

## Article

# What Is the Maximum Likelihood Estimate When the Initial Solution to the Optimization Problem Is Inadmissible? The Case of Negatively Estimated Variances

Steffen Zitzmann <sup>1,\*</sup>, Julia-Kim Walther <sup>1</sup>, Martin Hecht <sup>2</sup>  and Benjamin Nagengast <sup>1,3</sup> 

<sup>1</sup> Hector Research Institute of Education Sciences and Psychology, University of Tübingen, 72072 Tübingen, Germany; julia-kim.walther@uni-tuebingen.de (J.-K.W.); benjamin.nagengast@uni-tuebingen.de (B.N.)

<sup>2</sup> Faculty of Humanities and Social Sciences, Helmut Schmidt University, 22043 Hamburg, Germany; martin.hecht@hsu-hh.de

<sup>3</sup> Department of Education and Brain & Motivation Research Institute, Korea University, Seoul 02841, Korea

\* Correspondence: steffen.zitzmann@uni-tuebingen.de

**Abstract:** The default procedures of the software programs *Mplus* and *lavaan* tend to yield an inadmissible solution (also called a Heywood case) when the sample is small or the parameter is close to the boundary of the parameter space. In factor models, a negatively estimated variance does often occur. One strategy to deal with this is fixing the variance to zero and then estimating the model again in order to obtain the estimates of the remaining model parameters. In the present article, we present one possible approach for justifying this strategy. Specifically, using a simple one-factor model as an example, we show that the maximum likelihood (ML) estimate of the variance of the latent factor is zero when the initial solution to the optimization problem (i.e., the solution provided by the default procedure) is a negative value. The basis of our argument is the very definition of ML estimation, which requires that the log-likelihood be maximized over the parameter space. We present the results of a small simulation study, which was conducted to evaluate the proposed ML procedure and compare it with *Mplus'* default procedure. We found that the proposed ML procedure increased estimation accuracy compared to *Mplus'* procedure, rendering the ML procedure an attractive option to deal with inadmissible solutions.

**Keywords:** maximum likelihood; Heywood case; inadmissible solution; confirmatory factor analysis



**Citation:** Zitzmann, S.; Walther, J.-K.; Hecht, M.; Nagengast, B. What Is the Maximum Likelihood Estimate When the Initial Solution to the Optimization Problem Is Inadmissible? The Case of Negatively Estimated Variances. *Psych* **2022**, *4*, 343–356. <https://doi.org/10.3390/psych4030029>

Academic Editors: Philipp Doebler and Alexander Robitzsch

Received: 18 May 2022

Accepted: 26 June 2022

Published: 30 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Confirmatory factor analysis is a common tool to analyze data from questionnaires. A latent variable software such as *Mplus* [1] or *lavaan* [2] is often used to conduct this type of analysis. When using the default procedure of such software, which is an unconstrained estimator, one might encounter a Heywood case—a solution for a loading, a measurement error variance, or a (residual) variance of a latent factor that is impossible. Such solutions are also often called “improper” or “inadmissible” in the psychometric literature [3,4], and many possible causes for their occurrence have been identified; for example, the model may be misspecified [5–7], the sample may be too small, or the population parameter itself may be close to the boundary of the parameter space e.g., a variance close to 0 [8]. For relatively prototypical statistical models, the probability of such solutions can be determined analytically e.g., [9–11]. To address Heywood cases, one may want to check the modification indices for model improvement. Deleting items from the model might also eliminate Heywood cases. However, there is no guarantee that this strategy will lead to the desired goal and, for example, a negatively estimated variance might remain. A prominent strategy for addressing negatively estimated variances that is often used in research practice is equal to the variance to zero and fits the model again to obtain the

estimates of the remaining model parameters for examples of this practice in psychological research, see, e.g., [12–14]. Another strategy is to use a nonnegativity constraint and thus constrained estimation e.g., [15,16] or penalized/Bayesian estimation e.g., [17] to force the variance estimate to be equal to or greater than zero. All these strategies have in common that they lead to variance estimates that are “admissible” (i.e., nonnegative values for variances).

The present article focuses on the class of factor models and builds on the statistical framework of maximum likelihood (ML) estimation. As a negatively estimated variance does often occur, we ask what the ML estimate of the variance is when the initial solution for this variance is a negative value (i.e., when the default procedure provides a negative value). Specifically, we present a statistical argument for why the ML estimate of the variance must be zero in this case, which can also be considered a formal justification for fixing the variance to zero and then estimating the model again in order to deal with negatively estimated variances in practice. The basic premise for the argument is the notion that ML estimates must be admissible; that is, they can only take on possible values (e.g., a nonnegative value for a variance). Technically speaking, ML estimates must lie in the parameter space. This notion is also backed up by the statistical literature. For example, Searle et al. [11] pointed out that “the very definition of maximum likelihood demands that the likelihood be maximized over the parameter space” (p. 81). In other words, whereas the initial solution to the optimization problem can be inadmissible, an ML estimate can never be inadmissible because it can only take on values from the parameter space. Therefore, a distinction should be made between inadmissible solutions and the actual results from ML estimation; see [11,18–20]. The property of being admissible is also highly desirable from a practical point of view because improper solutions can hardly be interpreted, particularly when this improper solution has been tested as statistically significant. For example, one cannot interpret a negative value for a variance because the parameter space of a variance includes only nonnegative values. By contrast, the ML estimate of the variance—and also any function thereof that is itself an ML estimate—can be interpreted in a straightforward way.

Our article is organized as follows: We consider a simple one-factor model as an example, which allows us to clarify our main point without too many distracting details. Specifically, we show that the ML estimate of the variance of the latent factor must be zero when the initial solution for this variance is negative. In addition, we show how the ML estimate of the remaining model parameter in the simple factor model (i.e., the measurement error variance in this case) can be obtained, and we argue that our findings also hold for variances in more complex factor models and other classes of statistical models. Using simulations, we evaluate and compare the proposed ML procedure with Mplus’ default procedure.

In the following discussion, we will assume that the analysis model is not misspecified and that it fits the data well, but in the discussion, we will also elaborate a bit on inadmissible solutions due to model misspecification.

