Further results on optimal critical values of pre-test when estimating the regression error variance

ALAN T.K. WAN†,* , GUOHUA ZOU‡ AND KAZUHIRO OHTANI§

†Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong
E-mail: msawan@cityu.edu.hk
‡Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China
§Faculty of Economics, Kobe University, Rokkodai, Nada-ku, Kobe 657, Japan

Received: April 2004

Summary This paper enlarges on results of Wan and Zou [Journal of Econometrics 114 (2003), 165–96] on the choice of critical values for pre-test procedures based on the minimum risk criterion. We consider a modification of the general theorem given in Wan and Zou (2003) to obtain the optimal critical value that minimizes the risks of various inequality pre-test estimators of the regression error variance under a general class of first-order differentiable loss functions. Theoretical proofs of earlier numerical results are provided. This paper also presents results on the optimal pre-test critical values for the simultaneous estimation of the error variance and coefficient vector.

Keywords: Entropy loss, First-order differentiable, Inequality constraint, Lebesgue integrable, Regression variance.

1. INTRODUCTION

Econometricians frequently determine the ultimate specification of a model on the basis of preliminary tests on that model. The resulting estimators of the parameters in the new model are called pre-test estimators. Consequences of pre-test practice have been examined at length in the literature by Judge and Bock (1978), Giles and Giles (1993a), Magnus (1999), Danilov and Magnus (2004), among others. In all cases the sampling properties of pre-test estimators depend crucially on the critical value selected for the preliminary test. This suggests that in pre-testing one should choose a critical value that is optimal in some sense, rather than use the arbitrary 1% or 5% levels of significance. When the properties of pre-test estimators and in particular, issues on optimal pre-test significance levels have been studied, generally the analysis has been in terms of the

*Correspondence address: Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
†It is well known that pre-test estimators have poor properties. For example, they are not differentiable and hence inadmissible. The defence of continued research in this area is that practitioners persist on pre-testing despite the existence of such results because the analysis is easier to interpret if one takes an either–or stand on a model.

© Royal Economic Society 2006. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA, 02148, USA.
pre-test risk based on a particular loss function (e.g. Ohtani 1988). Standard in the pre-test literature is the quadratic loss, which leads to risk measures that incorporate the bias-variance tradeoff when assessing estimators’ performance. In recent years, many of the standard results in the pre-test literature have been reconsidered in the context of different loss functions. Some examples using absolute error loss, LINEX loss and balanced loss include the contributions of Giles (1993), Giles and Giles (1993b), Wan (1994a), Ohtani et al. (1997), among others. In a recent article, Wan and Zou (2003) proved a general result which suggests that for certain pre-testing problems, the same critical value may be optimal for all first-order differentiable loss functions provided that a mild technical condition is satisfied. Wan and Zou (2003) demonstrated the applicability of their general theorem in the context of the pre-test estimators of the error variance arising from a preliminary test of either exact linear restrictions on the regression coefficients or homogeneity of error variances. Necessary and sufficient conditions concerning the dominance of estimators for several specific loss functions have also been worked out.

In this paper, we extend the results of Wan and Zou (2003) by considering the pre-test of an inequality restriction on the coefficients, as opposed to equality restrictions as discussed in the previous paper. Optimal critical values minimizing the risks of various inequality pre-test estimators of the regression error variance under a general class of first-order differentiable loss functions are obtained. This particular pre-test problem considered has been examined in the literature previously under quadratic loss and LINEX loss (see Wan 1996; Geng and Wan 2000). It must be noted that while the technical arguments to be used here are similar to those developed in Wan and Zou (2003), their extension to the present context is however not straightforward due to the one-sided nature of the pre-test considered here. All the proofs are gathered in the Appendix.

2. MODEL FRAMEWORK AND ESTIMATORS

Consider the traditional linear statistical model
\[ y = X\beta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2 I_n\right), \] (1)

where \( y \) and \( \varepsilon \) are \( n \times 1 \) vectors, \( X \) is a non-stochastic \( n \times k \) matrix of full column rank, \( \beta \) is a \( k \times 1 \) vector of unknown coefficients and \( \sigma^2 \) is an unknown parameter. The (uncertain) prior information available to the investigator is represented by
\[ H_0 : C'\beta \geq r, \] (2)

2As stated in Zellner (1986), often a lot more thought should be given to the choice of a loss function, rather than to blindly trust in the traditional quadratic loss. Several authors have motivated other losses with reference to such applied problems as in dam construction, real estate appraisal, quality assurance and space travel. See Zellner (1973, 1986), Varian (1975) and Wan et al. (2000) for details. Zellner (1986) also gave an example to illustrate the risk consequence on using the optimal estimator relative to quadratic loss under LINEX loss.

3The overwhelming emphasis of the pre-test literature is on the estimation of the regression coefficients. In practice, the regression error variance is also of interest in respect of confidence interval construction or hypothesis testing. See Giles and Giles (1993a) for a convenient summary of the relevant literature on pre-test estimation of the regression error variance in linear regression.
where $C$ is a $k \times 1$ known vector and $r$ is a scalar. For purposes of analysis, (1) and (2) are reparameterized as

$$y = H\theta + \epsilon$$

and

$$H_o : \theta_1 \geq r_o,$$

respectively, where $H = XS^{-1/2} Q'$, $\theta = QS^{1/2} \beta$, $S = X'X$, $\theta_1$ is the first element of $\theta$, $r_o$ is a positive scalar multiple of $r$, and $Q$ is an orthogonal matrix such that $QS^{-1/2}C'(S^{-1}C)^{-1}C'S^{-1/2}Q' = \left( \begin{array}{c} 1 \\ 0 \end{array} \right)$.

The inequality restricted least squares estimators of $\theta$ and $\sigma^2$ are

$$\theta^{**} = I_{(-\infty,r_o)}(\tilde{\theta})\theta^* + I_{(r_o,\infty)}(\tilde{\theta})\tilde{\theta}$$

and

$$\sigma^{**2} = I_{(-\infty,r_o)}(\tilde{\theta})\sigma^{*2} + I_{(r_o,\infty)}(\tilde{\theta})\tilde{\sigma}^2,$$

respectively, where $\tilde{\theta} = H'y$ is the unrestricted least squares estimator of $\theta$, $\theta^* = (r_o, \tilde{\theta}_{(k-1)})$ is the equality restricted least squares estimator of $\theta'$, $\tilde{\theta}_1$ and $\tilde{\theta}_{(k-1)}$ are respectively the first and remaining $k - 1$ elements of $\tilde{\theta}$, $I(.)$ is an indicator function which equals one if the event inside the bracket occurs and zero otherwise, $\tilde{\sigma}^2 = \tilde{u}'\tilde{u}/(v + \gamma)$, $v = n - k$, $\tilde{u} = y - H\tilde{\theta}$, $\sigma^{*2} = u^*u^*/(v + 1 + \delta)$, and $u^* = y - H\theta^*$. Three common component estimators correspond to fixing $\delta = \gamma = 0$ for the least squares (LS) estimators, $\delta = k - 1$ and $\gamma = k$ for the maximum likelihood (ML) estimators and $\delta = \gamma = 2$ for the minimum mean square error (MM) estimators. Our subsequent analysis requires $\gamma \leq 1 + \delta$. Consider the test statistic

$$t = \frac{\sqrt{v}(\tilde{\theta}_1 - r_o)}{\tilde{\sigma} \sqrt{v + \gamma}} \sim t_{v,\tilde{\sigma}^2},$$

where $n_o^2 = \tau^2/(2\sigma^2)$ is the non-centrality parameter of the non-central $t$ distribution and $\tau = r_o - \theta_1$, which is non-positive if the inequality constraint is correct. Given the test statistic we may reject $H_o$ and use the unrestricted estimators if $t < c$, and not to reject $H_o$ and use the inequality restricted estimators if $t \geq c$, where $c \leq 0$ is the critical value of the central $t$ distribution for a given level of significance, $\alpha$. This leads to the inequality pre-test estimators

$$\hat{\theta} = I_{(-\infty,c)}(t)\tilde{\theta} + I_{(c,\infty)}(t)\theta^{**}$$

and

$$\hat{\sigma}^2 = I_{(-\infty,c)}(t)\tilde{\sigma}^2 + I_{(c,\infty)}(t)\sigma^{**2}.$$

