Statistical Inference, Econometric Analysis and Matrix Algebra

Bernhard Schipp • Walter Krämer
Editors

Statistical Inference, Econometric Analysis and Matrix Algebra

Festschrift in Honour of Götz Trenkler

Physica-Verlag
A Springer Company

Editors

Prof. Dr. Walter Krämer
Universität Dortmund
LS Statistik und Ökonometrie
Vogelpothsweg 78
44221 Dortmund
Germany
walterk@statistik.uni-dortmund.de

Prof. Dr. Bernhard Schipp
TU Dresden
Fak. Wirtschaftswissenschaften
Abt. Ökonometrie
Mommsenstr. 12
01062 Dresden
Germany
schipp@tu-dresden.de

ISBN: 978-3-7908-2120-8 e-ISBN: 978-3-7908-2121-5

Library of Congress Control Number: 2008936140

© 2009 Physica-Verlag Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Optimal Estimation in a Linear Regression Model using Incomplete Prior Information

Helge Toutenburg, Shalabh, and Christian Heumann

Abstract For the estimation of regression coefficients in a linear model when incomplete prior information is available, the optimal estimators in the classes of linear heterogeneous and linear homogeneous estimators are considered. As they involve some unknowns, they are operationalized by substituting unbiased estimators for the unknown quantities. The properties of resulting feasible estimators are analyzed and the effect of operationalization is studied. A comparison of the heterogeneous and homogeneous estimation techniques is also presented.

1 Introduction

Postulating the prior information in the form of a set of stochastic linear restrictions binding the coefficients in a linear regression model, Theil and Goldberger [3] have developed an interesting framework of the mixed regression estimation for the model parameters; see e.g., Srivastava [2] for an annotated bibliography of earlier developments and Rao et al. [1] for some recent advances. Such a framework assumes that the variance covariance matrix in the given prior information is known. This specification may not be accomplished in many practical situations where the variance covariance may not be available for one reason or the other. Even if available, its accuracy may be doubtful and consequently its credibility may be sufficiently low. One may then prefer to discard it and treat it as unknown. Appreciating such circumstances, Toutenburg et al. [4] have introduced the method of weakly unbiased estimation for the regression coefficients and have derived the optimal estimators in the classes of linear homogeneous as well as linear heterogeneous estimators through the minimization of risk function under a general quadratic loss structure. Unfortunately, the thus obtained optimal estimators are not functions of

Helge Toutenburg
Institut für Statistik, Universität München, D-80799 München, Germany
toutenb@stat.uni-muenchen.de

observations alone. They involve the coefficient vector itself, which is being estimated, besides the scaling factor of the disturbance variance covariance matrix. Consequently, as acknowledged by Toutenburg et al. [4], such estimators have no practical utility.

In this paper, we apply a simple operationalization technique for obtaining the feasible versions of the optimal estimators. The technique essentially involves replacement of unknown quantities by their unbiased and/or consistent estimators. Such a substitution generally destroys the optimality and superiority properties. A study of the damage done to the optimal properties is the subject matter of our investigations. It is found that the process of operationalization may often alter the conclusions that are drawn from the performance of optimal estimators that are not friendly with users due to involvement of unknown parameters.

The plan of presentation is as follows. In Sect. 2, we describe the model and present the estimators for the vector of regression coefficients. Their properties are discussed in Sect. 3. Some numerical results about the behaviour of estimators in finite samples are reported in Sect. 4. Some summarizing remarks are then presented in Sect. 5. In the last, the Appendix gives the derivation of main results.

2 Estimators for Regression Coefficients

Consider the following linear regression model:

$$y = X\beta + \varepsilon \,, \tag{1}$$

where y is a $n \times 1$ vector of n observations on the study variable, X is a $n \times p$ matrix of n observations on the p explanatory variables, β is a $p \times 1$ vector of regression coefficients and ε is a $n \times 1$ vector of disturbances.

In addition to the observations, let us be given some incomplete prior information in the form of a set of stochastic linear restrictions binding the regression coefficients:

$$r = R\beta + \phi \,, \tag{2}$$

where r is a $m \times 1$ vector, R is a full row rank matrix of order $m \times p$ and ϕ is a $m \times 1$ vector of disturbances.

It is assumed that ε and ϕ are stochastically independent. Further, ε has mean vector 0 and variance covariance matrix $\sigma^2 W$ in which the scalar σ is unknown but the matrix W is known. Similarly, ϕ has mean vector 0 and variance covariance matrix $\sigma^2 V$.

When V is available, the mixed regression estimator of β proposed by Theil and Goldberger [3] is given by

$$\begin{aligned} b_{MR} &= (S + R'V^{-1}R)^{-1}(X'W^{-1}y + R'V^{-1}r) \\ &= b + S^{-1}R'(RS^{-1}R' + V)^{-1}(r - Rb) \,, \end{aligned} \tag{3}$$

where S denotes the matrix $X'W^{-1}X$ and $b = S^{-1}X'W^{-1}y$ is the generalized least squares estimator of β.

