{RS} &= \boldsymbol{\mathcal{S}}'(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}^2_{ML}, \widetilde{\gamma}; \boldsymbol{Y}) \boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}^2_{ML}, \widetilde{\gamma}; \boldsymbol{Y}) \boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}^2_{ML}, \widetilde{\gamma}; \boldsymbol{Y}) \\
&= \begin{bmatrix} 0 \\ \mathbf{0} \\ \frac{1}{2\widetilde{\sigma}^4_{ML}} \mathbf{1}'\boldsymbol{V}\bar{\boldsymbol{U}} \end{bmatrix}' \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2_{ML}} \boldsymbol{X}'\boldsymbol{X} & 0 & 0 \\ 0 & \frac{n}{2\widetilde{\sigma}^4_{ML}} & \frac{1}{2\widetilde{\sigma}^4_{ML}} \mathbf{1}'\boldsymbol{V}\mathbf{1} \\ 0 & \frac{1}{2\widetilde{\sigma}^4_{ML}} \mathbf{1}'\boldsymbol{V}\mathbf{1} & \frac{1}{2\widetilde{\sigma}^4_{ML}} \mathbf{1}'\boldsymbol{V}\boldsymbol{V}\mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ 0 \\ \frac{1}{2\widetilde{\sigma}^4_{ML}} \mathbf{1}'\boldsymbol{V}\bar{\boldsymbol{U}} \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} \mathbf{0} & 0 & \bar{\boldsymbol{U}}'\boldsymbol{V}\mathbf{1} \end{bmatrix} \begin{bmatrix} 2\widetilde{\sigma}^2_{ML} \boldsymbol{X}'\boldsymbol{X} & 0 & 0 \\ 0 & n & \mathbf{1}'\boldsymbol{V}\mathbf{1} \\ 0 & \mathbf{1}'\boldsymbol{V}\mathbf{1} & \mathbf{1}'\boldsymbol{V}\boldsymbol{V}\mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ 0 \\ \mathbf{1}'\boldsymbol{V}\bar{\boldsymbol{U}} \end{bmatrix} \\
&= \frac{1}{2} \bar{\boldsymbol{U}}'\boldsymbol{V}\mathbf{1} \left(\mathbf{1}'\boldsymbol{V}\boldsymbol{V}\mathbf{1}\right)^{(-1)} \mathbf{1}'\boldsymbol{V}\bar{\boldsymbol{U}}.
\end{aligned}
$$

As all the components of log-likelihood score for the unrestricted parameters vanish, we only need to find the Schur complement of the block $\mathbf{1}'\boldsymbol{V}\boldsymbol{V}\mathbf{1}$, which is simple to compute due to the block-diagonality of the submatrix with respect to the parameter groups $\boldsymbol{\beta}$ and σ^2. We obtain

$$
\begin{aligned}
\left(\mathbf{1}'\boldsymbol{V}\boldsymbol{V}\mathbf{1}\right)^{(-1)} &= \left(\mathbf{1}'\boldsymbol{V}\boldsymbol{V}\mathbf{1} - \begin{bmatrix} \mathbf{0} & \mathbf{1}'\boldsymbol{V}\mathbf{1} \end{bmatrix} \begin{bmatrix} 2\widetilde{\sigma}^2_{ML} \boldsymbol{X}'\boldsymbol{X} & \mathbf{0} \\ \mathbf{0} & n \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{1}'\boldsymbol{V}\mathbf{1} \end{bmatrix} \right)^{-1} \\
&= \left(\mathbf{1}'\boldsymbol{V}\boldsymbol{V}\mathbf{1} - \frac{1}{n} (\mathbf{1}'\boldsymbol{V}\mathbf{1})(\mathbf{1}'\boldsymbol{V}\mathbf{1}) \right)^{-1} = \left(\mathbf{1}'\boldsymbol{V}\boldsymbol{V}\mathbf{1} - \mathbf{1}'\boldsymbol{V}\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{V}\mathbf{1} \right)^{-1} \\
&= \left(\mathbf{1}'\boldsymbol{V} \left[\boldsymbol{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' \right] \boldsymbol{V}\mathbf{1} \right)^{-1}.
\end{aligned}
$$

Rao's Score statistic for testing the significance of a single additive variance component γ finally reads

$$
T_{RS} = \frac{1}{2} \bar{\boldsymbol{U}}'\boldsymbol{V}\mathbf{1} \left(\mathbf{1}'\boldsymbol{V} \left[\boldsymbol{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' \right] \boldsymbol{V}\mathbf{1} \right)^{-1} \mathbf{1}'\boldsymbol{V}\bar{\boldsymbol{U}}. \tag{4.3-327}
$$

As this test statistic has an approximate $\chi^2(1)$-distribution by virtue of Proposition 4.1, Rao's Score test is given by

$$
\phi_{AVC}(\boldsymbol{y}) = \begin{cases} 1, \text{ if } T_{RS} > k_{1-\alpha}^{\chi^2(1)}, \\ 0, \text{ if } T_{RS} < k_{1-\alpha}^{\chi^2(1)}, \end{cases} \tag{4.3-328}
$$

Example 4.2: Significance testing of a distance-dependent variance component. Koch (1981) considered an additive heteroscedasticity model of the form

$$
\sigma_i^2 = a + b \cdot s_i^2 \tag{4.3-329}
$$

to explain the variances σ_i^2 of distance measurements $s_1, \ldots, s_n$ by a constant part a and a distance-dependent part γs_i^2. If we further assume the observations to be uncorrelated, then the stochastic model follows to be

$$
\boldsymbol{\Sigma}\{\boldsymbol{Y}\} = \sigma^2 \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \gamma \cdot \begin{bmatrix} s_1^2 & 0 & 0 & 0 \\ 0 & s_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & s_n^2 \end{bmatrix}.
$$

We might desire a test of the null hypothesis that measured distances have constant accuracy σ^2 against the alternative hypothesis that there is a significant distance-dependent variance component superposing the variance σ^2. These hypotheses take the form

$$
H_0 : \bar{\gamma} = 0 \text{ versus } H_1 : \bar{\gamma} \neq 0.
$$

Under the assumption of normally distributed observations, the resulting observation model reads

$$
\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} + \gamma \boldsymbol{V}).
$$

4.4 Application 7: Testing for non-normality of the observation errors

Let us consider the linear functional model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}, \tag{4.4-330}$$

where $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ denotes a known design matrix of full rank and $\boldsymbol{\beta} \in \mathbb{R}^{m \times 1}$ a vector of unknown functional parameters. We will assume for now that the errors $\boldsymbol{U}$ are uncorrelated and homoscedastic according to the stochastic model

$$\boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2 \boldsymbol{I}. \tag{4.4-331}$$

We have used such a Gauss-Markov model, for instance, in Section 3 to obtain the UMPI statistic M for testing linear restrictions $\boldsymbol{H}\bar{\boldsymbol{\beta}} = \boldsymbol{w}$. The exact χ^2- or F-distribution of this test statistic has been derived from the basic premise that the error variables are normally distributed. This normality assumption becomes even more evident if we recall that we used the normal density/likelihood function to derive of the Likelihood Ratio and Rao's Score statistic for that problem. If the errors do not follow a normal distribution, then these tests are not reliable anymore, because the exact distributions of these test statistics (and therefore the critical values) will be at least inaccurate, and the likelihood function will be misspecified.

