Department of Economics
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND

ISSN 1171-0705

THE SAMPLING PERFORMANCE OF INEQUALITY RESTRICTED AND PRE-TEST ESTIMATORS IN A MIS-SPECIFIED LINEAR MODEL

Alan T. K. Wan

Discussion Paper

No. 9207

Department of Economics, University of Canterbury
Christchurch, New Zealand

Discussion Paper No. 9207

September 1992

THE SAMPLING PERFORMANCE OF INEQUALITY RESTRICTED AND PRE-TEST ESTIMATORS IN A MIS-SPECIFIED LINEAR MODEL

Alan T. K. Wan

THE SAMPLING PERFORMANCE OF INEQUALITY RESTRICTED AND PRE-TEST ESTIMATORS IN A MIS-SPECIFIED LINEAR MODEL*

Alan T.K. Wan

Department of Economics

University of Canterbury

July, 1992

Abstract

We evaluate the sampling performance of the inequality restricted and pre-test estimators for the prediction vector and the scale parameter when the underlying model involves a single inequality constraint on the coefficents and a design matrix from which relevant regressors are possibly omitted. Within the context of this analysis, we also derive and compute the optimal critical values for the preliminary test using both the minimum average relative risk and mini-max regret principles. It is found that most of our results concur qualitatively with those that one obtains when the restriction is assumed to hold as a strict equality.

*This work forms part of the author's Ph.D research. The author would like to thank Professor David Giles and Dr. Judith Giles for their guidance and many useful suggestions. Helpful comments from Kevin Albertson and John Small are also gratefully acknowledged. An earlier version of this paper was presented at the Australasian Meeting of the Econometric Society in Melbourne, July 6-8, 1992.

Contact address : Department of Economics, University of Canterbury, Private Bag, Christchurch, NEW ZEALAND.

FAX : +64-3-3642635 ; **Phone** : +64-3-3642033
Internet : a.wan@csc.canterbury.ac.nz

1. INTRODUCTION

In applied econometric analysis, inequality restrictions are often imposed on the regression parameters as a result of the incompleteness of prior knowledge on the parameters of interest. Within the context of the linear regression model, the properties of the resulting inequality restricted estimator have received considerable attention in the literature in recent years. It has been shown, for example, in the case of simple regression, that the inequality restricted estimator has a truncated normal distribution (Zellner (1961)). In terms of the sampling performance of this estimator, it has been demonstrated, in a variety of circumstances, that the inequality restricted estimator can be better than the unrestricted estimator, if there is only one inequality constraint, the random disturbances in the model are normally distributed, and the prior information is either true or close enough to being true ((Lovell and Prescott (1970), Judge *et al.* (1980), Judge and Yancey (1981, 1986) and Thomson and Schmidt (1982)). When the *a priori* information involves more than one inequality constraint, Thomson (1982) shows that the performance of the inequality restricted estimator depends not only on the accuracy of the constraints involved, but also on the degree of correlation between the constrained estimates.[1]

Recently, attention has also been paid to the sampling properties of the pre-test estimator that chooses between the unrestricted and the inequality restricted estimators, based on the outcome of a test for the validity of the inequality restrictions. Assuming a known disturbance variance and a single linear inequality constraint, Judge and Yancey (1986) show that this inequality pre-test estimator is biased, and that, in terms of sampling performance, there is no region in the parameter space in which it is the best, being ranked between the inequality restricted and unrestricted estimators when the prior information is valid or nearly so. A related problem of estimating a normal mean subject to an inequality restriction is considered by Hasegawa (1989). Finally, Yancey *et al.* (1989) examine the risk characteristics of two multivariate inequality pre-test estimators that result from different inequality test structures, one with an equality null and the other with an inequality null. They find that although neither inequality pre-test estimator is uniformly superior, the one which corresponds to the equality null is superior only near the region of the parameter space in which the restrictions hold as strict equalities.

The results given in the literature show that the relative performance of various estimators do rely, to a large extent and among other things, upon the accuracy of the inequality constraints involved. Then the question naturally arises as to whether these comparisons are still valid if the underlying model is already mis-specified in some ways. This is of interest because econometricians in practice invariably work with mis-specified models due to ignorance, lack of data or the inability of economic theory to define the correct specification. However, it is interesting to note that with the exception of Ohtani (1991b), who considers inequality restricted estimation in a proxy variable model, the discussion on inequality restricted and pre-test estimation to date is based on the premise that the underlying data generating process is properly specified.[2]

Moreover, the literature on inequality restricted and pre-test estimation concentrates overwhelmingly on the estimation of the regression coefficient vector. Although the application of the linear regression model typically also involves the estimation of the disturbance variance, σ^2, the literature is totally silent on the properties of the estimators for σ^2 that take into account the *a priori* inequality restrictions imposed on the regression coefficients.[3]

As for other pre-test problems that have been investigated in the literature, the significance level of the pre-test also has an important bearing on the properties of the inequality pre-test estimator. The question of determining the optimum pre-test size for inequality restrictions is unresolved. A number of studies, however, consider the question of choosing optimal levels of significance for preliminary tests in other pre-test contexts, and various optimality criteria have been proposed. For instance, in the case of pre-testing of linear restrictions, Sawa and Hiromatsu (1973), Brook (1976), Toyoda and Wallace (1976) and Brook and Fletcher (1981) have obtained optimal critical values according to the criteria of mini-max regret and minimizing average relative risk. These criteria are also used to determine the optimal significance level for a pre-test in pooling variance (Toyoda and Wallace (1975) and Ohtani and Toyoda (1978)). The particular size chosen depends, as we would expect, on the pre-test problem being investigated and also on the adopted optimality criterion.

In this paper we add to the literature on the properties of inequality restricted and pre-test estimators by considering the problem of applying these estimators in a model which is mis-specified through the exclusion of relevant regressors. To keep our results tractable, we assume that the prior information is in the form of a single linear inequality constraint imposed on the regression coefficients. In the context of this analysis, which nests a properly specified model as a special case, we derive and evaluate the risk functions of the estimators for both the prediction vector and the error variance. Using the two commonly adopted criteria of mini-max regret and minimum average relative risk, we also address the issue of optimal critical values of pre-test for an inequality restriction.

The remainder of this paper is organised as follows: Section 2 presents the assumptions underlying the analysis and introduces the inequality restricted and pre-test estimators. In section 3, we derive and numerically evaluate the risks of various estimators for the prediction vector. The question of choosing an optimal critical value for the pre-test when estimating the prediction vector is addressed in section 4. In section 5, we derive, evaluate and compare the risk functions of several estimators for the scale parameter. Section 6 concludes the paper.

2. MODEL FRAMEWORK AND THE ESTIMATORS

Consider the data generating process :

$$y = X\beta + Z\eta + \varepsilon \quad ; \quad \varepsilon \sim N(0,\sigma^2 I) \tag{1}$$

where y and ε are $n \times 1$ vectors; X and Z are non-stochastic matrices of full column rank and are $n \times k$ and $n \times p$ respectively; β and η are unknown coefficient vectors and are $k \times 1$ and $p \times 1$ respectively.

Assume, however, that the model is incorrectly specified as:

$$y = X\beta + \mu \tag{2}$$

So, $\mu = Z\eta + \varepsilon$, and $\mu \sim N(Z\eta,\sigma^2 I)$, but it is assumed by the researcher that $\mu \sim N(0,\sigma^2 I)$. In addition to the sample information, there exists uncertain prior information about the coefficient vector β, in the form of a single linear inequality hypothesis:

$$C'\beta \geq r, \tag{3}$$

where C' is a $1 \times k$ known vector and r is a known scalar.

It is well known that the unrestricted estimators of β and σ^2, which utilize only sample information, are $\tilde{\beta} = S^{-1}X'y$, where $S = X'X$, and $\tilde{\sigma}^2 = (y-X\tilde{\beta})'(y-X\tilde{\beta})/(n+\delta)$ respectively. The equality restricted estimator of β, which uses both the sample information and the exact restriction $C'\beta = r$, is $\beta^* = \tilde{\beta} - S^{-1}C(C'S^{-1}C)^{-1}(C'\tilde{\beta}-r)$, and the corresponding estimator for σ^2 is $\sigma^{*2} = (y-X\beta^*)'(y-X\beta^*)/(n+\gamma)$. The least squares (LS) estimator for σ^2 corresponds to $\delta = -k$ and $\gamma = -k+1$, the maximum likelihood (ML) estimator corresponds to $\delta = \gamma = 0$, and the minimum mean square error (MM) estimator corresponds to $\delta = -k+2$ and $\gamma = -k+3$. Throughout this paper, the subscripts LS, ML and MM are used to characterize the estimators corresponding to these three components. If the researcher combines sample information with the inequality restriction (3) in estimating model (1), then either the unrestricted estimator does not violate the inequality constraint (3) and is chosen as the estimator for the model, or the unrestricted estimator violates (3) and the exact restriction $C'\beta = r$ is imposed on the parameters. The equality restricted estimator is used in this case.

