Tilburg University

Estimation of the Mean of a Univariate Normal Distribution When the Variance is not Known

Danilov, D.L.; Magnus, J.R.

Publication date:
2002

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Danilov, D. L., & Magnus, J. R. (2002). *Estimation of the Mean of a Univariate Normal Distribution When the Variance is not Known*. (CentER Discussion Paper; Vol. 2002-77). Econometrics.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2002-77

**ESTIMATION OF THE MEAN OF A UNIVARIATE
NORMAL DISTRIBUTION WHEN THE VARIANCE IS
NOT KNOWN**

By Dmitry Danilov

August 2002

ISSN 0924-7815

Discussion paper

Estimation of the mean of a univariate normal distribution when the variance is not known

Dmitry Danilov

CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

E-mail: dan@kub.nl

Abstract

Summary: We consider the problem of estimating the first k coefficients in a regression equation with $k + 1$ variables. For this problem with known variance of innovations, the neutral Laplace weighted-average least-squares estimator was introduced in Magnus (2002). We investigate properties of this estimator in the case where the unknown variance is estimated by least squares. We find that the optimality properties of the Laplace estimator only change marginally. Therefore we recommend the neutral Laplace estimator to be used in practice.

Key words: Regression, pretest, Laplace estimator.

JEL Codes: C11, C13.

1 Introduction

Let x be a single observation from a univariate normal distribution with unknown mean θ and unknown variance σ^2 , that is $x \sim N(\theta, \sigma^2)$. Suppose also, that an estimator of σ^2 is available, namely s^2 , that is independent of x and such that $s^2\nu/\sigma^2$ has a χ_ν^2 distribution, where ν is known. In this article we consider the problem of estimating θ in some optimal manner.

The stated problem arises from the following practical situation. Consider the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{z} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$. The difference between $\mathbf{X}$ and $\mathbf{z}$ is that $\mathbf{X}$ contains regressors that always have to be in the model, while $\mathbf{z}$ (also called the auxiliary regressor) may or may not be in the model. We want to estimate parameter $\boldsymbol{\beta}$ in the ‘best’ possible way.

The problem of optimal estimation of $\boldsymbol{\beta}$ has a long history. Early work on pretesting goes back to Berkson (1942) and Bankroft (1944). Huntsberger (1955) explicitly writes out the pretest estimator as a continuous weighted average of the restricted and unrestricted estimators, and proposes an alternative estimator. Feldstein (1973) is concerned with regression estimation when regressors are highly correlated. Admissibility issues are discussed in Blyth (1951), Farrell (1968), Brown (1971), and Berger (1976). A review of the early literature is provided in Judge and Bock (1978). Sawa and Hiromatsu (1973) consider the pretest estimator using the minimax regret criterion, while Toyoda and Wallace (1976) use the average minimum risk criterion. The minimax regret approach is

also used by Droge and Georg (1995) in obtaining adaptive least-squares regression estimates. Roehrig (1984) finds an expression for the mean squared error of the pretest estimator. In a similar way Magnus and Durbin (1999) derive moments of the general weighted-average least-squares estimator. Magnus (2002) introduces the neutral Laplace WALS estimator in a regression context. Recent developments in theory and practice of pretesting can be found in Giles and Giles (1993), Chatfield (1995), and Magnus (1999).

The rest of paper is organized as following. In section 2 we introduce notation and explain the basic setup. In section 3 we reconsider properties of WALS estimators, such as admissibility, risk and regret in the situation when σ is unknown. In section 4 we consider the performance of the neutral Laplace estimator relative to the usual pretest estimator. Section 4 concludes. An appendix contains proofs of all results.

2 Notation and setup

We consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{z} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (1)$$

where $\mathbf{y}(n \times 1)$ is the vector of observations, $\mathbf{X}(n \times k)$ and $\mathbf{z}(n \times 1)$ are matrices of nonrandom regressors, $\boldsymbol{\varepsilon}(n \times 1)$ is a vector of disturbances, and $\boldsymbol{\beta}$ and γ are unknown nonrandom parameters. Suppose that the matrix $(\mathbf{X} : \mathbf{z})$ has full column-rank. Let $\mathbf{b}_u$ and $\hat{\gamma}$ denote the ordinary least-squares estimators of $\boldsymbol{\beta}$ and γ in model (1), and $\mathbf{b}_r$ the ordinary least-squares estimator in model (1) under the restriction $\gamma = 0$. Denote $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\theta = \gamma/\sqrt{\mathbf{z}'\mathbf{M}\mathbf{z}}$ and $\hat{\theta} = \hat{\gamma}/\sqrt{\mathbf{z}'\mathbf{M}\mathbf{z}}$. The weighted-average least-squares estimator, introduced in Magnus and Durbin (1999), is defined as

$$\mathbf{b} = \lambda\mathbf{b}_u + (1 - \lambda)\mathbf{b}_r, \quad \lambda = \lambda(\hat{\theta}, s^2),$$

where $s^2 = \mathbf{y}'\mathbf{M}\mathbf{y}/(n - k - 1)$ is the least-squares estimator of σ^2 . The equivalence theorem in Magnus and Durbin (1999) states that the mean squared error of the WALS estimator of $\boldsymbol{\beta}$ can be represented as