Both $\hat{\theta}$ and $\hat{\sigma}^2$ have a discontinuity at $t = c$. Judge and Yancey (1981, 1986), Wan (1994b, 1996) and Geng and Wan (2000) discussed these estimators in detail. Briefly, in estimating $\theta$, pre-testing is sometimes the worst strategy in quadratic risk terms and is never the best; in contrast, there are estimators in $\hat{\sigma}^2$ that strictly dominate the unrestricted estimator in terms of both quadratic and LINEX risks; in some cases, the pre-test estimator of $\sigma^2$ can dominate both its components over the same regions of the parameter space.

© Royal Economic Society 2006
Clearly, by substituting (5) and (6) into (8) and (9) respectively, it follows that the inequality pre-test estimator can be written as a combination of the unrestricted and equality restricted estimators: if \( t < c \) or \( t \geq 0 \), the inequality pre-test estimator reduces to the unrestricted estimator, whereas if \( c \leq t < 0 \), the restricted estimator is used. Thus, the inequality pre-test estimator is logically similar to the equality pre-test estimator for the linear hypothesis \( \theta = r_0 \), but with a different decision rule. For the latter estimator, the decision rule is to use the unrestricted estimator if \( t < -c_1 \) or \( t > c_1 \), and the equality restricted estimator if \( -c_1 \leq t \leq c_1 \), where \( c_1 > 0 \) is the pre-test critical value for the equality hypothesis \( \theta = r_0 \). The paper by Wan and Zou (2003) has investigated the choice of optimal pre-test critical values for the case of the equality pre-test estimator. An important question is, given the close analogy between the equality and inequality restriction, can the general theorem established in Wan and Zou (2003, Section 2) be applied directly or through straightforward reparameterization to the case of an inequality pre-test estimator? The answer is no due to the one-sided nature of the hypothesis test for the inequality restriction. To bring out the salient points, let us rewrite the inequality pre-test estimator in (9) as

\[
\hat{\sigma}^2/\sigma^2 = \frac{\xi}{(v + \gamma)} + \left[\frac{(v + t^2)}{(v(v + 1 + \delta))} - 1/(v + \gamma)\right] I_{[0,c_1]}(t),
\]

where \( \xi = \hat{u}/\sigma^2 \sim X^2_v \), or

\[
\hat{\sigma}^2/\sigma^2 = \psi_1/(v + \gamma) + \left[\frac{(\psi_1 + \psi_2)}{(v + 1 + \delta)} - \psi_1/(v + \gamma)\right] I_{[0,c_1]}(t)
\]

by letting \( \psi_1 = \xi \) and \( \psi_2 = \xi t^2/v \). Equation (11) is in a form analogous to the following expression of the equality pre-test estimator (denoted as \( \hat{\sigma}^2_E \), here) given in Wan and Zou (2003, Equation 3.4) for the special case of a single equality restriction \( \theta = r_0 \):

\[
\hat{\sigma}^2_E/\sigma^2 = \psi_1/(v + \gamma) + \left[\frac{(\psi_1 + \psi_2)}{(v + 1 + \delta) - \psi_1/(v + \gamma)}\right] I_{[0,c_1]}(v \psi_2/\psi_1).
\]

Equation (12) is in fact crucial for Wan and Zou’s (2003) results. Their general theorem requires that the estimator is such that it can be written in the form of (12). Note that in the case of \( \hat{\sigma}^2_E \), because the underlying t-test is two sided, so we can write \( F = t^2 = v \psi_2/\psi_1 \) in (12), but in the case of \( \hat{\sigma}^2 \), the test is one sided, hence \( t \) in (11) cannot be expressed as a function of \( \psi_1 \) and \( \psi_2 \) because \( t \neq \sqrt{F} \) although \( F = t^2 \). In other words, in the current context, the results of Wan and Zou (2003) are not applicable directly or by straightforward reparameterization, and it is unclear how the properties of \( \hat{\sigma}^2 \) may be compared with those of \( \hat{\sigma}^2_E \).

3. OPTIMAL CRITICAL VALUES

3.1. A general theorem

Let \( \hat{\sigma}^q(q > 0) \) be an estimator of the parameter \( \sigma^q \). Consider the loss function,

\[
L(\hat{\sigma}^q/\sigma^q - 1),
\]

where \( L(.) \) is an arbitrary non-negative Lebesgue integrable function. Correspondingly, the risk of \( \hat{\sigma}^q \) is given by \( R(\hat{\sigma}^q) = E[ L(\hat{\sigma}^q/\sigma^q - 1) ] \). The following theorem provides a unified framework for obtaining the critical values that result in a minimum of the risks of some common pre-test estimators of \( \sigma^q \) under the loss structure (13).
Theorem 1  Let $W_1 > 0$ and $W_2$ be two continuous random variables, and $f_{W_2|W_1}(w_2)$ the conditional density function of $W_2$ given $W_1$. Consider a pre-test estimator of the form
\[
\hat{\sigma}^q / \sigma^q = l_1 W_1^{q/2} + \left[ l_2 \left( l_3 + W_2^{q/2} \right) - l_4 \right] W_1^{q/2} I_{(c,0)}(l_3 W_2),
\]
where $l_1$, $l_3$ and $l_4 \geq 0$, $l_2$ and $l_5 > 0$, $l_4/l_2 \geq l_5^{q/2}$ and $c \leq 0$ is the pre-test’s critical value. Suppose that
\[
\frac{\partial^j}{\partial c^j} \left\{ E_{W_1}[E_{W_2|W_1}(L(\hat{\sigma}^q / \sigma^q - 1))] \right\} = E_{W_1} \left\{ \frac{\partial^j}{\partial c^j} \left[ E_{W_2|W_1}(L(\hat{\sigma}^q / \sigma^q - 1)) \right] \right\}, \quad j = 1, 2
\]
then it can be shown that
\[
\frac{\partial R(\hat{\sigma}^q)}{\partial c} = E_{W_1} \left\{ [L(l_1 W_1^{q/2} - 1) - L(l_1 W_1^{q/2}) + (l_2 l_3 + c^2 l_5^{q/2}/l_4) W_1^{q/2} - 1)] f_{W_2|W_1}(c/l_5)/l_5 \right\}
\]
and
\[
\frac{\partial R(\hat{\sigma}^q)}{\partial c} \bigg|_{c=c^*} = 0,
\]
where $c^* = -l_5 \sqrt{(l_4/l_2)^{2/q} - I_3}$. Suppose also that $L(.)$ and $f_{W_2|W_1}(.)$ are first-order differentiable, then
\[
\frac{\partial^2 R(\hat{\sigma}^q)}{\partial c^2} \bigg|_{c=c^*} = q l_2^{2/q} l_5^{-2/q} \sqrt{(l_4/l_2)^{2/q} - I_3 E_{W_1} \left\{ L'(l_1 W_1^{q/2} - 1) W_1^{q/2} f_{W_2|W_1}(c^*/l_5) \right\}} / l_5^2,
\]
where $L'(.)$ is the derivative of $L(.)$.