Following Judge and Yancey (1986), (1), (2) and (3) can be reparameterized as :

$$y = H\theta + B\pi + \varepsilon, \tag{4}$$

$$y = H\theta + \mu, \tag{5}$$

and
$$\theta_1 \geq r_0 \tag{6}$$

respectively, where $H = XS^{-1/2}Q'$; $B = ZT^{-1/2}V'$; $\theta = QS^{1/2}\beta$; $\pi = VT^{1/2}\eta$; $T = Z'Z$; $r_0 = r/h_1$; h_1 is the first element of $h' = C'S^{-1/2}Q'$ and is assumed to be positive without loss of generality; θ_1 is the first element of θ; $V'V = I_p$, and Q is an orthogonal matrix such that

$$QS^{-1/2}C(C'S^{-1}C)^{-1}C'S^{-1/2}Q' = \begin{pmatrix} 1 & 0' \\ 0 & 0 \end{pmatrix}. \tag{7}$$

Now, the unrestricted and equality restricted estimators for θ are :

$$\tilde{\theta} = H'y \text{ and } \theta^* = \begin{pmatrix} r_0 \\ \tilde{\theta}_{(k-1)} \end{pmatrix} \text{ respectively,}$$

where $\tilde{\theta}_{(k-1)} = \left(0, I_{(k-1)}\right)\tilde{\theta}$. Similarly, the unrestricted and equality restricted estimators for σ^2 may be expressed as $\tilde{\sigma}^2 = (y-H\tilde{\theta})'(y-H\tilde{\theta})/(n+\delta)$ and $\sigma^{*2} = (y-H\theta^*)'(y-H\theta^*)/(n+\gamma)$ respectively.

According to the two-step procedure described earlier, the inequality restricted (IR) estimator for θ is

$$\theta^{**} = \begin{cases} \tilde{\theta} & \text{if } \tilde{\theta}_1 \geq r_0 \\ \theta^* & \text{if } \tilde{\theta}_1 < r_0 \end{cases} = I_{(-\infty, r_0)}(\tilde{\theta}_1)\theta^* + I_{[r_0, \infty)}(\tilde{\theta}_1)\tilde{\theta}, \tag{8}$$

and the corresponding estimator for σ^2 is

$$\sigma^{**2} = \begin{cases} \tilde{\sigma}^2 & \text{if } \tilde{\theta}_1 \geq r_0 \\ \sigma^{*2} & \text{if } \tilde{\theta}_1 < r_0 \end{cases} = I_{(-\infty, r_0)}(\tilde{\theta}_1)\sigma^{*2} + I_{[r_0, \infty)}(\tilde{\theta}_1)\tilde{\sigma}^2. \tag{9}$$

where $I_{(.)}(u)$ is an indicator function which takes the value unity if u falls in the subscripted interval and 0 otherwise. If we let $\tau = r_0 - \theta_1$ be the slack variable associated with (6) and $\beta^{**} = S^{-1/2}Q'\theta^{**}$ be the IR estimator for the coefficient vector, β, then (8) and (9) can be transformed to

$$\beta^{**} = \tilde{\beta} - S^{-1/2}Q' \begin{bmatrix} I_{(-\infty, \tau\sigma^{-1})}(u_1)(\sigma u_1 - \tau) \\ 0_{(k-1)} \end{bmatrix} \tag{10}$$

and

$$\sigma^{**2} = \tilde{\sigma}^2 + I_{(-\infty, \tau\sigma^{-1})}(u_1)\left[\left((\sigma u_1 - \tau)^2 - \tilde{\sigma}^2(\gamma - \delta)\right)/(n+\gamma)\right] \tag{11}$$

respectively, where $u_1 = (\tilde{\theta}_1 - \theta_1)/\sigma$ is a normal random variable with mean ξ/σ and variance 1, where $\xi = (H'B'\pi)_1$ is the first element in the vector $H'B'\pi$.

Typically the reseacher is uncertain of the validity of the inequality constraint and so may test for the inequality restriction (6). The test structure is given by

$$H_0\colon \theta_1 \geq r_0 \quad vs \quad H_1\colon \theta_1 < r_0,$$

and the usual test statistic is

$$t'' = \sqrt{v}\,(\tilde{\theta}_1 - r_0)\tilde{\sigma}^{-1}/\sqrt{n+\delta}\ .$$

t'' has a doubly non-central t distribution with v degrees of freedom and non-centrality parameters $\lambda_1^2 = (\tau-\xi)^2/2\sigma^2$ and $\lambda_2 = \pi' B'(I-HH')B\pi/2\sigma^2$. We write $t'' \sim t_{(v;\lambda_1,\lambda_2)}$. However, without realising the specification error in the model, the applied researcher believes t'' to have a central t distribution when θ=r and applies a t-test to test the null hypothesis. Hence, the decision rule is to reject the null if $t'' < c$, where c is the size - α critical value for the central t variate with v degrees of freedom. If the null is rejected, then the unrestricted estimator is chosen, otherwise the IR estimator is used in the estimation process. Accordingly, the inequality pre-test (IPT) estimators for θ and σ^2 are :

$$\hat{\theta} = \begin{cases} \tilde{\theta} & \text{if } t'' < c \\ \theta^{**} & \text{if } t'' \geq c \end{cases} = I_{(-\infty,c)}(t'')\tilde{\theta} + I_{[c,\infty)}(t'')\theta^{**} \tag{12}$$

and

$$\hat{\sigma}^2 = \begin{cases} \tilde{\sigma}^2 & \text{if } t'' < c \\ \sigma^{**2} & \text{if } t'' \geq c \end{cases} = I_{(-\infty,c)}(t'')\tilde{\sigma}^2 + I_{[c,\infty)}(t'')\sigma^{**2} \tag{13}$$

respectively. Again, if we define $\tau = r_0-\theta_1$ and let $\hat{\beta} = S^{-1/2}Q'\hat{\theta}$ be the IPT estimator for the coefficient vector, β, then (12) and (13) can be transformed to

$$\hat{\beta} = \beta^{**} + S^{-1/2}Q' \begin{bmatrix} I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)(\sigma u_1-\tau) \\ 0 \end{bmatrix} \tag{14}$$

and

$$\hat{\sigma}^2 = \sigma^{**2} - I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)\left[\left((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\right)/(n+\gamma)\right] \tag{15}$$

respectively,[4] where $c' = c\sqrt{(n+\delta)/v}$. The risk functions corresponding to these estimators are derived and analysed in the following sections.

3. THE RISK FUNCTIONS OF THE ESTIMATORS OF THE PREDICTION VECTOR

If b is any estimator of β in model (1), then the risk function of the prediction vector under squared error loss is defined as:

$$\rho(Xb,E(y)) = E\Big[(Xb - E(y))'(Xb - E(y))\Big]/\sigma^2,$$

From Mittelhammer (1984),

$$\rho(X\tilde{\beta},E(y)) = k + 2\lambda_2. \tag{16}$$

$$\rho(X\beta^*,E(y)) = k + 2\lambda_2 + 2\lambda_1^2 - 1. \tag{17}$$

Now, from (10) and (14), the risk functions of $X\beta^{**}$ and $X\hat{\beta}$ may be expressed as:

$$\rho(X\beta^{**},E(y)) = \rho(X\tilde{\beta},E(y)) + E\Big[I_{(-\infty,\tau\sigma^{-1})}(u_1)(\sigma u_1-\tau)(2\xi-\tau-\sigma u_1)\Big]/\sigma^2 \tag{18}$$

$$\rho(X\hat{\beta},E(y)) = \rho(X\tilde{\beta},E(y)) - E\Big[\Big(I_{(-\infty,\tau\sigma^{-1})}(u_1) - I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)\Big)$$
$$(\sigma^2u_1^2 - \tau^2 - 2\xi(\sigma u_1-\tau))\Big]/\sigma^2 \tag{19}$$

In order to evaluate these risk functions, we need supporting Lemmas 1 and 2 which are given and proved in Appendix A. Using these Lemmas, we establish the following Theorem:

THEOREM 1

If $\lambda_1 \leq 0$, then

$$\rho(X\beta^{**},E(y)) = k + 2\lambda_2 - P_3/2 + \lambda_1^2P_1, \tag{20}$$

$$\rho(X\hat{\beta},E(y)) = k + 2\lambda_2 + (E_{3,v} - P_3)/2 - \lambda_1^2(E_{1,v} - P_1) . \tag{21}$$

If $\lambda_1 > 0$, then

$$\rho(X\beta^{**},E(y)) = k + 2\lambda_2-1 + P_3/2 + 2\lambda_1^2 - \lambda_1^2P_1, \tag{22}$$

$$\rho(X\hat{\beta},E(y)) = k + 2\lambda_2 -1 + 2\lambda_1^2 + (E_{3,v} + P_3)/2 - \lambda_1^2(E_{1,v} + P_1)$$
$$- 2\lambda_1^2G_{1,v} + G_{3,v} , \tag{23}$$

where

$$P_1 = P(\chi_1^2 \geq 2\lambda_1^2),$$

$$E_{i,j} = e^{-\lambda_2}\sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!2^{v/2+t}\Gamma(\frac{v}{2}+t)} \int_0^{\infty} P(\chi_i^2 \geq (cq_j/\sqrt{v}+\sqrt{2}\lambda_1)^2) q_j^{2^{v/2+t-1}} e^{-q_j^2/2} dq_j^2,$$

$$G_{i,j} = e^{-\lambda_2}\sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!2^{v/2+t}\Gamma(\frac{v}{2}+t)} \int_0^{2v\lambda_1^2/c^2} P(\chi_i^2 < (cq_j/\sqrt{v}+\sqrt{2}\lambda_1)^2) q_j^{2^{v/2+t-1}} e^{-q_j^2/2} dq_j^2,$$

$i = 1, 3$ and q_j^2 is a non-central chi-square random variable with j degrees of freedom and non-centrality parameter λ_2, *i.e.* $q_j^2 \sim \chi'^2_{(j,\lambda_2)}$.

