

Article

Mean Shift versus Variance Inflation Approach for Outlier Detection—A Comparative Study

Rüdiger Lehmann ^{1,*} , Michael Lösler ²  and Frank Neitzel ³ ¹ Faculty of Spatial Information, University of Applied Sciences Dresden, 01069 Dresden, Germany² Faculty 1: Architecture—Civil Engineering—Geomatics, Frankfurt University of Applied Sciences, 60318 Frankfurt, Germany; michael.loesler@fb1.fra-uas.de³ Technische Universität Berlin, Institute of Geodesy and Geoinformation Science, 10623 Berlin, Germany; frank.neitzel@tu-berlin.de

* Correspondence: ruediger.lehmann@htw-dresden.de; Tel.: +49-351-462-3146

Received: 8 May 2020; Accepted: 12 June 2020; Published: 17 June 2020



Abstract: Outlier detection is one of the most important tasks in the analysis of measured quantities to ensure reliable results. In recent years, a variety of multi-sensor platforms has become available, which allow autonomous and continuous acquisition of large quantities of heterogeneous observations. Because the probability that such data sets contain outliers increases with the quantity of measured values, powerful methods are required to identify contaminated observations. In geodesy, the mean shift model (MS) is one of the most commonly used approaches for outlier detection. In addition to the MS model, there is an alternative approach with the model of variance inflation (VI). In this investigation the VI approach is derived in detail, truly maximizing the likelihood functions and examined for outlier detection of one or multiple outliers. In general, the variance inflation approach is non-linear, even if the null model is linear. Thus, an analytical solution does usually not exist, except in the case of repeated measurements. The test statistic is derived from the likelihood ratio (LR) of the models. The VI approach is compared with the MS model in terms of statistical power, identifiability of actual outliers, and numerical effort. The main purpose of this paper is to examine the performance of both approaches in order to derive recommendations for the practical application of outlier detection.

Keywords: mean shift model; variance inflation model; outlier detection; likelihood ratio test; Monte Carlo integration; data snooping

1. Introduction

Nowadays, outlier detection in geodetic observations is part of the daily business of modern geodesists. As Rofatto et al. [1] state, we have well established and practicable methods for outlier detection for half a century, which are also implemented in current standard geodetic software. The most important toolbox for outlier detection is the so-called data snooping, which is based on the pioneering work of Baarda [2]. A complete distribution theory of data snooping, also known as DIA (detection, identification, and adaptation) method, was developed by Teunissen [3].

In geodesy, methods for outlier detection can be characterised as statistical model selection problem. A null model is opposed to one or more extended or alternative models. While the null model describes the expected stochastic properties of the data, the alternative models deviate from such a situation in one way or another. For outlier detection, the alternative models relate to the situation, where the data are contaminated by one or more outliers. According to Lehmann [4], an outlier is defined by “an observation that is so probably caused by a gross error that it is better not used or not used as it is”.

From a statistical point of view, outliers can be interpreted as a small amount of data that have different stochastic properties than the rest of the data, usually a shift in the mean or an inflation of the variance of their statistical distribution. This situation is described by extra parameters in the functional or stochastic model, such as shifted means or inflated variances. Such an extended model is called an alternative model. Due to the additionally introduced parameters, the discrepancies between the observations and the related results of the model decrease w. r. t. the null model. It has to be decided whether such an improvement of the goodness of fit is statistically significant, which means that the alternative model describes the data better than the null model. This decision can be made by hypothesis testing, information criteria, or many other statistical decision approaches, as shown by Lehmann and Lösler [5,6].

The standard alternative model in geodesy is the mean shift (MS) model, in which the contamination of the observations by gross errors is modelled as a shift in the mean, i.e., by a systematic effect. This approach is described in a large number of articles and textbooks, for example, the contributions by Baarda [2], Teunissen [7], and Kargoll [8]. However, there are other options besides this standard procedure. The contamination may also be modelled as an inflation of the variance of the observations under consideration, i.e., by a random effect. This variance inflation (VI) model is rarely investigated in mathematical statistics or geodesy. Bhar and Gupta [9] propose a solution based on Cook's statistic [10]. Although this statistic was invented for the MS model, it can also be made applicable when the variance is inflated.

Thompson [11] uses the VI model for a single outlier in the framework of the restricted (or residual) maximum likelihood estimation, which is known as REML. In contrast to the true maximum likelihood estimation, REML can produce unbiased estimates of variance and covariance parameters and it causes less computational workload. Thompson [11] proposes that the observation with the largest log-likelihood value can be investigated as a possible outlier. Gumedze et al. [12] take up this development and set up a so-called variance shift outlier model (VSOM). Likelihood ratio (LR) and score test statistics are used to identify the outliers. The authors conclude that VSOM gives an objective compromise between including and omitting an observation, where its status as a correct or erroneous observation cannot be adequately resolved. Gumedze [13] review this approach and work out a one-step LR test, which is a computational simplification of the full-step LR test.

In geodesy, the VI model was introduced by Koch and Kargoll [14]. The estimation of the unknown parameters has been established as an iteratively reweighted least squares adjustment. The expectation maximization (EM) algorithm is used to detect the outliers. It is found that the EM algorithm for the VI model is very sensitive to outliers, due to its adaptive estimation, whereas the EM algorithm for the MS model provides the distinction between outliers and good observations. Koch [15] applies the method to fit a surface in three-dimensional space to the Cartesian coordinates of a point cloud obtained from measurements with a laser scanner.

The main goal of this contribution is a detailed derivation of the VI approach in the framework of outlier detection and compare it with the well-established MS model. The performance of both approaches is to be compared in order to derive recommendations for the practical application of outlier detection. This comprises the following objectives:

1. Definition of the generally accepted null model and specification of alternative MS and VI models (Section 2). Multiple outliers are allowed for in both alternative models to keep the models equivalent.
2. True maximization of the likelihood functions of the null and alternative models, not only for the common MS model, but also for the VI model. This means, we do not resort to the REML approach of Thompson [11], Gumedze et al. [12], and Gumedze [13]. This is important for the purpose of an insightful comparison of MS and VI (Section 3).
3. Application of likelihood ratio (LR) test for outlier detection by hypothesis testing and derivation of the test statistics for both the MS and the VI model. For this purpose, a completely new rigorous likelihood ratio test in the VI model is developed and an also completely new comparison with the equivalent test in the MS model is elaborated (Section 3).

4. Comparison of both approaches using the illustrative example of repeated observations, which is worked out in full detail (Section 4).

Section 5 briefly summarises the investigations that were carried out and critically reviews the results. Recommendations for the practical application of outlier detection conclude this paper.

2. Null Model, Mean Shift Model, and Variance Inflation Model

In mathematical statistics, a hypothesis H is a proposed explanation that the probability distribution of the random n -vector y of observations belongs to a certain parametric family W of probability distributions with parameter vector θ , e.g., Teunissen [7],

$$H : y \sim W(\theta), \quad \theta \in \Theta \tag{1}$$

The parameter vector θ might assume values from a set Θ of admissible parameter vectors. A model is then simply the formulation of the relationship between observations y and parameters θ based on H . In geodesy, the standard linear model is based on the hypothesis that the observations follow a normal distribution N , e.g., Koch [16], Teunissen [7], i.e.,

$$H_0 : y \sim N(Ax, \Sigma), \tag{2}$$

with u -vector of functional parameters x and covariance matrix Σ . The latter matrix might contain further stochastic parameters, like a variance factor σ^2 , according to

$$\Sigma = \sigma^2 Q, \tag{3}$$

with Q being the known cofactor matrix of y . In this case, θ is the union of x and σ^2 . The covariance matrix Σ might eventually contain more stochastic parameters, known as variance components, cf. Koch ([16], p. 225ff). Matrix A is said to be the $n \times u$ -matrix of design. The model that is based on H_0 is called the null model.

In outlier detection, we oppose H_0 with one or many alternative hypotheses, most often in the form of a mean shift (MS) hypothesis, e.g., Koch [16], Teunissen [7]

$$H_{MS} : y \sim N(Ax + C\nabla, \Sigma), \quad \nabla \neq 0, \tag{4}$$

where ∇ is a m -vector of additional functional bias parameters and matrix C extends the design. In this case, θ is extended by ∇ . This relationship gives rise to the MS model, where the mean of the observations is shifted from Ax to $Ax + C\nabla$ by the effect of gross errors, see Figure 1. $C\nabla$ can be interpreted as accounting for the systematic effect of gross observation errors superposing the effect of normal random observation errors already taken into account by Σ in (2). The great advantage of H_{MS} is that, if the null model is linear or linearized, so is the MS model. The determination of the model parameters is numerically easy and computationally efficient, cf. Lehmann and Lösler [5].

However, there are different possibilities to set up an alternative hypothesis. The most simple one is the variance inflation hypothesis

$$H_{VI} : y \sim N(Ax, \Sigma'), \tag{5}$$

where Σ' is a different covariance matrix, which consists of inflated variances. Σ' can be interpreted as accounting for the joint random effect of normal observation errors in all observations and zero mean gross errors in few outlying observations, see Figure 2.