The class of $L(.)$ contains commonly used loss functions such as quadratic loss, LINEX loss and entropy loss. Condition (15), which guarantees that limits and integrals can be interchanged, is crucial for the invariance of $c^*$. Depending on the sign of (18), $c^*$ may maximize or minimize the pre-test risk.

Clearly, (10) is in the form of (14) with $q = 2$, $l_1 = 1/(v + \gamma)$, $l_2 = 1/[v(v + 1 + \delta)]$, $l_3 = v$, $l_4 = 1/(v + \gamma)$, $l_5 = 1$, $W_1 = \xi$ and $W_2 = t$. Thus, the following result applies:

Corollary 1  The critical value $c^* = -\sqrt{v(1 + \delta - \gamma)/(v + \gamma)}$ (i.e., $c^* = -1$ for LS, $c^* = 0$ for ML and $c^* = -\sqrt{v/(v + 2)}$ for MM) gives rise to a stationary point in the risk of $\hat{\sigma}^2$ for any non-negative Lebesgue integrable loss function that satisfies (15) for $j = 1$.

Previously, this result was observed under quadratic loss and LINEX loss (Wan 1996; Geng and Wan 2000). In what follows the necessary and sufficient conditions for $c^*$ to minimize the risk of $\hat{\sigma}^2$ are given for three of the most commonly used loss functions in decision theoretic analysis.

Note that although dropping the requirement $W_2 > 0$ and allowing for a negative critical value is the main modification to the corresponding theorem given in Wan and Zou (2003), one cannot obtain Theorem 1 by mere reparameterization of the theorem in Wan and Zou (2003), as discussed in the previous section.
3.2. The case of quadratic loss

Under quadratic loss, $L(.) = (\hat{\sigma}^2/\sigma^2 - 1)^2$. It is readily seen that (15) is satisfied for quadratic loss, and accordingly $c^*$ is the critical value that results in a stationary point in the risk of $\hat{\sigma}^2$. In addition, we have the following theorem, the proof of which is omitted due to similarity to Theorem 1.

**Theorem 2** (a) Under quadratic loss, the condition $\delta < 2$ is necessary and sufficient for $c^*$ to result in a minimum in the risk of $\hat{\sigma}^2$ for all $\tau \geq 0$. (b) If $\delta \leq -(v - 1)/2$, then $\hat{\sigma}^2|_{c=c^*}$ dominates all other pre-test estimators in the region $\tau \geq 0$; if $-(v - 1)/2 < \delta \leq 2$, then $\hat{\sigma}^2|_{c=c^*}$ has the smallest risk in the class of $\hat{\sigma}^2$ with $c \in [c_Q, 0)$ over $\tau \geq 0$, where $c_Q = -\sqrt{v}[(v + 3)(v + \delta + 1)/(v + \gamma)(v + 2\delta - 1) - 1]$, which equals $c^*$ when $\delta = 2$ and is less than $c^*$ when $\delta < 2$.

The following remarks may be drawn from Theorem 2:

(i) The stationary point $c = c^*$ corresponds to a minimum turning point in the risk of $\hat{\sigma}^2$ over $\tau \geq 0$ for the LS component pre-test estimator, or for the ML component estimator when $k \leq 2$, but it cannot result in a minimum for the MM component, for which $\delta = 2$.

(ii) In the case of the LS component estimators, if $v = 1$, then $\hat{\sigma}^2_{\text{LS}}|_{c=c^*}$ is the minimum risk estimator and is preferred to both the LS component unrestricted and inequality restricted estimators in the region $\tau \geq 0$.

(iii) If $v > 1$, then $\hat{\sigma}^2_{\text{LS}}|_{c=c^*}$ has the smallest risk in the class of the LS component pre-test estimators with $c \in [c_Q, 0)$ for all $\tau \geq 0$.

(iv) When $k \leq 2$, there exists no ML component pre-test estimator with $c > c_Q$ that has smaller risk than the corresponding unrestricted estimator over $\tau \geq 0$.

(v) Note that $c_Q = c^*$ when $\delta = 2$. So, in the case of the MM component estimators (i.e., $c^* = -\sqrt{v}/(v + 2)$), $\hat{\sigma}^2_{\text{MM}}|_{c=c^*}$ has the minimum risk among the class of pre-test estimators with $c \in [-\sqrt{v}/(v + 2), 0)$ over the region $\tau \geq 0$.

In a broad sense, some of these results were noted earlier in the numerical work of Wan (1996), who showed for certain values of $c$, the risks of the LS and MM component pre-test estimators approach those of their corresponding unrestricted estimators from below, and among the pre-test estimators that dominate the unrestricted estimator, it is optimal to use $c = c^*$.

3.3. The case of LINEX loss

The LINEX loss function is given by $L(.) = \exp[a(\hat{\sigma}^2/\sigma^2 - 1)] - a(\hat{\sigma}^2/\sigma^2 - 1) - 1$. As in the case of the quadratic loss, it satisfies equation (15).

**Theorem 3** (a) Under the LINEX loss structure, provided that $v + \gamma - 2a > 0$, then $\delta < \delta_o = 2a(1 - \exp(-2a/(v + 3)))^{-1} - (v + 1)$ is both necessary and sufficient for $c^*$ to attain a minimum turning point in the risk of $\hat{\sigma}^2$ over the region of $\tau \geq 0$. (b) In the same region, if $\delta \leq \delta_o$, then $\hat{\sigma}^2|_{c=c^*}$ has the smallest risk in the class of pre-test estimators with $c \in [c_L, 0)$, where $c_L = -\sqrt{v}[2a(1 - \exp(-2a/(v + 3)))^{-1} - (v + \gamma)]/(v + \gamma)$, which is less than $c^*$ when $\delta < \delta_o$ and is equal to $c^*$ when $\delta = \delta_o$.

Theorem 3 leads to the following observations:

(i) Note that $-(v + 1) < \delta_o < 2$ when $a < 0$, and $\delta_o > 2$ when $a > 0$. It follows immediately that for $a > 0$, in the region $\tau \geq 0$, the LS and MM pre-test risks attain a relative minimum.
at $c = c^*$ and the estimators that use $c = c^*$ dominate other pre-test estimators with $c > c_L$; if $k < 1 + \delta_o$, then the same results hold also for the ML component estimators.

(ii) On the other hand, for $a < 0$, the outcomes in (i) are observed only when $\delta_o > 0$ and using the LS component estimators, or when using the ML components, $\delta_o > 0$ and $k = 1$, or $\delta_o > 1$ and $k = 2$. When using the MM components, however, since $\delta > \delta_o$, the pre-test risk cannot be minimized at $c = c^*$. These results corroborate the numerical findings of Geng and Wan (2000).