Therefore, if we have serious doubts about the normality of the error variables, then we should test this assumption. In this section, we will investigate a test of normality which fits into the framework of parametric testing problems, and which may be derived conveniently on the basis of Rao's Score statistic.

Let us start by recalling that the density of a univariate normal distribution is characterized by four parameters: a variable mean μ, a variable variance σ^2, a constant skewness $\gamma_1 = 0$ (reflecting symmetry about the mean), and a constant kurtosis $\gamma_2 = 0$ (indicating a mesokurtic shape). The mean is then identical to the first moment

$$\mu_1 = \int_{-\infty}^{\infty} x f(x) dx$$

and the variance identical to the second central moment

$$\mu_2 = \int_{-\infty}^{\infty} (x - \mu_1)^2 f(x) dx,$$

whereas the skewness and kurtosis are based on the third and fourth central moments

$$\mu_3 = \int_{-\infty}^{\infty} (x - \mu_1)^3 f(x) dx$$

and

$$\mu_4 = \int_{-\infty}^{\infty} (x - \mu_1)^4 f(x) dx$$

through the relations

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} \tag{4.4-332}$$

and

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 \tag{4.4-333}$$

(see Stuart and Ord, 2003, p. 74 and 109). A natural idea is now to estimate the skewness and kurtosis from the given data and to compare these estimates with the values ascribed to the normal distribution. On the one hand, if the empirical skewness turns out to be significantly smaller/larger than 0, then the errors will have a non-symmetrical distribution with a lower/upper tail that is heavier than for a normal distribution. On the other hand, if the empirical kurtosis is significantly smaller/larger than 0, then the errors will have a platykurtic/leptokurtic distribution with a flatter/sharper top (see Stuart and Ord, 2003, p. 109). Unfortunately, it is not clear how large these deviations from 0 must be to indicate significant non-normality, because we do not know the probability distribution of the estimators for γ_1 and γ_2.

Nevertheless, this problem of testing $H_0 : \gamma_1 = 0, \gamma_2 = 0$ versus $H_1 : \gamma_1 \neq 0, \gamma_2 \neq 0$ may be tackled in a quite convenient way by considering **Pearson's collection of distributions** $\mathcal{W}_P$. The density function of each univariate distribution within $\mathcal{W}_P$ satisfies the differential equation

$$\frac{d}{du} \ln f(u; c_0, c_1, c_2) = \frac{c_1 - u}{c_0 - c_1 u + c_2 u^2} \qquad (u \in \mathbb{R}), \tag{4.4-334}$$

where the parameters c_0, c_1, and c_2 determine the shape of the density function f, and where u is a quantity measured about its mean, such as an error $u_i = y_i - \boldsymbol{X}_i\boldsymbol{\beta}$ in (4.4-330). We will see later that the density function of the centered normal distribution (with $\mu = 0$) satisfies (4.4-334) for $c_0 = \sigma^2$, $c_1 = 0$, and $c_2 = 0$. Furthermore, the parameters c_1 and c_2 correspond to the skewness and kurtosis through the relations

$$c_1 = \frac{\gamma_1(\gamma_2 + 6)\sqrt{\mu_2}}{10\gamma_2 - 12\gamma_1^2 + 12} \tag{4.4-335}$$

and

$$c_2 = \frac{3\gamma_1^2 - 2\gamma_2}{10\gamma_2 - 12\gamma_1^2 + 12} \tag{4.4-336}$$

(see Equation 6.4 in Stuart and Ord, 2003, p. 217), so that the problem of testing of $H_0 : \gamma_1 = 0, \gamma_2 = 0$ versus $H_1 : \gamma_1 \neq 0, \gamma_2 \neq 0$ is equivalent to testing $H_0 : c_1 = 0, c_2 = 0$ versus $H_1 : c_1 \neq 0, c_2 \neq 0$.

Let us now determine the general solution of (4.4-334). Integration of (4.4-334) yields

$$\ln f(u; c_0, c_1, c_2) + k = \int \frac{c_1 - u}{c_0 - c_1 u + c_2 u^2}\, du,$$

which we may rewrite as

$$\ln f(u; c_0, c_1, c_2) + k = g(u; c_0, c_1, c_2) \qquad (k \in \mathbb{R}),$$

where k denotes the integration constant and

$$g(u; c_0, c_1, c_2) := \int \frac{c_1 - u}{c_0 - c_1 u + c_2 u^2}\, du \tag{4.4-337}$$

an antiderivative. Now, using $\exp(\ln u) = u$, the general solution of (4.4-334) follows to be

$$f(u; c_0, c_1, c_2) = \exp\{g(u; c_0, c_1, c_2) - k\} = \exp\{-k\} \cdot \exp\{g(u; c_0, c_1, c_2)\} =: k^* \cdot \exp\{g(u; c_0, c_1, c_2)\}. \tag{4.4-338}$$

The integration constant k^* is determined by standardizing the area under f to 1, which yields

$$1 = \int_{-\infty}^{+\infty} f(u; c_0, c_1, c_2) du = \int_{-\infty}^{+\infty} k^* \cdot \exp g(u; c_0, c_1, c_2) du = k^* \cdot \int_{-\infty}^{+\infty} \exp g(u; c_0, c_1, c_2) du.$$

Now, substituting $k^* = 1/\int_{-\infty}^{+\infty} \exp g(u; c_0, c_1, c_2)\, du$ into (4.4-338) leads to

$$f(u; c_0, c_1, c_2) = \frac{\exp g(u; c_0, c_1, c_2)}{\int_{-\infty}^{+\infty} \exp g(u; c_0, c_1, c_2)\, du} \tag{4.4-339}$$

as the standardized solution of (4.4-334).

Pearson's collection of distributions comprises a large number of standard distributions and, as indicated earlier, the normal distribution is its most prominent member. Some particularly useful members of $\mathcal{W}_P$ are summarized by the following proposition.

Proposition 4.2. *The following univariate distributions are members of Pearson's collection of distribution:*

1. *Centered normal distrubution* $N(0, \sigma^2)$,
2. *Gamma distribution* $G(b, p)$,
3. *Beta distribution* $B(\alpha, \beta)$,
4. *Student distribution* $t(k)$.

Proof. 1. To see that the density function $f(u; \sigma^2)$ of $N(0, \sigma^2)$ satisfies (4.4-334), set $c_0 = \sigma^2, c_1 = c_2 = 0$ in (4.4-337), for which (4.4-339) becomes

$$f(u; \sigma^2) = \frac{\exp\left(\int \frac{-u}{\sigma^2}\, du\right)}{\int_{-\infty}^{+\infty} \exp\left(\int \frac{-u}{\sigma^2}\, du\right) du} = \frac{\exp\left(\frac{-u^2}{2\sigma^2}\right)}{\int_{-\infty}^{+\infty} \exp\left(\frac{-u^2}{2\sigma^2}\right) du}. \tag{4.4-340}$$

The integral in the denominator is solved by using

$$\int_0^{+\infty} \exp\left(-a^2 x^2\right) dx = \frac{\sqrt{\pi}}{2a} \quad (a > 0) \tag{4.4-341}$$

(see Bronstein and Semendjajew, 1991, p. 66; Integral 3), where in the given case $a^2 = 1/2\sigma^2$ is positive as a consequence of the fact that $\sigma^2 > 0$ by definition. Also note that integration on $(-\infty, +\infty)$ doubles the value of (4.4-341) because of $\exp\left(-a^2(-x)^2\right) = \exp\left(-a^2x^2\right)$. Thus it follows that

$$\int_{-\infty}^{+\infty} \exp\left(\frac{-u^2}{2\sigma^2}\, du\right) du = \sqrt{\pi} \cdot \sqrt{2\sigma^2}, \tag{4.4-342}$$

from which the density function

$$f(u; \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2\right\} \tag{4.4-343}$$

of the centered normal distribution is obtained.