$$\text{MSE}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \text{MSE}(\lambda(\hat{\theta}, s^2)\hat{\theta}) \mathbf{q}\mathbf{q}',$$

where $\mathbf{q} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}(\mathbf{z}'\mathbf{M}\mathbf{z})^{-1/2}$. Therefore we can say that the problem of estimation of $\boldsymbol{\beta}$ in regression model (1) is equivalent to the problem of estimation of the parameter θ by one bivariate observation $(\hat{\theta}, s^2)$. We call the first problem *the regression problem* and second problem *the auxiliary problem*. For determining the optimal $\mathbf{b}$ (in the mean squared error sense) we need to find a function $\lambda(\hat{\theta}, s^2)$

which provides the optimal estimator of θ . Consider the mean squared error of $\lambda(\hat{\theta}, s^2)\hat{\theta}$, that is,

$$\text{MSE}(\lambda(\hat{\theta}, s^2)\hat{\theta}) = \text{E}(\lambda(\hat{\theta}, s^2)\hat{\theta} - \theta)^2,$$

where the expectation on the right hand side is taken with respect to the product of two independent distributions: $N(\theta, \sigma^2)$ and χ_ν . In Magnus (2002) the problem of finding the optimal λ was considered under the assumption that σ is known. He used a λ -function of particular form: $\lambda(\hat{\theta}, s^2) = \Lambda(\hat{\eta})$, where $\hat{\eta} = \hat{\theta}/\sigma$. Then,

$$\text{MSE}(\lambda(\hat{\theta}, s^2)\hat{\theta}) = \sigma^2 \text{E}(\Lambda(\frac{\hat{\theta}}{\sigma})\frac{\hat{\theta}}{\sigma} - \frac{\theta}{\sigma})^2 = \sigma^2 \text{E}(\Lambda(\hat{\eta})\hat{\eta} - \eta)^2,$$

where $\eta = \theta/\sigma$. Therefore the regression problem with known σ is equivalent to the estimation of the parameter η by one observation from the $N(\eta, 1)$ distribution.¹ In general, however, σ is not known. We estimate σ by the usual least-squares estimator, $s = \sqrt{y'My/(n-k-1)}$. The expression $\hat{\theta}/s$ is then interpreted as the usual regression t -statistic. The WALS estimator based on

$$\lambda(\hat{\theta}, s^2) = \Lambda(\hat{\theta}/s) \quad (2)$$

generalises in a natural way the usual pretest estimator. The mean squared error of $\lambda(\hat{\theta}, s^2)\hat{\theta}$ is now given by

$$\text{MSE}(\lambda(\hat{\theta}, s^2)\hat{\theta}) = \sigma^2 \text{E}(\Lambda(\frac{\hat{\theta}}{s})\frac{\hat{\theta}}{\sigma} - \frac{\theta}{\sigma})^2 = \sigma^2 \text{E}(\Lambda(\frac{\hat{\eta}}{v}\sqrt{\nu})\hat{\eta} - \eta)^2, \quad (3)$$

where $v = \frac{s}{\sigma}\sqrt{\nu}$, $\nu = n - k$. In expression (3), the statistics v and $\hat{\eta}$ are independent and distributed as χ_ν and $N(\eta, 1)$ respectively. We see that the risk function depends on the unknown parameter σ , which is only a scale parameter. Therefore we can write the mean squared error as

$$\text{MSE}(\lambda(\hat{\theta}, s^2)\hat{\theta}) = \sigma^2 R(\eta, \Lambda),$$

where the standardized risk function R is defined as

$$R(\eta, \Lambda) = \text{E}_\eta \text{E}_\nu (\Lambda(x\sqrt{\nu}/v)x - \eta)^2, \quad (4)$$

with $x \sim N(\eta, 1)$, $v \sim \chi_\nu$, independent of x , and E_η denotes expectation with respect to $N(\eta, 1)$, and E_ν with respect to the χ_ν distribution. The function R can be interpreted as the mean squared error of the WALS estimator in the case where $\sigma = 1$. Now denote

$$R_v(\eta, \Lambda) = \text{E}_\eta (\Lambda(x\sqrt{\nu}/v)x - \eta)^2, \quad (5)$$

so that then $R(\eta, \Lambda) = \text{E}_\nu R_v(\eta, \Lambda)$. The function $R_v(\eta, \Lambda)$ can be interpreted as the risk function for the problem with known variance but with different Λ , specifically $\Lambda_v(x) = \Lambda(x\sqrt{\nu}/v)$.

¹In Magnus (2002) this is called *the* $N(\eta, 1)$ *problem*.

3 WALS estimation in auxiliary problem with unknown variance

We wish to investigate properties of the estimator of η , based on one bivariate observation (x, s) , where $x \sim N(\eta, 1)$, $s = w\sigma/\sqrt{\nu}$, $w \sim \chi_\nu$, independent on x . The estimator takes the form

$$t(\Lambda, x, s) = \Lambda(x/s)x. \quad (6)$$

Note, that the weight $\Lambda(x/s)$ depends not only on x but also on the independent statistic s . This additional randomness is caused by the necessity to estimate the nuisance parameter σ . We will assume that Λ satisfies the regularity conditions R1, i.e.