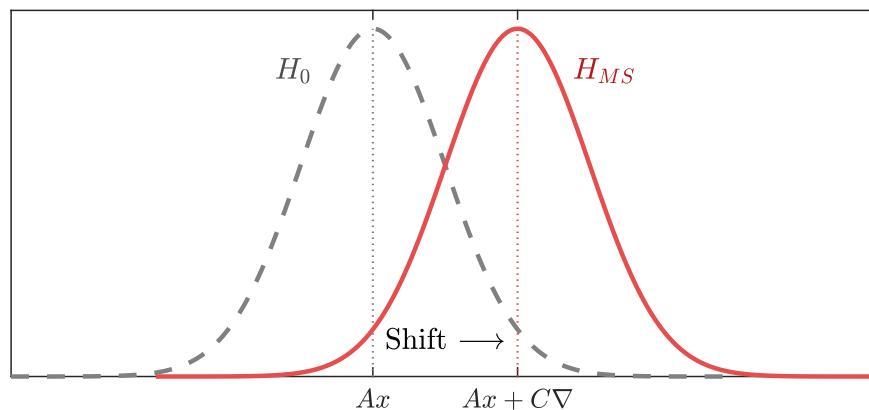


Figure 1. Schematic representation of the MS approach. The null model H_0 is depicted by a dashed grey line and the alternative model H_{MS} is shown as a solid red line.

The VI model might be considered to be more adequate to describe the outlier situation when the act of falsification of the outlying observations is thought of as being a random event, which might not be exactly reproduced in a virtual repetition of the observations. However, even if the VI model might be more adequate to describe the stochastics of the observations, this does not mean that it is possible to estimate parameters or to detect outliers better than with some less adequate model like MS. This point will be investigated below.

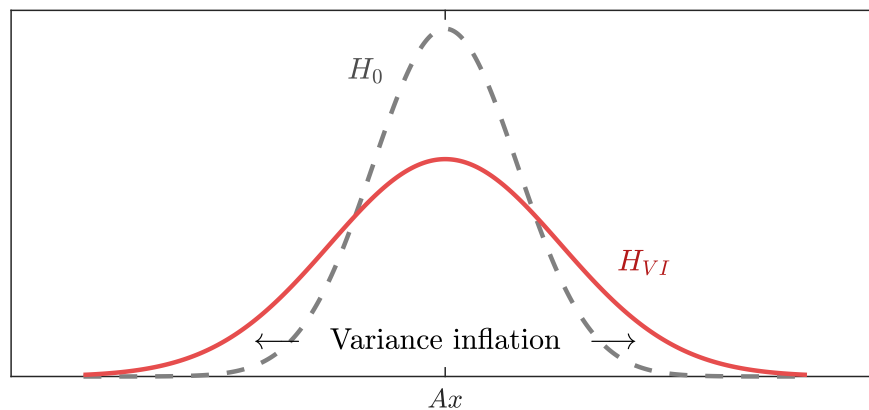


Figure 2. Schematic representation of the VI approach. The null model H_0 is depicted by a dashed grey line and the alternative model H_{VI} is shown as solid red line.

In the following we will only consider the case that y are uncorrelated observations, where both Σ and Σ' are diagonal matrices, such that the hypotheses read

$$H_0 : y \sim N(Ax, \sigma_1^2, \dots, \sigma_n^2), \tag{6a}$$

$$H_{VI} : y \sim N(Ax, \tau_1 \sigma_1^2, \dots, \tau_m \sigma_m^2, \sigma_{m+1}^2, \dots, \sigma_n^2), \quad \tau_1 > 1, \dots, \tau_m > 1. \tag{6b}$$

Here, H_{VI} accounts for m random zero mean gross errors in the observations y_1, \dots, y_m modeled by stochastic parameters τ_1, \dots, τ_m . In this case, θ is extended by τ_1, \dots, τ_m , which will be called variance inflation factors. Thus, τ_1, \dots, τ_m can be interpreted as a special case of extra variance components.

Note that the term variance inflation factors is used differently in multiple linear regression when dealing with multicollinearity, cf. James et al. ([17], p. 101f).

3. Outlier Detection by Hypothesis Tests

Outlier detection can be characterised as a statistical model selection problem. The null model, which describes the expected stochastic properties of the data, is opposed to one or more alternative models, which deviate from such properties. Usually, the decision, whether the null model is rejected in favour of a proper alternative model, is based on hypothesis testing. Multiple testing, consisting of a sequence of testings with one single alternative hypothesis only, is required if there are many possible alternative models. In this study, we first focus on such one single testing only. In the following subsections, the test statistics in the MS model as well as in the VI model are derived.

For the sake of simplicity, the scope of this contribution is restricted to cases, where all of the estimates to be computed are unique, such that all matrices to be inverted are regular.

3.1. Mean Shift Model

In the case of no stochastic parameters, i.e., Σ is known, the optimal test statistic $T_{MS}(y)$ for the test problem H_0 in (2) versus H_{MS} in (4) is well known and yields, cf. Teunissen ([7], p. 76):

$$T_{MS}(y) := \hat{\nabla}^T \Sigma_{\hat{\nabla}}^{-1} \hat{\nabla} \tag{7}$$

$$= \hat{e}_0^T \Sigma^{-1} C (C^T \Sigma^{-1} \Sigma_{\hat{e}_0}^{-1} \Sigma^{-1})^{-1} C^T \Sigma^{-1} \hat{e}_0,$$

where $\hat{\nabla}$ and $\Sigma_{\hat{\nabla}}$ are the vector of estimated bias parameters in the MS model and its covariance matrix, respectively. Furthermore, \hat{e}_0 and $\Sigma_{\hat{e}_0}$ are the vector of estimated residuals $e := Ax - y$ in the null model and its covariance matrix, respectively. This second expression offers the opportunity to perform the test purely based on the estimation in the null model, cf. Teunissen ([7], p. 75), i.e.,

$$\hat{x}_0 = (A^T \Sigma^{-1} A)^{-1} A^T \Sigma^{-1} y, \tag{8a}$$

$$\hat{e}_0 = (I - A(A^T \Sigma^{-1} A)^{-1} A^T \Sigma^{-1}) y, \tag{8b}$$

$$\Sigma_{\hat{e}_0} = \Sigma - A(A^T \Sigma^{-1} A)^{-1} A^T. \tag{8c}$$

Each estimation is such that the likelihood function of the model is maximized. The hat will indicate maximum likelihood estimates below.

The test statistic (7) follows the central or non-central χ^2 distributions

$$T_{MS}|H_0 \sim \chi^2(q, 0), \tag{9a}$$

$$T_{MS}|H_{MS} \sim \chi^2(q, \lambda), \tag{9b}$$

where $q = \text{rank}(\Sigma_{\hat{\nabla}})$ is the degree of freedom and $\lambda = \nabla^T \Sigma_{\hat{\nabla}}^{-1} \nabla$ is called the non-centrality parameter, depending on the true but unknown bias parameters ∇ , e.g., (Teunissen [7], p. 77).

The test statistic (7) has the remarkable property of being uniformly the most powerful invariant (UMPI). This means, given a probability of type 1 decision error (rejection of H_0 when it is true) α , (7)

- has the least probability of type 2 decision error (failure to reject H_0 when it is false) β (most powerful);
- is independent of ∇ (uniform); and,
- but only for some transformed test problem (invariant).

For the original test problem, no uniformly most powerful (UMP) test exists. For more details see Arnold [18], Kargoll [8].

Lehmann and Voß-Böhme [19] prove that (7) has the property of being a UMP χ^2 test, i.e., a UMP test in the class of all tests with test statistic following a χ^2 distribution. It can be shown that (7) belongs to the class of likelihood ratio (LR) tests, where the test statistic is equivalent to the ratio

$$\frac{\max L_0(x)}{\max L_{MS}(x, \nabla)}. \tag{10}$$

Here, L_0 and L_{MS} denote the likelihood functions of the null and alternative model, respectively, cf. Teunissen ([7], p. 53), Kargoll [8]. (Two test statistics are said to be equivalent, if they always define the same critical region and, therefore, bring about the same decision. A sufficient condition is that one test statistic is a monotone function of the other. In this case, either both or none exceed their critical value referring to the same α .) The LR test is very common in statistics for the definition of a test statistic, because only a few very simple test problems permit the construction of a UMP test. A justification for this definition is provided by the famous Neyman-Pearson lemma, cf. Neyman and Pearson [20].

3.2. Variance Inflation Model

For the test problem (6a) versus (6b), no UMP test exists. Therefore, we also resort to the LR test here. We start by setting up the likelihood functions of the null and alternative model. For the null model (6a), the likelihood function reads

$$L_0(x) = (2\pi)^{-\frac{n}{2}} \prod_{i=1}^n \sigma_i^{-1} \exp \left\{ -\frac{\Omega_0}{2} \right\}, \tag{11a}$$

$$\Omega_0 := \sum_{i=1}^n \frac{(y_i - a_i x)^2}{\sigma_i^2}, \tag{11b}$$

and for the alternative model (6b) the likelihood function is given by

$$L_{VI}(x, \tau_1, \dots, \tau_m) = (2\pi)^{-\frac{n}{2}} \prod_{i=1}^n \sigma_i^{-1} \prod_{i=1}^m \tau_i^{-\frac{1}{2}} \exp \left\{ -\frac{\Omega_{VI}}{2} \right\}, \tag{12a}$$

$$\Omega_{VI} := \sum_{i=1}^m \frac{(y_i - a_i x)^2}{\sigma_i^2 \tau_i} + \sum_{i=m+1}^n \frac{(y_i - a_i x)^2}{\sigma_i^2}, \tag{12b}$$

where $a_i, i = 1, \dots, n$ are the row vectors of A . According to (10), the likelihood ratio reads

$$\frac{\max L_0(x)}{\max L_{VI}(x, \tau_1, \dots, \tau_m)} = \frac{\max \exp \left\{ -\frac{\Omega_0}{2} \right\}}{\max \prod_{i=1}^m \tau_i^{-\frac{1}{2}} \exp \left\{ -\frac{\Omega_{VI}}{2} \right\}}. \tag{13}$$