Proofs for 2.-4. are found, for instance, in Stuart and Ord (2003, Chap. 6). □

Now, the log-likelihood function may be determined from the densities (4.4-339) with respect to the errors $u_i = y_i - \boldsymbol{X}_i\boldsymbol{\beta}$ $(i = 1, \ldots, n)$. If these error variables are assumed to be independently distributed, then the joint density (as a function of $\boldsymbol{y}$ with additional parameters $\boldsymbol{\beta}$) may be factorized as

$$f(\boldsymbol{y}; \boldsymbol{\beta}, c_0, c_1, c_2) = \prod_{i=1}^{n} \frac{\exp g(u_i; c_0, c_1, c_2)}{\int_{-\infty}^{+\infty} \exp g(u_i; c_0, c_1, c_2)\, du_i}. \tag{4.4-344}$$

Notice that the value of the integral in the denominator is only a function of c_0, c_1, and c_2, not of $\boldsymbol{\beta}$, because u_i acts there only as an integration variable. Defining the parameter vector as $\boldsymbol{\theta} := [\boldsymbol{\beta}', c_0, c_1, c_2]'$, the log-likelihood function follows to be

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \ln \prod_{i=1}^{n} \frac{\exp g(u_i; c_0, c_1, c_2)}{\int_{-\infty}^{+\infty} \exp g(u_i; c_0, c_1, c_2)\, du_i} = \sum_{i=1}^{n} \left(g(u_i; c_0, c_1, c_2) - \ln \int_{-\infty}^{+\infty} \exp g(u_i; c_0, c_1, c_2)\, du_i \right).$$

Taking the first partial derivatives with respect to the functional parameters β_j ($j = 1, \ldots, m$) yields

$$\begin{aligned}\mathcal{S}_{\beta_j}(\boldsymbol{\theta};\boldsymbol{y}) &:= \frac{\partial \mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial g(u_i;c_0,c_1,c_2)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \int \frac{c_1 - u_i}{c_0 - c_1 u_i + c_2 u_i^2}\, du_i \\ &= \sum_{i=1}^{n} \frac{\partial u_i}{\partial \beta_j} \frac{\partial}{\partial u_i} \int \frac{c_1 - u_i}{c_0 - c_1 u_i + c_2 u_i^2}\, du_i = -\sum_{i=1}^{n} X_{i,j} \frac{c_1 - u_i}{c_0 - c_1 u_i + c_2 u_i^2},\end{aligned}$$

which defines the random variable

$$\mathcal{S}_{\beta_j}(\boldsymbol{\theta};\boldsymbol{Y}) = -\sum_{i=1}^{n} X_{i,j} \frac{c_1 - U_i}{c_0 - c_1 U_i + c_2 U_i^2}. \tag{4.4-345}$$

Peracchi (2001, p. 365) points out an elegant way to derive the score with respect to the parameters c_j ($j = 0, \ldots, 2$), which shall be explained here in greater detail. Applying the partial derivative to both terms within the sum of the log-likelihood function gives

$$\begin{aligned}\mathcal{S}_{c_j}(\boldsymbol{\theta};\boldsymbol{y}) &:= \frac{\partial \mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})}{\partial c_j} = \sum_{i=1}^{n} \left(\frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_j} - \frac{\partial}{\partial c_j} \ln \int_{-\infty}^{+\infty} \exp g(u_i;c_0,c_1,c_2)\, du_i \right) \\ &= \sum_{i=1}^{n} \left(\frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_j} - \frac{1}{\int_{-\infty}^{+\infty} \exp g(u_i;c_0,c_1,c_2)\, du_i} \cdot \frac{\partial}{\partial c_j} \int_{-\infty}^{+\infty} \exp g(u_i;c_0,c_1,c_2)\, du_i \right).\end{aligned}$$

Here we may interchange the integral and derivative, which results in

$$\begin{aligned}\mathcal{S}_{c_j}(\boldsymbol{\theta};\boldsymbol{y}) &= \sum_{i=1}^{n} \left(\frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_j} - \frac{1}{\int_{-\infty}^{+\infty} \exp g(u_i;c_0,c_1,c_2)\, du_i} \cdot \int_{-\infty}^{+\infty} \frac{\partial}{\partial c_j} \exp g(u_i;c_0,c_1,c_2)\, du_i \right) \\ &= \sum_{i=1}^{n} \left(\frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_j} - \frac{\int_{-\infty}^{+\infty} \exp g(u_i;c_0,c_1,c_2) \cdot \frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_j}\, du_i}{\int_{-\infty}^{+\infty} \exp g(u_i;c_0,c_1,c_2)\, du_i} \right).\end{aligned}$$

The next step is to see that the integral in the denominator can be moved into the integral in the nominator, which allows us to apply (4.4-344), that is

$$\begin{aligned}\mathcal{S}_{c_j}(\boldsymbol{\theta};\boldsymbol{y}) &= \sum_{i=1}^{n} \left(\frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_j} - \int_{-\infty}^{+\infty} \frac{\exp g(u_i;c_0,c_1,c_2)}{\int_{-\infty}^{+\infty} \exp g(u_i;c_0,c_1,c_2)\, du_i} \cdot \frac{\partial g(u;c_0,c_1,c_2)}{\partial c_j}\, du \right) \\ &= \sum_{i=1}^{n} \left(\frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_j} - \int_{-\infty}^{+\infty} f(u;c_0,c_1,c_2) \frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_j}\, du_i \right).\end{aligned}$$

Finally, we may use the fact that the integral represents the expectation of the random variable $\partial g(U_i;c_0,c_1,c_2)/\partial c_j$, which leads to the result

$$\mathcal{S}_{c_j}(\boldsymbol{\theta};\boldsymbol{Y}) = \sum_{i=1}^{n} \left(\frac{\partial g(U_i;c_0,c_1,c_2)}{\partial c_j} - E\left\{ \frac{\partial g(U_i;c_0,c_1,c_2)}{\partial c_j} \right\} \right) \tag{4.4-346}$$

as given in Peracchi (2001). To compute the partial derivatives $\partial g(u_i;c_0,c_1,c_2)/\partial c_j$ regarding the antiderivative defined in (4.4-337), we may again interchange the derivative and the integral. Then we obtain

$$\begin{aligned}\frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_0} &= -\int \frac{c_1 - u_i}{(c_0 - c_1 u_i + c_2 u_i^2)^2}\, du_i, \\ \frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_1} &= \int \frac{1 \cdot (c_0 - c_1 u_i + c_2 u_i^2) - (c_1 - u_i)(-u_i)}{(c_0 - c_1 u_i + c_2 u_i^2)^2}\, du_i = \int \frac{c_0 + c_2 u_i^2 - u_i^2}{(c_0 - c_1 u_i + c_2 u_i^2)^2}\, du_i, \\ \frac{\partial g(u_i;c_0,c_1,c_2)}{\partial c_2} &= -\int \frac{(c_1 - u_i) u_i^2}{(c_0 - c_1 u_i + c_2 u_i^2)^2}\, du_i.\end{aligned}$$