1. $0 \leq \Lambda(x) \leq 1$, for all x ,
2. $\Lambda(-x) = \Lambda(x)$, for all x ,
3. Λ is nondecreasing on $[0, \infty]$,
4. Λ is continuous except possibly on a set of measure zero.

These conditions allow us to interpret $\Lambda(x/s)x$ as a shrinkage estimator. There is an apparent parallel in investigating the properties of (6) for known and unknown σ . In particular, all notions of admissibility of the estimators can be reformulated straightforwardly by considering the new risk function,

$$R(\eta, \Lambda) = E(t(\Lambda, x, s) - \eta)^2,$$

but the actual investigation of admissibility and regret requires some care. Our main question is whether the conclusions made for the WALS estimator with known variance are still true (or almost true) in the case with unknown variance. In Magnus (2002) several important one-parametric classes of WALS estimators were considered. The *normal Bayes* estimator is defined as $t^{(1)} = x/(1 + c)$. The usual pretest estimator is defined by choosing Λ as

$$\Lambda_c^{(2)}(x) = \begin{cases} 0 & \text{if } |x| \leq c, \\ 1 & \text{if } |x| > c, \end{cases} \quad (7)$$

and the class of all pretest estimators that have $0 < c < \infty$ is denoted as $\mathcal{L}^{(2)}$. The Laplace estimator corresponds to the Λ -function of the form

$$\Lambda_c^{(3)}(x) = 1 - \frac{h(x)}{x}c, \quad (8)$$

where

$$h(x) = \frac{1 - e^{2cx}\psi(x)}{1 + e^{2cx}\psi(x)}, \quad \psi(x) = \frac{\Phi(-x - c)}{\Phi(x + c)},$$

and the class of all Laplace estimators with positive finite c is denoted as $\mathcal{L}^{(3)}$. The 'neutral' Laplace estimator corresponds to $c = 0.6931$. Finally the Burr estimator was defined as

$$\Lambda_c^{(4)}(x) = \begin{cases} 0 & \text{if } |x| \leq c, \\ 1 - \frac{c}{|x|} & \text{if } |x| > c. \end{cases} \quad (9)$$

The corresponding class of Burr estimators is denoted by $\mathcal{L}^{(4)}$.

The natural starting point is admissibility. The usual estimator is defined by $\Lambda = 1$. Since this Λ does not depend on s at all, we can expect that little is changed in the properties of the resulting estimator. Indeed, Theorem 3 shows that the usual estimator is unbiased, admissible and has constant risk equal to 1. Similar considerations (Theorem 4) show that the normal Bayes estimators are admissible for any $0 \leq c < \infty$. In the auxiliary problem with known σ , the pretest estimator is proved to be $\mathcal{L}^{(2)}$ -admissible. Theorem 7 shows that the pretest estimator for each value of c is $\mathcal{L}^{(2)}$ -admissible also when σ is unknown. A similar result holds for the Burr estimator (Theorem 8). This simply reflects the fact that in each class no estimator dominates an other. Moreover their risks are bounded and converge pointwise to 1 as c tends to infinity. For the Laplace estimator it is possible to prove a stronger property, namely that the Laplace estimator is admissible (not just $\mathcal{L}^{(3)}$ -admissible) even when σ is unknown (Theorem 9). We can also establish $\mathcal{L}^{(1)}$ -admissibility of the pretest estimator and $\mathcal{L}^{(2)}$ -admissibility of the Burr estimator.

It is well known (see e.g. Judge and Bock (1978)) that the risk function of the pretest estimator depends on ν . Figure 1 shows that this is indeed the case. In the figure we plot the risk functions of the ordinary 5% pretest estimator², together with the risk functions of the restricted ($\Lambda = 0$) and the unrestricted ($\Lambda = 1$) estimators. First of all note that as ν grows, the risk profiles converge to some limiting function, say $R_\infty(\theta)$. This function is nothing more than the risk function of the pretest estimator in the auxiliary problem with known σ . Indeed, the conditional risk function (5) satisfies

$$R_\nu(\theta, \Lambda) = R_\nu(\theta, \Lambda_c) = R_1(\theta, \Lambda \frac{c\nu}{\sqrt{\nu}}).$$

²Recall that for 5% pretest estimator parameter c is determined as $c = T_\nu^{-1}(0.975)$, where T_ν is c.d.f. of t -distribution with ν degrees of freedom.

Figure 1. Risk function of the ordinary 5% pretest estimator for various values of ν .

Since $v/\sqrt{\nu} \rightarrow 1$ a.s., it follows that $R(\eta, \Lambda) = E_{\nu} R_{\nu}(\eta, \Lambda) \rightarrow R_1(\theta, \Lambda_c)$, uniformly in θ . It is also quite obvious (and Figure 1 confirms this) that the risk functions of the restricted and the unrestricted estimators do not depend on ν . We know that $c = c_{\nu} = T_{\nu}^{-1}(0.975)$ grows as ν decreases. Therefore for small values of ν the 5% pretest estimator behaves more and more like the restricted estimator. This explains the fact that as ν decreases, risk profiles become smaller for small η and larger for large η . It appears, however, that risks depend very little on ν around the point $\eta = 1.3370$, where the graphs $R_{\infty}(\theta)$ and θ^2 intersect.³ Similarly, there is very little dependence on ν around the point $\eta = 1$. Hence in this important region the pretest estimator still remains worse than both the restricted and the unrestricted estimator. The maximum of the risk for the ordinary pretest estimator becomes larger when ν decreases. For $\nu = 5$ the maximum risk is 45% higher than for the case of known σ . Therefore for small samples the usual 5% pretest estimator seems to perform even worse than for the case of known σ . In contrast, the neutral Laplace estimator reveals considerable robustness. In Figure 2 we plot the risk functions of the neutral Laplace estimator for different values of ν .