## 2. Example Model

In psychology, a widely used approach for assessing a person characteristic is to ask  $i = 1, \dots, n$  persons to rate this characteristic on  $k = 1, \dots, K$  items. Based on these ratings, a simple one-factor model can be formulated, which relates the ratings to an underlying latent factor  $\eta$ . To improve readability, we assume that the variables are mean-centered (i.e., intercepts equal to zero). The factor model reads:

$$y_{ki} = 1 \cdot \eta_i + \varepsilon_{ki} \quad (1)$$

where  $\varepsilon_{ki}$  are measurement errors, which are normally distributed with variance  $\theta$ . Notice that, in the model, the item loadings are all one and thus equal across items. In addition, the measurement error variances are equal across items, which we indicated by dropping the index  $k$  from the  $\theta$  symbol.

One efficient way to fit this factor model is optimizing the likelihood function. Because the model assumes parallel items (i.e., equal loadings and equal measurement error variances across items) and identification is achieved by setting the first item's loading to one, only two parameters need to be estimated, the variance of the latent factor ( $\psi$ ) and the measurement error variance ( $\theta$ )—, a task that can be easily solved analytically. To this end, we begin with the assumption that the  $i$ th person's responses  $y_{1i}, \dots, y_{Ki}$  follow a multivariate normal distribution with the following density function:

$$f(y_{1i}, \dots, y_{Ki}) = \frac{1}{(2\pi)^{K/2}} \det \begin{pmatrix} \psi + \theta & \dots & \psi \\ \vdots & \ddots & \vdots \\ \psi & \dots & \psi + \theta \end{pmatrix}^{-1/2} \cdot \exp \left\{ -\frac{1}{2} (y_{1i}, \dots, y_{Ki}) \begin{pmatrix} \psi + \theta & \dots & \psi \\ \vdots & \ddots & \vdots \\ \psi & \dots & \psi + \theta \end{pmatrix}^{-1} \begin{pmatrix} y_{1i} \\ \vdots \\ y_{Ki} \end{pmatrix} \right\}. \quad (2)$$

This density can be interpreted as the  $i$ th person's likelihood function  $L_i$ . The individual log-likelihood is then given by:

$$\log(L_i) = c - \frac{K-1}{2} \log(\theta) - \frac{1}{2} \log(g) - \frac{1}{2\theta(K\psi + \theta)} \left\{ (K-1)\psi \sum_{k=1}^K y_{ki}^2 + \theta \sum_{k=1}^K y_{ki}^2 - \psi \sum_{\substack{k'=1 \\ k' \neq k}}^K \sum_{\substack{k=1 \\ k \neq k'}}^K y_{ki} y_{k'i} \right\} \quad (3)$$

where  $c$  is a constant term. If we use  $g = K\psi + \theta$  and some algebraic tricks, the individual log-likelihood can further be simplified to:

$$\log(L_i) = c - \frac{K-1}{2} \log(\theta) - \frac{1}{2} \log(g) - \frac{1}{2} \left\{ \frac{1}{\theta} \sum_{k=1}^K y_{ki}^2 - \frac{K(g-\theta)}{\theta g} \bar{y}_{\bullet i}^2 \right\}. \quad (4)$$

with  $\bar{y}_{\bullet i} = \frac{1}{K} \sum_{k=1}^K y_{ki}$ . Because the  $n$  persons are independently drawn, the overall log-likelihood is simply the sum of the individual log-likelihoods:

$$\log(L) = c - \frac{n(K-1)}{2} \log(\theta) - \frac{n}{2} \log(g) - \frac{1}{2\theta} \sum_{i=1}^n \sum_{k=1}^K y_{ki}^2 + \frac{K(g-\theta)}{2\theta g} \sum_{i=1}^n \bar{y}_{\bullet i}^2. \quad (5)$$

In order to obtain the estimates of the parameters  $\theta$  and  $g$ , the log-likelihood needs to be maximized. To this end, we first calculate the partial derivatives with respect to the two parameters:

$$\frac{\partial \log(L)}{\partial \theta} = -\frac{n(K-1)}{2\theta} + \frac{1}{2\theta^2} \left( \sum_{i=1}^n \sum_{k=1}^K y_{ki}^2 - K \sum_{i=1}^n \bar{y}_{\bullet i}^2 \right) \quad (6)$$

$$\frac{\partial \log(L)}{\partial g} = -\frac{n}{2g} + \frac{K}{2g^2} \sum_{i=1}^n \bar{y}_{\bullet i}^2$$

Then, setting these equations equal to zero yields an equation system. Solving the system and indicating the solutions by a hat ( $\hat{\cdot}$ ) yields the following estimation equations:

$$\begin{aligned}\hat{\theta} &= \frac{1}{n(K-1)} \left( \sum_{i=1}^n \sum_{k=1}^K y_{ki}^2 - K \sum_{i=1}^n \bar{y}_{\bullet i}^2 \right) \\ \hat{g} &= \frac{K}{n} \sum_{i=1}^n \bar{y}_{\bullet i}^2\end{aligned}\quad (7)$$

Because  $\hat{g} = K\hat{\psi} + \hat{\theta}$ , we obtain:

$$\hat{\psi} = \frac{1}{n} \sum_{i=1}^n \bar{y}_{\bullet i}^2 - \frac{1}{K} \hat{\theta} \quad (8)$$

When we exclude the special case of  $y_{ki}$  equaling  $\bar{y}_{\bullet i}$  for all  $i$  and  $k$ , it is evident from Equation (7) that  $\hat{\theta}$  can take on only positive values. However, because  $\hat{\psi}$  is computed as the difference between two terms, and the latter term can be greater than the former,  $\hat{\psi}$  can become negative. However, as the ML estimate has to be nonnegative,  $\hat{\psi}$  cannot be the ML estimate when  $\hat{\psi} < 0$ . But what is the ML estimate in this case? In the next section, we will develop a formal argument for why the ML estimate of the variance of the latent factor is zero when the initial solution to the optimization problem is negative.