In practice, V may not be known all the time and then the mixed regression estimator cannot be used. Often, V may be given but its accuracy and credibility may be questionable. Consequently, one may be willing to assume V as unknown rather than known. In such circumstances, the mixed regression estimator (3) cannot be used.

For handling the case of unknown V, Toutenburg et al. [4] have pioneered the concept of weakly unbiasedness and utilized it for the estimation of β. Accordingly, an estimator $\hat{\beta}$ is said to be weakly–(R,r)–unbiased with respect to the stochastic linear restrictions (2) when the conditional expectation of $R\hat{\beta}$ given r is equal to r itself, i.e.,

$$E(R\hat{\beta} \mid r) = r \tag{4}$$

whence it follows that the unconditional expectation of $R\hat{\beta}$ is $R\beta$.

It may be observed that the unbiasedness of $\hat{\beta}$ for β implies weakly–(R,r)–unbiasedness of $\hat{\beta}$ but its converse may not be necessarily always true.

Taking the performance criterion as

$$R_A(\hat{\beta},\beta) = E(\hat{\beta}-\beta)'A(\hat{\beta}-\beta)\,, \tag{5}$$

that is, the risk associated with an estimator $\hat{\beta}$ of β under a general quadratic loss function with a positive definite loss matrix A, Toutenburg et al. [4] have discussed the minimum risk estimator of β; see also Rao et al. [1] for an expository account.

The optimal estimator in the class of linear and weakly unbiased heterogeneous estimators for β is given by

$$\hat{\beta}_1 = \beta + A^{-1}R'(RA^{-1}R')^{-1}(r - R\beta) \tag{6}$$

while the optimal estimator in the class of linear and weakly unbiased homogeneous estimators is

$$\hat{\beta}_2 = \frac{\beta'X'W^{-1}y}{\sigma^2+\beta'S\beta}\left[\beta + A^{-1}R'(RA^{-1}R')^{-1}\left(\frac{\sigma^2\beta'S\beta}{\beta'S\beta}r - R\beta\right)\right]. \tag{7}$$

Clearly, $\hat{\beta}_1$ and $\hat{\beta}_2$ are not estimators in true sense owing to involvement of β itself besides σ^2 which is also unknown. As a consequence, they have no practical utility.

A simple solution to operationalize $\hat{\beta}_1$ and $\hat{\beta}_2$ is to replace the unknown quantities by their estimators. Such a process of operationalization generally destroys the optimality of estimators.

If we replace β by its generalized least squares estimator b and σ^2 by its unbiased estimator

$$s^2 = \left(\frac{1}{n-p}\right)(y-Xb)'W^{-1}(y-Xb), \tag{8}$$

we obtain the following feasible versions of $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$\tilde{\beta}_1 = b + A^{-1}R'(RA^{-1}R')^{-1}(r - Rb) \tag{9}$$

$$\tilde{\beta}_2 = \frac{b'Sb}{s^2 + b'Sb}\left[b + A^{-1}R'(RA^{-1}R')^{-1}\left(\frac{s^2 + b'Sb}{b'Sb}r - Rb\right)\right]. \tag{10}$$

It may be remarked that Toutenburg, Toutenburg et al. ([4], Sect. 4) have derived a feasible and unbiased version of the estimator $\hat{\beta}_1$ such that it is optimal in the class of linear homogeneous estimators. This estimator is same as $\tilde{\beta}_1$. It is thus interesting to note that when the optimal estimator in the class of linear heterogeneous estimators is operationalized, it turns out to have optimal performance in the class of linear homogeneous estimators.

3 Comparison of Estimators

It may be observed that a comparison of the estimator $\tilde{\beta}_1$ with $\hat{\beta}_1$ and $\tilde{\beta}_2$ with $\hat{\beta}_2$ will furnish us an idea about the changes in the properties due to the process of operationalization. Similarly, if we compare $\hat{\beta}_1$ and $\hat{\beta}_2$ with $\tilde{\beta}_1$ and $\tilde{\beta}_2$, it will reveal the changes in the properties of the optimal estimator and its feasible version in the classes of linear heterogeneous and linear homogeneous estimators.

3.1 Linearity

First of all, we may observe that both the estimators $\hat{\beta}_1$ and $\tilde{\beta}_1$ are linear and thus the process of operationalization does not alter the linearity of estimator. This is not true when we consider the optimal estimator $\hat{\beta}_2$ in the class of linear homogeneous estimators and its feasible version $\tilde{\beta}_2$. Further, from (9) and (10), we notice that

$$\tilde{\beta}_2 = \frac{1}{s^2 + b'Sb}[b'Sb\tilde{\beta}_1 + s^2A^{-1}R'(RA^{-1}R')^{-1}r] \tag{11}$$

so that $\tilde{\beta}_2$ is a weighted average of $\tilde{\beta}_1$ and $A^{-1}R'(RA^{-1}R')^{-1}r$ while such a result does not hold in case of $\hat{\beta}_2$.