³Actually, the dependence of R on ν does not disappear completely but only becomes very small.

Figure 2. Risk of the neutral Laplace estimator as function of η for various values of ν .

Figure 2 shows that the risk of the Laplace estimator depends remarkably little on ν .⁴ Comparing with the case of known σ , the risk for small ν is slightly higher when $|\eta| < 2$, but outside this region the situation is reversed. Of course, the limiting value for $\eta \rightarrow \infty$ is the same and equal to $1 + c^2$.

One of the well-accepted approaches in finding an optimal Λ is the minimax regret approach (Savage (1951), Chernoff and Moses (1959), Sawa and Hiro-matsu (1973)). An optimal Burr estimator was obtained in Magnus (2000) as the minimax regret solution within a specific subclass. Magnus also used regret to characterise optimality of the Laplace estimator. In order to apply this approach for our case we need to find the regret function. Regret is defined as the difference between risk of the estimator and minimal risk in a given class of estimators. Theorem 11 establishes that the minimal risk of the WALs estimator within all $R1$ -regular $\mathcal{L}^{(0)}$ Λ -s is the same as in the problem with known σ , that is,

$$\inf_{\lambda \in \mathcal{L}^{(0)}} R(\eta, \lambda) = \frac{\eta^2}{1 + \eta^2}.$$

⁴Note that in Figure 2 the scale on vertical axis is larger than in Figure 1. If we would not do this, the reader would not see *any* difference in risk profiles for different ν .

Therefore, the regret function for a particular WALs estimator is defined as

$$r_0(\eta, \Lambda) = R(\eta, \Lambda) - \frac{\eta^2}{1 + \eta^2}.$$

Now, let us see if and how the regret properties of the neutral Laplace estimator change in the problem with unknown σ .

ν	5	10	15	20	40	∞
min r_0	0.1183	0.1099	0.1069	0.1053	0.1030	0.1006
max r_0	0.5036	0.5075	0.5091	0.5099	0.5112	0.5127
min R	0.4725	0.4582	0.4536	0.4513	0.4479	0.4446
max R	1.4805	1.4805	1.4805	1.4805	1.4805	1.4805

Table 1. Extreme values for risk and regret of the neutral Laplace estimator.

In Table 1 we gathered the extreme values for risk and regret of the Laplace estimator for different values of ν . The last column represents the case of known σ . We see that minimum risk is about 6% higher for small ν , but that the maximum risk is the same. In contrast, maximum regret is only about 2% lower for small ν , but minimum regret is about 18% higher. (This is a direct consequence of the fact that the lower bound of risk of the WALs estimator is the same for all η .) Each row converges quickly to their limits when ν grows. This limiting value coincides of course with the value for the case of known σ .

Summarising we can say that the neutral Laplace estimator, developed for the case of known σ , performs exceptionally good. The difference in properties caused by estimation of σ is small. In terms of mean squared error this difference does not exceed 5% for the relevant range of ν . Therefore we recommend to use the neutral Laplace estimator in practice.

Nevertheless one potential resource for improvement still remains. It is possible to apply, for example, the optimal minimax regret Burr estimator in the new setup. Since the regret function is different for each ν , this optimal solution will take into account this dependence. *A priori* it is not clear, how much we can gain by applying this approach. Figure 3 classifies the situation.

In the Figure 3 we plotted risk profiles of the minimax regret Burr estimator for various values of ν . Comparing Figure 3 with Figure 2 we can only conclude that the potential gain is marginal. Hence, adapting WALs estimators via the minimax regret principle to each ν does not seem to be a productive idea. The Laplace solution still appears to be the best.

Figure 3. Risk of the optimal minimax regret Burr estimator as a function of η for various values of ν .

4 Relative efficiency of the Laplace estimator

In previous sections we investigated properties of several WALS estimators in the auxiliary problem. We have established several important properties concerning admissibility and risk of these estimators. Our general conclusion was that we can use the optimal solution obtained in Magnus (2002), because there is only little difference in risk and regret values obtained for known and unknown σ . However, our original problem concerned the estimation of a focus regressor in the partitioned regression problem. We now compare the performance of the Laplace and the pretest estimator in a real regression problem. For this purpose we consider the ratio of the mean squared errors of the ordinary pretest estimator and the neutral Laplace estimator. More precisely, let us consider the problem of estimating a general linear combination of the parameters β , say $\omega'\beta$, where ω is a known $k \times 1$ vector. The mean squared error of $\omega'b$, according to the equivalence theorem, is