Proof: see Appendix A.

When there is no mis-specification in the model, $\lambda_2 = 0$ and (20) - (23) collapse to the expressions given by Hasegawa (1989).[5]

These expressions are difficult to evaluate analytically. Accordingly, numerical evaluations of the risks have been carried out for n = 10, 30, 50; k = 2, 5; α = 0, 0.01, 0.05, 0.10, 0.25, 0.40; λ_1= [-10,10] and various values of λ_2. The NAG (1991) subroutine D01AJF and subroutines from Press *et al.* (1986) are used to evaluate the integrals $E_{i,j}$ and $G_{i,j}$. Some representative risk diagrams are given in Appendix B. Notice from the definition of λ_1 that for a given level of the constraint specification error, τ, λ_1 changes with the magnitude of ξ, the model specification error. The case of $\xi = 0$ is represented in Figure 1, which illustrates the results given by Judge and Yancey (1981, 1986)[6]. Figure 2 illustrates a typical case for $\xi \neq 0$ (and, hence $\lambda_2 \neq 0$). If the horizontal axis of these diagrams measures τ, then the risk functions depicted in Figure 2 will shift either to the right or to the left of its correctly specified counterparts (Figure 1), depending on the sign of ξ. This implies that in an underfitted model, the use of valid prior information does not necessarily lead to a reduction in risk. Equally, pre-testing is not necessarily risk superior to ignoring prior information, even if it is perfectly correct. This is consistent with Mittelhammer's (1984) results for the case in which the prior information exists in an exact form.

From the diagrams and the analytical results, we observe that given λ_2, $\rho(X\beta^{**},E(y))$ is bounded and approaches $\rho(X\tilde{\beta},E(y))$ as $\lambda_1 \to -\infty$, but it is unbounded and approaches $\rho(X\beta^{*},E(y))$ as $\lambda_1 \to \infty$; $\rho(X\hat{\beta},E(y))$ is bounded and approaches $\rho(X\tilde{\beta},E(y))$ as $|\lambda_1| \to \infty$. Given λ_1, $\rho(X\beta^{**},E(y))$ is unbounded as $\lambda_2 \to \infty$; for any fixed λ_1, $\rho(X\hat{\beta},E(y))$ is unbounded and approaches $\rho(X\beta^{**},E(y))$ as $\lambda_2 \to \infty$.[7] $\left[\rho(X\hat{\beta},E(y))-\rho(X\tilde{\beta},E(y))\right]$ and $\left[\rho(X\beta^{**},E(y))-\rho(X\tilde{\beta},E(y))\right]$ are both bounded by $-P_3/2 + \lambda_1^2P_1$ (given $\lambda_1 \leq 0$), or by $P_3/2 + 2\lambda_1^2 - 1 - \lambda_1^2P_1$ (given $\lambda_1 > 0$) as $\lambda_2 \to \infty$. It is apparent from these results that if λ_1 and λ_2 are both large, $X\hat{\beta}$ could have significantly greater risk than the unrestricted predictor $X\tilde{\beta}$. This result is significant as it implies that pre-testing could be potentially dangerous when the errors associated with the constraint and model specification are unknown in practice. This obviously cannot occur if there is no mis-specification in the model, in which case λ_2 vanishes and $\rho(X\hat{\beta},E(y))$ approaches $\rho(X\tilde{\beta},E(y))$ as $\lambda_1 \to \infty$. When $c \to 0$ or $c \to -\infty$, $\rho(X\hat{\beta},E(y))$ approaches $\rho(X\tilde{\beta},E(y))$ and $\rho(X\beta^{**},E(y))$ respectively. Regardless of the size of the pre-test (and hence the level of c), there exists no region in the λ_1 space such that the risk of the pre-test predictor is smaller than the risks of the unrestricted and inequality restricted predictor simultaneously. However, there is always a region such that $X\hat{\beta}$ has higher risk than both $X\tilde{\beta}$ and $X\beta^{**}$. This suggests that if we want to pre-test, then we need to choose an appropriate critical value which brings the pre-test risk function as close as possible to the smallest risk that can be achieved.

4. THE CHOICE OF OPTIMAL CRITICAL VALUES FOR THE PRE-TEST WHEN ESTIMATING E(y)

Various criteria have been proposed for choosing an optimal critical value of a pre-test. One such criterion is that of mini-max regret, which aims to find a level of c such that the maximum regret of not being on the minimum risk boundary is minimized. Along the lines of Sawa and Hiromatsu (1973) and Brook (1976), the regret function of $X\hat{\beta}$ is defined as

$$REG(\lambda_1,c) = \rho(X\hat{\beta},\ E(y)) - \inf_c \rho(X\hat{\beta},\ E(y)), \tag{24}$$

where $\inf_c \rho(X\hat{\beta},\ E(y))$ is the infimum of $\rho(X\hat{\beta},\ E(y))$ over all values of λ_1.

It is observed that

$$\inf_c \rho(X\hat{\beta},\ E(y)) = \begin{cases} \rho(X\hat{\beta},\ E(y)|c=-\infty) = \rho(X\beta^{**},E(y)) & \text{if } \lambda_1 < \lambda_1^* \\ \rho(X\hat{\beta},\ E(y)|c=0) = \rho(X\tilde{\beta},E(y)) & \text{if } \lambda_1 \geq \lambda_1^* \end{cases}, \tag{25}$$

where λ_1^* is that value of $\lambda_1 > 0$ for which $\rho(X\beta^{**},E(y)) = \rho(X\tilde{\beta},E(y))$.

If we let d_L and d_U denote the least favourable values of the regret function for $\lambda_1 < \lambda_1^*$ and $\lambda_1 \geq \lambda_1^*$ respectively, then, analogous to the case in which the linear restriction is held as a strict equality, the mini-max regret procedure is to seek a critical value which makes both d_L and d_U as small as possible. However, it is found empirically that increasing $|c|$ decreases d_L but increases d_U. Therefore, to minimize the mini-max regret function over all values of c and λ_1, the procedure is to seek c^* such that

$$\sup_{\lambda_1<\lambda_1^*} REG(\lambda_1,c^*) = \sup_{\lambda_1\geq\lambda_1^*} REG(\lambda_1,c^*). \tag{26}$$

c^* is then the mini-max regret critical value.

As it seems impossible to derive the mini-max regret critical values analytically, we rely on numerical computations.[8] Table 1 in Appendix B reports the mini-max regret critical values and the corresponding level of α for λ_2 = 0, 2, 10, 25, 50 and various values of v. From these results, the following conclusions are drawn :

(i) When the model is properly specified (*i.e.* $\lambda_2=0$), the optimal critical values do not vary much from -1.12 regardless of the degrees of freedom. This is qualitatively consistent with the results of Sawa and Hiromatsu (1973) and Brook (1976) for the case in which the linear restriction is held as a strict equality. However, this does not imply a constant α, which varies from 19.1 % to 13.1 % as v varies from 2 to 80. The risk function of $X\hat{\beta}$ with the mini-max regret critical value is depicted in Figure 1.

(ii) Once we allow for the omission of relevant regressors in the model (*i.e.* $\lambda_2 > 0$), the optimal critical values vary with the degrees of freedom of the model and can differ considerably from -1.12. This is illustrated in Figure 2. The rate at which c^* changes with the degrees of freedom in the model increases as λ_2 increases. Furthermore, for given degrees of freedom, $|c^*|$ decreases as λ_2 increases. Again, these results are qualitatively consistent with those reported in the literature for the case in which the linear restriction is held as a strict equality (see, Giles *et al.* (1992)).

Alternatively, one may consider choosing an optimal critical value according to the criterion suggested by Toyoda and Wallace (1976) of minimizing the average relative risk over the range of λ_1 (which in the present context is the area between $\rho(X\hat{\beta},E(y))$ and min $[\rho(X\tilde{\beta},E(y)), \rho(X\beta^{**},E(y))]$). However, since the latter is independent of c, this criterion effectively amounts to choosing an optimal c such that the area under the pre-test risk function is minimized. If the area is expressed as a function of the critical value, it can be shown that regardless of the magnitude of λ_2, this function reaches a minimum at c=0[9]. Hence this criterion always leads to the choice of the unrestricted estimator. This conclusion is consistent with Toyoda and

Wallace's (1976) result for the case in which the prior restriction on β exists in the form of a single linear equality. Given their results it is unclear whether our findings would extend to more than one inequality restriction.[10]

5. THE RISK FUNCTIONS OF THE ESTIMATORS OF THE SCALE PARAMETER

If $\bar{\sigma}^2$ is an estimator of σ^2, then the risk function of $\bar{\sigma}^2$ under squared error loss is defined as:

$$\rho(\bar{\sigma}^2,\ \sigma^2) = E\left[(\bar{\sigma}^2 - \sigma^2)^2\right]/\sigma^4.$$

From Giles (1990, 1991b),

$$\rho(\tilde{\sigma}^2,\ \sigma^2) = \left[2(v+4\lambda_2) + (v+2\lambda_2-(n+\delta))^2\right]/(n+\delta)^2 \tag{27}$$

$$\rho(\sigma^{*2},\ \sigma^2) = \Big[2[1 + v + 4(\lambda_1^2 + \lambda_2)] + [1 - k + 2(\lambda_1^2 + \lambda_2) - \gamma]^2\Big]/(n+\gamma)^2. \tag{28}$$

(27) and (28) collapse to the expressions given in Giles and Clarke (1989) when $\gamma=\delta=0$ and to those in Clarke *et al.* (1987b) when $\lambda_2=0$.