Eventually, we will have to evaluate the scores at the ML estimates with the restrictions $H_0 : c_1 = c_2 = 0$. Furthermore, the parameter c_0 will be identical to the variance σ^2 under these restrictions, as mentioned above. Then, evaluation of the partial derivatives at $c_1 = c_2 = 0$ with the parameters $c_0 = \sigma^2$ and $\boldsymbol{\beta}$ remaining unspecified gives

$$\begin{aligned}
\frac{\partial g(u_i;\sigma^2,0,0)}{\partial c_0} &= \int \frac{u_i}{\sigma^4}\,du_i = \frac{1}{\sigma^4}\int u_i\,du_i = \frac{u_i^2}{2\sigma^4},\\
\frac{\partial g(u_i;\sigma^2,0,0)}{\partial c_1} &= \int \frac{\sigma^2-u_i^2}{\sigma^4}\,du_i = \frac{1}{\sigma^2}\int du_i - \frac{1}{\sigma^4}\int u_i^2\,du_i = \frac{u_i}{\sigma^2} - \frac{u_i^3}{3\sigma^4},\\
\frac{\partial g(u_i;\sigma^2,0,0)}{\partial c_2} &= \int \frac{u_i^3}{\sigma^4}\,du_i = \frac{1}{\sigma^4}\int u_i^3\,du_i = \frac{u_i^4}{4\sigma^4}.
\end{aligned}$$

These quantities define random variables whose expectations, under the restrictions H_0, are given by

$$\begin{aligned}
E\left\{\frac{\partial g(u_i;\sigma^2,0,0)}{\partial c_0}\right\} &= \frac{E\{U_i^2\}}{2\sigma^4} = \frac{1}{2\sigma^2},\\
E\left\{\frac{\partial g(u_i;\sigma^2,0,0)}{\partial c_1}\right\} &= \frac{E\{U_i\}}{\sigma^2} - \frac{E\{U_i^3\}}{3\sigma^4} = 0,\\
E\left\{\frac{\partial g(u_i;\sigma^2,0,0)}{\partial c_2}\right\} &= \frac{E\{U_i^4\}}{4\sigma^4} = \frac{3}{4},
\end{aligned}$$

where we used the following facts about the moments of U_i: (1) $E\{U_i\} = 0$ by virtue of the first Markov condition; (2) $E\{U_i^2\} = \sigma^2 = \mu_2$ in light of the second Markov condition; (3) $c_1 = \gamma_1 = 0$ implies $E\{U_i^3\}\mu_3 = 0$ because of (4.4-332); and (4) $c_2 = \gamma_2 = 0$ implies $E\{U_i^4\} = \mu_4 = 3\mu_2^2 = 3\sigma^4$ due to (4.4-333). This gives finally the components of the score (4.4-346)

$$\mathcal{S}_{\sigma^2}(\boldsymbol{\theta};\boldsymbol{Y}) = \sum_{i=1}^{n}\left(\frac{U_i^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right) = \frac{1}{2\sigma^4}\sum_{i=1}^{n}U_i^2 - \frac{n}{2\sigma^2}, \tag{4.4-347}$$

$$\mathcal{S}_{c_1}(\boldsymbol{\theta};\boldsymbol{Y}) = \sum_{i=1}^{n}\left(\frac{U_i}{\sigma^2} - \frac{U_i^3}{3\sigma^4} - 0\right) = \frac{1}{\sigma^2}\sum_{i=1}^{n}U_i - \frac{1}{3\sigma^4}\sum_{i=1}^{n}U_i^3 \tag{4.4-348}$$

$$\mathcal{S}_{c_2}(\boldsymbol{\theta};\boldsymbol{Y}) = \sum_{i=1}^{n}\left(\frac{U_i^4}{4\sigma^4} - \frac{3}{4}\right) = \frac{1}{4\sigma^4}\sum_{i=1}^{n}U_i^4 - \frac{3n}{4}. \tag{4.4-349}$$

To construct Rao's Score statistic, the scores (4.4-345) and (4.4-347) - (4.4-349) must be evaluated at the restricted ML estimates $\widetilde{\boldsymbol{\theta}}$. Under the restrictions $c_1 = c_2 = c_0$, the Gauss-Markov model (4.4-330) and (4.4-331) has normally distributed errors $\boldsymbol{U}$. Therefore, the restricted ML estimator for $\boldsymbol{\beta}$ is identical to the least squares estimator $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. The residuals are then estimated by $\widetilde{\boldsymbol{U}} = \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$. This leads to the restricted ML estimator for $c_0 = \sigma^2$, that is $\widetilde{\sigma}_{ML}^2 = \widetilde{\boldsymbol{U}}'\widetilde{\boldsymbol{U}}/n$, which differs from the least squares estimator only in using the factor $1/n$ instead of $1/(n-m)$. Finally, we will also use the estimators $\widetilde{\mu}_j = \sum_{i=1}^{n}\widetilde{U}_i^j/n$ $(j = 2,3,4)$ for the second, third, and fourth (central) moments (the first moment $\widetilde{\mu}_j = \sum_{i=1}^{n}\widetilde{U}_i$ is zero as we assumed the presence of an intercept parameter).

Exploiting the orthogonality between the j-th column of $\boldsymbol{X}$ (i.e. the j-th row of $\boldsymbol{X}'$) and the vector of estimated residuals, we obtain from (4.4-345)

$$\mathcal{S}_{\beta_j}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) = -\sum_{i=1}^{n}X_{i,j}\frac{-\widetilde{U}_i}{\widetilde{\sigma}_{ML}^4} = \frac{1}{\widetilde{\sigma}_{ML}^4}\sum_{i=1}^{n}X_{i,j}\widetilde{U}_i = 0, \tag{4.4-350}$$

and from (4.4-347) - (4.4-349)

$$\mathcal{S}_{\sigma^2}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) = \frac{1}{2\widetilde{\sigma}_{ML}^4}\sum_{i=1}^{n}\widetilde{U}_i^2 - \frac{n}{2\widetilde{\sigma}_{ML}^2} = \frac{n\widetilde{\sigma}_{ML}^2}{2\widetilde{\sigma}_{ML}^4} - \frac{n}{2\widetilde{\sigma}_{ML}^2} = 0, \tag{4.4-351}$$

$$\mathcal{S}_{c_1}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) = \frac{1}{\widetilde{\sigma}_{ML}^2}\sum_{i=1}^{n}\widetilde{U}_i - \frac{1}{3\widetilde{\sigma}_{ML}^4}\sum_{i=1}^{n}U_i^3 = -\frac{n\widetilde{\mu}_3}{3\widetilde{\sigma}_{ML}^4}, \tag{4.4-352}$$

$$\mathcal{S}_{c_2}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) = \frac{1}{4\widetilde{\sigma}_{ML}^4}\sum_{i=1}^{n}\widetilde{U}_i^4 - \frac{3n}{4} = \frac{n\widetilde{\mu}_4}{4\widetilde{\sigma}_{ML}^4} - \frac{3n}{4}. \tag{4.4-353}$$

As already mentioned in Section 2.5.6, we see that the score vanishes in the direction of the unrestricted parameters ($\boldsymbol{\beta}$ and σ^2), because the estimates for the unrestricted parameters are given the freedom to satisfy the corresponding likelihood equations exactly, i.e. to maximize the log-likelihood function in those directions.