3.4. The case of entropy loss

For entropy loss, $L(.) = \hat{\sigma}^2 / \sigma^2 - 1 - \log(\hat{\sigma}^2 / \sigma^2)$. The idea of entropy loss in pre-test risk analysis was first discussed in Wan and Zou (2003). Now, it is easily seen that condition (15) is satisfied for risk among the LS component pre-test estimators with $\hat{c}$.

Let $\gamma$ be given in the following theorem: $\gamma = \exp(-\tau^2/(2\sigma^2)) / \sqrt{2\pi}$. Theorem 4 (a) For entropy loss, the necessary and sufficient condition for $c^*$ to result in a minimum turning point in the risk of $\hat{\sigma}^2$ in the region $\tau \geq 0$ is $\delta < 0$. (b) In the same region, if $\delta \leq 0$, then $\hat{\sigma}^2|_{c=c^*}$ has the smallest risk in the class of pre-test estimators with $c \in [c_E, 0]$, where $c_E = -\sqrt{v(1 - \gamma)}/(v + \gamma)$.

Thus, no pre-test risks based on the LS, ML or MM components can achieve a minimum turning point at $c = c^*$ over the region $\tau \geq 0$. On the other hand, $\hat{\sigma}^2|_{c=-1}$ has the minimum risk among the LS component pre-test estimators with $c \in [-1, 0]$. No corresponding result is observed in the cases of the ML and MM component estimators. Wan and Zou (2003) derived the risks of $\hat{\sigma}^2$ and $\sigma^2$ under an entropy loss function. The corresponding risk expressions of $\sigma^{**2}$ and $\delta^2$ are given in the following theorem:

**Theorem 5**

$$
R(\sigma^{**2}) = -\frac{\gamma}{v + \gamma'} + \log(v + \gamma') + \frac{\tau \exp(-\tau^2/(2\sigma^2))}{\sqrt{2\pi} (v + 1 + \delta)\sigma} \left\{ \frac{v + 1 + \tau^2/\sigma^2}{v + 1 + \delta} - \frac{v}{v + \gamma'} + \log(v(v + 1 + \delta)) - \log(v + \gamma) \right\} Pr(z < \tau/\sigma) - \int_0^\infty \log(\psi_1) f_{\psi_1}(\psi_1)d\psi_1 - \int_{-\infty}^0 \log(\psi_3^2 + v) f_{\psi_3}(\psi_3)d\psi_3 \quad (19)
$$

and

$$
R(\hat{\delta}^2) = -\frac{\gamma}{v + \gamma'} + \log(v + \gamma') + \{\log[v(v + 1 + \delta)] - \log(v + \gamma')\} Pr(c \leq t \leq 0) + \frac{\exp(-\tau^2/(2\sigma^2))}{\sqrt{2\pi}} \sum_{i=0}^\infty \frac{1}{i!} \left( \frac{\tau}{\sigma} \right)^i \Gamma\left( \frac{i + 1}{2} \right) \left[ \frac{v(v - 1 - \delta)}{(v + \gamma)(v + 1 + \delta)} \right] I_{(i + 1)}\left( \frac{i + 1}{2}, \frac{v}{v + 1 + \delta} \right) + \frac{(i + 1)}{(v + 1 + \delta)} I_{(i + 3)}\left( \frac{i + 3}{2}, \frac{v}{2} \right) - \int_0^\infty \log(\psi_1) f_{\psi_1}(\psi_1)d\psi_1 - \int_0^0 \log(\psi_3^2 + v) f_{\psi_3}(\psi_3)d\psi_3, \quad (20)
$$
Figure 1. Risks of estimators under entropy loss (LS component), $v = 15, k = 5$.

where $\psi_1 = \tilde{u}/\tilde{u}/\langle 2^* \rangle$, $\psi_3 \sim t_{v, 0^2}$, $f_{\psi_1}(\psi_3)$ is the density function of $\psi_1$; and $f_{\psi_1}(\psi_3)$ is defined analogously; $z = (\tilde{\theta} - \theta) \sim N(0, 1)$; $c_1 = c^2/(v + c^2)$ and $I_{c_1}(a_1, a_2) = \int_0^{a_1} p^{a_1-1}(1-p)^{a_2-1} dp / \int_0^{1} p^{a_1-1}(1-p)^{a_2-1} dp$ is the Incomplete Beta function.

Numerical evaluations of (19) and (20) as well as the risks of $\tilde{\sigma}^2$ and $\sigma^*^2$ have been carried out for $k = 5, 10, 15, 20, \alpha = 0.01, 0.05, 0.3, 0.5$ and the value that corresponds to $c = c^*$. The NAG routines, D01AME and D01AJE are used to evaluate the values of the cumulative Non-Central $t$ distribution as well the integrals in the risk expressions, while the routine S15ABE evaluates the values of the cumulative Normal distribution function. The Gamma and Incomplete Beta functions are evaluated via the routines GAMMQ and BETAI of Press et al. (1993). Figures 1–3 depict the risk results for $v = 15$ and $k = 5$. For ease of comparison, the risk of the unrestricted estimator is scaled to one in each case. It may be noted that the actual risks of $\tilde{\sigma}^2_{LS}$, $\tilde{\sigma}^2_{ML}$ and $\tilde{\sigma}^2_{MM}$ are 0.06815, 0.10583 and 0.07566 respectively in the figures. In other words, under entropy loss, the minimum mean square error component no longer yields the minimum risk estimator.

Figure 1 corroborates our analytical results that in the case of the LS component, a class of pre-test estimators strictly dominates the unrestricted estimator over $\tau \geq 0$, and in this class the estimator with $c = c^* = -1$ has the smallest risk. Over the region of the parameter space in which the constraint is true or nearly true, the estimator $\hat{\sigma}^2_{LS} = c^*$ typically has greater risk than the pre-test estimators with $\alpha = 0.01$ or 0.05 or the inequality constrained estimator; for moderate to large values of $\tau/\sigma$, however, the estimator $\hat{\sigma}^2_{LS} = c^*$ dominates all other estimators. On the other hand, in the case of the ML component, there also exist certain levels of pre-test such that the inequality pre-test estimator strictly dominates the unrestricted estimator, but $\hat{\sigma}^2_{ML}$ cannot dominate both the unrestricted and inequality restricted estimators simultaneously, as shown in Figure 2. In contrast, Figure 3 shows that for the MM component, pre-testing with $c = c^*$ yields an estimator with greater risk uniformly than the unrestricted estimator, and no pre-test estimator strictly dominates the unrestricted estimator. Comparing (19) and (20) with the risk expressions of $\tilde{\sigma}^2$ and $\sigma^*^2$ given
in Wan and Zou (2003), it is seen that $R(\hat{\sigma}^2) \rightarrow R(\tilde{\sigma}^2)$ as $c \rightarrow 0$, and $R(\hat{\sigma}_{**}^2) \rightarrow R(\sigma_{**}^2)$ as $c \rightarrow -\infty$.

