AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

CANTER

9203

Department of Economics
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND

ISSN 1171-0705

GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS
LIBRARY
WITHDRAWN

MAY 01 1992

PRE-TEST ESTIMATION AND TESTING IN ECONOMETRICS: RECENT DEVELOPMENTS

Judith A. Giles
and
David E. A. Giles

Discussion Paper

No. 9203

This paper is circulated for discussion and comments. It should not be quoted without the prior approval of the author. It reflects the views of the author who is responsible for the facts and accuracy of the data presented. Responsibility for the application of material to specific cases, however, lies with any user of the paper and no responsibility in such cases will be attributed to the author or to the University of Canterbury.

Department of Economics, University of Canterbury
Christchurch, New Zealand

Discussion Paper No. 9203

March 1992

PRE-TEST ESTIMATION AND TESTING IN ECONOMETRICS: RECENT DEVELOPMENTS

Judith A. Giles
and
David E. A. Giles

PRE-TEST ESTIMATION AND TESTING IN ECONOMETRICS : RECENT DEVELOPMENTS*

Judith A. Giles

and

David E.A. Giles

Department of Economics

University of Canterbury

March, 1992

* This paper was prepared while the authors were receiving support in the form of a University of Canterbury Research Grant, # 1770901. We are grateful to Les Oxley and Alan Wan for their comments on an earlier draft, and to the latter for supplying us with material for Figure 4.

CONTENTS

1. INTRODUCTION

1.1 Background Discussion

1.2 Historical Setting

2. PRINCIPAL RESULTS

2.1 Pre-Testing Linear Restrictions in Regression

2.2 Homoscedasticity Pre-Test Regression Estimators

2.2.1 Estimation of the Scale Parameter
2.2.2 Estimation of the Coefficient Vector

2.3 Another Pre-Test Estimator of the Error Variance

2.4 The Choice of Significance Level

3. ROBUSTNESS OF PRE-TEST ESTIMATORS

3.1 Mis-Specification of the Regressors

3.2 Non-Normal Regression Errors

3.3 Other Forms of Model Mis-Specification

4. PRE-TESTING WITH INEQUALITY RESTRICTIONS

5. EXACT DISTRIBUTIONS OF PRE-TEST ESTIMATORS

6. PRE-TEST TESTING

6.1 Introduction

6.2 Testing the Coefficient Vector After a Pre-Test of the Error Variance

6.3 Testing the Error Variance After a Pre-Test on the Coefficient Vector

6.4 Autocorrelation Pre-Test Testing

7. OTHER DEVELOPMENTS

7.1 Autocorrelation Pre-Test Estimation

7.2 Multi-Stage Pre-Testing

7.3 Alternative Loss Functions

7.4 Other Pre-Test Problems

8. CONCLUSIONS

1. INTRODUCTION

1.1 Background Discussion

In applied econometrics it is generally apparent that the researcher has undertaken a "search" for the preferred specification of the model, or for the appropriate estimator to use. Sometimes this strategy is made explicit and it may have been undertaken in a systematic way. In other cases there is only a vague impression that the final results are not the only ones that were generated during the course of the analysis. Most economists who use econometric tools are aware that "mining" the data may be distortive in some sense and that the end results may not be what they appear to be.

Consider some simple but common examples of this sort of sequential econometric analysis. First, suppose that the following regression model has been fitted to the data by Ordinary Least Squares (OLS):

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \quad . \tag{1}$$

Then, to determine the "significance" of x_2 in the model, a t-test is conducted. If the usual t-ratio exceeds the tabulated critical value (for the chosen significance level) then x_2 is deemed to be a "significant" regressor and it is retained in the model. On the other hand, if x_2 is "insignificant" it is deleted from the model, which is then re-estimated (effectively by Restricted Least Squares (RLS)). So, the final specification of the model depends on the outcome of a prior test and the estimates of the coefficients of the other variables in the model also depend on the outcome of this test. In addition, the properties of any further tests that may be conducted are affected by the way in which the model's specification was determined.

As a second example, consider the estimation of (1) by OLS. Then, suppose the Durbin-Watson statistic is computed to test whether or not the model's errors are serially independent. If this hypothesis cannot be rejected, the OLS estimates of the coefficients are retained. However, if serial independence is rejected the model is re-estimated (perhaps using the Cochrane-Orcutt (CO) estimator), and different coefficient estimates are obtained. Again, the estimates that are finally reported depend on the outcome of a "preliminary test", and if (for example) one then tests the significance of a regressor in the usual way, the "t-statistic" is no longer t-distributed.