The derivation of the second partial derivatives of the log-likelihood function with respect to all the parameters in $\boldsymbol{\theta}$, or equivalently of the first partial derivatives of the scores (4.4-345) and (4.4-347) - (4.4-349), is very lengthy. Therefore, we will refer to Proposition 2 in Bera and Jarque (1982), from which the information matrix at $\widetilde{\boldsymbol{\theta}}$ is obtained as

$$\boldsymbol{\mathcal{I}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) = \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{n}{2\widetilde{\sigma}^4_{ML}} & 0 & \frac{3n}{2\widetilde{\sigma}^2_{ML}} \\ \boldsymbol{0} & 0 & \frac{2n}{3\widetilde{\sigma}^2_{ML}} & 0 \\ \boldsymbol{0} & \frac{3n}{2\widetilde{\sigma}^2_{ML}} & 0 & 6n \end{bmatrix} \tag{4.4-354}$$

Now we obtain for Rao's Score statistic

$$\begin{aligned} T_{RS} &= \boldsymbol{\mathcal{S}}'(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})\boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})\boldsymbol{\mathcal{S}}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y}) \\ &= \begin{bmatrix} \boldsymbol{0} \\ 0 \\ -\frac{n\widetilde{\mu}_3}{3\widetilde{\sigma}^4_{ML}} \\ \frac{n\widetilde{\mu}_4}{4\widetilde{\sigma}^4_{ML}} - \frac{3n}{4} \end{bmatrix}' \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{n}{2\widetilde{\sigma}^4_{ML}} & 0 & \frac{3n}{2\widetilde{\sigma}^2_{ML}} \\ \boldsymbol{0} & 0 & \frac{2n}{3\widetilde{\sigma}^2_{ML}} & 0 \\ \boldsymbol{0} & \frac{3n}{2\widetilde{\sigma}^2_{ML}} & 0 & 6n \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{0} \\ 0 \\ -\frac{n\widetilde{\mu}_3}{3\widetilde{\sigma}^4_{ML}} \\ \frac{n\widetilde{\mu}_4}{4\widetilde{\sigma}^4_{ML}} - \frac{3n}{4} \end{bmatrix} \end{aligned}$$

If we define the subvectors

$$\boldsymbol{\mathcal{S}}_1 := \begin{bmatrix} \boldsymbol{0} \\ 0 \end{bmatrix},\ \boldsymbol{\mathcal{S}}_2 := n\begin{bmatrix} -\frac{\widetilde{\mu}_3}{3\widetilde{\sigma}^4_{ML}} \\ \frac{\widetilde{\mu}_4}{4\widetilde{\sigma}^4_{ML}} - \frac{3}{4} \end{bmatrix},$$

the submatrices

$$\boldsymbol{\mathcal{I}}_{11} := \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{n}{2\widetilde{\sigma}^4_{ML}} \end{bmatrix},\ \boldsymbol{\mathcal{I}}_{12} = \boldsymbol{\mathcal{I}}'_{21} := \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ 0 & \frac{3n}{2\widetilde{\sigma}^2_{ML}} \end{bmatrix},\ \boldsymbol{\mathcal{I}}_{22} := \begin{bmatrix} \frac{2n}{3\widetilde{\sigma}^2_{ML}} & 0 \\ 0 & 6n \end{bmatrix},$$

and the Schur complements $\boldsymbol{\mathcal{I}}^{(-1)}_{11}$, $\boldsymbol{\mathcal{I}}^{(-1)}_{12}$, $\boldsymbol{\mathcal{I}}^{(-1)}_{21}$, $\boldsymbol{\mathcal{I}}^{(-1)}_{22}$ as the blocks of the total inverse $\boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\theta}};\boldsymbol{Y})$, then T_{RS} follows to be

$$\begin{aligned} T_{RS} &= \boldsymbol{\mathcal{S}}'_1\boldsymbol{\mathcal{I}}^{(-1)}_{11}\boldsymbol{\mathcal{S}}_1 + \boldsymbol{\mathcal{S}}'_2\boldsymbol{\mathcal{I}}^{(-1)}_{21}\boldsymbol{\mathcal{S}}_1 + \boldsymbol{\mathcal{S}}'_1\boldsymbol{\mathcal{I}}^{(-1)}_{12}\boldsymbol{\mathcal{S}}_2 + \boldsymbol{\mathcal{S}}'_2\boldsymbol{\mathcal{I}}^{(-1)}_{22}\boldsymbol{\mathcal{S}}_2 \\ &= \boldsymbol{\mathcal{S}}'_2\boldsymbol{\mathcal{I}}^{(-1)}_{22}\boldsymbol{\mathcal{S}}_2. \end{aligned}$$

With

$$\begin{aligned} \boldsymbol{\mathcal{I}}^{(-1)}_{22} &= (\boldsymbol{\mathcal{I}}_{22} - \boldsymbol{\mathcal{I}}_{21}\boldsymbol{\mathcal{I}}^{-1}_{11}\boldsymbol{\mathcal{I}}_{12})^{-1} \\ &= \left(\begin{bmatrix} \frac{2n}{3\widetilde{\sigma}^2_{ML}} & 0 \\ 0 & 6n \end{bmatrix} - \begin{bmatrix} \boldsymbol{0} & 0 \\ \boldsymbol{0} & \frac{3n}{2\widetilde{\sigma}^2_{ML}} \end{bmatrix} \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2_{ML}}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{n}{2\widetilde{\sigma}^4_{ML}} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ 0 & \frac{3n}{2\widetilde{\sigma}^2_{ML}} \end{bmatrix} \right)^{-1} \\ &= \left(\begin{bmatrix} \frac{2n}{3\widetilde{\sigma}^2_{ML}} & 0 \\ 0 & 6n \end{bmatrix} - \begin{bmatrix} \boldsymbol{0} & 0 \\ \boldsymbol{0} & \frac{3n}{2\widetilde{\sigma}^2_{ML}} \end{bmatrix} \begin{bmatrix} \widetilde{\sigma}^2_{ML}(\boldsymbol{X}'\boldsymbol{X})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 2\widetilde{\sigma}^4_{ML}/n \end{bmatrix} \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ 0 & \frac{3n}{2\widetilde{\sigma}^2_{ML}} \end{bmatrix} \right)^{-1} \\ &= \left(\begin{bmatrix} \frac{2n}{3\widetilde{\sigma}^2_{ML}} & 0 \\ 0 & 6n \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \frac{9n}{2} \end{bmatrix} \right)^{-1} = \frac{1}{n}\begin{bmatrix} \frac{3}{2}\widetilde{\sigma}^2_{ML} & 0 \\ 0 & \frac{2}{3} \end{bmatrix}, \end{aligned}$$