### 3. The Argument

We consider an argument similar to those of Herbach [18] and Searle et al. [11]. Suppose that the initial solution to the maximization problem is  $\hat{\psi} < 0$ . Our goal is to show that the ML estimate is zero in this case. Because ML estimates of variances can only be zero or greater than zero, it is sufficient to show that the ML estimate cannot be greater than zero. Let ML estimates be indicated by a dot ( $\dot{\cdot}$ ) in order to better differentiate it from the initial solutions to the optimization problem, which we indicated by a hat symbol in the previous section.

Assume for a moment that the ML estimate  $\dot{\psi}$  would be greater than zero. Then, it follows that:

$$\dot{g} > \dot{\theta} \quad (9)$$

Moreover, from  $\hat{\psi} < 0$  and  $\hat{g} = K\hat{\psi} + \hat{\theta}$ , we infer:

$$\hat{\theta} > \hat{g} \quad (10)$$

Now, consider the following two cases. Case 1: The first case is that of  $\dot{\theta} \geq \hat{\theta}$ . Because of  $\dot{g} \geq \dot{\theta} \geq \hat{\theta} > \hat{g}$  (using Equations (9) and (10)), it holds that  $\dot{g} > \hat{g}$  or, equivalently:

$$\dot{g} - \hat{g} > 0 \quad (11)$$

Calculating the partial derivative with respect to  $g$  at the ML estimate  $\dot{g}$  yields:

$$\begin{aligned}\left. \frac{\partial \log(L)}{\partial g} \right|_{g=\dot{g}} &= -\frac{n}{2\dot{g}} + \frac{K}{2\dot{g}^2} \sum_{i=1}^n \bar{y}_{\bullet i}^2 \\ &= -\frac{n}{2\dot{g}^2} \left( \dot{g} - \frac{K}{n} \sum_{i=1}^n \bar{y}_{\bullet i}^2 \right) \\ &= -\frac{n}{2\dot{g}^2} (\dot{g} - \hat{g})\end{aligned}\quad (12)$$

Because of  $\dot{g} > 0$  (see Appendix A for the proof), the factor  $-n/(2\dot{g}^2)$  is negative. Along with Equation (11), we obtain:

$$\left. \frac{\partial \log(L)}{\partial g} \right|_{g=\dot{g}} < 0 \quad (13)$$

which contradicts the condition that the partial derivative at the ML estimate is zero and thus that  $\dot{g}$  is the ML estimate. To complete the proof, we need to show that the second case also leads to a contradiction. Case 2:  $\dot{\theta} < \hat{\theta}$ . Then:

$$\dot{\theta} - \hat{\theta} < 0. \quad (14)$$

The partial derivative with respect to  $\theta$  at  $\hat{\theta}$  is:

$$\begin{aligned} \left. \frac{\partial \log(L)}{\partial \theta} \right|_{\theta=\hat{\theta}} &= -\frac{n(K-1)}{2\theta} + \frac{1}{2\theta^2} \left( \sum_{i=1}^n \sum_{k=1}^K y_{ki}^2 + K \sum_{i=1}^n \bar{y}_{\bullet i}^2 \right) \\ &= -\frac{n(K-1)}{2\hat{\theta}^2} \left\{ \hat{\theta} - \frac{1}{n(K-1)} \left( \sum_{i=1}^n \sum_{k=1}^K y_{ki}^2 + K \sum_{i=1}^n \bar{y}_{\bullet i}^2 \right) \right\} \\ &= -\frac{n(K-1)}{2\hat{\theta}^2} \{ \hat{\theta} - \hat{\theta} \}. \end{aligned} \quad (15)$$

Generally,  $K > 1$ . Thus, we have  $-n(K-1)/(2\hat{\theta}^2) < 0$ , and because of Equation (14), we obtain:

$$\left. \frac{\partial \log(L)}{\partial \theta} \right|_{\theta=\hat{\theta}} > 0 \quad (16)$$

which is a contradiction to  $\hat{\theta}$  being the ML estimate.

It can be concluded that  $\dot{\psi}$  cannot be greater than zero, and as ML estimates of variances can only be equal to or greater than zero,  $\dot{\psi}$  must thus be zero, which proves that the ML estimate is indeed zero when the initial solution to the optimization problem is negative.  $\square$

Having shown that  $\dot{\psi}$  is zero, it follows that  $\dot{g} = K\dot{\psi} + \dot{\theta} = \dot{\theta}$ . Thus, to obtain  $\hat{\theta}$ , the log-likelihood needs to be optimized subject to the constraint  $g = \theta$ . To this end, we modify the log-likelihood in Equation (5) by setting  $g$  to  $\theta$ :

$$\log(L^*) = c - \frac{n(K-1)}{2} \log(\theta) - \frac{n}{2} \log(\theta) - \frac{1}{2\theta} \sum_{i=1}^n \sum_{k=1}^K y_{ki}^2. \quad (17)$$

The modified log-likelihood is optimized by first computing the following derivative:

$$\frac{\partial \log(L^*)}{\partial \theta} = -\frac{n(K-1)}{2\theta} - \frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n \sum_{k=1}^K y_{ki}^2 \quad (18)$$

and then equating it to zero. Finally, solving the equation for  $\hat{\theta}$  and making use of Equations (7) and (8) yields the ML estimate:

$$\hat{\theta} = \frac{1}{nK} \sum_{i=1}^n \sum_{k=1}^K y_{ki}^2 = \hat{\psi} + \hat{\theta}. \quad (19)$$

It is interesting to note that, because we assumed that the initial solution for the variance of the latent factor is negative ( $\hat{\psi} < 0$ ), it follows from Equation (19) that the ML estimate of the measurement error variance is smaller than the initial solution for the measurement error variance ( $\hat{\theta} < \hat{\theta}$ ).