Both of these examples are realistic, though they over-simplify the situation because in practice a sequence of such tests might be adopted. However, they capture the essential feature of "pre-test" strategies in econometrics - the choice of estimator, and the final estimates, depend on the outcome of a *random event*. The probability of choosing OLS or RLS estimation in the first example, or of choosing OLS or CO estimation in the second example, depends on the significance level for the pre-test, as well as on the test's power. In effect, the estimator that generates the reported estimates is a stochastic mixture of two (in these examples) "component" estimators. In general this Pre-Test Estimator (PTE) will differ from each of its components in the sense that it will have a different sampling distribution, and so generally its bias and precision will also differ. Pre-testing generally affects the sampling properties of the estimators that we use. For instance, in the first example given above the PTE is biased unless $\beta_2=0$. These effects are often complicated and depend on the unknown parameters in the model.

Pre-test *testing* is also widely practised in econometrics, and the consequences of such strategies are of considerable interest. In certain rather special cases, two successive econometric tests may be independent. Then, the effect of the first test on the properties of the second can be controlled, and this may have implications for the extent to which there is a pre-test testing "problem". Two further examples may be helpful.

Consider a sequence of "nested" models or hypotheses, such as when we take a multiple regression model, which successively delete regressors in such a way that each model in the sequence can be obtained from its predecessor by the deletion of one or more regressors, and we attempt to find the most parsimonious "significant" model specification. It is well known (*e.g.*, Mizon (1977)) that, with such a nesting, the size of the test of restrictions on the coefficients at any stage can be controlled (against the overall maintained hypothesis) if each null is tested against the previously accepted null in the sequence. Without such nesting, size control is not generally possible, and the resulting size distortion is due to the inherent pre-test nature of the analysis. Similarly, if the hypotheses are nested but the researcher tests each successive null against any alternative other than the immediately preceding null, then a pre-test problem arises.

Finally, consider the familiar Chow test for the structural stability of a regression coefficient vector. It is well known that the validity of

the standard form of this test relies, among other things, on the homoscedasticity of the error variance across the two subsamples. When this assumption is violated, we have a form of the famous Behrens-Fisher problem and alternative approximate tests are available. So, there is a strong motivation to pre-test for homogeneity of the variance prior to conducting the test of primary interest (the test for structural stability). In contrast to the preceding example, however, in this case the form of the second test depends on the outcome of the pre-test. The standard pre-test statistic in this case is an F-ratio, and it is readily shown (*e.g.*, Phillips and McCabe (1983)) that this statistic is independent of the Chow test statistic. Consequently, the *true* size of the second stage Chow test is the product of the *nominal* sizes chosen for this test and for the pre-test. Judicious choices of these nominal sizes can be used to control the true second stage size for the Chow test itself, as desired. However, even with this independence there remains a pre-test testing issue as an alternative test would be used only if the pre-test null was rejected. We explain this further in Section 6.1.

Typically non-independence of successive test statistics is the norm in sequential hypothesis testing in econometrics, and in these cases there are pre-test testing issues to be addressed, as we shall see in Section 6. Specifically, how are the (true) size and the power of a second test affected by the presence of a pre-test?

Although sequential inference alters the properties of estimators and tests, the commonly held view, that pre-testing is intrinsically "bad", need not be true, as we shall see. It should also be emphasised that these effects arise purely because of the introduction of a randomisation process, and not because the same data are being used more than once. Indeed, even if the sample is split and one part is used for preliminary testing and the other part for subsequent estimation, the pre-test effects that we have described still arise[1]. Also, if a sequence of pre-tests is used, the randomisation process becomes increasingly complicated, especially if the various tests are not independent of each other. In such cases it becomes difficult to determine the properties of the resulting estimators.

The above examples will be familiar to anyone who has undertaken applied econometric research. Indeed, pre-testing is probably the norm,

---

[1]For example, see Toyoda and Wallace (1979).

rather than the exception, in applied econometrics and so it not surprising that a sizeable literature on this topic has developed. While much of this literature does not attempt to offer *prescriptions* about what researchers should or should not do in their empirical work, it does provide a considerable amount of information about the likely *consequences* of pre-testing in econometrics. Accordingly, the recent contributions in this field are of practical interest to many economists, and the purpose of this paper is to highlight some of the major themes that have emerged recently in the pre-testing literature, at least with respect to econometric (as opposed to purely statistical) analysis. We also attempt to extract practical recommendations as far as possible to assist applied workers.

1.2 Historical Setting

The seminal contribution of Bancroft (1944) can be taken as the starting point for the analysis of pre-testing problems that are of direct interest to economists. Motivated in part by the earlier remarks of Berkson (1942), that there was a need to investigate the statistical properties of sequential estimation strategies, Bancroft analysed two pre-testing problems which set the scene for much of the subsequent work in this field.