3.2 Unbiasedness

From (9) and (10), we observe that

$$R\tilde{\beta}_1 = R\tilde{\beta}_2 = r \tag{12}$$

whence it is obvious that both the estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are weakly–(R,r)–unbiased like $\hat{\beta}_1$ and $\hat{\beta}_2$. Thus the operationalization does not disturb the property of weakly unbiasedness.

Next, let us consider the traditional unbiasedness property. It is easy to see that the optimal estimator $\hat{\beta}_1$ and its feasible version $\tilde{\beta}_1$ in the class of linear heterogeneous estimators are unbiased while the optimal estimators $\hat{\beta}_2$ and its feasible version $\tilde{\beta}_2$ in the class of homogeneous estimators are generally not unbiased. This may serve as an interesting example to demonstrate that weakly unbiasedness does not necessarily imply unbiasedness. Thus, with respect to the criterion of unbiasedness, no change arises due to operationalization.

3.3 Bias Vector

Let us examine the bias vectors of the estimators $\hat{\beta}_2$ and $\tilde{\beta}_2$.

It is easy to see that the bias vector of $\hat{\beta}_2$ is given by

$$\begin{aligned} B(\hat{\beta}_2) &= E(\hat{\beta}_2 - \beta) \\ &= -\frac{\sigma^2}{\sigma^2 + \beta' S \beta} A^{-1} M \beta \,, \end{aligned} \tag{13}$$

where

$$M = A - R'(RA^{-1}R')^{-1}R. \tag{14}$$

The exact expression for the bias vector of $\tilde{\beta}_2$ is impossible to derive without assuming any specific distribution for the elements of disturbance vector ε. It may be further observed that even under the specification of distribution like normality, the exact expression will be sufficiently intricate and any clear inference will be hard to deduce. We therefore consider its approximate expression using the large sample asymptotic theory. For this purpose, it is assumed that explanatory variables in the model are at least asymptotically cooperative, i.e., the limiting form of the matrix $n^{-1}X'W^{-1}X$ as n tends to infinity is a finite and nonsingular matrix. We also assume that ε follows a multivariate normal distribution.

Theorem I: If we write $Q = n^{-1}S$, the bias vector of $\tilde{\beta}_2$ to order $O(n^{-1})$ is given by

$$\begin{aligned} B(\tilde{\beta}_2) &= E(\tilde{\beta}_2 - \beta) \\ &= -\frac{\sigma^2}{n\beta' Q \beta} A^{-1} M \beta + \frac{\sigma^2}{n^2 \beta' Q \beta}\left[p + (p-1)\frac{\sigma^2}{\beta' Q \beta}\right] A^{-1} M \beta \end{aligned} \tag{15}$$

which is derived in the Appendix.

A similar expression for the optimal estimator to order $O(n^{-2})$ can be straightforwardly obtained from (13) as follows:

$$
\begin{aligned}
B(\hat{\beta}_2) &= -\frac{\sigma^2}{n\beta'Q\beta}\left(1+\frac{\sigma^2}{n\beta'Q\beta}\right)^{-1} A^{-1}M\beta \\
&= -\frac{\sigma^2}{n\beta'Q\beta}A^{-1}M\beta + \frac{\sigma^4}{n^2(\beta'Q\beta)^2}A^{-1}M\beta. \qquad (16)
\end{aligned}
$$

If we compare the optimal estimators $\hat{\beta}_2$ and its feasible version $\tilde{\beta}_2$ with respect to the criterion of bias to order $O(n^{-1})$ only, it follows from (15) and (16) that both the estimators are equally good. This implies that operationalization does not alter the asymptotic bias to order $O(n^{-1})$.

When we retain the term of order $O(n^{-2})$ also in the bias vector, the two estimators are found to have different bias vectors and the effect of operationalization precipitates.

Let us now compare the estimators $\hat{\beta}_2$ and $\tilde{\beta}_2$ according to the length of their bias vectors. If we consider terms upto order $O(n^{-3})$ only, we observe from (15) and (16) that

$$
[B(\hat{\beta}_2)]'[B(\hat{\beta}_2)] - [B(\tilde{\beta}_2)]'[B(\tilde{\beta}_2)] = \frac{2\sigma^2}{n^3\beta'Q\beta}\left[p+(p-2)\frac{\sigma^2}{\beta'Q\beta}\right]\beta'MA^{-2}M\beta.
$$

It is thus surprising that the feasible estimator $\tilde{\beta}_2$ is preferable to the optimal estimator with respect to the criterion of the bias vector length to the given order of approximation in the case of two or more explanatory variables in the model. If $p=1$, this result continues to hold true provided that $\beta'Q\beta$ is greater than σ^2. Thus it is interesting to note that operationalization of optimal estimator improves the performance with respect to the bias vector length criterion.