$$\begin{aligned} \text{MSE}(\omega'b) &= \sigma^2(\omega'(X'X)^{-1}\omega + \omega'qq'\omega R(\eta)) \\ &= \sigma^2\omega'(X'X)^{-1}\omega(1 + q_0^2 R(\eta)), \end{aligned}$$

where

$$q_0^2 = \frac{(\omega'(X'X)^{-1}Xz)^2}{\omega'(X'X)^{-1}\omega \ z'Mz},$$

and $R(\eta)$ is a generic notation for the risk function of the WALS estimator in the auxiliary problem. Therefore the ratio of mean squared errors of the neutral Laplace and pretest estimators is

$$G = \frac{1 + q_0^2 R^P(\eta)}{1 + q_0^2 R^L(\eta)},$$

where $R^L(\eta)$ denotes the risk of the neutral Laplace estimator, and $R^P(\eta)$ the risk of the ordinary pretest estimator. Values of G larger than 1 correspond to the region where the Laplace estimator performs better than the ordinary pretest, and vice versa.

Figure 4. Relative efficiency of the neutral Laplace estimator compared with the usual 5% pretest estimator, $\nu = 20$.

Figure 4 illustrates the behavior of the gain profile G for different values of q_0^2 as a function of η . We set ν equal to 20. First, we see that for values of q_0^2 close or equal to 0 we have little or no gain compared with the usual pretest case. The Laplace estimator performs better in the interval $[0.46, 3.93]$. For very small

and very large values of η , the Laplace estimator is slightly worse than the usual pretest estimator. Actually, this superiority interval will be different for different ν . Table 2 shows the nature of this dependence.

ν	1	5	10	15	20	30	40	∞
x_*	0.7732	0.5622	0.4985	0.4743	0.4615	0.4483	0.4416	0.4207
x^*	28.4806	5.0971	4.2447	4.0204	3.9174	3.8200	3.7732	3.6405

Table 2. Superiority interval of Laplace estimator as function of ν .

We see from the Table 2 that the superiority interval shrinks and moves to the right as ν grows. Both lower and upper boundaries of the interval tend monotonically to their limiting values 0.4207 and 3.6405 respectively.⁵ Note again, that the Laplace estimator always performs better in the region of moderately large values of η , including the important neighborhood of point $\eta = 1$. Moreover, for small ν the superiority interval for the Laplace estimator is larger than for large ν .

5 Conclusion

In the current article we discussed issues connected with the practical application of the neutral Laplace weighted-average least-squares estimator, introduced in Magnus (2002), relaxing the assumption that the variance of innovation is known. We found that properties of the Laplace estimator are surprisingly similar whether or not σ is known. Important properties of the Laplace solution such as admissibility, bounded risk and small regret values still hold in the new setup. Moreover, a comparison of the Laplace and ordinary pretest estimators shows that there is a large interval where the Laplace estimator performs significantly better than the ordinary pretest estimator. The superiority interval is larger for small sample size (more exactly, for small values of ν), but remains rather large even asymptotically. On the base of these results we recommend the neutral Laplace solution for use in practical situations.

Applicability of the Laplace estimator is not restricted to the estimation case. Simple calculations (not reproduced here) show that a relation, similar to the Magnus and Durbin (1999) Equivalence Theorem, holds also for the mean squared error of the out-of-sample regression prediction, and therefore the same optimal λ function can be used for prediction. Applicability of the Laplace estimator is also not restricted to one auxiliary regressor. Our preliminary investigations show

⁵These limiting values coincide of course with the lower and upper bounds for the superiority interval in the case when σ is known.

that there are no difficulties in applying optimal WALS estimator in a regression problem with several auxiliary regressors, if these regressors are orthogonal in some sense. If not, then a non-degenerate linear transformation of the auxiliary regressors is necessary to achieve orthogonality.

6 Appeddix

This appendix contains a set of results about risk, admissibility and regret of various WALS estimators in the auxiliary problem with unknown σ . Admissibility is investigated by application of standard methods of decision theory (see e.g. Berger (1985)). First, let us clarify our notation. Our decision rule δ is a function of two arguments $\delta = \delta(x, v) = \Lambda(x\sqrt{v}/v)x$. The loss function is quadratic $L(\delta, \eta) = (\delta - \eta)^2$, and the risk function is $R(\eta, \delta) = \mathbb{E}_\eta \mathbb{E}_v (\delta(x, v) - \eta)^2$, where expectation on the right hand side is taken with respect to the product of two independent distributions: $x \sim N(\theta, \sigma^2)$ and $v \sim \chi_\nu$. We need the following auxiliary result that establishes continuity of the risk function.

Lemma 1 The risk function

$$R(\eta, \delta) = \mathbb{E}_\eta \mathbb{E}_v (\Lambda(x\sqrt{v}/v)x - \eta)^2$$

is continuous in the region $C = \{\eta : |\eta| < 1\}$.