(which is a part of the result given in Proposition 3 by Bera and Jarque, 1982), we obtain

$$\begin{aligned} T_{RS} &= n\begin{bmatrix} -\frac{\widetilde{\mu}_3}{3\widetilde{\sigma}^4_{ML}} \\ \frac{\widetilde{\mu}_4}{4\widetilde{\sigma}^4_{ML}} - \frac{3}{4} \end{bmatrix}' \begin{bmatrix} \frac{3}{2}\widetilde{\sigma}^2_{ML} & 0 \\ 0 & \frac{2}{3} \end{bmatrix} \begin{bmatrix} -\frac{\widetilde{\mu}_3}{3\widetilde{\sigma}^4_{ML}} \\ \frac{\widetilde{\mu}_4}{4\widetilde{\sigma}^4_{ML}} - \frac{3}{4} \end{bmatrix} = n\begin{bmatrix} -\frac{\widetilde{\mu}_3}{3\widetilde{\sigma}^4_{ML}} \\ \frac{\widetilde{\mu}_4}{4\widetilde{\sigma}^4_{ML}} - \frac{3}{4} \end{bmatrix}' \begin{bmatrix} -\frac{\widetilde{\mu}_3}{2\widetilde{\sigma}^2_{ML}} \\ \frac{\widetilde{\mu}_4}{6\widetilde{\sigma}^4_{ML}} - \frac{1}{2} \end{bmatrix} \\ &= n\frac{\widetilde{\mu}_3^2}{6\widetilde{\sigma}^6_{ML}} + n\frac{\widetilde{\mu}_4^2}{24\widetilde{\sigma}^8_{ML}} - n\frac{\widetilde{\mu}_4}{8\widetilde{\sigma}^4_{ML}} - n\frac{\widetilde{\mu}_4}{8\widetilde{\sigma}^4_{ML}} + 3n/8 \\ &= \frac{n}{6}\cdot\frac{\widetilde{\mu}_3^2}{\widetilde{\sigma}^6_{ML}} + \frac{n}{24}\left(\frac{\widetilde{\mu}_4^2}{\widetilde{\sigma}^8_{ML}} - 6\frac{\widetilde{\mu}_4}{\widetilde{\sigma}^4_{ML}} + 9\right) \\ &= \frac{n}{6}\cdot\frac{\widetilde{\mu}_3^2}{\widetilde{\sigma}^6_{ML}} + \frac{n}{24}\left(\frac{\widetilde{\mu}_4}{\widetilde{\sigma}^4_{ML}} - 3\right)^2. \end{aligned}$$

Observe now that the definition of $\widetilde{\sigma}^2_{ML}$ is identical to that of the empirical second central moment $\widetilde{\mu}_2$. Then, if we define the empirical skewness

$$\widetilde{\gamma}_1 = \frac{\widetilde{\mu}_3}{\widetilde{\mu}_2^{3/2}}, \tag{4.4-355}$$

and the empirical kurtosis

$$\widetilde{\gamma}_2 = \frac{\widetilde{\mu}_4}{\widetilde{\mu}_2^2} - 3, \tag{4.4-356}$$

which depend on the estimated residuals of the Gauss-Markov model through the empirical moments $\widetilde{\mu}_j = \sum_{i=1}^n \widetilde{U}_i^j$ $(j = 2, \ldots, 4)$, then Rao's Score statistic takes its final form

$$T_{RS} = \frac{n}{6}\widetilde{\gamma}_1^2 + \frac{n}{24}\widetilde{\gamma}_2^2. \tag{4.4-357}$$

Evidently, this statistic measures the absolute deviations of the data's skewness and kurtosis from the values 0, thus compares how far the distribution of the estimated residuals differs from a normal distribution. This test of normality is also called the **Jarque-Bera test** (Bera and Jarque, 1982).

Example 4.3: Testing the Gravity Dataset for non-normality. In Example 3.2 we considered a two-dimensional polynomial model of degree 1 with additional mean shift parameters. To check whether the errors $\boldsymbol{U}$ in the model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\nabla} + \boldsymbol{U}$$

follow a normal distribution, we first compute the residuals

$$\widetilde{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{Z}\widetilde{\boldsymbol{\nabla}}$$

based on the least squares estimates

$$\begin{bmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\nabla}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{y} \end{bmatrix}.$$

Contrary to Example 3.2, we use tildes instead of hats on top of the estimate $\widetilde{\boldsymbol{\nabla}}$, because here the mean shift parameters $\boldsymbol{\nabla}$, which were proven to be significant, naturally belong to the functional model. The tildes indicate that the estimates have been determined under the restrictions $c_1 = c_2 = 0$. From these residuals we the obtain $\widetilde{\gamma}_1 = -0.25$ for the empirical skewness (4.4-355) and $\widetilde{\gamma}_2 = -0.29$ for the kurtosis measure (4.4-356). With these values, Rao's Score statistic (4.4-357) becomes $T_{RS} = 1.28$, which is insignificant in light of the critical value $k_{0.95}^{\chi^2(2)} = 5.99$. Therefore, we may assume the errors to be normally distributed, which is also roughly reflected by the following histogram plot. □

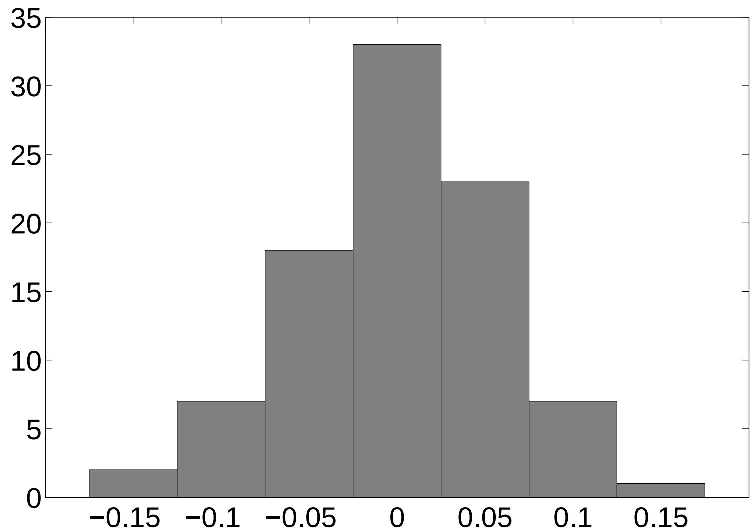

Figure 4.2. Histogram of the estimated residuals.

5 Conclusion and Outlook

In the framework of the Gauss-Markov model with normally distributed errors, uniformly most powerful invariant tests generally exist. These tests have three equivalent formulations: (1) the form obtained from a direct application of invariance principles, (2) the likelihood ratio test, and (3) Rao's score test. Of the three, Rao's score test is easiest to compute for problems where significance testing is required. If the testing problem involves unknown parameters within the weight matrix, or if the errors do not follow a normal distribution, then no uniformly most powerful invariant tests exist. In these cases too, Rao's score test offers an attractive method that is both powerful and computationally convenient. This thesis has demonstrated that hypothesis testing by applying Rao's score method is an effective - and in many cases optimal - approach for resolving a wide range of problems faced in geodetic model analysis.