3.4 Conditional Risk Function

From Toutenburg et al. ([4], p. 530), the conditional risk function of $\hat{\beta}_1$, given r is

$$
\begin{aligned}
R_A(\hat{\beta}_1,\beta \mid r) &= E[(\hat{\beta}_1-\beta)'A(\hat{\beta}_1-\beta)\mid r] \\
&= (r-R\beta)'(RA^{-1}R')^{-1}(r-R\beta). \qquad (17)
\end{aligned}
$$

Similarly, the conditional risk function of $\hat{\beta}_2$ given r can be easily obtained:

$$
\begin{aligned}
R_A(\hat{\beta}_2,\beta \mid r) &= E[(\hat{\beta}_2-\beta)'A(\hat{\beta}_2-\beta)\mid r] \\
&= (r-R\beta)'(RA^{-1}R')^{-1}(r-R\beta) \\
&\quad + \frac{\sigma^2}{\sigma^2+n\beta'Q\beta}\left[\beta'M\beta+\left(1+\frac{\sigma^2}{n\beta'Q\beta}\right)r'(RA^{-1}R')^{-1}r\right]. \qquad (18)
\end{aligned}
$$

Using the result

$$\begin{aligned}\frac{\sigma^2}{\sigma^2+n\beta'Q\beta} &= \frac{\sigma^2}{n\beta'Q\beta}\left(1+\frac{\sigma^2}{n\beta'Q\beta}\right)^{-1}\\ &= \frac{\sigma^2}{n\beta'Q\beta}-\frac{\sigma^4}{n^2(\beta'Q\beta)^2}+O(n^{-3}),\end{aligned} \tag{19}$$

we can express

$$\begin{aligned}R_A(\hat{\beta}_2,\beta \mid r) &= (r-R\beta)'(RA^{-1}R')^{-1}(r-R\beta)\\ &\quad+\frac{\sigma^2}{n\beta'Q\beta}[\beta'M\beta+r'(RA^{-1}R')^{-1}r]-\frac{\sigma^4\beta'M\beta}{n^2(\beta'Q\beta)^2}+O(n^{-3}).\end{aligned} \tag{20}$$

For the feasible estimator $\tilde{\beta}_1$, it can be easily seen that the conditional risk function of $\tilde{\beta}_1$ given r is given by

$$\begin{aligned}R_A(\tilde{\beta}_1,\beta \mid r) &= E[(\tilde{\beta}_1-\beta)'A(\tilde{\beta}_1-\beta)\mid r]\\ &= (r-R\beta)'(RA^{-1}R')^{-1}(r-R\beta)+\frac{\sigma^2}{n}trMQ^{-1}.\end{aligned} \tag{21}$$

As the exact expression for the conditional risk of the estimator $\tilde{\beta}_2$ is too complex to permit the deduction of any clear inference regarding the performance relative to other estimators, we consider its asymptotic approximation under the normality of disturbances. This is derived in Appendix.

Theorem II: The conditional risk function of the estimator $\tilde{\beta}_2$ given r to order $O(n^{-2})$ is given by

$$\begin{aligned}R_A(\tilde{\beta}_2,\beta \mid r) &= E[(\tilde{\beta}_2-\beta)'A(\tilde{\beta}_2-\beta)\mid r]\\ &= (r-R\beta)'(RA^{-1}R')^{-1}(r-R\beta)+\frac{\sigma^2}{n}trMQ^{-1}\\ &\quad-\frac{\sigma^4}{n^2\beta'Q\beta}\left[2trMQ^{-1}-5\left(\frac{\beta'M\beta}{\beta'Q\beta}\right)\right].\end{aligned} \tag{22}$$

It is obvious from (17) and (21) that the operationalization process leads to an increase in the conditional risk. Similarly, comparing $\hat{\beta}_2$ and $\tilde{\beta}_2$ with respect to the criterion of the conditional risk given r to order $O(n^{-1})$, we observe from (20) and (22) that the operationalization process results in an increase in the conditional risk when

$$trMQ^{-1} > \frac{\beta'M\beta}{\beta'Q\beta}+\frac{r'(RA^{-1}R')r}{\beta'Q\beta}. \tag{23}$$

The opposite is true, i.e., operationalization reduces the conditional risk when the inequality (23) holds true with a reversed sign.