To summarize so far, we have seen that the ML estimates can deviate from the initial solutions to the optimization problem. Taking the simple one-factor model as an example,

the ML estimation procedure for estimating the two variance parameters can be stated as follows:

$$\psi = \begin{cases} \hat{\psi} & \text{if } \hat{\psi} \geq 0 \\ 0 & \text{if } \hat{\psi} < 0 \end{cases} \quad (20)$$

$$\hat{\theta} = \begin{cases} \hat{\theta} & \text{if } \hat{\psi} \geq 0 \\ \hat{\psi} + \hat{\theta} & \text{if } \hat{\psi} < 0 \end{cases} . \quad (21)$$

When the initial solution  $\hat{\psi}$  for the variance of the latent factor is equal to or greater than zero and is thus admissible, the ML estimate  $\psi$  of the variance of the latent factor and the ML estimate  $\hat{\theta}$  of the measurement error variance are simply the initial solutions  $\hat{\psi}$  and  $\hat{\theta}$  to the optimization problem. However, when  $\hat{\psi}$  is smaller than zero and thus a Heywood case,  $\psi$  is zero, and  $\hat{\theta}$  takes on the value  $\hat{\psi} + \hat{\theta}$ .

To describe the statistical properties of this ML procedure, a simulation was run, which we will report in the following section.

#### 4. Simulation Study

The primary goal of the study was to evaluate the proposed ML procedure and to compare it with the default procedure of the prominent latent variable software *Mplus*. We did not include lavaan as another comparison standard because *Mplus* and lavaan were assumed to yield identical results.

##### 4.1. Method and Evaluation Criteria

We simulated data from a simple one-factor model with three items, each with an unstandardized loading of 1 because the one from the first item was set equal to 1, and the items were assumed to be parallel (i.e., equal loadings and equal measurement error variances across items). The variance of the latent factor was 1. As a negatively estimated variance will likely occur when only a little information is available for assessing the latent factor (e.g., when the standardized loadings are weak or the sample is small), the standardized loadings were varied to be all either 0.3, 0.4, or 0.6 and thus rather small (see Appendix B). In addition, the sample size was varied between 25, 50, 100, and 200 persons. Thus, the study design included  $3 \cdot 4 = 12$  conditions. For each condition, we generated a total of 1000 data sets. To analyze them, a factor model with one latent factor was specified in *Mplus*, with three items loading on this factor. The first item's loading was fixed to 1 in order to identify the factor. Moreover, the loadings and measurement error variances were constrained to be equal across items (i.e., parallel items). Thus, the analysis model corresponded with the data-generating model. Differences in the results could thus be attributed to the use of different estimation procedures. The first procedure was *Mplus'* default procedure, which is an example of an unconstrained estimator that tends to exhibit inadmissible solutions when the standardized loadings are weak or the sample is small. The second approach was our proposed ML procedure, which we developed based on the premise that ML estimates cannot be inadmissible because such estimates must lie in the parameter space. In the simple one-factor model, this procedure yields values that are identical to those from *Mplus'* default procedure when the initial solution for the variance of the latent factor is nonnegative. Otherwise, the procedure yields zero as the value for the variance of the latent factor, and it yields the sum of the initial solutions for the two variance parameters in the model as the value for the measurement error variance (see Equation (20)). In the latter case, we fixed the variance of the latent factor to zero by using the at operator (@) in *Mplus* (see the *Mplus* code in Appendix B). Note that this is equivalent to constraining  $g$  in Equation (5) to be equal to the measurement error variance, which we did in the previous section in order to derive the ML estimates. *Mplus'* default procedure and the proposed ML procedure were compared with regard to the relative bias, the overall estimation accuracy as assessed by the relative root mean squared error (RMSE), and coverage rate for estimating both the variance of the latent variable and the measurement error variance. The relative bias is the deviation of the expected value of the

estimates from the population parameter divided by that parameter. The relative RMSE can be expressed as the square root of the sum of the squared bias and the variability of the estimates divided by the population parameter e.g., [21,22], and the coverage rate is defined as the probability that the 95% confidence interval captures the population parameter.

#### 4.2. Results

In the simulation, the simple one-factor model always converged with *Mplus'* default procedure as indicated by 0% nonconverged solutions in all conditions.

##### 4.2.1. Percentage Inadmissible Solutions

Table 1 offers an overview of the percentages of inadmissible solutions that *Mplus'* default procedure provided. An inadmissible solution was indicated either by WARNING: THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IS NOT POSITIVE DEFINITE or the message RESIDUAL COVARIANCE MATRIX (THETA) IS NOT POSITIVE DEFINITE. As can be seen, *Mplus'* default procedure yielded inadmissible solutions when the sample was rather small (i.e.,  $n < 100$ ). Notably, all inadmissible solutions were negatively estimated variances of the latent factor. The percentage of inadmissible solutions was particularly high in these situations when the standardized loadings were very weak ( $\lambda_s = 0.3$ ). However, when the sample size increased or the standardized loadings became stronger, inadmissible solutions became less likely.

**Table 1.** Simulation study results: Percentages of inadmissible solutions for *Mplus'* default procedure.

No. of Persons	Standardized Loadings	Inadmissible Solutions
$n = 25$	$\lambda_s = 0.3$	23.7
	$\lambda_s = 0.4$	11.5
	$\lambda_s = 0.6$	0.4
$n = 50$	$\lambda_s = 0.3$	15.6
	$\lambda_s = 0.4$	4.0
	$\lambda_s = 0.6$	0.0
$n = 100$	$\lambda_s = 0.3$	6.9
	$\lambda_s = 0.4$	0.4
	$\lambda_s = 0.6$	0.0
$n = 200$	$\lambda_s = 0.3$	1.6
	$\lambda_s = 0.4$	0.0
	$\lambda_s = 0.6$	0.0

Note.  $n$  = number of persons;  $\lambda_s$  = standardized loadings.

##### 4.2.2. Statistical Properties

Before we proceed with the statistical properties, note that we computed those of *Mplus'* default procedure in two ways. First, we computed them based on the full set of data sets per condition, including also data sets for which *Mplus'* default procedure provided an inadmissible solution. Second, the properties were computed only from the data sets for which *Mplus'* default procedure provided an admissible solution because it is very common to exclude data sets that caused estimation problems e.g., [23,24]. The statistical properties of the proposed ML procedure were based on the full set of data sets per condition.