New satellite missions such as GOCE (Gravity Field and steady-state Ocean Circulation Explorer) require powerful and computationally feasible tests for diagnosing functional and stochastic models that are far more complex than the models considered in this thesis (cf. Lackner, 2006). To find convincing solutions to these challenges, it will be necessary for geodesists to further elaborate their understanding of statistical testing theory. Looking at the methodology currently offered by mathematical statistics, some directions of further research are particularly promising. Rao's score approach can be applied to a full range of testing problem fields such as deformation analysis, time series analysis, or geostatistics - applications that have not yet been explored in modern geodetic literature. It is crucial that geodesists develop a stronger expertise in the asymptotic behaviour of statistical theories, such as given in Lehmann and Romano (2005, Part II). This will be a necessary step towards assessing the quality of geodetic hypothesis tests, such as those presented in Section 4, for which no strict optimality criteria are applicable.

The scope of the theory presented in this thesis is restricted to a specific minimization problem regarding Type I and Type II error probabilities within the class of invariant tests. However, minimizing error probability does not correspond to a minimization of costs when one considers losses in work time, computational time, or even accuracy of estimated parameters. To overcome this limitation, hypothesis tests could be derived within the framework of decision theory, by minimizing a loss function which represents the expected loss/cost due to an erroneous test decision (cf. Lehmann and Romano, 2005, p. 59).

Finally, it is often argued that classical testing theory is too limited in that the test decision is always made on the premise of a true null hypothesis, and that a priori information with respect to the unknown parameters may not be used (see, for instance, Jaynes, 2003, Chapter 16). A theory which does allow the treatment of the null and the alternative hypothesis on equal grounds and incorporation of a priori information is offered by *Bayes statistics*. Bayesian tests may be viewed as generalizations of likelihood ratio tests in that the likelihood ratio is extended by an a priori density with respect to the unknown parameters, which are treated as random variables (cf. Koch, 2000, Section 3.4). It would be highly instructive to formulate the model misspecification tests developed in this thesis within the Bayesian framework and to compare them in terms of testing power, applicability to a wide range of problems, and computational convenience.

6 Appendix: Datasets

6.1 Dam Dataset

The numerical values of the observation model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \boldsymbol{\Sigma}\{\boldsymbol{U}\} = \sigma^2 \boldsymbol{I},$$

for the Dam dataset used in Application 3 of Sect. 3.5 are given by

$$\left[\begin{array}{r} 3.433 \\ 80.001 \\ 80.000 \\ 76.568 \\ 76.569 \\ 76.570 \\ -0.477 \\ 77.047 \\ 77.046 \\ 67.705 \\ 67.706 \\ 1.444 \\ 66.262 \\ 66.262 \\ 66.260 \\ 1.162 \\ 65.097 \\ 65.098 \\ -0.002 \\ -0.002 \\ \hline 80.000 \\ 79.999 \\ 76.570 \\ 76.569 \\ 76.569 \\ 77.046 \\ 77.047 \\ 67.704 \\ 67.706 \\ 66.261 \\ 66.261 \\ 66.260 \\ 65.098 \\ 65.099 \end{array}\right] = \left[\begin{array}{rrrrrrrr|rrr} +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & +1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & +1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & +1 \end{array}\right] \left[\begin{array}{c} H_2 \\ H_3 \\ H_4 \\ H_5 \\ H_6 \\ H_7 \\ H_8 \\ H_9 \\ H_{7'} \\ H_{8'} \\ H_{9'} \end{array}\right] + \left[\begin{array}{c} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \\ u_{11} \\ u_{12} \\ u_{13} \\ u_{14} \\ u_{15} \\ u_{16} \\ u_{17} \\ u_{18} \\ u_{19} \\ u_{20} \\ \hline u_{21} \\ u_{22} \\ u_{23} \\ u_{24} \\ u_{25} \\ u_{26} \\ u_{27} \\ u_{28} \\ u_{29} \\ u_{30} \\ u_{31} \\ u_{32} \\ u_{33} \\ u_{34} \end{array}\right] \tag{6.1-358}$$

6.2 Gravity Dataset

Index	Anomaly	Latitude		Longitude		Index	Anomaly	Latitude		Longitude	
i	dg	°	min	°	min	i	dg	°	min	°	min
1	15.11	48	10.26	16	19.01	47	14.91	46	41.99	13	40.00
2	15.13	48	13.97	16	20.24	48	15.03	46	37.09	13	50.04
3	14.80	48	18.72	16	25.85	49	15.01	46	39.45	14	18.82
4	15.02	48	43.15	16	18.20	50	15.07	46	33.27	14	02.63
5	15.11	48	30.75	16	37.30	51	15.03	47	46.26	12	56.56
6	15.25	48	41.56	16	52.24	52	15.09	47	36.09	12	42.24
7	15.02	48	21.72	15	24.23	53	15.07	47	16.94	12	29.00
8	15.07	48	11.88	14	31.68	54	15.06	47	17.75	12	46.07
9	15.07	48	07.54	14	52.73	55	15.09	47	25.13	13	13.15
10	14.91	48	09.56	15	06.06	56	15.01	47	08.05	13	40.82
11	15.12	48	13.45	15	21.48	57	15.00	48	34.22	13	59.69
12	15.09	48	10.32	15	37.15	58	15.09	48	15.41	13	02.34
13	15.10	47	48.42	16	12.18	59	15.07	48	27.50	13	26.07
14	15.06	47	39.61	15	53.51	60	15.10	48	18.52	14	14.85
15	15.09	48	08.43	16	54.04	61	15.05	48	04.40	14	03.43
16	15.04	47	45.98	16	27.31	62	14.99	47	31.66	11	42.49
17	14.97	47	35.08	16	25.54	63	15.11	47	35.06	12	09.98
18	15.04	47	42.41	16	54.87	64	15.00	47	29.02	10	43.38
19	15.11	47	15.10	16	14.65	65	14.91	47	15.63	10	45.56
20	15.15	47	24.82	16	29.78	66	14.95	47	18.51	11	04.50
21	14.98	46	59.34	16	15.65	67	14.95	47	16.83	10	59.00
22	15.10	47	53.35	16	39.69	68	15.05	47	16.07	11	16.24
23	15.03	47	44.49	15	18.64	69	14.91	47	25.52	11	13.90
24	14.93	47	36.95	15	46.11	70	14.94	47	25.76	11	14.81
25	15.10	47	23.54	13	41.47	71	15.08	47	23.78	11	49.35
26	15.20	47	23.16	15	05.63	72	15.06	47	29.19	12	03.96
27	15.09	47	24.93	16	01.25	73	14.98	47	07.76	10	15.93
28	14.98	47	04.68	13	55.63	74	14.93	47	08.95	10	34.70
29	15.07	47	02.35	15	10.19	75	14.89	47	08.05	10	30.99
30	15.13	47	03.41	15	25.04	76	14.92	47	14.22	10	44.33
31	15.12	47	01.59	15	29.14	77	14.92	47	10.92	10	37.03
32	15.16	47	04.59	15	34.97	78	15.13	47	11.84	10	39.41
33	15.11	47	04.19	15	44.22	79	14.90	47	13.30	10	45.35
34	15.13	46	53.11	15	29.91	80	14.92	47	14.08	10	51.24
35	15.06	47	40.55	15	29.64	81	15.06	47	03.85	11	29.28
36	15.01	47	36.45	15	40.38	82	15.03	47	07.63	11	27.17
37	15.11	47	04.77	15	59.64	83	15.03	47	13.92	11	23.17
38	14.93	47	00.05	12	32.54	84	15.05	47	14.04	11	52.87
39	14.90	46	56.21	12	34.42	85	14.85	46	58.09	10	32.12
40	15.00	46	49.03	12	47.28	86	14.82	46	50.92	10	30.33
41	15.01	46	49.70	12	45.45	87	14.69	47	21.65	10	49.87
42	14.82	46	54.75	13	15.50	88	14.89	47	31.80	9	51.80
43	15.04	46	50.30	13	22.21	89	15.13	47	27.44	9	38.44
44	14.92	46	48.97	13	25.04	90	14.98	47	25.84	9	45.38
45	14.83	46	49.89	14	26.78	91	14.95	47	07.95	10	07.31
46	15.11	46	39.85	12	59.82						

References

Arnold, S.F. (1981). *The theory of linear models and multivariate analysis.* Wiley, New York.