From (11) and (15), it is straight forward to show that the risk functions of the inequality restricted and pre-test estimators may be written as:

$$\begin{aligned}\rho(\sigma^{**2},\sigma^2) = \rho(\tilde{\sigma}^2,\sigma^2) + E\Big\{I_{(-\infty,\tau\sigma^{-1})}(u_1)\Big[\Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big]\\ \times\Big[2(\tilde{\sigma}^2-\sigma^2) + \Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4\end{aligned} \tag{29}$$

$$\begin{aligned}\rho(\hat{\sigma}^2,\sigma^2) = \rho(\sigma^{**2},\sigma^2) - E\Big\{I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)\Big[\Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big]\\ \times\Big[2(\tilde{\sigma}^2-\sigma^2) + \Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4\end{aligned} \tag{30}$$

Now, using Lemmas 1 and 2 in Appendix A and the Lemmas given in Judge and Bock (1978) or Clarke *et al.* (1987a), we establish the following theorem which defines the risk functions for σ^{**2} and $\hat{\sigma}^2$:

THEOREM 2

If $\lambda_1 \leq 0$, then

$$\rho(\sigma^{**2},\sigma^2) = \rho(\tilde{\sigma}^2,\sigma^2) + \Big[\lambda_1^2(n+\delta)^2[\lambda_1^2+v+2\lambda_2-(n+\delta)]/2 + v(\gamma-\delta)[2(n+\delta)(n+\gamma)-(n+\delta)(v+2)-(v+2)(n+\delta)]/2 + 2\lambda_2(\gamma-\delta)[(n+\delta)(n+\gamma)+(\delta-\gamma)(v+2)-\lambda_2(2n+\delta+\gamma)-2(n+\gamma)(v+2)]\Big]P_1/((n+\delta)(n+\gamma))^2 + \Big[8\lambda_1\lambda_2+4v\lambda_1+8\lambda_1^3-4\lambda_1(n+\delta)\Big]P_2/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + \Big[v+2\lambda_2+6\lambda_1^2-(n+\gamma)\Big]P_3/(n+\gamma)^2 + 8\lambda_1 P_4/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + 3P_5/(2(n+\gamma)^2) \tag{31}$$

$$\rho(\hat{\sigma}^2,\sigma^2) = \rho(\sigma^{**2},\sigma^2) - 3E_{5,v}/2(n+\gamma)^2 - 8\lambda_1E_{4,v}/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + (n+\gamma-6\lambda_1^2)E_{3,v}/(n+\gamma)^2 + (4\lambda_1(n+\gamma)-8\lambda_1^3)E_{2,v}/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + (2\lambda_1^2(n+\gamma)-2\lambda_1^4)E_{1,v}/(n+\gamma)^2 - vE_{3,v+2}/(n+\delta)(n+\gamma)^2 - 4v\lambda_1E_{2,v+2}/\Gamma(\tfrac{1}{2})(n+\gamma)^2 - v(2\lambda_1^2+(\gamma-\delta))E_{1,v+2}/(n+\gamma)^2 - 2\lambda_2E_{3,v+4}/(n+\gamma)^2 - 8\lambda_1\lambda_2E_{2,v+4}/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + \Big\{\Big[(\gamma-\delta)v(v+2)(2n+\gamma+\delta) - 4\lambda_2(\gamma-\delta)(n+\delta)(n+\gamma) - 8\lambda_1^2\lambda_2(n+\delta)^2\Big]E_{1,v+4} + 2(\gamma-\delta)(2v+4)\lambda_2(\gamma+\delta+2n)E_{1,v+6} + 4\lambda_2^2(\gamma-\delta)(2n+\gamma+\delta)E_{1,v+8}\Big\}/(2(n+\delta)^2(n+\gamma)^2) \tag{32}$$

If $\lambda_1 > 0$, then

$$\rho(\sigma^{**2},\sigma^2) = \rho(\tilde{\sigma}^2,\sigma^2) + \Big\{\lambda_1^2(n+\delta)^2[\lambda_1^2+v+2\lambda_2-(n+\delta)] + v(\gamma-\delta)[2(n+\delta)(n+\gamma) - (n+\delta)(v+2) - (v+2)(n+\delta)] + 4\lambda_2(\gamma-\delta)[(n+\delta)(n+\gamma) + (\delta-\gamma)(v+2) - \lambda_2(2n+\delta+\gamma) - 2(n+\gamma)(v+2)] - \Big[\lambda_1^2(n+\delta)^2[\lambda_1^2+v+2\lambda_2 - (n+\delta)]/2 + v(\gamma-\delta)[2(n+\delta)(n+\gamma) - (n+\delta)(v+2) - (v+2)(n+\delta)]/2 + 2\lambda_2(\gamma-\delta)[(n+\delta)$$

$$(n+\gamma) + (\delta-\gamma)(v+2) - \lambda_2(2n+\delta+\gamma) - 2(n+\gamma)(v+2)]\Big]P_1\Big\}/((n+\delta)(n+\gamma))^2$$

$$+ \Big[8\lambda_1\lambda_2+4v\lambda_1+8\lambda_1^3-4\lambda_1(n+\delta)\Big]P_2/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + 2\Big\{\Big[v+2\lambda_2+6\lambda_1^2-(n+\gamma)\Big] -$$

$$\Big[v+2\lambda_2+6\lambda_1^2-(n+\gamma)\Big]P_3\Big\}/(n+\gamma)^2 + 8\lambda_1P_4/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + (6-3P_5)$$

$$/(2(n+\gamma)^2) \tag{33}$$

$$\rho(\hat{\sigma}^2,\sigma^2) = \rho(\sigma^{**2},\sigma^2) - 3E_{5,v}/2(n+\gamma)^2 - 8\lambda_1E_{4,v}/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + (n+\gamma-6\lambda_1^2)$$

$$E_{3,v}/(n+\gamma)^2 + (4\lambda_1(n+\gamma)-8\lambda_1^3)E_{2,v}/\Gamma(\tfrac{1}{2})(n+\gamma)^2 + (2\lambda_1^2(n+\gamma) -2\lambda_1^4)$$

$$E_{1,v}/(n+\gamma)^2 - vE_{3,v+2}/(n+\delta)(n+\gamma)^2 - 4v\lambda_1E_{2,v+2}/\Gamma(\tfrac{1}{2})(n+\gamma)^2$$

$$- v(2\lambda_1^2+(\gamma-\delta))E_{1,v+2}/(n+\gamma)^2 - 2\lambda_2E_{3,v+4}/(n+\gamma)^2 - 8\lambda_1\lambda_2E_{2,v+4}/$$

$$\Gamma(\tfrac{1}{2})(n+\gamma)^2 + \Big\{\Big[(\gamma-\delta)v(v+2)(2n+\gamma+\delta) - 4\lambda_2(\gamma-\delta)(n+\delta)(n+\gamma) -$$

$$8\lambda_1^2\lambda_2(n+\delta)^2\Big]E_{1,v+4} + 2(\gamma-\delta)(2v+4)\lambda_2(\gamma+\delta+2n)E_{1,v+6} + 4\lambda_2^2(\gamma-\delta)$$

$$(2n+\gamma+\delta)E_{1,v+8}\Big\}/(2(n+\delta)^2(n+\gamma)^2) - 3G_{5,v}/(n+\gamma)^2 + 2(n+\gamma-6\lambda_1^2)$$

$$G_{3,v}/(n+\gamma)^2 + 4(\lambda_1^2(n+\gamma)-\lambda_1^4)G_{1,v}/(n+\gamma)^2 - 2vG_{3,v+2}/(n+\delta)(n+\gamma)^2$$

$$-2v (2\lambda_1^2 + (\gamma-\delta))G_{1,v+2}/(n+\gamma)^2 - 4\lambda_2G_{3,v+4}/(n+\gamma)^2 + \Big\{\Big[(\gamma-\delta)$$

$$v(v+2)(2n+\gamma+\delta) - 4\lambda_2(\gamma-\delta)(n+\delta)(n+\gamma) - 8\lambda_1^2\lambda_2(n+\delta)^2\Big]G_{1,v+4}$$

$$+ 2(\gamma-\delta)(2v+4)\lambda_2(\gamma+\delta+2n)G_{1,v+6} + 4\lambda_2^2(\gamma-\delta)(2n+\gamma+\delta)G_{1,v+8}\Big\}$$

$$/((n+\delta)^2(n+\gamma)^2) \tag{34}$$

Proof: see Appendix A.