Equivalently, we might use the double negative logarithm of the likelihood ratio as test statistic, because it brings about the same decision as the likelihood ratio itself, i.e.,

$$\begin{aligned} T_{VI} &:= -2 \log \frac{\max \exp \left\{ -\frac{\Omega_0}{2} \right\}}{\max \prod_{i=1}^m \tau_i^{-\frac{1}{2}} \exp \left\{ -\frac{\Omega_{VI}}{2} \right\}} \\ &= \min \Omega_0 - \min \left\{ \Omega_{VI} + \sum_{i=1}^m \log \tau_i \right\}. \end{aligned} \tag{14}$$

The first minimization result is the well-known least squares solution (8). The second minimization must be performed not only with respect to x , but also with respect to the unknown variance inflation factors τ_1, \dots, τ_m . The latter yield the necessary conditions

$$\hat{\tau}_i = \frac{(y_i - a_i \hat{x}_{VI})^2}{\sigma_i^2} = \frac{\hat{e}_{VI,i}^2}{\sigma_i^2}, \quad i = 1, \dots, m. \tag{15}$$

This means that τ_1, \dots, τ_m are estimated, such that the first m residuals in the VI model $\hat{e}_{VI,i}$ equal in magnitude their inflated standard deviations $\sigma_i\sqrt{\hat{\tau}_i}$, and the subtrahend in (14) is obtained by

$$\begin{aligned} \min \left\{ \Omega_{VI} + \sum_{i=1}^m \log \tau_i \right\} &= \min \left\{ m + \sum_{i=1}^m \log \frac{(y_i - a_i x_{VI})^2}{\sigma_i^2} + \sum_{i=m+1}^n \frac{(y_i - a_i x_{VI})^2}{\sigma_i^2} \right\} \\ &= \min \left\{ m + \sum_{i=1}^m \log \frac{e_{VI,i}^2}{\sigma_i^2} + \sum_{i=m+1}^n \frac{e_{VI,i}^2}{\sigma_i^2} \right\}. \end{aligned} \tag{16}$$

In the latter expression the minimum is to be found only with respect to the free parameter vector x_{VI} . This expression differs from $\min \Omega_0$ essentially by the logarithm of the first m normalized residuals. This means that those summands are down-weighted, whenever the residuals $\hat{e}_{VI,i}$ are larger in magnitude than their non-inflated standard deviations σ_i . The necessary conditions for x_{VI} are obtained by nullifying the first derivatives of (16) and read

$$0 = \sum_{i=1}^m \frac{a_{ij}}{y_i - a_i \hat{x}_{VI}} + \sum_{i=m+1}^n \frac{y_i - a_i \hat{x}_{VI}}{\sigma_i^2} a_{ij}, \quad j = 1, \dots, u, \tag{17}$$

where a_{ij} denotes the j -th element of a_i .

This system of equations can be rewritten as a system of polynomials of degree $m + 1$ in the parameters $\hat{x}_{VI,i}$. In general, the solution for \hat{x}_{VI} must be executed by a numerical procedure. This extra effort is certainly a disadvantage of the VI model.

Another disadvantage is that (14) does not follow a well known probability distribution, which complicates the computation of the critical value, being the quantile of this distribution. Such a computation is best performed by Monte Carlo integration, according to Lehmann [21].

Note that the likelihood function L_{VI} in (12) has poles at $\tau_i = 0, i = 1, \dots, m$. These solutions must be excluded from consideration, because they belong to minima of (12) or equivalently to maxima of (16). (Note that $\log \tau_i$ is dominated by $1/\tau_i$ at $\tau_i \rightarrow 0$).

A special issue in the VI model is what to do if $\max \hat{\tau}_i \leq 1$ is found in (15). In this case, the variance is not inflated, such that H_0 must not be rejected in favour of H_{VI} , see (6b). However, it might happen that, nonetheless, T_{VI} in (14) exceeds its critical value, especially if α is large. In order to prevent this behaviour, we modify (14) by

$$T_{VI}(y) := \begin{cases} 0 & \text{if } \max \hat{\tau}_i \leq 1, \\ \min \Omega_0 - \min \{ \Omega_{VI} + \sum_{i=1}^m \log \tau_i \} & \text{otherwise.} \end{cases} \tag{18}$$

If H_0 is true, then there is a small probability that $\max \hat{\tau}_i > 1$ and, consequently, $T_{VI} > 0$ arises, i.e.,

$$\Pr(T_{VI} > 0 | H_0) =: \alpha_{\max}. \tag{19}$$

We see that a type 1 error cannot be required more probable than this α_{\max} , i.e., contrary to the MS model, there is an upper limit for the choice of α .

Even more a problem is what to do, if $\min \hat{\tau}_i < 1 < \max \hat{\tau}_i$ is found in (15). Our argument is that, in this case, H_0 should be rejected, but possibly not in favour of H_{VI} in (6b). A more suitable alternative hypothesis should be found in the framework of a multiple test.

4. Repeated Observations

There is one case, which permits an analytical treatment, even of the VI model, i.e., when one scalar parameter x is observed directly n times, such that we obtain $A = (1, \dots, 1)^T =: \mathbf{1}$. By transformation

of the observations, also all other models with $u = 1$ can be mapped to this case. For compact notation, we define the weighted means of all observations and of only the last $n - m$ inlying observations, i.e.,

$$w := \frac{\sum_{i=1}^n y_i \sigma_i^{-2}}{\sum_{i=1}^n \sigma_i^{-2}}, \tag{20a}$$

$$W := \frac{\sum_{i=m+1}^n y_i \sigma_i^{-2}}{\sum_{i=m+1}^n \sigma_i^{-2}}. \tag{20b}$$

By covariance propagation, the related variances of those expressions are obtained, i.e.,

$$\sigma_w^2 = \frac{1}{\sum_{i=1}^n \sigma_i^{-2}}, \tag{21a}$$

$$\sigma_W^2 = \frac{1}{\sum_{i=m+1}^n \sigma_i^{-2}}. \tag{21b}$$

Having the following useful identities

$$\frac{w}{\sigma_w^2} = \frac{W}{\sigma_W^2} + \sum_{i=1}^m \frac{y_i}{\sigma_i^2}, \tag{22a}$$

$$\frac{1}{\sigma_w^2} = \frac{1}{\sigma_W^2} + \sum_{i=1}^m \frac{1}{\sigma_i^2}, \tag{22b}$$

$$\begin{aligned} w - W &= \sigma_w^2 \left(\frac{W}{\sigma_W^2} - \frac{W}{\sigma_w^2} \right) + \sigma_w^2 \sum_{i=1}^m \frac{y_i}{\sigma_i^2} \\ &= \sigma_w^2 \sum_{i=1}^m \frac{y_i - W}{\sigma_i^2}, \end{aligned} \tag{22c}$$

the estimates in the null model (8) can be expressed as

$$\hat{x}_0 = w, \tag{23a}$$

$$\hat{e}_0 = y - \mathbf{1}w, \tag{23b}$$

$$\Sigma_{\hat{e}_0} = \Sigma - \mathbf{1}\mathbf{1}^T \sigma_w^2, \tag{23c}$$

and the minimum of the sum of the squared residuals is

$$\min \Omega_0 = \sum_{i=1}^n \frac{(y_i - w)^2}{\sigma_i^2}. \tag{23d}$$

4.1. Mean Shift Model

In the MS model, the first m observations are falsified by bias parameters $\nabla_1, \dots, \nabla_m$. Matrix

$$C = \begin{pmatrix} I \\ 0 \end{pmatrix} \tag{24}$$

in (4) is a block matrix of the $m \times m$ identity matrix and a $(n - m) \times m$ null matrix. Maximizing the likelihood function yields the estimated parameters and residuals, i.e.,

$$\hat{x}_{MS} = W, \tag{25a}$$

$$\hat{e}_{MS} = y - \mathbf{1}W, \tag{25b}$$

respectively, as well as the estimated bias parameters and their related covariance matrix, i.e.,

$$\hat{\nabla} = \begin{pmatrix} y_1 - W \\ \vdots \\ y_m - W \end{pmatrix} = \begin{pmatrix} \hat{\epsilon}_{MS,1} \\ \vdots \\ \hat{\epsilon}_{MS,m} \end{pmatrix}, \tag{25c}$$

$$\Sigma_{\hat{\nabla}} = \begin{pmatrix} \sigma_1^2 + \sigma_W^2 & \sigma_W^2 & \dots & \sigma_W^2 \\ \sigma_W^2 & \sigma_2^2 + \sigma_W^2 & \dots & \sigma_W^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_W^2 & \sigma_W^2 & \dots & \sigma_m^2 + \sigma_W^2 \end{pmatrix}, \tag{25d}$$

respectively. Note that (25d) is obtained by covariance propagation that was applied to (25c). By applying the Sherman—Morrison formula, cf. Sherman and Morrison [22], the inverse matrix of $\Sigma_{\hat{\nabla}}$ is obtained,