If we compare the exact expressions (17) and (18) for the conditional risk function given r, it is seen that the estimator $\hat{\beta}_1$ is uniformly superior to $\hat{\beta}_2$. This result remains true, as is evident from (21) and (22), for their feasible versions also when the criterion is the conditional risk given r to order $O(n^{-2})$ and

$$trMQ^{-1} < 2.5\left(\frac{\beta' M\beta}{\beta' Q\beta}\right) \tag{24}$$

while the opposite is true, i.e., $\tilde{\beta}_2$ has smaller risk than $\tilde{\beta}_1$ when

$$trMQ^{-1} > 2.5\left(\frac{\beta' M\beta}{\beta' Q\beta}\right). \tag{25}$$

The conditions (24) and (25) have little usefulness in actual practice because they cannot be verified due to involvement of β. However, we can deduce sufficient conditions that are simple and easy to check.

Let $\lambda_{\min}$ and $\lambda_{\max}$ be the minimum and maximum eigen values of the matrix M in the metric of Q, and T be the total of all the eigenvalues. Now, it is seen that the condition (24) is satisfied so long as

$$T < 2.5\lambda_{\min} \tag{26}$$

which is a sufficient condition for the superiority of $\tilde{\beta}_1$ over $\tilde{\beta}_2$.

Similarly, for the superiority of $\tilde{\beta}_2$ over $\tilde{\beta}_1$, the following sufficient condition can be deduced from (25):

$$T > 2.5\lambda_{\max}\,. \tag{27}$$

We thus observe that the optimal estimator $\hat{\beta}_1$ is uniformly superior to $\hat{\beta}_2$ with respect to both the criteria of conditional and unconditional risks. The property of uniform superiority is lost when they are operationalized for obtaining feasible estimators. So much so that the superiority result may take an opposite turn at times.

Further, we notice that the reduction in the conditional risk of $\hat{\beta}_1$ over $\hat{\beta}_2$ is generally different in comparison to the corresponding reduction in the conditional risk when their feasible versions are considered. The change in the conditional risk performance of the optimal estimators starts appearing in the term of order $O(n^{-1})$. When their feasible versions are compared, the leading term of the change in risk is of order $O(n^{-2})$. This can be attributed to the process of operationalization.

3.5 Unconditional Risk Function

Now let us compare the estimators under the criterion of the unconditional risk function.

It can be easily seen from (17), (18), (20), (21) and (22) that the unconditional risk functions of the four estimators are given by

$$\begin{aligned}
R_A(\hat{\beta}_1,\beta) &= E(\hat{\beta}_1-\beta)'A(\hat{\beta}_1-\beta)\\
&= \sigma^2 trV(RA^{-1}R')^{-1} \qquad (28)\\
R_A(\hat{\beta}_2,\beta) &= E(\hat{\beta}_2-\beta)'A(\hat{\beta}_2-\beta)\\
&= \sigma^2 trV(RA^{-1}R')^{-1}\\
&\quad +\frac{\sigma^2}{n\beta'Q\beta}\left[\beta'A\beta+\sigma^2 trV(RA^{-1}R')^{-1}-\frac{\sigma^2}{\sigma^2+n\beta'Q\beta}\beta'M\beta\right]\\
&= \sigma^2 trV(RA^{-1}R')^{-1}+\frac{\sigma^2}{n\beta'Q\beta}\left[\beta'A\beta+\sigma^2 trV(RA^{-1}R')^{-1}\right]\\
&\quad -\frac{\sigma^4\beta'M\beta}{n^2(\beta'Q\beta)^2}+O(n^{-3}) \qquad (29)\\
R_A(\tilde{\beta}_1,\beta) &= E(\tilde{\beta}_1-\beta)'A(\tilde{\beta}_1-\beta)\\
&= \sigma^2 trV(RA^{-1}R')^{-1}+\frac{\sigma^2}{n}trMQ^{-1} \qquad (30)\\
R_A(\tilde{\beta}_2,\beta) &= E(\tilde{\beta}_2-\beta)'A(\tilde{\beta}_2-\beta)\\
&= \sigma^2 trV(RA^{-1}R')^{-1}+\frac{\sigma^2}{n}trMQ^{-1}\\
&\quad -\frac{\sigma^4}{n^2\beta'Q\beta}\left[2trMQ^{-1}-5\left(\frac{\beta'M\beta}{\beta'Q\beta}\right)\right]+O(n^{-3}). \qquad (31)
\end{aligned}$$

Looking at the above expressions, it is interesting to note that the relative performance of one estimator over the other is same as observed under the criterion of the conditional risk given r.