Proof: By definition,

$$\begin{aligned} R(\eta, \delta) &= \int_{-\infty}^{\infty} \int_0^{\infty} (\Lambda(x\sqrt{v}/v)x - \eta)^2 \phi_{\eta,1}(x) f_\nu(v) dx dv \\ &= \int_{-\infty}^{\infty} \int_0^{\infty} (\Lambda(\frac{(y+\eta)\sqrt{v}}{v})(y+\eta) - \eta)^2 \phi_{0,1}(y) f_\nu(v) dy dv. \end{aligned}$$

Therefore it is enough to prove that

$$\int_{-\infty}^{\infty} \int_0^{\infty} \Lambda^n(\frac{(y+\eta)\sqrt{v}}{v}) y^m \phi_{0,1}(y) f_\nu(v) dy dv, \quad (10)$$

for $n = 0, 1, 2$ and $m = 0, \dots, n$ are continuous in C . This is the classical case of continuity of Lebegue integral with respect to a parameter (see e.g. Kolmogorov and Fomin (1957) or Apostol (1974)). The only nontrivial thing to show is the existence of an integrable majorant function, that is $|y|^m \phi_{0,1}(y) f_\nu(v)$. This function is integrable because the normal distribution has moments of all orders, and it is a majorant for the integrand in (10) because $|\Lambda(u)| \leq 1$, by regularity condition R1. ■

Theorem 2 (Brown-Farrell) The decision rule δ is admissible if there exists a sequence of finite nonnegative measures $\{G_n\}$ such that $G_n(C) \geq 1$ and

$$r(G_n, \delta) - r(G_n, \delta_{G_n}) \rightarrow 0, \text{ as } n \rightarrow \infty,$$

where δ_{G_n} is Bayes rule with respect to G_n , and $C = \{\eta : |\eta| < 1\}$.

Proof: Suppose the conditions of the theorem are satisfied but δ is inadmissible. Then we can find another estimator, say δ' , whose risk doesn't uniformly exceed the risk of δ , i.e. $R(\eta, \delta') \leq R(\eta, \delta)$. At the same time δ' must differ from δ on the set of nonzero measure, i.e.

$$\int (\delta'(x, v) - \delta(x, v))^2 dx dv > 0.$$

Define $\delta''(x, v) = \frac{1}{2}(\delta(x, v) - \delta'(x, v))$. Then,

$$\int (\delta'(x, v) - \delta(x, v))^2 \phi_{\eta,1}(x) f_\nu(v) dx ds = E(\delta' - \delta)^2 > 0,$$

because $\phi_{\eta,1}(x) f_\nu(v) > 0$. However,

$$\begin{aligned} E(\delta' - \delta)^2 &= E((\eta - \delta) - (\eta - \delta'))^2 \\ &= E(\eta - \delta)^2 + E(\eta - \delta')^2 - 2E(\eta - \delta)(\eta - \delta') > 0, \end{aligned}$$

and therefore

$$2E(\eta - \delta)(\eta - \delta') < E(\eta - \delta)^2 + E(\eta - \delta')^2. \quad (11)$$

By inequality (11) we see that

$$\begin{aligned} R(\eta, \delta'') &= E(\eta - \delta'')^2 = \frac{1}{4}\{E(\eta - \delta)^2 + E(\eta - \delta')^2 + 2E(\eta - \delta)(\eta - \delta')\} \\ &< \frac{1}{2}(E(\eta - \delta)^2 + E(\eta - \delta')^2) = \frac{1}{2}(R(\eta, \delta) + R(\eta, \delta')) \leq R(\eta, \delta). \end{aligned}$$

Since $R(\eta, \delta)$ and $R(\eta, \delta'')$ are both continuous (by Lemma 1), there exists an $\varepsilon > 0$ such that $R(\eta, \delta'') < R(\eta, \delta) - \varepsilon$ for $|\eta| < 1$. Hence,

$$\int_C (R(\eta, \delta) - R(\eta, \delta'')) G(d\eta) \geq \int_C \varepsilon G(d\eta) \geq \varepsilon, \quad (12)$$

and also

$$\varepsilon \leq r(G_n, \delta) - r(G_n, \delta'') \leq r(G_n, \delta) - r(G_n, \delta_{G_n}). \quad (13)$$

Inequality (13) contradicts (11) and thus proves the theorem. ■

Theorem 3 The usual estimator for η , $t(x, \Lambda, s) = x$ is

1. unbiased,
2. has constant risk equal 1,
3. admissible.

Proof: Clauses 1 and 2 are straightforward. The proof of 3 is based on Theorem 5.6.1 from Brown (1971) and in fact is nothing more than an extension of Stein's sufficient condition of admissibility (see Stein (1955)). This condition is formulated by us as Theorem 2. To apply Theorem 2 we need to check that there exists a sequence of nonnegative measures (generalized priors) G_i such that

- a) $G_n(C) \geq 1$.
- b) $r(G_n, \delta) - r(G_n, \delta_{G_n}) \rightarrow 0$, as $n \rightarrow \infty$.

Now let g_n be the rescaled normal density

$$\pi_n(\theta) = \exp\left(-\frac{\theta^2}{2n}\right).$$

Then claim a) follows because $G_n(C) = \sqrt{2\pi n} \operatorname{Erf}\left(\frac{1}{\sqrt{2n}}\right)$, where $\operatorname{Erf}\left(\frac{y}{\sqrt{2}}\right) = 2\Phi(y) - 1$, $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution, and because $G_{n+1}(C) \geq G_n(C)$ and $G_1(C) \approx 1.71$. Then straightforward Bayes calculations give us

$$r(G_n, \delta) = \sqrt{2\pi}\sqrt{n}, \quad r(G_n, \delta_{G_n}) = \sqrt{2\pi} \frac{\sqrt{n^3}}{n+1},$$

that satisfy both a) and b). Therefore the usual estimator is admissible. ■

Theorem 4 The estimator $t^{(1)} = x/(1+c)$ is admissible for any $c > 0$.