$$\begin{aligned} \Sigma_{\hat{\nabla}}^{-1} &= \left[\begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_m^2 \end{pmatrix} + \sigma_W^2 \mathbf{1}\mathbf{1}^T \right]^{-1} \\ &= \begin{pmatrix} \sigma_1^{-2} & 0 & \dots & 0 \\ 0 & \sigma_2^{-2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_m^{-2} \end{pmatrix} - \frac{\sigma_W^2}{1 + \sigma_W^2 \sum_{i=1}^m \sigma_i^{-2}} \begin{pmatrix} \sigma_1^{-2} \\ \sigma_2^{-2} \\ \vdots \\ \sigma_m^{-2} \end{pmatrix} \begin{pmatrix} \sigma_1^{-2} \\ \sigma_2^{-2} \\ \vdots \\ \sigma_m^{-2} \end{pmatrix}^T \\ &= \begin{pmatrix} \sigma_1^{-2} & 0 & \dots & 0 \\ 0 & \sigma_2^{-2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_m^{-2} \end{pmatrix} - \sigma_w^2 \begin{pmatrix} \sigma_1^{-4} & \sigma_1^{-2}\sigma_2^{-2} & \dots & \sigma_1^{-2}\sigma_m^{-2} \\ \sigma_1^{-2}\sigma_2^{-2} & \sigma_2^{-4} & \dots & \sigma_2^{-2}\sigma_m^{-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1^{-2}\sigma_m^{-2} & \sigma_2^{-2}\sigma_m^{-2} & \dots & \sigma_m^{-4} \end{pmatrix}, \tag{26} \end{aligned}$$

and the test statistic (7) in the MS model becomes

$$T_{MS}(y) = \sum_{i=1}^m \frac{\hat{\epsilon}_{MS,i}^2}{\sigma_i^2} - \sigma_w^2 \sum_{i=1}^m \sum_{j=1}^m \frac{\hat{\epsilon}_{MS,i}\hat{\epsilon}_{MS,j}}{\sigma_i^2\sigma_j^2}. \tag{27}$$

According to (9), the distributions of the null model and the alternative model are given by

$$T_{MS}|H_0 \sim \chi^2(m, 0), \tag{28a}$$

$$T_{MS}|H_{MS} \sim \chi^2(m, \lambda), \tag{28b}$$

respectively, where the non-centrality parameter reads

$$\lambda = \sum_{i=1}^m \frac{\nabla_i^2}{\sigma_i^2} - \sigma_w^2 \sum_{i=1}^m \sum_{j=1}^m \frac{\nabla_i \nabla_j}{\sigma_i^2 \sigma_j^2}. \tag{29}$$

For the special cases of $m = 1$ and $m = 2$ extra bias parameters, as well as the case of independent and identically distributed random observation errors, the related test statistics (27) are given by

Case $m = 1$:

$$T_{MS}(y) = \frac{\hat{e}_{MS,1}^2}{\sigma_1^2 + \sigma_w^2}, \tag{30}$$

Case $m = 2$:

$$T_{MS}(y) = \hat{e}_{MS,1}^2 \frac{\sigma_1^2 - \sigma_w^2}{\sigma_1^4} + \hat{e}_{MS,2}^2 \frac{\sigma_2^2 - \sigma_w^2}{\sigma_2^4} - 2\sigma_w^2 \frac{\hat{e}_{MS,1}\hat{e}_{MS,2}}{\sigma_1^2\sigma_2^2}, \tag{31}$$

Case $\sigma_1 = \sigma_2 = \dots = \sigma_m =: \sigma$:

$$T_{MS}(y) = \frac{1}{\sigma^2} \left(\sum_{i=1}^m \hat{e}_{MS,i}^2 - \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^m \hat{e}_{MS,i}\hat{e}_{MS,j} \right). \tag{32}$$

In the case of $\sigma_w \ll \min \sigma_i$, which often arises when $m \ll n$, the test statistic tends to

$$T_{MS}(y) \rightarrow \sum_{i=1}^m \frac{\hat{e}_{MS,i}^2}{\sigma_i^2}. \tag{33}$$

4.2. Variance Inflation Model—General Considerations

In the VI model, the first m observations are falsified by variance inflation factors τ_1, \dots, τ_m . The necessary condition (17) reads

$$0 = \sum_{i=1}^m \frac{1}{y_i - \hat{x}_{VI}} + \sum_{i=m+1}^n \frac{y_i - \hat{x}_{VI}}{\sigma_i^2}. \tag{34}$$

Using (20b) and (21b), this can be rewritten to

$$\hat{x}_{VI} - W = \sum_{i=1}^m \frac{\sigma_w^2}{y_i - \hat{x}_{VI}}. \tag{35}$$

This solution \hat{x}_{VI} is obtained as the real root of a polynomial of degree $m + 1$, which might have, at most $m + 1$, real solutions. In the model, it is easy to exclude the case that $y_i = y_j, i \neq j$, because they are either both outliers or both good observations. They should be merged into one observation. Let us index the observations, as follows: $y_1 < y_2 < \dots < y_m$. We see that

- in the interval $-\infty \dots y_1$ of \hat{x}_{VI} the right hand side of (35) goes from 0 to $+\infty$,
- in each interval $y_{i-1} \dots y_i$ it goes from $-\infty$ to $+\infty$, and
- in the interval $y_m \dots +\infty$ it goes from $-\infty$ to 0.
- The left hand side of (35) is a straight line.

Therefore, (35) has always at least one real solution \hat{x}_{VI} in each interval $y_{i-1} \dots y_i$, where one of them must be a maximum of (16), because (16) goes from $-\infty$ up to some maximum and then down again to $-\infty$ in this interval. Besides these $m - 1$ uninteresting solutions, (35) can have no more or two more real solutions, except in rare cases, where it might have one more real solution. If $W < y_1$, then there are no solutions above y_m . If $W > y_m$, then there are no solutions below y_1 .

From these considerations it becomes clear that (35) can have, at most, one solution that is a minimum of (16), see also Figure 3.

The second-order sufficient condition for a strict local minimum of (16) is that the Hessian matrix H of (16),

$$H(x_{VI}, \tau_1, \dots, \tau_m) = 2 \begin{pmatrix} \frac{1}{\sigma_W^2} + \sum_{i=1}^m \frac{1}{\tau_i \sigma_i^2} & \frac{y_1 - x_{VI}}{\tau_1^2 \sigma_1^2} & \dots & \frac{y_m - x_{VI}}{\tau_m^2 \sigma_m^2} \\ \frac{y_1 - x_{VI}}{\tau_1^2 \sigma_1^2} & \frac{(y_1 - x_{VI})^2}{\tau_1^3 \sigma_1^2} - \frac{1}{2\tau_1^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{y_m - x_{VI}}{\tau_m^2 \sigma_m^2} & 0 & \dots & \frac{(y_m - x_{VI})^2}{\tau_m^3 \sigma_m^2} - \frac{1}{2\tau_m^2} \end{pmatrix}, \tag{36}$$

must be positive-definite at $\hat{x}_{VI}, \hat{\tau}_1, \dots, \hat{\tau}_m$, cf. Nocedal and Wright ([23] p.16), i.e.,

$$H(\hat{x}_{VI}, \hat{\tau}_1, \dots, \hat{\tau}_m) = 2 \begin{pmatrix} \frac{1}{\sigma_W^2} + \sum_{i=1}^m \frac{1}{\hat{\tau}_i \sigma_i^2} & \frac{y_1 - \hat{x}_{VI}}{\hat{\tau}_1^2 \sigma_1^2} & \dots & \frac{y_m - \hat{x}_{VI}}{\hat{\tau}_m^2 \sigma_m^2} \\ \frac{y_1 - \hat{x}_{VI}}{\hat{\tau}_1^2 \sigma_1^2} & \frac{1}{2\hat{\tau}_1^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{y_m - \hat{x}_{VI}}{\hat{\tau}_m^2 \sigma_m^2} & 0 & \dots & \frac{1}{2\hat{\tau}_m^2} \end{pmatrix}. \tag{37}$$

must be a positive definite matrix. A practical test for positive definiteness that does not require explicit calculation of the eigenvalues is the principal minor test, also known as Sylvester’s criterion. The k -th leading principal minor is the determinant that is formed by deleting the last $n - k$ rows and columns of the matrix. A necessary and sufficient condition that a symmetric $n \times n$ matrix is positive definite is that all n leading principal minors are positive, cf. Prussing [24], Gilbert [25]. Invoking Schur’s determinant identity, i.e.,

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(D) \det(A - BD^{-1}C) \tag{38}$$

and in combination with (15), we see that the k -th leading principal minor of H in (37) is

$$\prod_{i=1}^k \left(\frac{1}{2\hat{\tau}_k^2} \right) \left(\frac{1}{\sigma_W^2} + \sum_{i=1}^m \frac{1}{\hat{\tau}_i \sigma_i^2} - \sum_{i=1}^k \frac{2}{\hat{\tau}_i \sigma_i^2} \right). \tag{39}$$

To be positive, the second factor must be ensured to be positive for each k . Obviously, if this is true for $k = m$, it is also true for all other k . Therefore, the necessary and sufficient condition for a local minimum of (16) reads

$$\sum_{i=1}^m \frac{\sigma_W^2}{\hat{\tau}_i \sigma_i^2} = \sum_{i=1}^m \frac{\sigma_W^2}{(y_i - \hat{x}_{VI})^2} < 1. \tag{40}$$

In other words, if and only if \hat{x}_{VI} is sufficiently far away from all outlying observations, it belongs to a strict local minimum of (16).