4 Simulation Study

We conducted a simulation experiment to study the performance of the estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ with respect to the ordinary least squares estimator b. The sample size was fixed at $n = 30$. The design matrix X contained an intercept term and six covariates which were generated from multivariate normal distribution with variance 1 and equal correlation of 0.4. The mean vector of the covariates was $(-2,-2,-2,2,2,2)$. The true response vector (without the error term ε) was then calculated as $\tilde{y} = X\beta$ with the 7×1 true parameter vector $\beta = (10,10,10,10,-1,-1,-1)$. The restriction matrix R was generated as a 3×7 matrix containing uniform random numbers. The true restriction vector (without the error term ϕ) was calculated as $\tilde{r} = R\beta$. Then in a loop with 5,000 replications, in every replication, new error terms ε and ϕ were added in $\tilde{y}$ and $\tilde{r}$ to get y and r respectively. The errors were generated independently from normal random variables with variances $\sigma^2 = 40$ for ε_i, $i = 1,\dots,n$ and σ^2/c for ϕ_j, $j = 1,2,3$. The factor c controls the accuracy of the prior information compared to the noise in the data. If c is high, the prior information is more accurate than the case when c is low. Note that $c < 1$ means that the prior information is more noisy than the data which indicates that it is probably useless in practice. In fact we only expect the proposed estimators to be better than b if c is considerably larger than 1. For comparison of the estimators, we calculated the measure

$$\mathrm{MRMSE} = \frac{1}{5000} \sum_{k=1}^{5000} \sqrt{\frac{1}{7}(\hat{\beta} - \beta)'(\hat{\beta} - \beta)}\,,$$

(mean of root mean squared errors) where $\hat{\beta}$ stands for one of the estimators b, $\tilde{\beta}_1$ or $\tilde{\beta}_2$. Figure 1 shows the distribution of the root mean squared errors $\sqrt{\frac{1}{7}(\hat{\beta} - \beta)'(\hat{\beta} - \beta)}$ for each estimator based on 5,000 replications with $c = 100$. This means that the prior information was not perfect but very reliable ($\sigma^2/c = 40/100 = 0.4$). A considerable gain can be observed by using one of the new proposed estimators while there is no noticeable difference between $\tilde{\beta}_1$ and $\tilde{\beta}_2$. The *MRMSE*s in that run were 2.64 for b and 1.85 for $\tilde{\beta}_1$ and $\tilde{\beta}_2$. The picture changes when we decrease c. For example when $c = 4$ (which means that the standard error of ϕ_j is half of the standard error of the noise in the data), then the *MRMSE* were 3.09 for b and 2.65 for $\tilde{\beta}_1$ and $\tilde{\beta}_2$. Figure 2 shows the corresponding boxplots. But a general conclusion is not possible since the results clearly also depend on the matrices X, R and vector β itself.

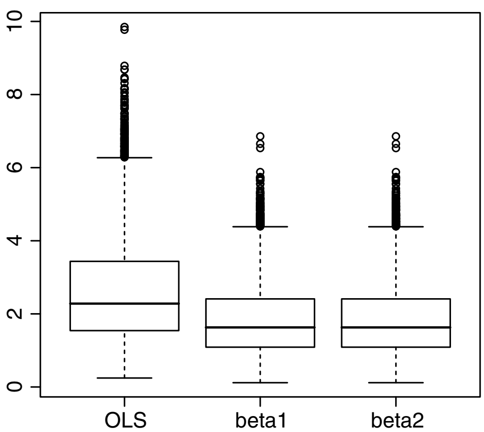

Fig. 1 Boxplot of root mean squared errors of the three estimators with $c = 100$

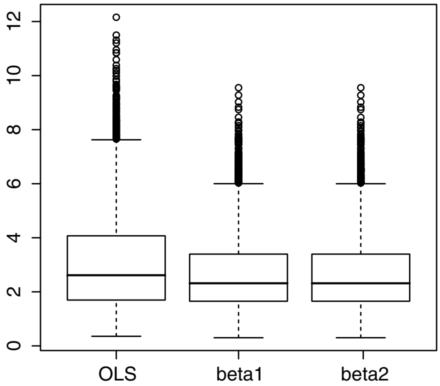

Fig. 2 Boxplot of root mean squared errors of the three estimators with $c = 4$

5 Some Summarizing Remarks

We have considered the minimum risk approach for the estimation of coefficients in a linear regression model when incomplete prior information specifying a set of linear stochastic restrictions with unknown variance covariance matrix is available. In the linear and weakly unbiased heterogeneous and homogeneous classes of estimators, the optimal estimators obtained by Toutenburg et al. [4] as well as their feasible versions are presented. Properties of these four estimators are then discussed.

Analyzing the effect of operationalizing the optimal estimators, we have observed that the property of linearity is retained only in case of heterogeneous estimation. So far as the property of weakly unbiasedness is concerned, the process of operationalization has no influence. But when the traditional unbiasedness is considered, it is seen that the optimal heterogeneous estimator remains unbiased while the optimal homogeneous estimator is generally biased. This remains true when their feasible versions are considered. In other words, the process of operationalizations does not bring any change in the performance of estimators.

Looking at the direction and magnitude of bias, we have found that the optimal estimator and its feasible version in the case of homogeneous estimation have identical bias vectors to order $O(n^{-1})$ implying that the operationalization process has no effect on the bias vector in large samples. But when the sample size is not large enough and the term of order $O(n^{-2})$ is no more negligible, the effect of operationalization appears. If we compare the optimal estimator and its feasible version with respect to the criterion of the length of the bias vector to order $O(n^{-3})$, it is seen that the operationalization improves the performance provided that there are two or more explanatory variables in the model. This result remains true in the case of one explanatory variable also under a certain condition.