Given the difficulty in analyzing these expressions, we numerically evaluate them for n = 10, 30, 50, k = 2, 5, λ_1 = [-10,10] and various choices of

λ_2 and α in the same way as for the prediction vector. From the analytical and numerical results, we observe the following. At least for the cases that we have considered, when the direction of the constraint is correct (i.e. $\tau \le 0$), and there is no specification error in the model, (*i.e.* $\xi=0$ such that $\lambda_1=\tau/\sqrt{2}$ and $\lambda_2=0$), the inequality $\rho(\sigma^{**2}, \sigma^2) \le \rho(\hat{\sigma}^2, \sigma^2) \le \rho(\tilde{\sigma}^2, \sigma^2)$ always holds. However, this is not necessarily true when $\xi \ne 0$ (and hence $\lambda_2 \ne 0$). This result is analogous to the corresponding result when estimating E(y).

For a given value of λ_2, both $\left[\rho(\sigma^{**2}, \sigma^2) - \rho(\tilde{\sigma}^2, \sigma^2)\right]$ and $\left[\rho(\hat{\sigma}^2, \sigma^2) - \rho(\tilde{\sigma}^2, \sigma^2)\right]$ are bounded and approach zero as $\lambda_1 \to -\infty$; when $\lambda_1 \to \infty$, $\rho(\sigma^{**2}, \sigma^2)$ is unbounded and approaches $\rho(\sigma^{*2}, \sigma^2)$, while $\rho(\hat{\sigma}^2, \sigma^2)$ is bounded and approaches $\rho(\tilde{\sigma}^2, \sigma^2)$. For any fixed λ_1, $\rho(\hat{\sigma}^2, \sigma^2) \to \rho(\sigma^{**2}, \sigma^2)$ as $\lambda_2 \to \infty$, but $\left[\rho(\sigma^{**2}, \sigma^2) - \rho(\tilde{\sigma}^2,\sigma^2)\right]$ and $\left[\rho(\hat{\sigma}^2, \sigma^2) - \rho(\tilde{\sigma}^2, \sigma^2)\right]$ are both unbounded as $\lambda_2 \to \infty$. This result contrasts with what we have observed when estimating the prediction vector, in which case the corresponding differences are both bounded as $\lambda_2 \to \infty$, for given λ_1. Other things being equal, the difference between $\rho(\hat{\sigma}^2, \sigma^2)$ and $\rho(\tilde{\sigma}^2, \sigma^2)$ can be significant when λ_1 is relatively large and λ_2 increases without limit.

Furthermore, with a relatively large λ_2, $\tilde{\sigma}^2_{ML}$ can be uniformly risk superior to the other estimators corresponding to the ML component that we have considered, while this situation does not emerge if the component estimator is the LS or MM estimator instead. By constrast, with the LS or MM estimators, there always exists a family of inequality pre-test estimators which strictly dominates the unrestricted estimator, regardless of the value of λ_2. Over certain regions in the λ_1 space, this family of pre-test estimators also dominate the inequality restricted estimator. These results concur qualitatively with those of Giles (1990, 1991b) for the case in which the restriction holds as a strict equality and the disturbances are spherically symmetric.

From our numerical evaluations, it is also apparent that for sufficiently large λ_2, the inequality pre-test estimator with $c = -1$ for the LS component and $c = -\sqrt{v/(v+2)}$ for the MM component may strictly dominate their respective unrestricted and inequality restricted estimators. This constrasts with what we observed when estimating the prediction vector. At least for the cases that we have examined, out of the three component estimators, in terms of minimizing estimator risk, it is always preferable to use the estimators based on the minimum mean square error principle, other things being equal. Figures 3 to 6 in Appendix B illustrate some of these results.

The choice of optimal levels of significance when estimating σ^2 is currently being investigated. However, it is irrelevant to consider optimal pre-test size when λ_2 is large, as the inequality pre-test estimators with $c = -1$ (for LS), $c = -\sqrt{v/(v+2)}$ (for MM) and $c = 0$ (for ML) strictly dominate the family of IPT estimators. Our preliminary results show that when λ_2 is small, $c = -1$ (for LS) and $c = -\sqrt{v/(v+2)}$ (for MM) are the optimal critical values under both the criteria of mini-max regret and minimum average relative risk for the cases that we have considered.[11] When the maximum likelihood method is applied, then the optimal critical value varies not only with the model's degrees of freedom, but also with the number of observations in the sample. Further analysis of this case is currently being undertaken by the author.

6. CONCLUSIONS

In this paper we have considered the sampling performance of the inequality restricted and inequality pre-test estimators for both the prediction vector and the scale parameter in an omitted variable model. This was also the first analysis of the estimation of the scale parameter that takes into account the inequality nature of non-sample information imposed on the regression coefficients. The risk functions of these estimators were derived and numerically evaluated.

Common to both the problems of estimating the prediction vector and scale parameter, we found that the use of perfectly correct valid information does not ensure a reduction in risk in an underfitted model. Both the inequality restricted and pre-test estimators for the prediction vector were found to be risk inferior to the unrestricted estimator over a large portion of the (λ_1,λ_2) space. In particular, the degree of inferiority increases with λ_2. If the method of maximum likelihood is applied to estimating the scale parameter, it was found that the unrestricted estimator can uniformly dominate all other estimators under consideration when the degree of model mis-specification is serious. Thus, when estimating E(y) or σ^2 using the principle of maximum likelihood in an underfitted model, the application of the unrestricted estimator may be preferable to pre-testing or imposing restrictions naively in terms of minimizing estimator risk.

If the least square or minimum mean square error estimator is chosen as the component estimator for the scale parameter, then our results show that, for any finite λ_2, there exists a family of pre-test estimators which are uniformly superior to the estimator that ignores the prior information. When λ_2 is sufficently large, the inequality pre-test estimator, with an appropriate choice of critical value, may strictly dominate both its components. These results clearly indicate that, potentially, there is much to gain from pre-testing when estimating the scale parameter using the least square or minimum mean square error principles, and specification error tends to make the IPT estimator more favourable relative to its components.

However, in practice, researchers rarely estimate the scale parameter and prediction vector separately. Given our results, it is unclear whether the risk gain from pre-testing in estimating the scale parameter can compensate the corresponding potential risk loss when estimating the prediction vector. This suggests that one should perhaps consider a joint risk function for estimating both the scale parameter and the prediction vector. This remains an interesting point of departure for future research.

We have also shown that, if the non-sample information is a single linear inequality constraint, then a critical value of zero for the pre-test (*i.e.* always ignore the restriction) is the best strategy for estimating the prediction vector according to the criterion of minimizing the average relative risk. If the alternative mini-max regret criterion is used, then we have shown that, under the maintained assumption of a properly specified model, the optimal critical values are invariant to the model's degrees of freedom. However, this property no longer holds once we allow for possible mis-specification in the regressor matrix. Accordingly, any attempt to apply a mini-max critical value obtained under the assumption that $\lambda_2=0$ will not necessarily lead to an optimal pre-test risk when the model is in fact underfitted. An investigation on the choice of the optimal critical value for the pre-test when estimating the scale parameter is currently being undertaken by the author.

References

Albertson, K.V. (1991), "Pre-test estimation in a regression model with a mis-specified error covariance matrix", Discussion paper no. 9115, Department of Economics, University of Canterbury.

Brent, R.P. (1974), *Algorithms for minimization without derivatives*, Englewood Cliffs, N.J. : Prentice Hall.

Brook, R.J. (1976), "On the use of a regret function to set significane points in prior tests of estimation", *Journal of the American Statistical Association* 71, 126-131.

Brook, R.J. and R.H. Fletcher (1981), "Optimal significance levels of prior tests in the presence of multicollinearity", *Communications in Statistics : Theory and Methods* A10, 1401-1413.

Clarke, J.A., D.E.A. Giles and T.D. Wallace (1987a), "Estimating the error variance in regression after a preliminary test of restrictions on the coefficients", *Journal of Econometrics* 34, 293-304.

Clarke, J.A., D.E.A. Giles and T.D. Wallace (1987b), "Preliminary test estimation of the error variance in linear regression", *Econometric Theory* 3, 299-304.

Giles, D.E.A. (1986), "Preliminary test estimation in mis-specified regressions", *Economics Letters* 21, 325-328.

Giles, D.E.A. and J.A. Clarke (1989), "Preliminary test estimation of the scale parameter in a mis-specified regression model", *Economics Letters* 30, 201-205.

Giles, D.E.A., O. Lieberman, and J.A. Giles (1992), "The optimal size of a preliminary test of linear restrictions in a mis-specified regression model", *Journal of the American Statistical Association* forthcoming.

Giles, J.A. (1990), "Preliminary test estimation of a mis-specified linear model with spherically symmetric disturbances", Ph.D. thesis, University of Canterbury.

Giles, J.A. (1991a), "Pre-testing for linear restrictions in a regression with spherically symmetric disturbances", *Journal of Econometrics* 50, 377-398.

Giles, J.A. (1991b), "Pre-testing in a mis-specified regression model", *Communications in Statistics: Theory and Methods* A20, 3221-3238.

Hasegawa, H (1989), "On some comparisons between Bayesian and sampling theoretic estimators of a normal mean subject to an inequality constraint" *Journal of the Japan Statistical Society* 19, 167-177.

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., W.E. Griffiths, R. Carter Hill, H. Lutkepol and T.C. Lee (1980), *The theory and practice of econometrics* (1st ed.), John Wiley and Sons, New York.