Using (23d),(16), the test statistic (18) in the VI model for $\max \hat{\tau}_i > 1$ reads

$$\begin{aligned} T_{VI} &= \min \Omega_0 - \min \left\{ \Omega_{VI} + \sum_{i=1}^m \log \tau_i \right\} \\ &= \sum_{i=1}^n \frac{(y_i - w)^2}{\sigma_i^2} - m - \sum_{i=1}^m \log \frac{(y_i - \hat{x}_{VI})^2}{\sigma_i^2} - \sum_{i=m+1}^n \frac{(y_i - \hat{x}_{VI})^2}{\sigma_i^2} \\ &= \sum_{i=1}^n \left(\frac{(y_i - w)^2}{\sigma_i^2} - \frac{(y_i - \hat{x}_{VI})^2}{\sigma_i^2} \right) - m + \sum_{i=1}^m \left(\frac{(y_i - \hat{x}_{VI})^2}{\sigma_i^2} - \log \frac{(y_i - \hat{x}_{VI})^2}{\sigma_i^2} \right) \\ &= (\hat{x}_{VI} - w) \sum_{i=1}^n \frac{2y_i - \hat{x}_{VI} - w}{\sigma_i^2} - m + \sum_{i=1}^m (\hat{\tau}_i - \log \hat{\tau}_i) \\ &= -\frac{(\hat{x}_{VI} - w)^2}{\sigma_w^2} - m + \sum_{i=1}^m (\hat{\tau}_i - \log \hat{\tau}_i). \end{aligned} \tag{41}$$

Tracing back the flow sheet of computations, it becomes clear that T_{VI} depends on the observations only through y_1, \dots, y_m and w or equivalently through y_1, \dots, y_m and W . These $m + 1$ quantities represent a so-called "sufficient statistic" for the outlier test.

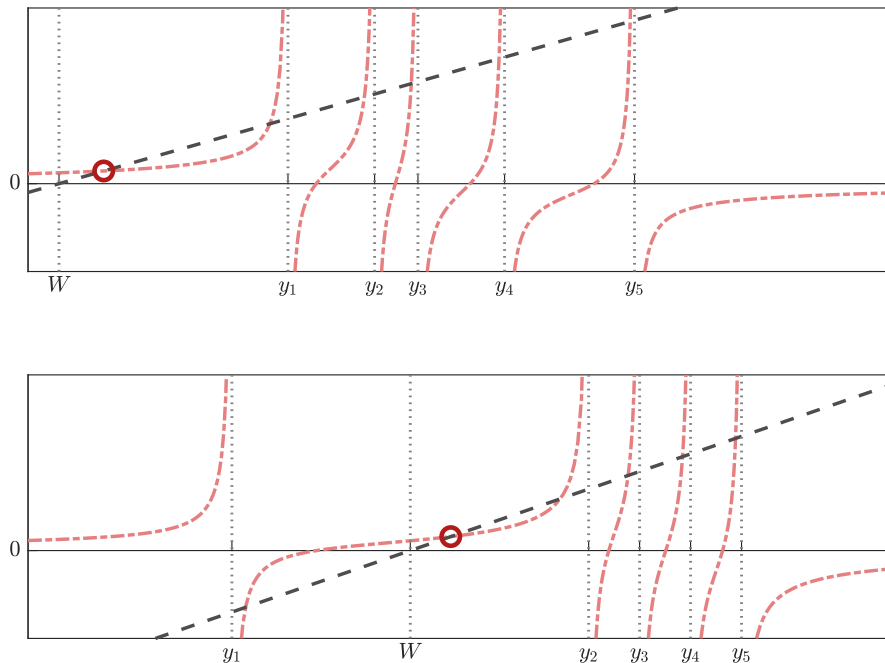


Figure 3. Illustration of two exemplary cases of the solution of (35). Whereas the dash-dot red styled lines indicate the right hand side of the function, the black colored dashed line depicts the left hand side of (35). A dark red circle symbolises the desired solution out of five.

An interesting result is obtained, when we consider the case $\sigma_W \rightarrow 0$, which occurs if the $n - m$ good observations contain much more information than the m suspected outliers. In this case, (35) can be rewritten as

$$(\hat{x}_{VI} - W) \prod_{i=1}^m (y_i - \hat{x}_{VI}) = \sigma_W^2 P(y_1, \dots, y_m, \hat{x}_{VI}). \tag{42}$$

where P is some polynomial. If the right hand side goes to zero, at least one factor of the left hand side must also go to zero. As $\sigma_W \rightarrow 0$, we obtain $m + 1$ solutions for \hat{x}_{VI} , approaching W, y_1, \dots, y_m . The first solution can be a valid VI solution, the others are invalid as $\hat{\tau}_i \rightarrow 0$. Note that we have $\hat{x}_{VI} \rightarrow \hat{x}_{MS}$ in this case, and also $\hat{e}_{VI,i} \rightarrow \hat{e}_{MS,i}$ for all $i = 1, \dots, m$. Having in mind (22c), we see that (41) becomes

$$T_{VI} \rightarrow - \frac{(W - w)^2}{\sigma_w^2} - m + \sum_{i=1}^m (\hat{\tau}_i - \log \hat{\tau}_i) \tag{43}$$

$$= \sigma_w^2 \left(\sum_{i=1}^m \frac{y_i - W}{\sigma_i^2} \right)^2 - m + \sum_{i=1}^m (\hat{\tau}_i - \log \hat{\tau}_i) \tag{44}$$

and also noting that with (22b) we find $\sigma_w \rightarrow 0$, such that

$$T_{VI} \rightarrow -m + \sum_{i=1}^m (\hat{\tau}_i - \log \hat{\tau}_i) \tag{45}$$

$$= -m + \sum_{i=1}^m \frac{\hat{e}_{MS,i}^2}{\sigma_i^2} - \log \frac{\hat{e}_{MS,i}^2}{\sigma_i^2}. \tag{46}$$

When comparing this to the equivalent result in the MS model (33), we see that T_{VI} and T_{MS} are equivalent test statistics under the sufficient condition $\min \hat{\tau}_i > 1$, because $\tau - \log \tau$ is a monotonic function for $\tau > 1$. This means that, in this case, the decision on H_0 is the same, both in the MS and in the VI model. However, $\max \hat{\tau}_i > 1$ might not be sufficient for this property.

4.3. Variance Inflation Model—Test for One Outlier

In the case $m = 1$ (35) reads

$$\hat{x}_{VI} = \frac{\sigma_W^2}{y_1 - \hat{x}_{VI}} + W \tag{47}$$

Rewriting this to a quadratic equation yields up to two solutions, i.e.,

$$\hat{x}_{VI} = \frac{y_1 + W}{2} \pm \sqrt{\frac{(y_1 - W)^2}{4} - \sigma_W^2}. \tag{48}$$

With (15), we find

$$\hat{\tau}_1 = \frac{1}{\sigma_1^2} \left(\frac{y_1 - W}{2} \pm \sqrt{\frac{(y_1 - W)^2}{4} - \sigma_W^2} \right)^2 = \frac{1}{\sigma_1^2} \left(\frac{\hat{e}_{MS,1}}{2} \pm \sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_W^2} \right)^2. \tag{49}$$

For a solution to exist at all, we must have $|\hat{e}_{MS,1}| \geq 2\sigma_W$. This means that y_1 must be sufficiently outlying, otherwise H_0 is to be accepted.

The condition for a strict local maximum (40) reads here

$$\sigma_W^2 < \hat{\tau}_1 \sigma_1^2 = \left(\frac{\hat{e}_{MS,1}}{2} \pm \sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_W^2} \right)^2. \tag{50}$$

For the sign of the square root equal to the sign of $\hat{e}_{MS,1}$, this inequality is trivially fulfilled. For the opposite sign, we require

$$\sigma_W < \frac{|\hat{e}_{MS,1}|}{2} - \sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_W^2}. \tag{51}$$

Rewriting this expression yields

$$\sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_W^2} < \frac{|\hat{e}_{MS,1}|}{2} - \sigma_W \tag{52}$$

and squaring both sides, which can be done because they are both positive, we readily arrive at $|\hat{e}_{MS,1}| < 2\sigma_w$, which is the case that no solution exists. Therefore, we have exactly one minimum of (16), i.e.,

$$\begin{aligned} \hat{x}_{VI} &= \frac{y_1 + W}{2} - \text{sign}(y_1 - W) \sqrt{\frac{(y_1 - W)^2}{4} - \sigma_w^2} \\ &= \frac{y_1 + W}{2} - \text{sign}(\hat{e}_{MS,1}) \sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_w^2}, \end{aligned} \tag{53a}$$

$$\hat{\tau}_1 = \frac{1}{\sigma_1^2} \left(\frac{\hat{e}_{MS,1}}{2} + \text{sign}(\hat{e}_{MS,1}) \sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_w^2} \right)^2, \tag{53b}$$

and the test statistic (18) becomes

$$T_{VI} = \begin{cases} 0 & \text{if } \hat{\tau}_1 \leq 1, \\ -\frac{(\hat{x}_{VI} - w)^2}{\sigma_w^2} - 1 + \hat{\tau}_1 - \log \hat{\tau}_1 & \text{otherwise.} \end{cases} \tag{54}$$

The condition $\hat{\tau}_1 > 1$ is equivalent to

$$\sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_w^2} > \sigma_1 - \frac{|\hat{e}_{MS,1}|}{2}, \tag{55}$$

which is trivially fulfilled, if the right hand side is negative. If it is non-negative, both sides can be squared and rearranged to

$$|\hat{e}_{MS,1}| > \frac{\sigma_1^2 + \sigma_w^2}{\sigma_1}. \tag{56}$$

Since this condition also covers the case that $|\hat{e}_{MS,1}| > 2\sigma_1$, it can be used exclusively as an equivalent of $\hat{\tau}_1 > 1$.