Examining the risk functions, it is observed that the relative performance of one estimator over the other remains unaltered whether the criterion is conditional risk given r or the unconditional risk.

When we compare the risk functions of the optimal heterogeneous estimator and its feasible version, it is found that the process of operationalization invariably increases the risk. Such is not the case when we compare the optimal homogeneous estimator and its feasible version. Here the operationalization may lead to a reduction in risk at times; see the condition (23).

Next, it is observed that the optimal heterogeneous estimator has always smaller risk in comparison to the optimal homogeneous estimator. When they are operationalized in a bid to obtain feasible estimators, the property of uniform superiority is lost. We have therefore obtained sufficient conditions for the superiority of one feasible estimator over the other. An important aspect of these conditions is that they are simple and easy to check in practice.

Further, we have observed the magnitude of change in the risk of one optimal estimator over the other optimal estimator is generally different when their feasible versions are considered. In case of optimal estimators, the change occurs at the level of order $O(n^{-1})$ but when the feasible estimators are compared, this level is of order $O(n^{-2})$. This brings out the impact of operationalization process.

Finally, it may be remarked that if we consider the asymptotic distribution of the estimation error, i.e., the difference between the estimator and the coefficient vector, both the optimal estimators as well as their feasible versions have same asymptotic distribution. Thus the process of operationalization does not show any impact on the asymptotic properties of estimators. It may alter the performance of estimators when the number of observations is not sufficiently large. The difference in the performance of estimators is clear in finite samples through simulation experiment.

Appendix

If we define

$$\begin{aligned} z &= \frac{1}{n^{1/2}} X'W^{-1}\varepsilon\,, \\ u &= \frac{1}{\sigma^2 n^{1/2}} \varepsilon' W^{-1}\varepsilon - n^{1/2}\,, \\ v &= \frac{1}{\sigma^2} \varepsilon' W^{-1} X S^{-1} X' W^{-1}\varepsilon\,, \end{aligned}$$

we can write

$$\begin{aligned} b'Sb &= \beta'S\beta + 2\beta'X'W^{-1}\varepsilon + \varepsilon'W^{-1}XS^{-1}X'W^{-1}\varepsilon \\ &= n\beta'Q\beta + 2n^{1/2}\beta'z + \sigma^2 v \qquad (32) \\ s^2 &= \frac{1}{(n-p)}(y-Xb)'W^{-1}(y-Xb) \\ &= \sigma^2\left[1 + \frac{u}{n^{1/2}} - \frac{v}{n}\right] + O_p(n^{-3/2}). \qquad (33) \end{aligned}$$

Using these, we can express

$$
\begin{aligned}
\frac{s^2}{s^2+b'Sb} &= \frac{\sigma^2}{n\beta'Q\beta}\left[1+\frac{u}{n^{1/2}}-\frac{v}{n}+O_p(n^{-3/2})\right] \\
&\quad * \left[1+\frac{2\beta'z}{n^{1/2}\beta'Q\beta}+\frac{\sigma^2(1+v)}{n\beta'Q\beta}+O_p(n^{-3/2})\right]^{-1} \\
&= \frac{\sigma^2}{n\beta'Q\beta}\left[1+\frac{u}{n^{1/2}}-\frac{v}{n}+O_p(n^{-3/2})\right] \\
&\quad * \left[1-\frac{2\beta'z}{n^{1/2}\beta'Q\beta}-\sigma^2\left(1+v-\frac{4\beta'zz'\beta}{\sigma^2\beta'Q\beta}\right)+O_p(n^{-3/2})\right] \\
&= \frac{\sigma^2}{n\beta'Q\beta}+\frac{\sigma^2}{n^{3/2}\beta'Q\beta}\left(u-\frac{2\beta'z}{\beta'Q\beta}\right) \\
&\quad -\frac{\sigma^2}{n^2\beta'Q\beta}\left(v+\frac{2u\beta'z+\sigma^2+\sigma^2 v}{\beta'Q\beta}-\frac{4\beta'zz'\beta}{(\beta'Q\beta)^2}\right)+O_p(n^{-5/2}).
\end{aligned}
\tag{34}
$$