4. IMPLICATIONS FOR THE COEFFICIENT PRE-TEST ESTIMATORS AND THE $t$-RATIO

Optimal critical values derived in the previous sections are for the error variance pre-test estimators. However, any investigator working with the linear regression model would typically

© Royal Economic Society 2006
be interested in the estimation of the coefficient vector also. Wan (1995) showed that if \( X \beta \) is estimated under quadratic loss and a mini–max regret criterion (Sawa and Hiromatsu 1973) is adopted then the optimal pre-test critical value equals approximately \(-1.12\) for moderate to large degrees of freedom. On the other hand, if the average relative risk criterion (Toyoda and Wallace 1976) is adopted, then the optimal critical value is exactly 0. These choices of critical values, of course, are not the same as the optimal critical values for the error variance pre-test estimators derived in Section 3. It would seem strange that one would want to use one pre-test critical value for estimating the coefficients, then another pre-test critical value for estimating the error variance. This raises the question of whether one can find a critical value that simultaneously optimizes the risk properties of both the coefficient vector and error variance pre-test estimators.

To address this question, one could consider varying \( \gamma \) and \( \delta \) such that the optimal critical values for estimating \( X \beta \) and \( \sigma^2 \) coincide. Specifically, let \( c_{o} \leq 0 \) be the optimal critical value for estimating \( X \beta \). Now, from the expression of \( c^* \) given in Corollary 1, by setting \( c^* = c_{o} \), we have

\[
\delta = (c_{o}^2 - 1) + \left(\frac{c_{o}^2}{v} + 1\right) \gamma, \tag{21}
\]

under which the optimal critical value for estimating \( X \beta \) is the same as that for estimating \( \sigma^2 \). One choice of \( \gamma \) and \( \delta \) that satisfies (21) is \( \gamma = 1 - c_{o}^2 \) and \( \delta = c_{o}^2 (1 - c_{o}^2)/v \). Now, if Wan’s (1995) mini–max regret criterion is adopted, then \( c_{o} = -1.12 \). Accordingly, \( \gamma = -0.2544 \), \( \delta = -0.3191/v \) and \( c^* = -1.12 \). Wan (1995) also showed that \( c_{o} = 0 \) minimizes the average relative risk of the pre-test estimator of \( X \beta \). One can conveniently choose \( \gamma = 0 \) and \( \delta = -1 \) to satisfy (21) for \( c_{o} = 0 \). Note that both \( \delta = -0.3191/v \) and \( \delta = -1 \) satisfy the conditions of Theorem 2 (i.e. \( \delta < 2 \)) and Theorem 4 (i.e. \( \delta < 0 \)) for \( c^* \) to result in a minimum in the risk of \( \delta^2 \) over \( \tau \geq 0 \) under quadratic loss and entropy loss. For LINEX loss, provided that \( v + \gamma > 2a \), the corresponding condition in Theorem 3 is also satisfied for \( a > 0 \). It should be noted that these choices of \( \gamma \) and \( \delta \) do not result in any of the three component estimators discussed previously, but when \( v \gg 1 \), they do lead to an estimator which approximately corresponds to the LS component estimator.

As suggested by a referee it is also interesting to relate the above results to a function of the two estimators, say the \( t \)-ratio. Let us consider the class of estimators \( \{a'\hat{\theta}/\hat{\sigma}, -\infty < c < \infty\} \) for \( a'\hat{\theta}/\sigma \), where \( a \) is a known vector. For purposes of analysis, we write \( \hat{\theta} = H'\gamma = \theta + H'\varepsilon \) and \( \theta^* = \hat{\theta} + (r_{o} - \hat{\theta})e_{1} \), where \( e_{1} = (1, 0, \ldots, 0)' \). From (5) and (8) we have,

\[
\hat{\theta} = \hat{\theta} + (\theta^* - \hat{\theta})I_{e_{1},0}(t)
= \theta + H'\varepsilon - (\sigma t\sqrt{\xi/(v)})e_{1}I_{e_{1},0}(t). \tag{22}
\]

Thus,

\[
a'\hat{\theta}/\hat{\sigma} = \frac{(a'\hat{\theta} + a'H'\varepsilon)/\sigma - (t\sqrt{\xi/(v)})a'I_{e_{1},0}(t)}{\sqrt{\xi/(v + \gamma) + [(v + \tau^2)/(v(v + 1 + \delta))] - 1/(v + \gamma)]\xi I_{e_{1},0}(t)}. \tag{23}
\]

Due to similarity, we shall only consider the case of \( a = e_{1} \), for which \( a'\hat{\theta}/\hat{\sigma} = \hat{\theta}_{1}/\hat{\sigma} \) and

\[
\hat{\theta}_{1}/\hat{\sigma} = \frac{r_{o}/\sigma + (t\sqrt{\xi/(v)})(1 - I_{e_{1},0}(t))}{\sqrt{\xi/(v + \gamma) + [(v + \tau^2)/(v(v + 1 + \delta))] - 1/(v + \gamma)]\xi I_{e_{1},0}(t)}. \tag{24}
\]

© Royal Economic Society 2006
The risk of $\hat{\theta}_1/\hat{\sigma}$ is given by

$$ R(\hat{\theta}_1/\hat{\sigma}) = E_\xi\{E_{t_{\xi}}[L(\hat{\theta}_1/\hat{\sigma} - \theta_1/\sigma)]\} $$

$$ = E_\xi\left[\int_{-\infty}^{c} L\left(r_o/\sigma + t\sqrt{1/v}/\sqrt{\xi}/(v + \gamma) - \theta_1/\sigma\right) f_{t_{\xi}}(t)dt + \int_{c}^{0} L\left(r_o/\sigma\right)/\sqrt{(v + t^2\xi)/(v(v + 1 + \delta)) - \theta_1/\sigma\right) f_{t_{\xi}}(t)dt + \int_{\infty}^{c} L\left(r_o/\sigma + t\sqrt{1/v}/\sqrt{\xi}/(v + \gamma) - \theta_1/\sigma\right) f_{t_{\xi}}(t)dt\right], \quad (25) $$

where $f_{t_{\xi}}(.)$ is the density of $t$ given $\xi$. Therefore,

$$ \partial R(\hat{\theta}_1/\hat{\sigma})/\partial c = E_\xi\left[L\left((r_o/\sigma + c\sqrt{1/v}/\sqrt{\xi}/(v + \gamma) - \theta_1/\sigma\right) - L\left(r_o/\sigma\right)/\sqrt{(v + c^2\xi)/(v(v + 1 + \delta)) - \theta_1/\sigma\right]\right] f_{t_{\xi}}(c) \cdot \quad \quad (26) $$

From the expression in (26) it is clear that when $\gamma = 1 + \delta$, a stationary point of the risk would result at $c = 0$. This will not happen, however, when $\gamma \neq 1 + \delta$, then there exists no value of $c$ independent of the unknowns that would lead to a stationary point in the risk. Focusing now on the case of $\gamma = 1 + \delta$, we investigate if $c = 0$ leads to a minimum in the risk of $\hat{\theta}_1/\hat{\sigma}$ along the same lines as in previous sections of the paper. In the case of quadratic loss we have

$$ \partial^2 R(\hat{\theta}_1/\hat{\sigma})/\partial c^2|_{c=0} = 2e^{-v^2/2\sqrt{v + \gamma}/(\sqrt{2\pi v}\sigma)}\left[r_o\sqrt{v + \gamma} - \theta_1\sqrt{2\Gamma((v + 1)/2)}/\Gamma(v/2)\right]. \quad \quad (27) $$

Denote $\gamma_o = 2\Gamma^2((v + 1)/2)/\Gamma^2(v/2) - v$. Then $\gamma_o < 0$. So from (27), if $r_o > (\ell)0$, then by choosing $\gamma > (\ell)\gamma_o$, $c = 0$ will minimize the risk of $\hat{\theta}_1/\hat{\sigma}$ over $\tau \geq 0$. If $r_o = 0$, then $c = 0$ will minimize the risk of $\hat{\theta}_1/\hat{\sigma}$ over $\tau > 0$.