With (22c) we see that both

$$\begin{aligned} \hat{x}_{VI} - w &= \frac{y_1 - W}{2} - \text{sign}(\hat{e}_{MS,1}) \sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_w^2} - \sigma_w \frac{y_1 - W}{\sigma_1^2} \\ &= \hat{e}_{MS,1} \left(\frac{1}{2} - \frac{\sigma_w^2}{\sigma_1^2} \right) - \text{sign}(\hat{e}_{MS,1}) \sqrt{\frac{\hat{e}_{MS,1}^2}{4} - \sigma_w^2} \end{aligned} \tag{57}$$

as well as $\hat{\tau}_1$ through (53b) depend on the observations only through $\hat{e}_{MS,1}$, and so does T_{VI} in (54). On closer examination, we see that T_{VI} in (54) depends even only on $|\hat{e}_{MS,1}|$. This clearly holds as well for T_{MS} in (30). Therefore, both test statistics are equivalent if T_{VI} can be shown to be a strictly monotone function of T_{MS} .

Figure 4 shows that (54) as a function of $\hat{e}_{MS,1}$ is monotone. A mathematical proof of monotony is given in the Appendix A. Thus, it is also monotone as a function of T_{MS} and even strictly monotone for $\hat{\tau}_1 > 1$, which is the case that we are interested in. Therefore, the MS model and the VI model are fully equivalent for repeated observations to be tested for $m = 1$ outlier.

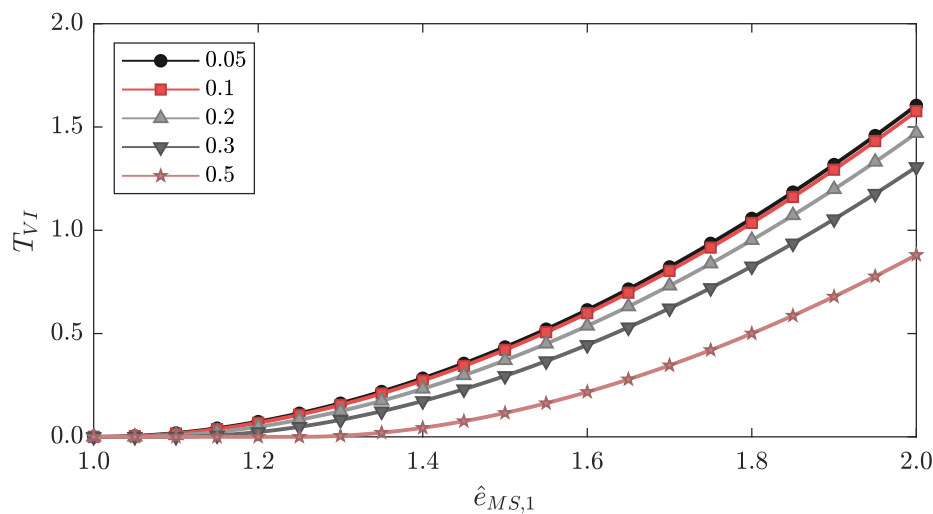


Figure 4. Resulting test statistic T_{VI} of the variance inflation (VI) approach (54) as a function of the residual $\hat{e}_{MS,1}$ for various ratios of σ_w/σ_1 .

Finally, we numerically determine the probability distribution of test statistic (54) using Monte Carlo integration by

- defining the ratio σ_w/σ_1 ,
- generating normally distributed pseudo random numbers for $e_{MS,1}$,
- evaluating (53) and (54), and
- taking the histogram of (54),

using 10^7 pseudo random samples of $\hat{e}_{MS,1}$. In Figure 5, the positive branch of the symmetric probability density function (PDF) is given in logarithmic scale for various ratios of σ_w/σ_1 . However, only about 30% of the probability mass is located under this curve, the rest is concentrated at $T_{VI} = 0$, and is not displayed. The quantiles of this distribution determine critical values and also α_{max} in (19). The results are summarized in Table 1.

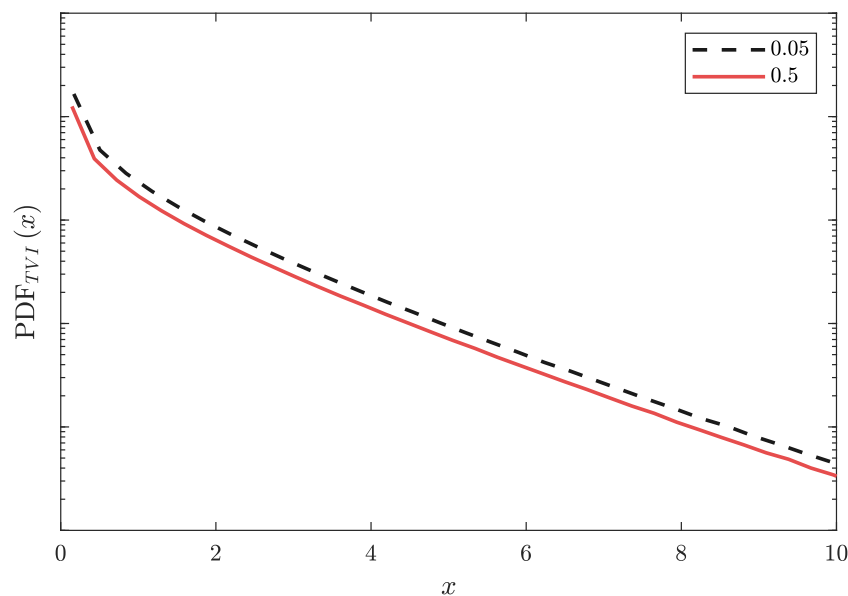


Figure 5. Probability density function of T_{VI} (54) as a function of the residual $\hat{e}_{MS,1}$ under H_0 , approximated by MCM. Vertical axis is in logarithmic scale. The strong peaks at 0 for $\hat{\tau}_1 \leq 1$ are not displayed. The dashed black and the solid red curve relates to $\sigma_w/\sigma_1 = 0.05$ and $\sigma_w/\sigma_1 = 0.5$, respectively.

Table 1. Maximum selectable type 1 error probability α_{\max} and critical value c_α for T_{VI} in (54) for various ratios of σ_w/σ_1 and $\alpha = 0.05$.

σ_w/σ_1	α_{\max}	c_α
0.05	0.317	1.50
0.10	0.315	1.49
0.20	0.308	1.47
0.30	0.297	1.43
0.50	0.264	1.33

4.4. Variance Inflation Model—Test for Two Outliers

In the case $m = 2$ outliers, (35) reads

$$\hat{x}_{VI} - W = \frac{\sigma_w^2}{y_1 - \hat{x}_{VI}} + \frac{\sigma_w^2}{y_2 - \hat{x}_{VI}}. \tag{58}$$

The solution for \hat{x}_{VI} can be expressed in terms of a cubic equation, which permits an analytical solution. One real solution must be in the interval $y_1 \dots y_2$, but there may be two more solutions

- both below y_1 , if $W < y_1$, or
- both above y_2 , if $W > y_2$, or
- both between y_1 and y_2 , if $y_1 > W > y_2$.

In rare cases, solutions may also coincide. The analytical expressions are very complicated and they do not permit a treatment analogous to the preceding subsection. Therefore, we have to fully rely on numerical methods, which, in our case, is the Monte Carlo method (MCM).

First, we compute the critical values of the test statistic (18)

$$T_{VI} = \begin{cases} 0 & \text{if } \max(\hat{\tau}_1, \hat{\tau}_2) \leq 1, \\ -\frac{(\hat{x}_{VI}-w)^2}{\sigma_w^2} - 2 + \hat{\tau}_1 + \hat{\tau}_2 - \log \hat{\tau}_1 - \log \hat{\tau}_2 & \text{otherwise.} \end{cases} \tag{59}$$

The MCM is preformed, using 10^7 pseudo random samples. We restrict ourselves to the case $\sigma_1 = \sigma_2$. The maximum selectable type 1 error probabilities α_{\max} are summarized in Table 2. It is shown that α_{\max} is mostly larger than for $m = 1$. The reason is that, more often, we obtain $\hat{\tau}_i > 1$, even if H_0 is true, which makes it easier to define critical values in a meaningful way. Moreover, Table 2 indicates the probabilities that under H_0

- (16) has no local minimum, and if it has, that
- $\max(\hat{\tau}_1, \hat{\tau}_2) \leq 1$
- $\hat{\tau}_1 \leq 1, \hat{\tau}_2 > 1$ or vice versa
- $\min(\hat{\tau}_1, \hat{\tau}_2) > 1$

i.e., none, one, or both variances are inflated. It is shown that, if the good observations contain the majority of the information, a minimum exists, but, contrary to our expectation, the case $\max(\hat{\tau}_1, \hat{\tau}_2) \leq 1$ is not typically the dominating case.

The important result is what happens, if H_0 is false, because variances are truly inflated. The probability that H_0 is rejected is known as the power $1 - \beta$ of the test, where β is the probability of a type 2 decision error. It is computed both with T_{MS} in (31) as well as with T_{VI} in (59). Table 3 provides the results. It is shown that the power of T_{MS} is always better than of T_{VI} . This is unexpected, because T_{MS} is not equivalent to the likelihood ratio of the VI model.

A possible explanation of the low performance of T_{VI} in (59) is that, in many cases, the likelihood function L_{VI} has no local maximum, such that (16) has no local minimum. Even for an extreme variance inflation of $\tau_1 = \tau_2 = 5$ this occurs with remarkable probability of 0.14. Moreover, the probability that $\max(\hat{\tau}_1, \hat{\tau}_2) \leq 1$ is hardly less than that. In both cases, H_0 cannot be rejected.