Utilizing these results, we can express

$$
\begin{aligned}
(\tilde{\beta}_2-\beta) &= (\tilde{\beta}_1-\beta)-\frac{s^2}{s^2+b'Sb}A^{-1}Mb \\
&= \xi_0+\frac{1}{n^{1/2}}\xi_{1/2}+\frac{1}{n}\xi_1+\frac{1}{n^{3/2}}\xi_{3/2}+\frac{1}{n^2}\xi_2+O_p(n^{-5/2})\,,
\end{aligned}
\tag{35}
$$

where

$$
\begin{aligned}
\xi_0 &= A^{-1}R'(R'A^{-1}R')^{-1}(r-R\beta) \\
\xi_{1/2} &= A^{-1}MQ^{-1}z \\
\xi_1 &= -\frac{\sigma^2}{\beta'Q\beta}A^{-1}M\beta \\
\xi_{3/2} &= -\frac{\sigma^2}{\beta'Q\beta}\left[\left(u-\frac{2\beta'z}{\beta'Q\beta}\right)A^{-1}M\beta+A^{-1}MQ^{-1}z\right] \\
\xi_2 &= \frac{\sigma^2}{\beta'Q\beta}\left[\left(v+\frac{2u\beta'z+\sigma^2+\sigma^2 v}{\beta'Q\beta}-\frac{4\beta'zz'\beta}{(\beta'Q\beta)^2}\right)A^{-1}M\beta\right. \\
&\qquad \left.-\left(u-\frac{2\beta'z}{\beta'Q\beta}\right)A^{-1}MQ^{-1}z\right].
\end{aligned}
$$

By virtue of normality of ε, it is easy to see that

$$
\begin{aligned}
E(\xi_0 \mid r) &= \xi_0\,,\;\; E(\xi_0)=0\,, \\
E(\xi_{1/2} \mid r) &= E(\xi_{1/2})=0\,, \\
E(\xi_1 \mid r) &= E(\xi_1)=\xi_1\,,
\end{aligned}
$$

$$E(\xi_{3/2} \mid r) = E(\xi_{3/2}) = 0 \,,$$
$$E(\xi_2 \mid r) = E(\xi_2) = \frac{\sigma^2}{\beta' Q \beta}\left[p + (p-1)\frac{\sigma^2}{\beta' Q \beta}\right] A^{-1} M \beta.$$

Using these results, we obtain from (35) the expression (15) of Theorem I.

Next, we observe from (35) that the conditional risk function of $\tilde{\beta}_2$ to order $O(n^{-2})$ is given by

$$\begin{aligned}
R_A(\tilde{\beta}_2, \beta \mid r) &= E[(\tilde{\beta}_2 - \beta)' A (\tilde{\beta}_2 - \beta) \mid r] \\
&= \xi_0' A \xi_0 + \frac{2}{n^{1/2}} E(\xi_0' A \xi_{1/2}) \\
&\quad + \frac{1}{n} E(\xi_{1/2}' A \xi_{1/2} + 2\xi_0' A \xi_1) + \frac{2}{n^{3/2}} E[\xi_0' A \xi_{3/2} + \xi_{1/2}' A \xi_1] \\
&\quad + \frac{1}{n^2} E(\xi_1' A \xi_1 + 2\xi_0' A \xi_2 + 2\xi_{1/2}' A \xi_{3/2}) + O_p(n^{-5/2}).
\end{aligned} \tag{36}$$

Now it can be easily seen that

$$\begin{aligned}
E(\xi_0' A \xi_{1/2} \mid r) &= 0 \,, \\
E(\xi_{1/2}' A \xi_{1/2} \mid r) &= \sigma^2 tr M Q^{-1} \,, \\
E(\xi_0' A \xi_1 \mid r) &= 0 \,, \\
E(\xi_0' A \xi_{3/2} \mid r) &= 0 \,, \\
E(\xi_{1/2}' A \xi_1 \mid r) &= 0 \,, \\
E(\xi_1' A \xi_1 \mid r) &= \frac{\sigma^4}{(\beta' Q \beta)^2} \beta' M \beta \,, \\
E(\xi_0' A \xi_2 \mid r) &= 0 \,, \\
E(\xi_{1/2}' A \xi_{3/2} \mid r) &= \frac{\sigma^4}{\beta' Q \beta}\left[-tr M Q^{-1} + 2\left(\frac{\beta' M \beta}{\beta' Q \beta}\right)\right] \,,
\end{aligned}$$

where repeated use has been made of the results $RA^{-1}M = 0$ and $MA^{-1}M = M$.

Substituting these results in (36), we obtain the result stated in Theorem II.

References

[1] Rao, C.R., Toutenburg, H., Shalabh, Heumann, C.: Linear Models and Generalizations: Least Squares and Alternatives (3rd ed.). Springer, New York (2008)

[2] Srivastava, V.K.: Estimation of linear single-equation and simultaneous equation models under stochastic linear constraints: An annotated bibliography. Int. Stat. Rev. **48**, 79–82 (1980)

[3] Theil, H., Goldberger, A.S.: On pure and mixed estimation in econometrics. Int. Econ. Rev. **2**, 65–78 (1961)

[4] Toutenburg, H., Trenkler, G., Liski, E.P.: Optimal estimation methods under weakened linear restrictions. Comput. Stat. Data An. **14**, 527–536 (1992)