5. CONCLUSIONS

Our paper has provided the proofs of some of the earlier results in the relevant literature of inequality pre-test estimation. It develops a unified framework of analysing the risk properties of the inequality pre-test estimator of the regression variance under a general class of loss functions. Our proofs show that pre-testing can be the preferred strategy (as compared with using its component estimators) under some situations. At best, it dominates both its constituent estimators in the same region of the parameter space under a range of commonly adopted loss functions. Situations are also found in which pre-test estimators can dominate the unrestricted estimator in a wide region of the parameter space. Our proofs are important since pre-testing is often regarded as undesirable in econometric and statistical practice because of the procedures’ discontinuity. Extension to other pre-test problems is clearly desirable since the framework developed here

\[5\text{Of the three losses considered, only the quadratic loss is appropriate for assessing the risk of } \hat{\theta}_1/\hat{\sigma}. \text{ The entropy loss cannot be used as } \theta_1/\hat{\sigma} \text{ can take on zero value; the LINEX loss is also not appropriate as it can lead to infinite risk of } \hat{\theta}_1/\hat{\sigma} \text{ when } r_o \neq 0. \]
is general, and enables analysis under all first-order differentiable losses. This paper has also considered the $t$-ratio and presented results that make it coincide for the optimal pre-test critical values for simultaneous estimation of the error variance and coefficient vector.

Finally, there are other smoother pre-test estimators, which are continuous functions of the decision rule. One example is the weighted average least squares (WALS) estimator (Magnus and Durbin 1999; Magnus 2002; Danilov and Magnus 2004) defined by a weighted average of the unrestricted and restricted estimators. So far, most interest has been focused on the use of random weights in the estimator. If one assumes non-random weight then it is possible to find the weight that minimizes the risk of the WALS estimator using an approach similar to the general theorem presented in Section 2. It is found, however, that the optimal weight is dependent on unknown parameters. The precise details can be obtained from the authors.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the helpful comments from Professor Stéphane Grégoir and the referee, and financial support from the Hong Kong Research Grant Council and the National Natural Science Foundation of China (Grant Nos. 10471043, 70221001).

REFERENCES


**APPENDIX**

**Proof of Theorem 1**

Note that from (14),

\[
R(\hat{\sigma}^q) = E[L(\hat{\sigma}^q/\sigma^q - 1)]
\]

\[
eq E_{W_1}\left[E_{W_2|W_1}\left[L\left(l_1W_1^{q/2} + \left[l_2(l_3 + W_2^{q/2} - l_4W_1^{q/2}I_{[c,0]}(l_5W_2) - 1\right]\right]\right]\right]
\]

\[
eq E_{W_1}\left[\int_{l_5}^{c/l_5} L\left(l_1W_1^{q/2} + \left[l_2(l_3 + w_2^{q/2} - l_4W_1^{q/2} - 1\right]W_1^{q/2} - 1\right)f_{w_2|W_1}(w_2)dw_2 + \int_{-\infty}^{c/l_5} L\left(l_1W_1^{q/2} - 1\right)f_{w_2|W_1}(w_2)dw_2 + \int_{0}^{\infty} L\left(l_1W_1^{q/2} - 1\right)f_{w_2|W_1}(w_2)dw_2\right].
\] (A.1)
Therefore,

\[ \frac{\partial R(\hat{\theta})}{\partial c} = E_{W_1}\left\{ -L\left(l_1W_1^{q/2} + l_2(l_3 + c^2/l_3)^{q/2} - l_3\right)W_1^{q/2} - 1\right\} f_{W_2|W_1}(c/l_3)/l_5 + L\left(l_1W_1^{q/2} - 1\right) f_{W_2|W_1}(c/l_3)/l_5 \right\}, \]  

(A.2)

which equals zero when \( c = c^* \). Straightforwardly, equation (18) can be verified using (A.2).

**Proof of Theorem 3**

It is seen from (10) and (16) that

\[ \frac{\partial R(\hat{\theta})}{\partial c} = E_t \left\{ \left[ L\left(\frac{\xi}{\tau(v + \gamma)} - 1\right) - L\left(\frac{(v + c^2)\xi}{\tau(v + 1 + \delta)}\right) - 1\right]\right\} f_{t|g}(c) \].  

(A.3)

Making use of (A.3), it is straightforward to show that

\[ \frac{\partial R(\hat{\theta})}{\partial c} = E_t \left\{ \exp(-a)\left(\exp\left(\frac{a\xi}{v + \gamma}\right) - \exp\left(\frac{a(v + c^2)\xi}{\tau(v + 1 + \delta)}\right)\right) \right\} \]

\[ - a\left(\frac{1}{v + \gamma} - \frac{v + c^2}{\tau(v + 1 + \delta)}\right)\xi f_{t|g}(c) \}

(A.4)

for LINEX loss. Using the mean value theorem in calculus, we can write

\[ \exp\left(\frac{a\xi}{v + \gamma}\right) - \exp\left(\frac{a(v + c^2)\xi}{\tau(v + 1 + \delta)}\right) = a\xi \exp(\theta_0a\xi)\left(\frac{1}{v + \gamma} - \frac{v + c^2}{\tau(v + 1 + \delta)}\right), \]

where \( \theta_0 \) lies between \( 1/(v + \gamma) \) and \( (v + c^2)/(\tau(v + 1 + \delta)) \). Now, if \( c^* < c < 0 \), then \( c^2 < c^*^2 \). So,

\[ \frac{v + c^2}{\tau(v + 1 + \delta)} < \theta_0 < \frac{1}{v + \gamma} \]

(A.6)

when \( c^* < c \). Thus,

\[ \exp\left(\frac{a\xi}{v + \gamma}\right) - \exp\left(\frac{a(v + c^2)\xi}{\tau(v + 1 + \delta)}\right) > a\xi \exp\left(\frac{a(v + c^2)\xi}{\tau(v + 1 + \delta)}\right)\left(\frac{1}{v + \gamma} - \frac{v + c^2}{\tau(v + 1 + \delta)}\right). \]

(A.7)

Making use of (A.4) and (A.7), it is clear that

\[ \frac{\partial R(\hat{\theta})}{\partial c} > a\left[\frac{1}{v + \gamma} - \frac{v + c^2}{\tau(v + 1 + \delta)}\right] E_t \left\{ \exp(-a)\xi \exp\left(\frac{(v + c^2)a\xi}{\tau(v + 1 + \delta)}\right) f_{t|g}(c) - \xi f_{t|g}(c) \right\}. \]

(A.8)