Table 2 shows the comparisons of the relative bias, the relative RMSE, and the coverage rate for the variance of the latent factor and the measurement measurement error variance between *Mplus'* default procedure (computed in the two mentioned ways) and the proposed ML procedure. Both estimation procedures provided approximately unbiased estimates of the two model parameters in large samples without any noticeable difference between the procedures. These findings reflect the fact that both estimators are asymptotically unbiased (i.e., unbiased when the sample size approaches infinity). However, differences existed for the variance of the latent factor in rather small samples, particularly when the standardized

loadings were weak. Whereas the bias was absent for *Mplus'* default procedure when its computation was based on the full set of data sets per condition, it was positive and substantial when computed from only the data sets for which the procedure provided admissible solutions. Notably, there was no bias for the proposed ML procedure except in the most extreme condition (i.e.,  $n = 25$ ,  $\lambda_s = 0.3$ ). However, this bias was only small. Both procedures provided negligibly biased estimates for the measurement error variance, without any noticeable difference.

In extreme conditions, the relative RMSE for the variance of the latent factor differed between the estimation procedures. *Mplus'* default procedure exhibited the largest relative RMSE when it was computed from the full set of data sets per condition, meaning that this procedure provided less accurate estimates than the proposed ML procedure. However, *Mplus'* default procedure became more accurate when based on only the data sets for which the procedure provided admissible solutions. However, when the sample size increased, the differences between the estimation procedures vanished. The relative RMSE for the measurement error variance was approximately equal between the procedures.

The coverage rate for the variance of the latent factor was close to the nominal 95% in all conditions when *Mplus'* default procedure was used, and the coverage rate was computed from the full set of data sets per condition. However, the coverage rate for this procedure tended to be too high when it was computed from only the data sets for which the procedure provided admissible solutions, particularly when the loadings were rather weak. By contrast, the coverage rate for the proposed ML procedure tended to be too low. The coverage rate for the measurement error variance was acceptable in all conditions except in those with a very small sample size (i.e.,  $n = 25$ ), where it tended to be too small. However, it did not differ between the estimation procedures.

**Table 2.** Simulation study results: Relative bias, relative root mean squared error (RMSE), and coverage rate from estimating the variance of the latent factor and the measurement error variance with *Mplus'* default procedure and the maximum likelihood (ML) procedure.

No. of Persons	Standardized Loadings	Relative Bias			Relative RMSE			Coverage Rate		
		Default †	Default ‡	ML †	Default †	Default ‡	ML †	Default †	Default ‡	ML †
Variance of Latent Factor										
$n = 25$	$\lambda_s = 0.3$	0.00	0.51	0.15	1.34	1.19	1.14	95.3	99.6	76.0
	$\lambda_s = 0.4$	-0.02	0.16	0.02	0.85	0.74	0.78	93.2	99.8	88.3
	$\lambda_s = 0.6$	-0.03	-0.03	-0.03	0.46	0.45	0.46	91.0	91.4	91.0
$n = 50$	$\lambda_s = 0.3$	-0.02	0.24	0.05	0.97	0.83	0.86	94.8	99.4	83.9
	$\lambda_s = 0.4$	-0.01	0.04	0.00	0.62	0.58	0.61	93.5	97.1	93.2
	$\lambda_s = 0.6$	-0.01	-0.01	-0.01	0.33	0.33	0.33	92.4	92.4	92.4
$n = 100$	$\lambda_s = 0.3$	0.03	0.12	0.04	0.70	0.64	0.67	95.7	98.9	92.1
	$\lambda_s = 0.4$	0.01	0.01	0.01	0.42	0.42	0.42	95.1	95.5	95.1
	$\lambda_s = 0.6$	-0.02	-0.02	-0.02	0.23	0.23	0.23	93.2	93.2	93.2
$n = 200$	$\lambda_s = 0.3$	-0.01	0.01	-0.01	0.48	0.46	0.48	95.4	97.0	95.4
	$\lambda_s = 0.4$	0.02	0.02	0.02	0.31	0.31	0.31	94.9	94.9	94.9
	$\lambda_s = 0.6$	-0.01	-0.01	-0.01	0.16	0.16	0.16	95.5	95.5	95.5
Measurement Error Variance										
$n = 25$	$\lambda_s = 0.3$	-0.05	-0.09	-0.06	0.20	0.20	0.19	89.0	86.4	88.1
	$\lambda_s = 0.4$	-0.04	-0.07	-0.05	0.20	0.19	0.20	90.1	88.9	89.5
	$\lambda_s = 0.6$	-0.05	-0.05	-0.05	0.20	0.20	0.20	89.2	89.2	89.2
$n = 50$	$\lambda_s = 0.3$	-0.02	-0.04	-0.02	0.14	0.14	0.14	91.9	90.5	91.1
	$\lambda_s = 0.4$	-0.03	-0.04	-0.03	0.14	0.14	0.14	91.7	91.7	91.5
	$\lambda_s = 0.6$	-0.02	-0.02	-0.02	0.14	0.14	0.14	93.3	93.3	93.3
$n = 100$	$\lambda_s = 0.3$	-0.01	-0.02	-0.01	0.10	0.10	0.10	93.3	93.0	93.3
	$\lambda_s = 0.4$	-0.01	-0.01	-0.01	0.10	0.10	0.10	94.8	94.9	94.7
	$\lambda_s = 0.6$	-0.01	-0.01	-0.01	0.10	0.10	0.10	94.5	94.5	94.5
$n = 200$	$\lambda_s = 0.3$	0.00	0.00	0.00	0.07	0.07	0.07	94.7	95.0	94.6
	$\lambda_s = 0.4$	-0.01	-0.01	-0.01	0.07	0.07	0.07	92.4	92.4	92.4
	$\lambda_s = 0.6$	-0.01	-0.01	-0.01	0.07	0.07	0.07	94.6	94.6	94.6

Note. † Results are based on the full set of simulated data sets per condition. ‡ These results are based on only those data sets for which *Mplus'* default estimator provided an admissible converged solution.  $n$  = number of persons;  $\lambda_s$  = standardized loadings.