Judge, G.G. and T. Takayama (1966), "Inequality restrictions in regression analysis", *Journal of the American Statistical Association* 61, 166-181.

Judge, G.G. and T.A. Yancey (1981), "Sampling properties of an inequality restricted estimator", *Economics Letters* 7, 327-333.

Judge, G.G. and T.A. Yancey (1986), *Improved methods of inference in econometrics*, Amsterdam, North-Holland.

Liew, C.K. (1976), "Inequality constrained least squares estimation", *Journal of the American Statistical Association* 71, 746-751.

Lovell, M.C. and E. Prescott (1970), "Multiple regression with inequality constraints: Pre-testing bias, hypothesis testing and efficiency", *Journal of the American Statistical Association* 65, 913-925.

Mittelhammer, R.C. (1984), "Restricted least squares, pre-test, OLS and Stein rule estimators: Risk comparisons under model mis-specification", *Journal of Econometrics* 25, 151-164.

Numerical Algorithm Group Ltd. (1991), "NAG Fortran Library Manual, Mark 15, Vol. 1", NAG Inc., U.S.A.

Ohtani, K. and T. Toyoda (1978), "Mini-max regret critical values for a preliminary test in pooling variance", *Journal of the Japan Statistical Society* 8, 15-20.

Ohtani, K. (1983), "Preliminary test predictor in the linear regression model including a proxy variable", *Journal of the Japan Statistical Society* 17, 81-89.

Ohtani, K. (1991a),"Estimation of the variance in a normal population after the one sided pre-test for the mean", *Communications in Statistics: Theory and Methods* A20, 219-234.

Ohtani, K. (1991b),"Some sampling properties of the inequality constrained least square estimator in a linear regression model with a proxy variable", mimeo., Kobe University.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1986), *Numerical recipes: The art of scientific computing,* New York, Cambridge University Press.

Sawa, T. and T. Hiromatsu (1973), "Mini-max regret significance points for a preliminary test in regression analysis", *Econometrica* 45, 1293-1298.

Thomson, M. (1982), "Some results on the statistical properties of an inequality constrained least squares estimator in a linear model with two regressors", *Journal of Econometrics* 19, 215-231.

Thomson, M. and P. Schmidt (1982), "A note on the comparison of the mean square error of inequality constrained least squares and other related estimators", *Review of Economics and Statistics* 64, 174-176.

Toyoda, T and T.D. Wallace (1975), "Estimation of variance after a preliminary test of homogeneity and optimal levels of significance for the pre-test", *Journal of Econometrics* 3, 395-404.

Toyoda, T and T.D. Wallace (1976), "Optimal critical values for pre-testing in regression", *Econometrica* 44, 365-375.

Yancey, T.A., G.G. Judge and R. Bohrer (1989), "Sampling performance of some joint one-sided preliminary test estimators under squared error loss", *Econometrica* 57, 1221-1228.

Yancey, T.A., G.G. Judge and D.M. Mandy (1983), "The sampling performance of pre-test estimators of the scale parameter under squared error loss", *Economics Letters* 12, 181-186.

Zellner, A. (1961), "Linear regression with inequality constraints on the coefficients", mimeographed report 6109 of the International Centre for Management Science, Rotterdam.

Footnotes

1. See also Liew (1976) and Judge and Yancey (1986).

2. By constrast, in cases for which the prior information exists as exact equalities, the sampling properties of the resulting equality restricted and pre-test estimators have been investigated rather extensively. [See, for example, Ohtani (1983), Mittelhammer (1984), Giles (1986), Giles and Clarke (1989), Giles (1990, 1991b) and Albertson (1991).]

3. Although Ohtani (1991a) has considered pre-testing for the variance in a normal population after a one sided pre-test of the mean, his pre-test estimator is a choice between the unrestricted and the equality restricted estimators. Yancey *et al.* (1983) discuss the sampling properties of the inequality restricted and pre-test estimators for σ^2 when an inequality constraint exists in the form of $\sigma^2 \geq \sigma_0^2$, which is a different problem to the one that we are investigating here.

4. It must be noted that these expressions are valid only for $c < 0$ (i.e. $\alpha < 0.5$). For $c \geq 0$, it can be easily shown that the unrestricted estimator is always used as the estimator for the model regardless of the outcome of the hypothesis test.

5. As is noted in the introduction section, Hasegawa (1989) considers the estimation of the mean in a normal population, which is related but not identical to the problem investigated here.

6. Judge and Yancey (1981, 1986) assume that σ^2 is known in their analysis. However, the results are qualitatively the same as for the σ^2 unknown case.

7. For any non-zero c and finite λ_1, G_{ij} and E_{ij} both approach zero as $\lambda_2 \rightarrow \infty$. Hence the risk of the pre-test estimator approaches that of the inequality restricted estimator as $\lambda_2 \rightarrow \infty$. When $\lambda_1 \rightarrow \infty$, $G_{ij} \rightarrow 1$ whereas $E_{ij} \rightarrow 0$, hence the risk of the pre-test estimator approaches that of the unrestricted

estimator.

8. Brent's (1974) algorithm is used to search for the value of λ_r^*. The Golden Section Search routine given in Press *et al.* (1986) is used to compute mini-max regret critical values.

9. Details are available on request.

10. Toyoda and Wallace (1976) find that when the number equality restrictions is less than 5, c=0 is the optimal critical value regardless of the model's degrees of freedom. When there are more than 5 restrictions, the optimal critical value increases with both the degrees of freedom and the number of restrictions, and is approximately 2 for the central F distribution when the number of restrictions is more than 60.

11. It can also be shown analytically that these pre-test risk functions achieve stationary points at $c = 0$ (for ML), $c= -1$ (for LS) and $c = -\sqrt{v/(v+2)}$ (for MM), which coincides with the results obtained when the *a priori* restriction holds as a strict equality, as shown by Giles (1990, 1991a, 1991b). Details are available on request.

Appendix A

Proof of Theorems 1 and 2

We need the following supporting lemmas in order to derive the risk functions of $X\beta^{**}$, $X\hat{\beta}$, σ^{**2} and $\hat{\sigma}^2$:

Lemma 1:

If w is a normal random variable with mean ϑ and variance 1, and $d \in R$, then

$$E\left[I_{(-\infty,d)}(w)w^j\right] = \begin{cases} \sum_{t=0}^{j} \binom{j}{t} (-1)^t \vartheta^{j-t} \Omega_t P(\chi^2_{t+1} \geq f^2)/2 & \text{if } f \leq 0 \quad \text{(A.1)} \\ \sum_{t=0}^{j} \binom{j}{t} \left[I(t) - P(\chi^2_{t+1} \geq f^2)/2\right] \Omega_t \vartheta^{j-t} & \text{if } f > 0 \quad \text{(A.2)} \end{cases},$$

where $\Omega_t = 2^{t/2}\Gamma((t+1)/2)/\Gamma(\frac{1}{2})$, $f = d-\vartheta$ and $I(t) = 0$ if t is odd, 1 otherwise. This Lemma generalises Theorem 1 of Judge and Yancey (1986, 72-73) to a normal variable with non-zero mean.

Proof:

Let $z = (w - \vartheta) \sim N(0,1)$, then $E\left[I_{(-\infty,d)}(w)w^j\right]$ can be written as $E\left[I_{(-\infty,d-\vartheta)}(z)(z+\vartheta)^j\right]$. Now $E\left[I_{(-\infty,d-\vartheta)}(z)(z^t\vartheta^{j-t})\right]$, $t=0,..j$, can be evaluated using Theorem 1 of Judge and Yancey(1986). Lemma 1 then follows.

Lemma 2:

Let w, d, f and I(t) be defined as in Lemma 1. Let $\psi \sim \chi^2_{(v,\lambda)}$ and $c \in R^-$, then for $f \leq 0$,

$$E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right]$$
$$= \sum_{t=0}^{j} \binom{j}{t} (-1)^t \vartheta^{j-t} \Omega_t e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!2^{v/2+i}\Gamma(\frac{v}{2}+i)} \int_0^{\infty} P\left(\chi^2_{t+1} \geq (c\sqrt{\psi}+f)^2\right) \psi^{\frac{\nu}{2}+i-1} e^{-\frac{\psi}{2}}/2 \, d\psi \quad \text{(A.3)}$$

and for $f > 0$,

$$E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right]$$

$$= \sum_{t=0}^{j}\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t e^{-\lambda}\sum_{l=0}^{\infty}\frac{\lambda^l}{l!2^{\nu/2+l}\Gamma(\frac{\nu}{2}+l)}\left[\int_0^{\infty}P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)\psi^{\frac{\nu}{2}+l-1}e^{-\frac{\psi}{2}}/2\,d\psi\right.$$

$$\left.+I(t)\int_0^{f^2/c^2}P\left(\chi^2_{t+1}<(c\sqrt{\psi}+f)^2\right)\psi^{\frac{\nu}{2}+l-1}e^{-\frac{\psi}{2}}d\psi\right] \qquad (A.4)$$

Proof:

Since $c \in R^-$, hence $f \leq 0 \Rightarrow c\sqrt{\psi}+f \leq 0$. Using Lemma 1, $E_{w|c\sqrt{\psi}+d\leq 0}\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = \sum_{t=0}^{j}\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)/2$. Therefore, $E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = \sum_{t=0}^{j}\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t E_{\psi}\left[P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)\right]/2$ when $f \leq 0$, which leads to (A.3). Now, when $f > 0$, the sign of $c\sqrt{\psi} + f$ is undetermined. When $c\sqrt{\psi} + f \leq 0$, the range of ψ is restricted to $\psi \geq f^2/c^2$, while $c\sqrt{\psi} + f > 0 \Rightarrow 0 < \psi < f^2/c^2$. Hence when $f > 0$, $E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = E\left\{I_{(f^2/c^2,\infty)}(\psi)E_{w|c\sqrt{\psi}+f\leq 0}\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] + I_{(0,f^2/c^2)}(\psi)E_{w|c\sqrt{\psi}+f>0}\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right]\right\}$. The two inner expectations may be evaluated using Lemma 1. Noting that $I_{(f^2/c^2,\infty)}(\psi) = 1 - I_{(0,f^2/c^2)}(\psi)$, $(-1)^t + 1 = 2I(t)$, $E_{\psi}\left[I_{(0,f^2/c^2)}(\psi)F(\psi)\right] = \int_0^{f^2/c^2}F(\psi)\ \text{p.d.f.}(\psi)\ d\psi$ and $1 - P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right) = P\left(\chi^2_{t+1}<(c\sqrt{\psi}+f)^2\right)$ for any given ψ, (A.4) follows directly.

Proof of Theorem 1

The expectations in (18) may be evaluated directly using Lemma 1, for j = 1, 2, 3, and by recognising that $\lambda_1^2 = (\tau-\xi)^2/2\sigma^2$ enables the derivation of $\rho(X\beta^{**},E(y))$.

To establish $\rho(X\hat{\beta},E(y))$, we require Lemma 2 to evaluate the expectation terms of the form $E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(.)\right]$ in (19). The remainder of the expectations may be evaluated using Lemma 1. Hence $\rho(X\hat{\beta},E(y))$ follows directly.

Proof of Theorem 2

The evaluation of $\rho(\sigma^{2**},\sigma^2)$ involves the straight forward application of Lemma 1 for j = 1, 2,..,5.

Finally, the derivation of $\rho(\hat{\sigma}^2,\sigma^2)$ involoves the evaluations of $E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)\tilde{\sigma}^2\right]$ and $E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)\tilde{\sigma}^4\right]$, among others. Now $I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)|u_1$ may be regarded as a function of $\tilde{\sigma}^2$, as $\tilde{\sigma}$ is defined only on the non-negative horizon and each $\tilde{\sigma}^2$ corresponds to a unique $\tilde{\sigma}$. Therefore, using the lemmas given in Clarke *et al.* (1987a) or Judge and Bock (1978, pg 319-321),

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)\tilde{\sigma}^2\right] = \frac{\sigma^2}{n+\delta}\left\{v\,E\left[I_{\left(-\infty,c\sqrt{\frac{\chi'^2_{(v+2,\lambda_2)}}{v}}+\tau\sigma^{-1}\right)}(u_1)\right] + 2\lambda_2\right.$$
$$\left. E\left[I_{\left(-\infty,c\sqrt{\frac{\chi'^2_{(v+4,\lambda_2)}}{v}}+\tau\sigma^{-1}\right)}(u_1)\right]\right\}. \tag{A.5}$$

The two inner expectations can be evaluated using Lemma 2. The evaluation of $E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)\sigma^{-1})}(u_1)\tilde{\sigma}^4\right]$ involves the repeated use of (A.5). Using these results and the two Lemmas given above yields $\rho(\hat{\sigma}^2,\sigma^2)$ directly.

Appendix B

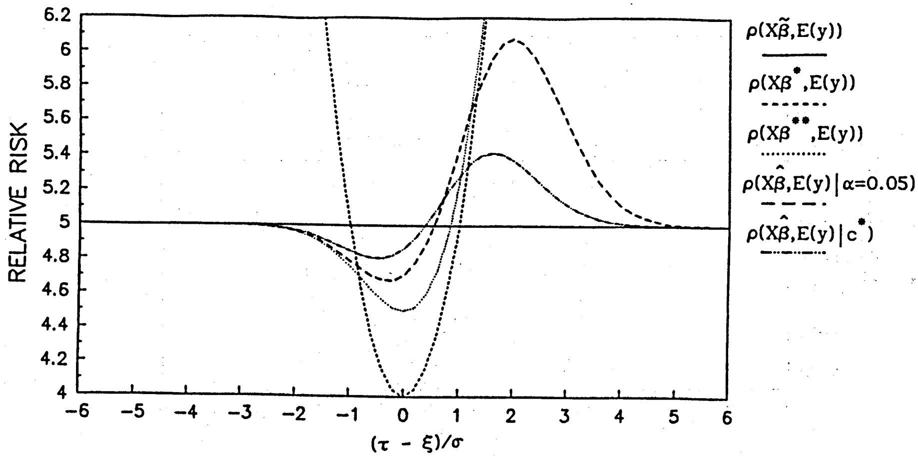


Figure 1. The risk functions for $X\tilde{\beta}$, $X\beta^{*}$, $X\beta^{**}$ and $X\hat{\beta}$ when n=30, k=5 and $\lambda_2 = 0$.

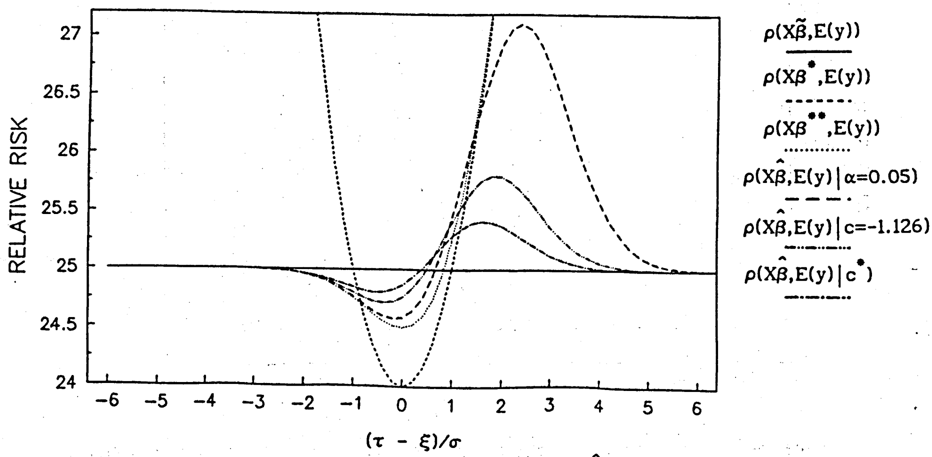


Figure 2. The risk functions for $X\tilde{\beta}$, $X\beta^{*}$, $X\beta^{**}$ and $X\hat{\beta}$ when n=30, k=5 and $\lambda_2 = 10$.

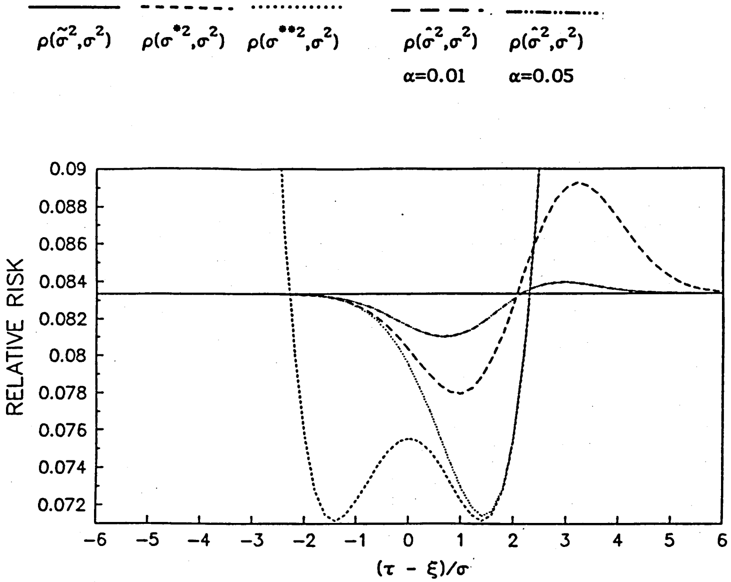


Figure 3. The risk functions for $\tilde{\sigma}^2$, σ^{*2}, σ^{**2} and $\hat{\sigma}^2$ (maximum likelihood component) when n=30, k=5 and λ_2= 0.