Note that the joint density function of \((t, \xi)\) is

\[ f(t, \xi) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)\sqrt{2\pi}\nu} \xi^{(\nu-1)/2} \exp\left\{ -\frac{\xi t^2}{2\nu} - \frac{\nu t^2}{\sigma^2} \right\}, -\infty < t < \infty, \xi > 0. \]

(A.9)
Hence,
\[ f_{\xi|c}(c) = \sqrt{\frac{\xi}{2\pi v}} \exp \left[ -\frac{1}{2} \left( \frac{\xi c + \tau}{\sqrt{v} \sigma} \right)^2 \right] . \] (A.10)

So for \( p > 0 \) and \( \theta \) satisfying \( a \theta < (v + c^2)/2v \), we can write
\[ E_\xi \left[ \xi^p \exp(\partial \alpha \xi) f_{\xi|c}(c) \right] = E_\xi \left[ \xi^p \exp(\partial \alpha \xi) \sqrt{\frac{\xi}{2\pi v}} \exp \left( -\frac{1}{2} \left( \frac{\xi c + \tau}{\sqrt{v} \sigma} \right)^2 \right) \right] \]
\[ = \frac{\exp(-\tau^2/(2\sigma^2))}{2^{v/2} \Gamma(v/2) \sqrt{2\pi v}} \int_0^\infty t^{p+(v-1)/2} \exp \left( -\left( \frac{v + c^2}{2v} - a \theta \right) t \right) \sum_{i=0}^\infty \frac{1}{i!} \left( -\frac{\tau}{\sigma} \right)^i t^{i/2} dt \]
\[ = \frac{v^{p+v/2} \exp(-\tau^2/(2\sigma^2))}{\Gamma(v/2) \sqrt{2\pi}} \sum_{i=0}^\infty \frac{1}{i!} \left( -\frac{\tau}{\sigma} \right)^i \frac{2^{p+(v+1)/2} \Gamma(p+(v+1)/2)}{(v+c^2-2v\sigma \theta)^{p+(v+1)/2}} . \] (A.11)

Applying (A.11) in (A.8), we obtain
\[ \frac{\partial R(\hat{\sigma}^2)}{\partial c} > a \left[ \frac{1}{v + \gamma} - \frac{v + c^2}{v(v+1+\delta)} \right] \frac{v^{1+v/2}}{\Gamma(v/2) \sqrt{2\pi}} \exp \left( -\tau^2/(2\sigma^2) \right) \sum_{i=0}^\infty \frac{1}{i!} \left( -\frac{\tau}{\sigma} \right)^i \]
\[ \times \frac{2^{(v+3)/2} \Gamma((v+3+i)/2)}{(v+c^2)(v+3+i)/2} \left[ \exp(-a) \left( 1 - \frac{2a}{v+1+\delta} \right)^{-(v+3+i)/2} \right] . \] (A.12)

Now, note that for \( i = 1, 2, 3, \ldots \)
\[ a \left[ \exp(-a) \left( 1 - \frac{2a}{v+1+\delta} \right)^{-(v+3+i)/2} \right] > a \left[ \exp(-a) \left( 1 - \frac{2a}{v+1+\delta} \right)^{-(v+3)/2} - 1 \right] \] (A.13)
and the right-hand side of (A.13) is non-negative when \( \delta \leq \delta_0 \). Making use of these results and (A.12), we observe that when \( \delta \leq \delta_0 \),
\[ \frac{\partial R(\hat{\sigma}^2)}{\partial c} > 0 \quad \text{for} \quad c^* < c < 0 . \] (A.14)

On the other hand, if \( c_L < c < c^* \), then \( c^2 > c^* \). Hence,
\[ \frac{1}{v + \gamma} < \theta_o < \frac{v + c^2}{v(v+1+\delta)} . \] (A.15)

Using (A.15) in conjunction with (A.5), we have,
\[ \exp \left( \frac{a \xi}{v + \gamma} \right) - \exp \left( \frac{a(v+c^2)\xi}{v(v+1+\delta)} \right) < a \xi \exp \left( \frac{a \xi}{v + \gamma} \right) \left( \frac{1}{v + \gamma} - \frac{v + c^2}{v(v+1+\delta)} \right) . \] (A.16)
Thus, from (A.4), (A.11) and (A.16), we obtain,
\[
\frac{\partial R(\hat{\delta}^2)}{\partial c} < a \left( \frac{1}{v + \gamma} - \frac{v + c^2}{v(v + 1 + \delta)} \right) E_\xi \left[ \exp(-a) \frac{a \xi}{v + \gamma} f_{\tilde{f}}(c) - \xi f_{\tilde{f}}(c) \right] \\
= a \left( \frac{1}{v + \gamma} - \frac{v + c^2}{v(v + 1 + \delta)} \right) \frac{v^{1+v/2}}{\Gamma(v/2)\sqrt{2\pi}} \exp(-\tau^2/(2\sigma^2)) \sum_{i=0}^{\infty} \frac{1}{i!} \left( -\frac{\tau c}{\sigma} \right)^i \\
\times 2^{(\nu+3i)/2} \Gamma \left( \frac{\nu+3i+1}{2} \right) \left( \frac{v+c^2}{v(v+1+\delta)} \right) \frac{v^{1+v/2}}{\Gamma(v/2)\sqrt{2\pi}} \exp(-\tau^2/(2\sigma^2)) \sum_{i=0}^{\infty} \frac{1}{i!} \left( -\frac{\tau c}{\sigma} \right)^i \\
\leq a \left( \frac{1}{v + \gamma} - \frac{v + c^2}{v(v + 1 + \delta)} \right) \frac{v^{1+v/2}}{\Gamma(v/2)\sqrt{2\pi}} \exp(-\tau^2/(2\sigma^2)) \sum_{i=0}^{\infty} \frac{1}{i!} \left( -\frac{\tau c}{\sigma} \right)^i \\
\times 2^{(\nu+3i+1)/2} \Gamma \left( \frac{\nu+3i+1}{2} \right) \left( \frac{v+c^2}{v(v+1+\delta)} \right) \frac{v^{1+v/2}}{\Gamma(v/2)\sqrt{2\pi}} \exp(-\tau^2/(2\sigma^2)) \sum_{i=0}^{\infty} \frac{1}{i!} \left( -\frac{\tau c}{\sigma} \right)^i. \\
\] (A.17)

Clearly, the last expression in (A.17) is smaller than or equal to zero for $c_L < c < c^*$, which implies that,
\[
\frac{\partial R(\hat{\delta}^2)}{\partial c} < 0 \quad \text{for} \ c_L < c < c^*. \\
(A.18)
\]

Recognizing (A.14) and (A.18), it becomes immediately obvious that if $\delta \leq \delta_0$, then $\hat{\delta}^2|_{c=c^*}$ is the minimum risk estimator in the class of pre-test estimators with $c \in (c_L, 0)$ when $\tau \geq 0$.