### 4.3. Summary

To sum up, our presentation revealed that, in the simple one-factor model, only the initial solution for the variance of the latent factor can become inadmissible (i.e., a negative value for that variance), and we argued that this will likely occur when there is only little information in the data. This was confirmed by the findings that all inadmissible solutions were indeed negatively estimated variances of the latent factor and that they occurred when the sample was small or the loadings were weak. Moreover, we found that *Mplus*' default procedure led to less bias than the proposed ML procedure in extreme conditions but that *Mplus*' default procedure was less accurate (i.e., it exhibited a larger RMSE) than the proposed ML procedure in these conditions when data sets for which *Mplus*' default procedure provided an inadmissible solution were not excluded. However, the ML procedure provided standard errors that were not very accurate as indicated by coverage rates that were too low. This could have been expected because the standard error is zero when the variance of the latent factor is fixed in *Mplus*. In the next section, we will offer a pragmatic approach to the standard error, and we will discuss a more sophisticated resampling technique.

## 5. Discussion and Recommendations

With the default procedures in latent variable software, one might encounter inadmissible solutions—for example, a negative value for a variance. They are dealt with in different ways. Some scholars have even argued that such solutions are completely useless e.g., [23]. We do not wish to go so far because the researcher's focus may be on a derived quantity, and such a quantity can be meaningfully interpreted even when it is computed from an inadmissible solution. See, for example, Molenberghs and Verbeke [25], who argued that, in the context of random effects modeling, the negative intraclass correlation obtained from a negatively estimated random effects variance can simply be interpreted as a negative within-cluster correlation. However, when the focus is directly on the variance, a negative value can hardly be interpreted. Therefore, one strategy to deal with this is fixing the negatively estimated variance to zero and estimating the model again to obtain the ML estimates of the remaining parameters. Using a simple one-factor model as an example, we showed that the ML estimate of the variance of the latent factor must indeed be zero when the initial solution for this variance is a negative value. We built on the statistical literature in which an inadmissible solution is distinguished from the actual result of the ML estimation. For example, Searle et al. [11] pointed out that the exact definition of ML estimation involves that the likelihood is maximized over the parameter space, and the parameter space includes only possible values. Therefore, an ML estimate must be admissible; see also [16]. An alternative to our proposed ML procedure is using constrained estimation e.g., [26]. As Gerbing and Anderson [27] noted, using a nonnegativity constraint for a variance is similar to equating the negatively estimated variance to zero. Some software programs follow this strategy. For example, EQS [28] uses it as the default. A similar strategy is penalized or Bayesian estimation, where prior distributions are often specified in such a way that results will be admissible [23,29,30]. For example, inverse gamma distributions are often specified for variances [31–34]. A variance is said to be inverse-gamma distributed when its inverse (i.e., the precision) follows a gamma distribution [35–37]. When properly specified, the resulting posterior distribution supports only nonnegative values, meaning that a negatively estimated variance cannot occur; see Figure 1 in [38]. Examples of software that use inverse gamma priors or, in the multivariate case, inverse Wishart priors are WinBUGS [39] and JAGS [40]. One may also parameterize a variance as the square of a standard deviation and then specify a prior for the standard deviation. Even if the prior also supports negative values (e.g., a uniform distribution over a wide range of values, including negative values), squaring leads to nonnegative values and thus to an admissible result for the variance. It is interesting to note that Hill and Thompson [10] pointed out that whether an admissible solution is useful or not depends on the specific question at hand and the requirements to address this question. Savalei and

Kolenikov [6]; see also Kolenikov and Bollen [7], who argued that an inadmissible solution can help to reveal model misspecifications when the statistical test is significant (e.g., when the null hypothesis that the variance is zero in the population can be rejected). This procedure works even if the model is saturated with zero degrees of freedom [7]. In addition, Savalei and Kolenikov [6] argued that it would suffice to require that parameters (not ML estimates!) lie in the parameter space. In other words, according to their view, any solution to the optimization problem is possible. However, Gerbing and Anderson [27] argued that statistical properties of an estimator can be poor in this case; see also [23,41]—an argument that is also supported by our findings from the simulation study. We found that *Mplus'* default procedure was less accurate than our proposed ML procedure, particularly when the data provided only a little information.

Before moving on to the practical recommendations, some possible limitations of our presentations should be mentioned. The example model is a simple one-factor model with only two parameters (mean structure ignored), and more realistic models are significantly more complex. Thus, the question arises whether our findings would generalize beyond the simple one-factor model. To address this question, more complex models could be investigated to develop ML estimators with similar properties. In less restrictive factor models (e.g., congeneric items) or in models with multiple factors, initial solutions for many more parameters can be inadmissible when certain conditions are met; see [5]. However, formal accounts would probably become tedious because they involve an increasing number of case distinctions. A general account for arbitrary factor models would certainly be preferable. As long as there is no such account, our findings should be applied with caution. An anonymous reviewer suspected that our findings would generalize to estimates of variances that are (approximately) independent as indicated by the information matrix (or the asymptotic covariance matrix). It would be interesting to test this assumption empirically by manipulating the dependency of variance estimates. This question can best be addressed in future simulation work. Another limitation is that the extreme conditions of the simulation were rather unrealistic because the data contained only a little information (i.e., a small sample or weak loadings), and thus, negatively estimated variance was likely to occur with *Mplus'* default procedure. However, this does not mean that, in settings with larger samples and loadings, inadmissible solutions will be less likely because whether an inadmissible solution occurs depends also on many other factors including model complexity and misspecification.

We recommend that researchers should use the proposed ML procedure rather than *Mplus'* default procedure to obtain admissible and thus easily interpretable results. One way for *Mplus* users to implement this ML procedure is first fitting the model using the software's default procedure. When the initial solution for a variance is a nonnegative value and thus admissible, this solution is the ML estimate, and one can interpret the estimate accordingly. However, when the initial solution is negative (i.e., a Heywood case), the actual ML estimate is zero, and the model needs to be estimated again with the variance fixed to zero in order to obtain the ML estimates of the remaining model parameters. The usual procedure is to interpret the ML estimate as what the data tells us about the population parameter. Thus, an ML estimate of zero can be interpreted as "the data suggested that persons did not differ much". As the standard error is zero when a variance is fixed to zero, we suggest that the standard error from *Mplus'* default procedure should be used for inferential purposes; see [17] for a similar recommendation in the context of penalized estimation. Alternatively, one can adopt a resampling technique, such as the jackknife procedure. The jackknife first computes estimates of the variance of the latent factor from  $R$  subsamples, each omitting  $d$  persons, and then these estimates are entered into a formula in order to obtain the standard error. One may divide the indices  $(1, \dots, n)$  into  $R = n/d$  non-overlapping subsets and use these subsets to create the subsamples see [42], who argued that this procedure can be computationally very efficient.