Arnold, S.F. (1985). *Sufficiency and invariance.* Statistics & Probability Letters 3: 275-279.

Arnold, S.F. (1990). *Mathematical statistics.* Prentice Hall, Englewood Cliffs, New Jersey.

Baarda, W. (1967). *Statistical concepts in geodesy.* Publications on Geodesy (New Series), Vol. 1, Number 4, Netherlands Geodetic Commission, Delft.

Baarda, W. (1968). *A testing procedure for use in geodetic networks.* Publications on Geodesy (New Series), Vol. 2, Number 5, Netherlands Geodetic Commission, Delft.

Bera, A.K. and C.M. Jarque (1982). *Model specification tests. A simultaneous approach.* Journal of Econometrics, 20:59-82.

Birkes, D. (1990). *Generalized likelihood ratio tests and uniformly most powerful tests.* The American Statistician, 44(2):163-166.

Bronstein, I.N. and K.A. Semendjajew (1991). *Taschenbuch der Mathematik.* Teubner, Stuttgart.

Casella, G. and R.L. Berger (2002). *Statistical inference.* Second edition. Duxbury, Pacific Grove, California.

Cox, D.R. and D.V. Hinkley (1974). *Theoretical statistics.* Chapman and Hall, London.

Dudewicz, E.J. and S. Mishra (1988). *Modern mathematical statistics.* Wiley, New York.

Engle, R.F. (1984). *Wald, likelihood ratio, and Lagrange multiplier tests in econometrics.* Z. Griliches and M.D. Intriligator (eds.): Handbook of Econometrics, Vol. 2:775-826.

Godfrey, L.G. (1988). *Misspecification tests in econometrics.* Cambridge University Press, Cambridge.

Hamilton, J.D. (1994). *Time series analysis.* Princeton University Press, Princeton.

Jaeger, R. et al. (2005). *Klassische und robuste Ausgleichungsverfahren.* Wichmann, Heidelberg.

Jaynes, E.T. (2003). *Probability theory. The logic of science.* Cambridge University Press, Cambridge.

Johnson, N.L. and S. Kotz (1970). *Distributions in statistics: continuous univariate distributions, Vol. 1.* Wiley, New York.

Johnson, N.L. and S. Kotz (1970). *Distributions in statistics: continuous univariate distributions, Vol. 2.* Wiley, New York.

Koch, K.-R. (1981). *Varianz- und Kovarianzkomponentenschätzung für Streckenmessungen auf Eichlinien.* Allgemeine Vermessungs-Nachrichten, 88:125-132.

Koch, K.-R. (1999). *Parameter estimation and hypothesis testing in linear models.* Springer, Heidelberg.

Koch, K.-R. (2000). *Einführung in die Bayes-Statistik.* Springer, Heidelberg.

Krämer, W. and H. Sonnberger (1986). *The linear regression model under test.* Physica, Heidelberg.

Kreyszig, E. (1998). *Statistische Methoden und ihre Anwendungen.* Vandenhoeck and Ruprecht, Göttingen.

Lackner, B. (2006). *Datainspection and hypothesis tests of very long time series applied to GOCE satellite gravity gradiometry data.* Dissertation, Graz University of Technology.

Lehmann, E.L. (1959a). *Testing statistical hypotheses.* First edition. Wiley, New York.

Lehmann, E.L. (1959b). *Optimum invariant tests.* Annals of Mathematical Statistics, 30:881-884.

Lehmann, E.L. and J.P. Romano (2005). *Testing statistical hypotheses.* Third edition. Springer, New York.

Meissl, P. (1982). *Least squares adjustment - a modern approach.* Mitteilungen der geodätischen Institute der Technischen Universität Graz, Vol. 43.

Meyer, C.D. (2000). *Matrix analysis and applied linear algebra*. SIAM.

Neyman, J. and E.S. Pearson (1928). *On the use and interpretation of certain test criteria for purposes of statistical inference.* Biometrika, 20A:175-240,263-295.

Neyman, J. and E.S. Pearson (1933). *On the problem of the most efficient tests of statistical hypotheses.* Philosophical Transactions of the Royal Society of London, Series A, 231:289-337.

Olive, D.J. (2006). *A course in statistical theory*. Online document, Department of Mathematics, Southern Illinois University. www.math.siu.edu/olive/infbook.htm.

Peracchi, F. (2001). *Econometrics.* Wiley, New York.

Pope, A.J. (1976). *The statistics of residuals and the detection of outliers.* NOAA Technical Report NOS65 NGS1, US Department of Commerce, National Geodetic Survey, Rockville, Maryland.

Rao, C.R. (1948). *Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation.* Proceedings of Cambridge Philosophical Society, 44:50-57.

Rao, C.R. (1973). *Linear statistical inference and its applications.* Wiley, New York.

Rousseeuw, P.J. and A.M. Leroy (2003). *Robust regression and outlier detection*. Wiley, New York.

Schuh, W.D. (1996). *Tailored numerical strategies for the global determination of the Earth's gravity field.* Mitteilungen der geodätischen Institute der Technischen Universität Graz, Vol. 81.

Schuh, W.-D. (2006). *Ausgleichungsrechnung und Statistik III.* Lecture notes, Institute of Geodesy and Geoinformation, University of Bonn.

Schuh, W.-D. (2006). *Seminar Robuste Parameterschätzung*. Lecture notes, Institute of Geodesy and Geoinformation, University of Bonn.

Silvey, S.D. (1959). *The Lagrangian multiplier test.* Annals of Mathematical Statistics, 30:389-407.

Stuart, A., Ord, J.K., and S. Arnold (1999). *The advanced theory of statistics, Vol. 2A: Classical inference and the linear model*. Arnold, London.

Stuart, A. and J.K. Ord (2003). *The advanced theory of statistics, Vol. 1: Distribution theory*. Arnold, London.

Teunissen, P.J.G. (2000). *Testing theory*. Delft University Press, Delft.

Acknowledgment

This thesis is the result of my studies in the field of adjustment theory and statistics at the Institute of Geodesy and Geoinformation at the University of Bonn. This opportunity was granted to me by Prof. Dr. techn. W.-D. Schuh, whose supervision was the perfect mix of non-restrictive guidance and open-minded dialogue. I am deeply indebted to him for the time and energy he invested into supporting and discussing my ideas. I also want to thank Prof. Dr. rer. nat. H.-P. Helfrich for serving as second referee. Further, I acknowledge the support by the BMBF Geotechnologien programmes (Grants 03F0329C and 03F0421B).

My very special thanks goes to Jeramy Flora. Without her love, patience, and ideas throughout the years, this thesis (and my own nervous system) would not exist.