Now, to prove the remainder of the theorem, we consider the case of $\delta \geq \delta_0$. If $\delta = \delta_0$, then it can be shown that
\[
\frac{\partial R(\hat{\delta}^2)}{\partial c} \bigg|_{\tau=0} > 0 \quad \text{for} \ c < c^*. \\
(A.19)
\]

In fact, $c^2 > c^{*2}$ when $c < c^*$. So, when $c < c^*$,
\[
\frac{1}{v + \gamma} < \theta_0 < \frac{v + c^2}{v(v + 1 + \delta)}, \\
(A.20)
\]

which implies (A.7) also holds when $c < c^*$. Thus,
\[
\frac{\partial R(\hat{\delta}^2)}{\partial c} \bigg|_{\tau=0} > a \left[ \frac{1}{v + \gamma} - \frac{v + c^2}{v(v + 1 + \delta)} \right] \frac{v^{1+v/2}}{\Gamma(v/2)\sqrt{2\pi}} \frac{2^{3/2} \Gamma((v+3)/2)}{(v+c^2)^{(v+3)/2}} \\
\times \left[ \exp(-a) \left( 1 - \frac{2a}{v + 1 + \delta} \right)^{-(v+3)/2} - 1 \right] = 0. \\
(A.21)
\]

Similarly, one can show that if $\delta > \delta_0$, then,
\[
\frac{\partial R(\hat{\delta}^2)}{\partial c} \bigg|_{\tau=0} > 0 \quad \text{for} \ c < c^* \quad \text{for} \ c^* < c < c_L. \\
(A.22)
\]

(Naturally, for the second inequality, $c_L$ is assumed to be well defined). In other words, the pre-test risk is maximized at $c = c^*$ when $\tau = 0$. It is clear from (A.19) and (A.22) (or the first inequality in (A.22)) that
\( \delta < \delta_o \) is the necessary and sufficient condition for \( c^* \) to result in a minimum in the risk of \( \hat{\sigma}^2 \) for all non-negative \( \tau \) values. This completes the proof of Theorem 3.

**Proof of Theorem 5**

The risks of \( \tilde{\sigma}^2 \) and \( \sigma^{**} \) are derived in Wan and Zou (2003). To show the derivation of the risk of \( \sigma^{**} \), note that \( u^* u^* = \bar{u} \bar{u} + (\sigma z - \tau)^2 \). So we can write,

\[
\frac{\sigma^{**}}{\sigma^2} = \frac{\tilde{\sigma}^2}{\sigma^2} + \left( \frac{\sigma^{**}}{\sigma^2} - \frac{\tilde{\sigma}^2}{\sigma^2} \right) I(\tau < \tau/(\sigma^2)) \tag{A.23}
\]

or

\[
\frac{\sigma^{**}}{\sigma^2} = \frac{\tilde{\sigma}^2}{\sigma^2} + \left( \frac{\sigma^{**}}{\sigma^2} - \frac{\tilde{\sigma}^2}{\sigma^2} \right) I(\tau < \tau/(\sigma^2)) \tag{A.24}
\]

Now,

\[
\int_{-\infty}^{\tau/(\sigma)} (z - \tau/(\sigma))^2 \frac{\exp(-z^2/2)}{\sqrt{2\pi}} dz = \frac{\tau \exp(-\tau^2/(2\sigma^2))}{\sqrt{2\pi} \sigma} + (1 + \tau^2/\sigma^2) \Pr(z < \tau/\sigma). \tag{A.25}
\]

Making use of (A.23)–(A.25), we obtain,

\[
E\left( \frac{\sigma^{**}}{\sigma^2} \right) = \frac{\nu}{\nu + \gamma} + \frac{\nu(\gamma - 1 - \delta)}{(\nu + \gamma)(\nu + 1 + \delta)} \Pr(z < \tau/\sigma) + \frac{\tau \exp(-\tau^2/(2\sigma^2))}{(\nu + 1 + \delta)\sqrt{2\pi} \sigma} \\
+ \frac{(1 + \tau^2/\sigma^2)}{\nu + 1 + \delta} \times \Pr(z < \tau/\sigma) \\
= \frac{\nu}{\nu + \gamma} + \left( \frac{\nu + \tau^2/\sigma^2}{\nu + 1 + \delta} - \frac{\nu}{\nu + \gamma} \right) \Pr(z < \tau/\sigma) + \frac{\tau \exp(-\tau^2/(2\sigma^2))}{(\nu + 1 + \delta)\sqrt{2\pi} \sigma} \tag{A.26}
\]

and

\[
E\left[ \log \left( \frac{\sigma^{**}}{\sigma^2} \right) \right] = E(\log(\xi)) + E\left\{ \log \left( \frac{1}{\nu + \gamma} + \frac{(\gamma - 1 - \delta)}{(\nu + \gamma)(\nu + 1 + \delta)} + \frac{\tau^2}{\nu(v + 1 + \delta)} \right) I(\tau < \tau/(\sigma)) \right\} \\
= E(\log(\xi)) - \log(\nu + \gamma) \Pr(\tau > 0) + \int_{-\infty}^{0} \log \left( \psi_3^2 + v \right) f_{\psi_3}(\psi_3)d\psi_3 \\
- \log [\nu(\nu + 1 + \delta)] \Pr(\tau < 0) \\
= - \log(\nu + \gamma) \Pr(z > \tau/\sigma) - \log [\nu(\nu + 1 + \delta)] \Pr(z < \tau/\sigma) \\
+ \int_{-\infty}^{\infty} \log(\psi_1) f_{\psi_1}(\psi_1) d\psi_1 + \int_{-\infty}^{0} \log \left( \psi_1^2 + v \right) f_{\psi_1}(\psi_1)d\psi_1. \tag{A.27}
\]

Equation (19) follows by combining (A.26) and (A.27).
Next, we derive the risk of \( \hat{\sigma}^2 \). Using (10), it is straightforward to show that,

\[
E \left[ \log \left( \frac{\hat{\sigma}^2}{\sigma^2} \right) \right] = -\log(v + \gamma) \left[ 1 - \Pr(c \leq t \leq 0) \right] - \log(v(v + 1 + \delta)) \Pr(c \leq t \leq 0) \\
+ \int_{0}^{\infty} \log(\psi_1) f_{\psi_1}(\psi_1) d\psi_1 + \int_{c}^{0} \log(\psi_3^2 + v) f_{\psi_3}(\psi_3) d\psi_3.
\]

(A.28)

Also, making use of (10) and (A.9), and along the lines of Geng and Wan (2000), it can be shown that,

\[
E \left( \frac{\hat{\sigma}^2}{\sigma^2} \right) = \frac{v}{v + \gamma} + \frac{\exp(-\tau^2/(2\sigma^2))}{2^{v/2} \Gamma(v/2) \sqrt{2\pi v}} \int_{0}^{\infty} \frac{\phi(t)}{i!} \left[ \frac{v^i (\gamma - (1 + \delta)) (v + \gamma) (v + 1 + \delta)}{(v + \gamma)(v + 1 + \delta)} \right] \phi^{(i)}(t) d\phi dt
\]

\[
= \frac{v}{v + \gamma} + \frac{\exp(-\tau^2/(2\sigma^2))}{\sqrt{2\pi}} \sum_{i=0}^{\infty} \left( \frac{\tau/\sigma}{} \right)^i \frac{1}{i!} \left\{ \frac{v^i (\gamma - 1 - \delta)}{(v + \gamma)(v + 1 + \delta)} \right\} \]

\[
\times I_{c1} \left( \frac{i + 1}{2}, \frac{v + 2}{2} \right) + \frac{i + 1}{v + 1 + \delta} I_{c1} \left( \frac{i + 3}{2}, \frac{v}{2} \right). \tag{A.29}
\]

Equation (20) thus follows and the proof is completed.