Proof: The modified risk function (4) in this case boils down to the risk function for known σ and therefore the proof from Magnus (2000, Theorem A.2), can be applied without changes. That is: for $0 < c < \infty$ the normal Bayes estimators are Bayes with respect to prior $\pi(\eta) \sim N(0, 1/c)$ and therefore admissible. For $c = \infty$, we have $\Lambda = 0$ and $R(0, 0) = 0$, which implies admissibility. ■

Note, that $\operatorname{BIAS}(\eta, \Lambda) = \mathbf{E}_\eta \mathbf{E}_\nu(\Lambda(x\sqrt{\nu}/w)x - \eta)$ is an antisymmetric function, and that the risk $R(\eta, \Lambda)$ is a symmetric function with respect to η . This allows us to consider only positive η when investigating properties of the estimator.

Theorem 5 Suppose Λ is $R1$ -regular function. Then $R(\eta, \Lambda)$ is bounded if there is a K , $0 \leq K \leq \infty$ such that $|\zeta(x)| \leq K$ for all x , where $\zeta(x) = (1 - \Lambda(x))x$.

Proof: Change of variable $x = u + \eta$, $u \sim N(0, 1)$ leads to

$$\begin{aligned} R(\eta, \Lambda) &= \mathbf{E}_0 \mathbf{E}_\nu (u - v\varepsilon(x\sqrt{\nu}/v))^2 \\ &\leq 2\mathbf{E}_0 \mathbf{E}_\nu (u^2 + v^2\varepsilon(x\sqrt{\nu}/v)^2) \leq 2(1 + K^2 A), \end{aligned}$$

where A is a second moment of the χ -distribution, that is a finite number. ■

The following result is useful when investigating properties of the WALS estimator in the auxiliary problem with unknown σ .

Theorem 6 Consider the estimation problem with known σ and suppose that the one-parametric class $\mathcal{L}^{(a)} = \{\Lambda_c : 0 \leq c \leq \infty\}$ consists of $R1$ -regular estimators that are $\mathcal{L}^{(a)}$ -admissible. Assume also that

$$\Lambda_c\left(\frac{x}{s}\right) = \Lambda_{cs}(x), \quad (14)$$

for any $s \geq 0$. Then the estimator $\Lambda_c(\frac{x}{s})x$ with risk function $R(\eta, \Lambda_c) = \mathbb{E}_s R_s(\eta, \Lambda_c(\frac{\cdot}{s}))$ is $\mathcal{L}^{(a)}$ -admissible.

Proof: Suppose that the opposite is true and that the estimator $\Lambda_c(\frac{x}{s})x$ is not admissible for some $c \geq 0$. Then we can find c_o such that $R(\eta, \Lambda_{c_o}) \leq R(\eta, \Lambda_c)$, and such that for at least one η , say η_o , this inequality is strict. Then,

$$\mathbb{E}_s(R_s(\eta_o, \Lambda_{c_o}, s) - R_s(\eta_o, \Lambda_c, s)) < 0,$$

so there is an $s > 0$ such that $R_s(\eta_o, \Lambda_{c_o}, s) < R_s(\eta_o, \Lambda_c, s)$. Therefore $R(\eta_o, \Lambda_{c_o}) < R(\eta_o, \Lambda_c)$ by property (14). Hence inadmissibility follows for $\Lambda_{c'}(x)x$, $c' = cs$. This contradicts the definition of $\mathcal{L}^{(a)}$. ■

The following two propositions are direct consequences of Theorem 6.

Proposition 7 The pretest estimator (7) in the auxiliary problem with unknown variance is $\mathcal{L}^{(1)}$ -admissible.

Proposition 8 The Burr estimator (9) in the auxiliary problem with unknown variance is $\mathcal{L}^{(2)}$ -admissible.

Theorem 9 The Laplace estimator (8) in the auxiliary problem with unknown variance is admissible.

Proof: The main idea of proof is close to the proof of Theorem 6. Suppose that the Laplace estimator is inadmissible. Then there exists an estimator $\delta(x, s)$ such that $R(\eta, \delta) \leq R(\eta, \Lambda^{(4)})$ for any η and there is at least one point η_o where this inequality is strict. Then,

$$\mathbb{E}_s(R_s(\eta_o, \delta) - R_s(\eta_o, \Lambda^{(4)})) < 0,$$

and therefore $R_s(\eta, \delta(\cdot, s)) - R_s(\eta, \Lambda^{(4)}(\frac{\cdot}{s})) < 0$ for some $s > 0$. This implies inadmissibility for the estimator

$$\Lambda_{c,s}^{(4)}(x) = \Lambda^{(4)}\left(\frac{x}{s}\right) \quad (15)$$

in the problem with known σ . However, the estimator (15) is a Bayes estimator by

Lemma 10 and therefore admissible. ■

Lemma 10 The estimator (15) is a Bayes estimator with prior $\pi(\eta, c_1) = \frac{c_1}{2} e^{-c_1|\eta|}$, $c_1 = c/s$ and $x|\eta \sim N(\eta, s^2)$.