Table 2. Maximum selectable type 1 error probability α_{\max} and critical value c_α for T_{VI} in (59) for various ratios of $\sigma_W/\sigma_1 = \sigma_W/\sigma_2$ and $\alpha = 0.05$ as well as probabilities that (16) has no local minimum or that 0 or 1 or 2 variances are inflated.

		Probabilities for					
		no		0	1	2	
$\sigma_W/\sigma_1 = \sigma_W/\sigma_2$	α_{\max}	c_α	min	Inflated Variances			
0.01	0.52	4.68	0.03	0.45	0.42	0.10	
0.02	0.51	4.36	0.06	0.42	0.41	0.10	
0.03	0.50	4.11	0.09	0.40	0.40	0.10	
0.05	0.48	3.69	0.15	0.36	0.38	0.10	
0.10	0.43	3.00	0.29	0.27	0.33	0.10	
0.20	0.33	2.32	0.53	0.14	0.24	0.09	
0.30	0.24	1.82	0.69	0.06	0.16	0.09	
0.50	0.12	0.97	0.88	0.01	0.05	0.07	

Table 3. Test power $1 - \beta_{MS}$ for test using T_{MS} in (31) and test power $1 - \beta_{VI}$ for test using T_{VI} in (59) for various true values of the variance inflation factors τ_1, τ_2 for $\sigma_W = 0.1 \cdot \sigma_1 = 0.1 \cdot \sigma_2$ and $\alpha = 0.05$, as well as probabilities that (16) has no local minimum or that 0 or one or two variances are inflated.

		Probabilities for					
		Test Power		no	0	1	2
τ_1	τ_2	$1 - \beta_{MS}$	$1 - \beta_{VI}$	min	Inflated Variances		
1.0	1.0	0.05	0.05	0.29	0.27	0.33	0.10
1.2	1.2	0.08	0.08	0.27	0.24	0.35	0.13
1.5	1.5	0.13	0.12	0.24	0.21	0.38	0.17
2.0	2.0	0.22	0.20	0.21	0.17	0.39	0.23
3.0	3.0	0.36	0.32	0.18	0.12	0.39	0.31
5.0	5.0	0.55	0.49	0.14	0.08	0.36	0.42
1.0	1.5	0.09	0.09	0.27	0.24	0.36	0.13
1.0	3.0	0.21	0.19	0.24	0.18	0.41	0.18
2.0	3.0	0.30	0.26	0.19	0.14	0.40	0.27
2.0	5.0	0.40	0.36	0.18	0.11	0.40	0.31

4.5. Outlier Identification

If it is not known which observations are outlier-suspected, a multiple test must be set up. If the critical values are identical in all tests, then we simply have to look for the largest test statistic. This is the case for T_{MS} when considering the same number m of outlier-suspected observations, see (9). If we even consider different numbers m in the same multiple test, we have to apply the p -value approach, cf. Lehmann and Lösler [5].

In the VI model, the requirement of identical critical values is rarely met. It is, in general, not met for repeated observations, not even for $m = 1$, as can be seen in Figure 5. However, in this case, it is no problem, because the test with T_{VI} in (54) is equivalent to T_{MS} in (30), as demonstrated. This also means that the same outlier is identified with both test statistics.

For repeated observations, we find identical critical values only for identical variances of the outlier-suspected observations, such that those observations are fully indistinguishable from each other. For example, for $n = 27, m = 2, \alpha = 0.05$, and $\sigma_1 = \dots = \sigma_n$, we find $\sigma_W/\sigma_i = 0.20$ and $c_\alpha = 2.32$ for all 351 pairs of outlier-suspected observations, see Table 2.

We evaluate the identifiability of two outliers in $n = 10$ and $n = 20$ repeated observations with $m = 2$ outliers while using the MCM. In each of the 10^6 repetitions, random observations are generated having equal variances. Two cases are considered. Whereas, in the first case, two variances are inflated by τ according to the VI model, in the second case, two observation values are shifted by ∇ according to the MS model. Using (31) and (59), the test statistics T_{MS} and T_{VI} are computed

for all $n(n - 1)/2 = 45$ or 190 pairs of observations. If the maximum of the test statistic is attained for the actually modified pair of observations, the test statistic correctly identifies the outliers. Here, we assume that α is large enough for the critical value to be exceeded, but otherwise the results are independent of the choice of α . The success probabilities are given in Table 4.

Table 4. Success probabilities for outlier identification in repeated observations of equal variance σ^2 with two outliers.

$\tau_1 = \tau_2$	Success Probabilities for $n = 10$		Success Probabilities for $n = 20$	
	of T_{MS} in (31)	of T_{VI} in (59)	of T_{MS} in (31)	of T_{VI} in (59)
1.0	0.022	0.021	0.005	0.005
2.0	0.065	0.061	0.028	0.026
3.0	0.108	0.104	0.061	0.057
4.0	0.150	0.144	0.095	0.089
5.0	0.185	0.180	0.128	0.121
6.0	0.218	0.212	0.156	0.151
$\nabla_1 = \nabla_2$	of T_{MS} in (31)	of T_{VI} in (59)	of T_{MS} in (31)	of T_{VI} in (59)
0.0	0.022	0.021	0.005	0.005
1σ	0.081	0.065	0.035	0.028
2σ	0.325	0.286	0.226	0.202
3σ	0.683	0.652	0.606	0.586
4σ	0.912	0.902	0.892	0.886
5σ	0.985	0.984	0.984	0.984

As expected, the success probabilities increase as τ or ∇ gets large. However, for both cases, T_{MS} outperforms T_{VI} . In Figure 6, the ratio r_T of the success probabilities between the VI and the MS approach is depicted, having $n = 10$ repeated observations. If $r_T > 1$, the success rate of T_{VI} is higher than for T_{MS} and vice versa. The ratio is always $r_T < 1$ and tends to 1, as shown in Figure 6. Therefore, the success probability of the MS is higher than for the VI approach, even if the outliers are caused by an inflation of the variances.

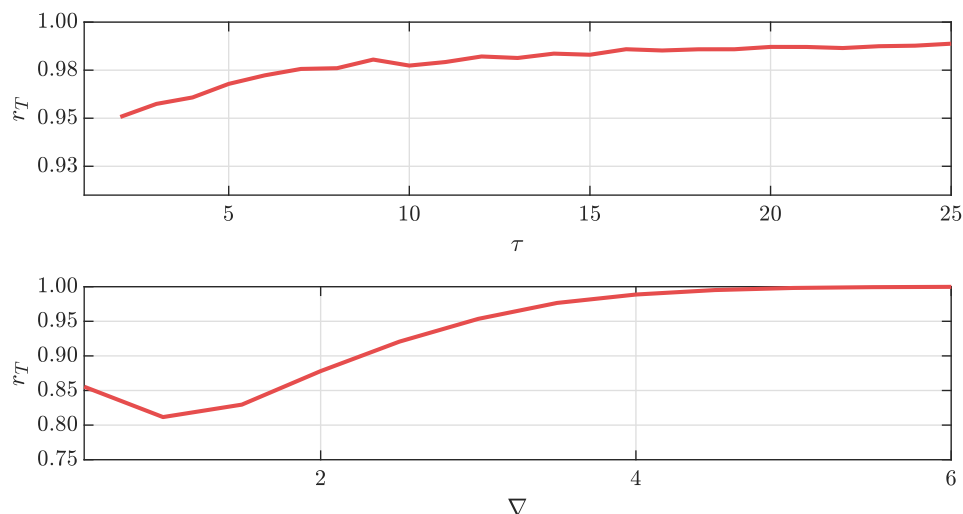


Figure 6. Ratio r_T of the success probabilities between the VI and the MS approach using $n = 10$ repeated observations. The top figure depicts r_T for the case of inflated variances. The bottom figure depicts r_T for the case of shifted mean values.

5. Conclusions

We have studied the detection of outliers in the framework of statistical hypothesis testing. We have investigated two types of alternative hypotheses: the mean shift (MS) hypothesis, where the probability distributions of the outliers are thought of as having a shifted mean, and the variance inflation (VI) model, where they are thought of as having an inflated variance. This corresponds to an outlier-generating process thought of as being deterministic or random, respectively. While the first type of alternative hypothesis is routinely applied in geodesy and in many other disciplines, the second is not. However, even if the VI model might be more adequate to describe the stochastics of the observations, this does not mean that it is possible to estimate parameters or detect outliers better than with some less adequate model, like MS.

The test statistic has been derived by the likelihood ratio of null and alternative hypothesis. This was motivated by the famous Neyman–Pearson lemma, cf. Neyman and Pearson [20], even though that this lemma does not directly apply to this test. Therefore, the performance of the test must be numerically evaluated.

When compared to existing VI approaches, we

- strived for a true (non-restricted) maximization of the likelihood function;
- allowed for multiple outliers;
- fully worked out the case of repeated observations; and,
- computed the corresponding test power by MC method for the first time.

We newly found out that the VI stochastic model has some critical disadvantages:

- the maximization of the likelihood function requires the solution of a system of u polynomial equations of degree $m + 1$, where u is the number of model parameters and m is the number of suspected outliers;
- it is neither guaranteed that the likelihood function actually has such a local maximum, nor that it is unique;
- the maximum might be at a point where some variance is deflated rather than inflated. It is debatable, what the result of the test should be in such a case;
- the critical value of this test must be computed numerically by Monte Carlo integration. This must even be done for each model separately; and,
- there is an upper limit (19) for the choice of the critical value, which may become small in some cases.