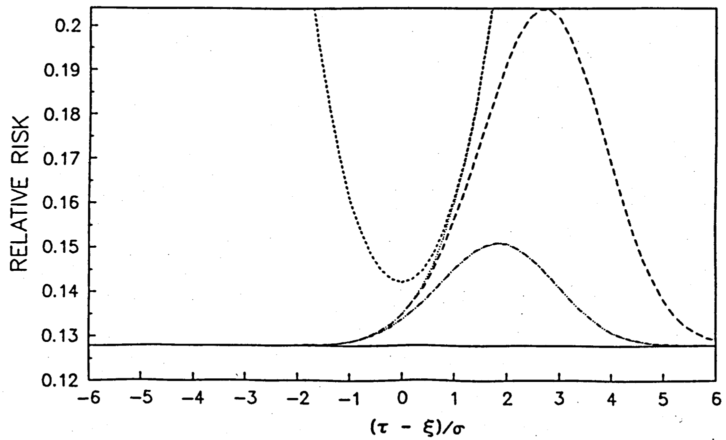
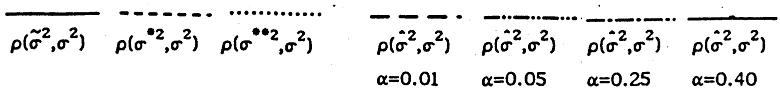


Figure 4. The risk functions for $\tilde{\sigma}^2$, σ^{*2}, σ^{**2} and $\hat{\sigma}^2$ (maximum likelihood component) when n=30, k=5 and λ_2= 5.



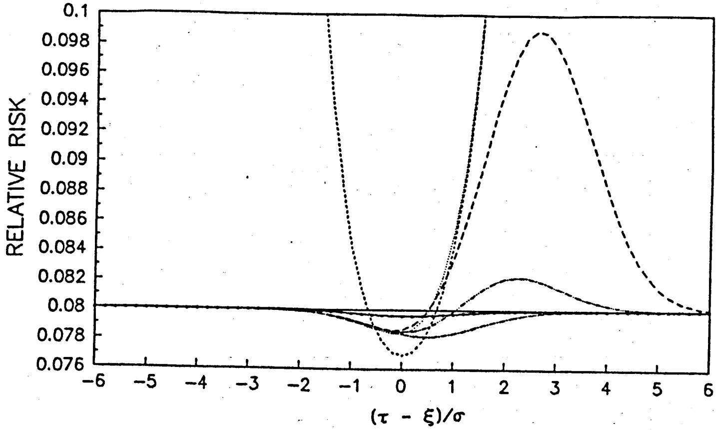


Figure 5. The risk functions for $\tilde{\sigma}^2$, σ^{*2}, σ^{**2} and $\hat{\sigma}^2$ (least squares component) when n=30, k=5 and λ_2= 0.

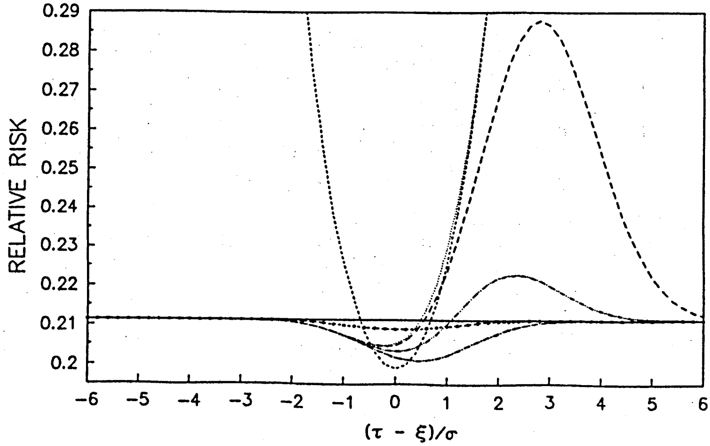


Figure 6. The risk functions for $\tilde{\sigma}^2$, σ^{*2}, σ^{**2} and $\hat{\sigma}^2$ (minimum mean square error component) when n=30, k=5 and λ_2= 5.

Table 1: Mini-max regret critical values for the pre-test of an inequality restriction when estimating E(y)

	λ_2									
	0		2		10		25		50	
v	c^*	α(%)	c^*	α(%)	c^*	α(%)	c^*	α(%)	c^*	α(%)
2	-1.112	19.096	-0.642	29.322	-0.338	38.363	-0.221	42.287	-0.158	44.455
5	-1.118	15.717	-0.834	22.115	-0.502	31.834	-0.340	37.401	-0.246	40.774
10	-1.122	14.395	-0.949	18.253	-0.649	26.538	-0.460	32.774	-0.340	37.049
15	-1.124	13.929	-0.999	16.682	-0.737	23.632	-0.541	29.816	-0.407	34.482
20	-1.125	13.691	-1.027	15.829	-0.796	21.765	-0.602	27.687	-0.460	32.514
25	-1.126	13.547	-1.045	15.293	-0.839	20.458	-0.651	26.061	-0.504	30.926
30	-1.127	13.449	-1.058	14.926	-0.873	19.490	-0.690	24.769	-0.542	29.603
35	-1.127	13.38	-1.067	14.658	-0.899	18.705	-0.723	23.715	-0.574	28.477
40	-1.127	13.327	-1.074	14.455	-0.92	18.150	-0.752	22.835	-0.603	27.504
45	-1.127	13.286	-1.08	14.295	-0.938	17.667	-0.776	22.093	-0.628	26.652
50	-1.127	13.253	-1.085	14.165	-0.953	17.267	-0.797	21.454	-0.660	25.617
55	-1.127	13.227	-1.088	14.059	-0.965	16.929	-0.816	20.900	-0.672	25.227
60	-1.127	13.204	-1.092	13.970	-0.976	16.640	-0.833	20.414	-0.691	24.623
65	-1.127	13.185	-1.094	13.894	-0.986	16.390	-0.848	19.984	-0.708	24.078
70	-1.128	13.169	-1.097	13.828	-0.994	16.172	-0.861	19.601	-0.724	23.582
75	-1.128	13.155	-1.099	13.771	-1.002	15.980	-0.874	19.258	-0.738	23.130
80	-1.128	13.142	-1.1	13.722	-1.009	15.810	-0.885	18.949	-0.752	22.715

LIST OF DISCUSSION PAPERS*

(Continued on next page)

No. 9012 Some Consequences of Applying the Goldfeld-Quandt Test to Mis-Specified Regression Models, by David E. A. Giles and Guy N. Saxton.

No. 9013 Pre-testing in a Mis-specified Regression Model, by Judith A. Giles.

No. 9014 Two Results in Balanced-Growth Educational Policy, by Alan E. Woodfield.

No. 9101 Bounds on the Effect of Heteroscedasticity on the Chow Test for Structural Change, by David Giles and Offer Lieberman.

No. 9102 The Optimal Size of a Preliminary Test for Linear Restrictions when Estimating the Regression Scale Parameter, by Judith A. Giles and Offer Lieberman.

No. 9103 Some Properties of the Durbin-Watson Test After a Preliminary t-Test, by David Giles and Offer Lieberman.

No. 9104 Preliminary-Test Estimation of the Regression Scale Parameter when the Loss Function is Asymmetric, by Judith A. Giles and David E. A. Giles.

No. 9105 On an Index of Poverty, by Manimay Sengupta and Prasanta K. Pattanaik.

No. 9106 Cartels May Be Good For You, by Michael Carter and Julian Wright.

No. 9107 L_p-Norm Consistencies of Nonparametric Estimates of Regression, Heteroskedasticity and Variance of Regression Estimate when Distribution of Regression is Known, by Radhey S. Singh.

No. 9108 Optimal Telecommunications Tariffs and the CCITT, by Michael Carter and Julian Wright.

No. 9109 Price Indices : Systems Estimation and Tests, by David Giles and Ewen McCann.

No. 9110 The Limiting Power of Point Optimal Autocorrelation Tests, by John P. Small.

No. 9111 The Exact Power of Some Autocorrelation Tests When the Disturbances are Heteroscedastic, by John P. Small.

No. 9112 Some Consequences of Using the Chow Test in the Context of Autocorrelated Disturbances, by David Giles and Murray Scott.

No. 9113 The Exact Distribution of R^2 when the Disturbances are Autocorrelated, by Mark L. Carrodus and David E. A. Giles.

No. 9114 Optimal Critical Values of a Preliminary Test for Linear Restrictions in a Regression Model with Multivariate Student-t Disturbances, by Jason K. Wong and Judith A. Giles.

No. 9115 Pre-Test Estimation in a Regression Model with a Misspecified Error Covariance Matrix, by K. V. Albertson.

No. 9116 Estimation of the Scale Parameter After a Pre-test for Homogeneity in a Mis-specified Regression Model, by Judith A. Giles.

No. 9201 Testing for Arch-Garch Errors in a Mis-specified Regression, by David E. A. Giles, Judith A. Giles, and Jason K. Wong.

No. 9202 Quasi Rational Consumer Demand – Some Positive and Normative Surprises, by John Fountain.

No. 9203 Pre-test Estimation and Testing in Econometrics: Recent Developments, by Judith A. Giles and David E. A. Giles.

No. 9204 Optimal Immigration in a Model of Education and Growth, by K-L. Shea and A. E. Woodfield.

No. 9205 Optimal Capital Requirements for Admission of Business Immigrants in the Long Run, by K-L. Shea and A. E. Woodfield.

No. 9206 Causality, Unit Roots and Export-Led Growth: The New Zealand Experience, by David E. A. Giles, Judith A. Giles and Ewen McCann.

No. 9207 The Sampling Performance of Inequality Restricted and Pre-Test Estimators in a Mis-specified Linear Model, by Alan T. K. Wan.

* Copies of these Discussion Papers may be obtained for $4 (including postage, price changes occasionally) each by writing to the Secretary, Department of Economics, University of Canterbury, Christchurch, New Zealand. A list of the Discussion Papers prior to 1988 is available on request.