Proof: Performing Bayes calculations we get $t(x; c_1, s) = x - h_1(c_1, x)c_1s^2$, where

$$h_1(c_1, x) = \frac{1 - e^{2c_1x}d_1(c_1, x)}{1 + e^{2c_1x}d_1(c_1, x)}, \quad d_1(c_1, x) = \frac{\Phi\left(\frac{-x - c_1s^2}{s}\right)}{\Phi\left(\frac{x + c_1s^2}{s}\right)}.$$

But

$$t(x; c_1, s) = x\left(1 - \frac{h_1(c_1, x)}{x/s}c_1s\right) = x\Lambda_{c_1s}^{(4)}(x/s) = x\Lambda_c^{(4)}(x/s),$$

and the result follows. ■

Theorem 11

$$\inf_{\lambda \in \mathcal{L}^{(0)}} R(\eta, \lambda) = \frac{\eta^2}{1 + \eta^2},$$

where $\mathcal{L}^{(0)}$ includes all $R1$ -regular Λ -s.

Proof: We need to check

- a) for any $\Lambda \in \mathcal{L}^{(0)}$ the inequality $R(\eta, \Lambda) \geq \frac{\eta^2}{1 + \eta^2}$ holds.
- b) There exist at least one $\Lambda' \in \mathcal{L}^{(0)}$ such that $R(\eta, \Lambda') = \frac{\eta^2}{1 + \eta^2}$.

To prove a) consider the definition of $R(\eta, \Lambda)$, that is

$$R(\eta, \Lambda) = \mathbb{E}_\eta \mathbb{E}_\nu (\Lambda(x\sqrt{\nu}/v)x - \eta)^2 = \mathbb{E}_\nu R_v(\eta, \Lambda).$$

Due to (5), $R_v(\eta, \Lambda)$ can be interpreted as the risk function for the problem with known σ but with modified Λ , i.e. $\Lambda_v = \Lambda(x\sqrt{\nu}/v)$. However, *any* $R_v(\eta, \Lambda)$ must satisfy $R_v(\eta, \Lambda) \geq \frac{\eta^2}{1 + \eta^2}$ by Theorem A.7 in Magnus(2000). Therefore,

$$R(\eta, \Lambda) = \mathbb{E}_\nu R_v(\eta, \Lambda) \geq \mathbb{E}_\nu \frac{\eta^2}{1 + \eta^2} = \frac{\eta^2}{1 + \eta^2}.$$

To prove b) just consider $\Lambda = \frac{1}{1 + 1/\eta^2}$. ■

References

- Bancroft, T.A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Annals of Mathematical Statistics* 15, 190–204.
- Berkson, J. (1942). Tests of significance considered as evidence. *Journal of American Statistical Association* 37, 325–335.
- Blyth, C.R. (1951). On minimax statistical decision procedures and their admissibility. *Annals of Mathematical Statistics* 22, 22–42.
- Brown, L.D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems. *Annals of Mathematical Statistics* 42(3), 855–903.
- Brook, R.J. (1976). On the use of a regret function to set significance points in prior tests of estimation. *Journal of American Statistical Association* 71, 126–131.
- Berger, J.O. (1976). Admissibility results for generalised bayes estimators of coordinates of a location vector. *The Annals of Statistics* 4(2), 334–356.
- Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.
- Chatfield, C. (1995). Model uncertainty, data mining and statistical inference (with discussion). *Journal of the Royal Statistical Society, Series A*, 158, 419–466.
- Droge B. and T. Georg (1995). On selecting the smoothing parameter of least squares regression estimates using the minimax regret approach. *Statistics and Decisions* 13, 1–20.
- Farrell, R.H. (1968). On a necessary and sufficient condition for admissibility of estimators when strictly convex loss is used. *The Annals of Mathematical Statistics* 38(1), 23–28.
- Giles, J.A. and D.E.A. Giles (1993). Pre-test estimation and testing in econometrics: recent developments. *Journal of Economic Surveys* 7, 145–197.
- Judge, G.G. and M.E. Bock (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. Amsterdam: North-Holland.
- Judge, G.G. and M.E. Bock (1983). Biased estimation. In Z. Griliches, & M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. I, Chapter 10, Amsterdam: North-Holland.

- Magnus, J.R. and J. Durbin (1999). Estimation of regression coefficient of interest when other regression coefficients are of no interest. *Econometrica* 67(3), 639–643.
- Magnus, J.R. (1999). The traditional pretest estimator. *Theory of Probability and Its Applications* 44, 293–308.
- Magnus, J.R. (2002). Estimation of the mean of a univariate normal distribution with known variance. *The Econometrics Journal* 5, 225–236.
- Sawa, T. and T. Hiromatsu (1973). Minimax regret significance points for a preliminary test in regression analysis. *Econometrica* 41, 1093–1101.
- Toyoda, T. and T. Wallace (1976). Optimal critical values for pre-testing in regression. *Econometrica* 44, 365–375.