To conclude, we presented a formal approach for justifying the practice of constraining a negatively estimated variance to zero and estimating the model again in order to obtain

the remaining estimates, which is very common in confirmatory factor analysis and thus in the analysis of data from questionnaires. However, there are other strategies to deal with inadmissible solutions, and we briefly discussed them. Future research could extend our argument to other models and parameters and conduct extensive simulations to compare the different strategies for dealing with inadmissible solutions with one another.

**Author Contributions:** S.Z.: writing and mathematical derivations. J.-K.W.: writing and conducting the simulation. M.H.: writing. B.N.: writing and lead. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** All data are contained within the article.

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

### Appendix A. Proof of $\dot{g} > 0$

**Proof of  $\dot{g} > 0$ .** This follows from the relation  $\dot{g} = K\dot{\psi} + \dot{\theta}$ ,  $\dot{\theta} \geq \hat{\theta}$ , our observation that, in general,  $\hat{\theta} > 0$ , and the assumption  $\dot{\psi} > 0$ , which we want to prove wrong. More specifically, rearranging  $\dot{g} = K\dot{\psi} + \dot{\theta}$  yields  $\dot{g} - K\dot{\psi} = \dot{\theta}$ . From  $\dot{\theta} \geq \hat{\theta}$  and  $\hat{\theta} > 0$ , it follows that  $\dot{\theta} \geq \hat{\theta} > 0$  and thus  $\dot{\theta} > 0$ . If we use  $\dot{\theta} > 0$ , we obtain  $\dot{g} - K\dot{\psi} > 0$ , which can be rearranged to  $\dot{g} > K\dot{\psi}$ . As  $K\dot{\psi} > \dot{\psi}$  and we assumed  $\dot{\psi} > 0$ , we have  $K\dot{\psi} > \dot{\psi} > 0$  and thus  $K\dot{\psi} > 0$ . Finally, using  $K\dot{\psi} > 0$  yields  $\dot{g} > 0$ .  $\square$

### Appendix B. R & Mplus Code

The following R function can be used to generate data according to the simple one-factor model.

```
generateSimpleOneFactorModelData <- function( n, sl ) {

  v <- 3 # number of items
  sl <- rep( sl, v ) # standardized loadings
  l <- rep( 1, v ) # loadings

  m.eta <- 0 # mean of latent factor
  m.yy <- rep( 0, times = v ) # means of items

  var.eta <- 1.0 # variance of latent factor
  # Measurement error variances of items
  var.me.yy <- rep( NA, v )
  for ( j in 1 : v ) {
    var.me.yy[ j ] <- ( ( 1 - sl[ j ]^2 ) / sl[ j ]^2 ) * l[ j ]^2 * var.
      eta
  }

  # Latent factor
  eta <- rep( NA, n )
  for ( i in 1 : n ) {
    eta[ i ] <- rnorm( 1, m.eta, sqrt( var.eta ) )
  }
}
```

```

# Items
yy <- array( rep( NA, n*v ), dim=c(n,v) )
for ( i in 1 : n ) {
  for ( jj in 1 : v ) {
    yy[ i, jj ] <- rnorm( 1, m.yy[jj] + 1[jj]*eta[i], sqrt( var.me.yy[jj]
      ) )
  }
}

dat = data.frame( yy )
return( dat )

}

```

The following code fits the simple one-factor model with *Mplus*.

```

Title:          Simple one-factor model

Data:          File is
filename.dat;

Variable:      Names are y_1 y_2 y_3;
Usevariables are y_1 y_2 y_3;

Model:        eta by y_1 y_2@1 y_3@1;

eta(vareta);
!eta@0(vareta); ! use this line to set the variance of the latent
                factor to zero when the initial solution for this variance is a
                negative value

y_1(vare);
y_2(vare);
y_3(vare);

```

## References

1. Muthén, L.K.; Muthén, B.O. *Mplus User's Guide*, 7th ed.; Muthén & Muthén: Los Angeles, CA, USA, 2012.
2. Rosseel, Y. lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [[CrossRef](#)]
3. van Driel, O.P. On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika* **1978**, *43*, 225–243. [[CrossRef](#)]
4. Wothke, W. Nonpositive definite matrices in structural modeling. In *Testing Structural Equation Models*; Bollen, K.A., Long, J.S., Eds.; Sage: Newbury Park, CA, USA, 1993; pp. 256–293.
5. Chen, F.; Bollen, K.A.; Paxton, P.; Curran, P.J.; Kirby, J.B. Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociol. Methods Res.* **2001**, *29*, 468–508. [[CrossRef](#)]
6. Savalei, V.; Kolenikov, S. Constrained versus unconstrained estimation in structural equation modeling. *Psychol. Methods* **2008**, *13*, 150–170. [[CrossRef](#)] [[PubMed](#)]
7. Kolenikov, S.; Bollen, K.A. Testing negative error variances: Is a Heywood case a symptom of misspecification? *Sociol. Methods Res.* **2012**, *41*, 124–167. [[CrossRef](#)]
8. Jak, S.; Jorgensen, T.D.; Rosseel, Y. Evaluating cluster-level factor models with lavaan and Mplus. *Psych* **2021**, *3*, 134–152. [[CrossRef](#)]
9. Bhargava, A.K.; Disch, D. Exact probabilities of obtaining estimated non-positive definite between-group covariance matrices. *J. Stat. Comput. Simul.* **1982**, *15*, 27–32. [[CrossRef](#)]
10. Hill, W.G.; Thompson, R. Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics* **1978**, *34*, 429–439. [[CrossRef](#)]
11. Searle, S.R.; Casella, G.; McCulloch, C.E. *Variance Components*; Wiley: New York, NY, USA, 1992.
12. Baird, R.; Maxwell, S.E. Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychol. Methods* **2016**, *21*, 175–188. [[CrossRef](#)]
13. Lüdtke, O.; Marsh, H.W.; Robitzsch, A.; Trautwein, U. A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychol. Methods* **2011**, *16*, 444–467. [[CrossRef](#)]