For the first time, the application of the VI model has been investigated for the most simple model of repeated observations. It is shown that here the likelihood function admits at most one local maximum, and it does so, if the outliers are strong enough. Moreover, in the limiting case that the suspected outliers represent an almost negligible amount of information, the VI test statistic and the MS test statistic have been demonstrated to be almost equivalent.

For $m = 1$ outlier in the repeated observations, there is even a closed formula (54) for the test statistic, and the existence and uniqueness of a local maximum is equivalent to a simple checkable inequality condition. Additionally, here the VI test statistic and the MS test statistic are equivalent.

In our numerical investigations, we newly found out that for $m > 1$ outliers in the repeated observations the power of the VI test is worse than using the classical MS test statistic. The reason is the lack of a maximum of the likelihood function, even for sizable outliers. Our numerical investigations also show that the identifiability of the outliers is worse for the VI test statistic. This is clearly seen in the case that the outliers are truly caused by shifted means, but also in the other case the identifiability is slightly worse. This means that the correct outliers are more often identified with the MS test statistic.

In the considered cases, we did not find real advantages of the VI model, but this does not prove that they do not exist. As long as such cases are not found, we therefore recommend practically performing outlier detection by the MS model.

Author Contributions: Conceptualization, R.L.; Methodology, R.L.; Software, R.L. and M.L.; Validation, M.L.; Investigation, R.L. and M.L.; Writing—Original Draft Preparation, R.L. and M.L.; Writing—Review & Editing, R.L., M.L., F.N.; Visualization, M.L.; Supervision, F.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

To prove the monotony of the tails of the function $T_{VI}(e_{MS,1})$ in (54), the extreme values of T_{VI} have to be determined. Since T_{VI} is symmetric, it is sufficient to restrict the proof for positive values of $e_{MS,1}$. The first derivation of T_{VI} is given by

$$T'_{VI}(e_{MS,1}) = -\frac{4\sigma_W^2(\sigma_1^2 + 1) - e_{MS,1}^2(\sigma_1^2 + \sigma_W^2) + e_{MS,1}(\sigma_1^2 - \sigma_W^2)\sqrt{e_{MS,1}^2 - 4\sigma_W^2}}{\sigma_W^2(\sigma_1^2 + \sigma_W^2)\sqrt{e_{MS,1}^2 - 4\sigma_W^2}}. \tag{A1}$$

Setting $T'_{VI}(e_{MS,1}) = 0$ yields two roots

$$e_{\pm} = \pm \frac{(\sigma_1^2 + \sigma_W^2)}{\sigma_1}. \tag{A2}$$

Inserting the positive extreme value e_+ into the second derivation of T_{VI} , i.e.,

$$T''_{VI}(e_{MS,1}) = \frac{e_{MS,1}(\sigma_1^2 + \sigma_W^2) + \sqrt{e_{MS,1}^2 - 4\sigma_W^2}(\sigma_W^2 - \sigma_1^2)}{\sigma_W^2(\sigma_1^2 + \sigma_W^2)\sqrt{e_{MS,1}^2 - 4\sigma_W^2}}, \tag{A3}$$

identifies e_+ as a minimum value, because $T''_{VI}(e_+)$ is always positive for $\tau_1 > 1$, cf. (55). For that reason, T_{VI} is a monotonically increasing function on the interval $(e_+, +\infty)$. Figure A1 depicts the positive tail of T_{VI} and T'_{VI} , respectively, as well as the minimum e_+ .

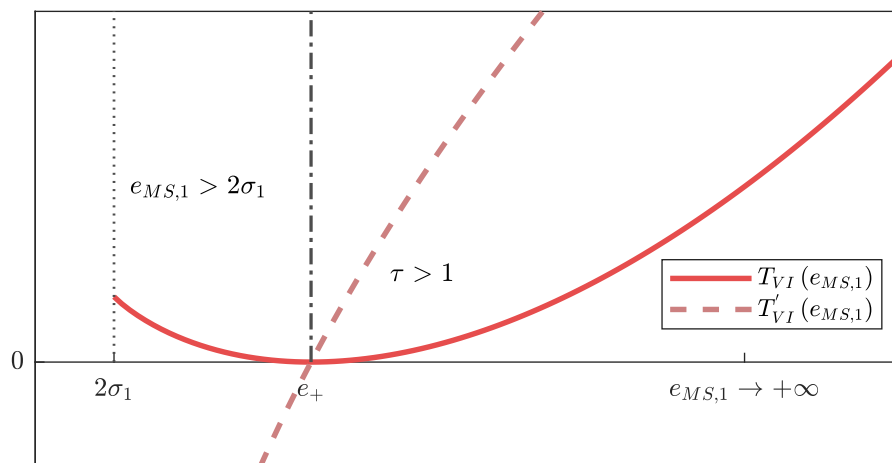


Figure A1. Tail of the function T_{VI} and its first derivation T'_{VI} for positive values of $e_{MS,1}$ as well as the minimum e_+ .

References

1. Rofatto, V.F.; Matsuoka, M.T.; Klein, I.; Veronez, M.R.; Bonimani, M.L.; Lehmann, R. A half-century of Baarda's concept of reliability: a review, new perspectives, and applications. *Surv. Rev.* **2018**, 1–17. [[CrossRef](#)]
2. Baarda, W. *A Testing Procedure for Use in Geodetic Networks*, 2nd ed.; Netherlands Geodetic Commission, Publication on Geodesy: Delft, The Netherlands, 1968.
3. Teunissen, P.J.G. Distributional theory for the DIA method. *J. Geod.* **2018**, 92, 59–80. [[CrossRef](#)]
4. Lehmann, R. On the formulation of the alternative hypothesis for geodetic outlier detection. *J. Geod.* **2013**, 87, 373–386. [[CrossRef](#)]
5. Lehmann, R.; Lösler, M. Multiple outlier detection: hypothesis tests versus model selection by information criteria. *J. Surv. Eng.* **2016**, 142, 04016017. [[CrossRef](#)]
6. Lehmann, R.; Lösler, M. Congruence analysis of geodetic networks – hypothesis tests versus model selection by information criteria. *J. Appl. Geod.* **2017**, 11. [[CrossRef](#)]
7. Teunissen, P.J.G. *Testing Theory—An Introduction*, 2nd ed.; Series of Mathematical Geodesy and Positioning; VSSD: Delft, The Netherlands, 2006.
8. Kargoll, B. *On the Theory and Application of Model Misspecification Tests in Geodesy*; C 674; German Geodetic Commission: Munich, Germany, 2012.
9. Bhar, L.; Gupta, V. Study of outliers under variance-inflation model in experimental designs. *J. Indian Soc. Agric. Stat.* **2003**, 56, 142–154.
10. Cook, R.D. Influential observations in linear regression. *J. Am. Stat. Assoc.* **1979**, 74, 169–174. [[CrossRef](#)]
11. Thompson, R. A note on restricted maximum likelihood estimation with an alternative outlier model. *J. R. Stat. Soc. Ser. B (Methodological)* **1985**, 47, 53–55. [[CrossRef](#)]
12. Gumedze, F.N.; Welhamb, S.J.; Gogel, B.J.; Thompson, R. A variance shift model for detection of outliers in the linear mixed model. *Comput. Stat. Data Anal.* **2010**, 54, 2128–2144. [[CrossRef](#)]
13. Gumedze, F.N. Use of likelihood ratio tests to detect outliers under the variance shift outlier model. *J. Appl. Stat.* **2018**. [[CrossRef](#)]
14. Koch, K.R.; Kargoll, B. Expectation maximization algorithm for the variance-inflation model by applying the t-distribution. *J. Appl. Geod.* **2013**, 7, 217–225. [[CrossRef](#)]
15. Koch, K.R. Outlier detection for the nonlinear Gauss Helmert model with variance components by the expectation maximization algorithm. *J. Appl. Geod.* **2014**, 8, 185–194. [[CrossRef](#)]
16. Koch, K.R. *Parameter Estimation and Hypothesis Testing in Linear Models*; Springer: Berlin/Heidelberg, Germany, 1999. [[CrossRef](#)]
17. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer Texts in Statistics; Springer: New York, NY, USA, 2013. [[CrossRef](#)]
18. Arnold, S.F. *The Theory of Linear Models and Multivariate Analysis*; Wiley Series in Probability and Statistics; John Wiley & Sons Inc.: New York, NY, USA, 1981.
19. Lehmann, R.; Voß-Böhme, A. On the statistical power of Baarda's outlier test and some alternative. *Surv. Rev.* **2017**, 7, 68–78. [[CrossRef](#)]
20. Neyman, J.; Pearson, E.S. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character* **1933**, 231, 289–337. [[CrossRef](#)]
21. Lehmann, R. Improved critical values for extreme normalized and studentized residuals in Gauss-Markov models. *J. Geod.* **2012**, 86, 1137–1146. [[CrossRef](#)]
22. Sherman, J.; Morrison, W.J. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann. Math. Stat.* **1950**, 21, 124–127. [[CrossRef](#)]
23. Nocedal, J.; Wright, S.J. *Numerical Optimization*, 2nd ed.; Springer: New York, NY, USA, 2006. [[CrossRef](#)]
24. Prussing, J.E. The principal minor test for semidefinite matrices. *J. Guid. Control Dyn.* **1986**, 9, 121–122. [[CrossRef](#)]
25. Gilbert, G.T. Positive definite matrices and Sylvester's criterion. *Am. Math. Mon.* **1991**, 98, 44. [[CrossRef](#)]