14. Zitzmann, S.; Lüdtke, O.; Robitzsch, A.; Marsh, H.W. A Bayesian approach for estimating multilevel latent contextual models. *Struct. Equ. Model.* **2016**, *23*, 661–679. [CrossRef]
15. Dijkstra, T.K. On statistical inference with parameter estimates on the boundary of the parameter space. *Br. J. Math. Stat. Psychol.* **1992**, *45*, 289–309. [CrossRef]
16. Schoenberg, R. Constrained maximum likelihood. *Comput. Econ.* **1997**, *10*, 251–266. [CrossRef]
17. Lüdtke, O.; Ulitzsch, E.; Robitzsch, A. A comparison of penalized maximum likelihood estimation and Markov chain Monte Carlo techniques for estimating confirmatory factor analysis models with small sample sizes. *Front. Psychol.* **2021**, *12*, 615162. [CrossRef]
18. Herbach, L.H. Properties of Model II-type analysis of variance tests, A: Optimum nature of the F-test for Model II in the balanced case. *Ann. Math. Stat.* **1959**, *30*, 939–959. [CrossRef]
19. Arnold, S.F. *The Theory of Linear Models and Multivariate Analysis*; Wiley: New York, NY, USA, 1981.
20. Yuan, K.H.; Bentler, P.M. On normal theory based inference for multilevel models with distributional violations. *Psychometrika* **2002**, *67*, 539–562. [CrossRef]
21. Greenland, S. Principles of multilevel modelling. *Int. J. Epidemiol.* **2000**, *29*, 158–167. [CrossRef]
22. Zitzmann, S.; Helm, C. Multilevel analysis of mediation, moderation, and nonlinear effects in small samples, using expected a posteriori estimates of factor scores. *Struct. Equ. Model.* **2021**, *28*, 529–546. [CrossRef]
23. Depaoli, S.; Clifton, J.P. A Bayesian approach to multilevel structural equation modeling With continuous and dichotomous outcomes. *Struct. Equ. Model.* **2015**, *22*, 327–351. [CrossRef]
24. Zitzmann, S. A computationally more efficient and more accurate stepwise approach for correcting for sampling error and measurement error. *Multivar. Behav. Res.* **2018**, *53*, 612–632. [CrossRef]
25. Molenberghs, G.; Verbeke, G. A note on a hierarchical interpretation for negative variance components. *Stat. Model.* **2011**, *11*, 389–408. [CrossRef]
26. Stoel, R.D.; Garre, F.G.; Dolan, C.; van den Wittenboer, G. On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychol. Methods* **2006**, *11*, 439–455. [CrossRef]
27. Gerbing, D.W.; Anderson, J.C. Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika* **1987**, *52*, 99–111. [CrossRef]
28. Bentler, P.M. *EQS 6 Structural Equations Program Manual*, 6th ed.; Multivariate Software, Inc.: Encino, CA, USA, 2006.
29. Martin, J.K.; McDonald, R.P. Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika* **1975**, *40*, 505–517. [CrossRef]
30. Lee, S.Y. A Bayesian approach to confirmatory factor analysis. *Psychometrika* **1981**, *46*, 153–160. [CrossRef]
31. Browne, W.J.; Draper, D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal.* **2006**, *1*, 473–514. [CrossRef]
32. Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 515–534. [CrossRef]
33. Lambert, P.C.; Sutton, A.J.; Burton, P.R.; Abrams, K.R.; Jones, D.R. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat. Med.* **2005**, *24*, 2401–2428. [CrossRef]
34. Zitzmann, S.; Helm, C.; Hecht, M. Prior specification for more stable Bayesian estimation of multilevel latent variable models in small samples: A comparative investigation of two different approaches. *Front. Psychol.* **2021**, *11*, 611267. [CrossRef]
35. Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*; Analytical Methods for Social Research, Cambridge University Press: New York, NY, USA, 2007.
36. Hoff, P.D. *A First Course in Bayesian Statistical Methods*; Springer: New York, NY, USA, 2009.
37. Zitzmann, S.; Lüdtke, O.; Robitzsch, A.; Hecht, M. On the performance of Bayesian approaches in small samples: A comment on Smid, McNeish, Miočević, and van de Schoot (2020). *Struct. Equ. Model.* **2021**, *28*, 40–50. [CrossRef]
38. Zitzmann, S.; Lüdtke, O.; Robitzsch, A. A Bayesian approach to more stable estimates of group-level effects in contextual studies. *Multivar. Behav. Res.* **2015**, *50*, 688–705. [CrossRef] [PubMed]
39. Spiegelhalter, D.; Thomas, A.; Best, N.; Lunn, D. WinBUGS User Manual. 2003. Available online: <http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf> (accessed on 17 May 2022).
40. Plummer, M. JAGS Version 3.4.0 User Manual. 2013. Available online: [http://sourceforge.net/projects/mcmc-jags/files/Manuals/3.x/jags\\_user\\_manual.pdf](http://sourceforge.net/projects/mcmc-jags/files/Manuals/3.x/jags_user_manual.pdf) (accessed on 17 May 2022).
41. Lüdtke, O.; Robitzsch, A.; Wagner, J. More stable estimation of the STARTS model: A Bayesian approach using Markov chain Monte Carlo techniques. *Psychol. Methods* **2018**, *23*, 570–593. [CrossRef]
42. Zitzmann, S.; Lohmann, J.F.; Krammer, G.; Helm, C.; Aydin, B.; Hecht, M. A Bayesian EAP-based nonlinear extension of Croon and van Veldhoven’s model for analyzing data from micro-macro multilevel designs. *Mathematics* **2022**, *10*, 842. [CrossRef]