The first is as follows. Suppose that we draw two independent samples from two Normal populations, each with possibly different variances. These variances can be estimated separately from the respective samples. We can also test for variance homogeneity, and if this hypothesis is accepted then there is good reason to "pool" the two samples and estimate the (common) population variance from the combined data. In particular this "pooled" estimator may be more efficient than one based on a single sample. This suggests a pre-test strategy : test for variance homogeneity and either pool the data, or don't, according to the outcome of the prior test. Bancroft derived exact analytic expressions for the bias and variance (and hence Mean Squared Error (MSE)) of this PTE, for the case where the pre-test involves a one-sided alternative. For economists, the interest in this result is that it can be translated into a regression problem where one is testing for a specific type of heteroscedasticity of the errors, and pooling subsamples of data prior to estimating the error variance only if the errors are thought to be homoscedastic. This in turn suggests another related pre-test problem to which we return later : after pre-testing for this type of homoscedasticity, what are the sampling properties of the regression coefficient estimator?

The second problem considered by Bancroft was explicitly regression oriented, and was essentially that discussed above in relation to equation (1). To reiterate, suppose a regression model with two regressors has been fitted by OLS, and the model is either retained or simplified by deleting one of the regressors, depending on the outcome of a t-test. Bancroft derived an exact analytic expression for the bias of this pre-test estimator. Subsequent authors considered the estimator's second moment, and extended the analysis to the more realistic case of the multiple regression model.

Much of the basic work on pre-test problems of this type, or problems of "inference based on conditional specification", as it was referred to, was undertaken by Bancroft in collaboration with his colleagues and students. Bancroft and Han (1977) provide an annotated bibliography of this early work, though many of the problems considered are not of direct interest to econometricians. A second bibliography is given by Han *et al.* (1988).

Pre-test problems with an explicit econometric content were taken up subsequently by a series of researchers. In particular, work by Dudley Wallace and a series of graduate students at the University of North Carolina led to several seminal developments, and helped to raise interest in this field among a number of Japanese researchers. Other path-breaking contributions came from George Judge, Thomas Yancey, Mary Ellen Bock and their associates and students at the University of Illinois, Purdue University, and later at Berkeley. The range of econometric pre-test problems that has now been analysed is extensive, but in many cases they are essentially variants of those first considered by Bancroft (1944).

In the next section the key results from the econometrics pre-testing literature are stated and summarised briefly, as a background to a more systematic discussion of the major themes that have emerged recently in research in this field. These are taken up in Sections 3 to 7, and some concluding remarks appear in Section 8. Each sub-section includes some summary comments to assist the reader who is concerned primarily with the practical implications of the results under discussion.

2. PRINCIPAL RESULTS

2.1 Pre-Testing Linear Restrictions in Regression

As in the rest of this paper, we concentrate here on "conventional"

PTE's that is, ones whose components are the traditional ones that have been used in empirical econometric analysis. In particular, while recognising their interest and importance, we do not consider Bayesian PTE's or ones whose components are of Stein-like form. For a discussion of these related alternatives, the reader is referred to Judge and Bock (1978, 1983), Vinod and Ullah (1981) and Judge *et al.* (1985).

Much of the econometrics pre-test literature is based on a natural generalisation of Bancroft's second problem. This involves the linear multiple regression model,

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

where y and e are (T×1); X is (T×k), non-stochastic and of rank k; and β is (k×1). In addition, prior information suggests m (<k) exact independent linear constraints on the regression coefficients

$$R\beta = r \tag{3}$$

where R is (m×k), r is (m×1), and both are known and non-stochastic. We will consider the estimation of both β and σ^2. Let $\delta = R\beta - r$ be the error in the prior information. This situation is commonly encountered in applied econometrics, except usually the researcher is uncertain of the accuracy of the prior beliefs. Accordingly, the procedure usually followed in practice is to (pre-)test the validity of the restrictions and if the outcome of the pre-test suggests that they are correct then the model's parameters are estimated incorporating the restrictions. If the pre-test rejects the prior information then the parameters are estimated from the sample information alone. Prior to considering the properties of such pre-test estimators of β and σ^2 we will briefly review the estimators which ignore the restrictions (the "unrestricted" estimators) and those which assume the restrictions are correct (the "restricted" estimators).

The unrestricted OLS (and maximum likelihood) estimator of β is well known to be $b = S^{-1}X'y$, where $S = (X'X)$ and $b \sim N\left(\beta, \sigma^2 S^{-1}\right)$. Consequently its risk under squared error loss (the sum of the MSE's of each individual element of b) is $\rho(\beta,b) = E\left((b-\beta)'(b-\beta)\right) = \sigma^2 tr(S^{-1})$. From the Gauss-Markov theorem we know that b is the best linear unbiased estimator (BLUE). It is minimax and, among the class of unbiased estimators, minimises risk under quadratic loss.

A best (minimum variance) quadratic unbiased estimator of σ^2 is the usual least squares estimator, given by $\tilde{\sigma}^2_L=(y-Xb)'(y-Xb)/v$, where $v=(T-k)$, and $\rho(\sigma^2,\tilde{\sigma}^2_L)=2\sigma^4/v$. If we allow the estimator of σ^2 to be biased then the estimator of σ^2 with smallest MSE is $\tilde{\sigma}^2_M=(y-Xb)'(y-Xb)/(v+2)$, and its risk is

$\rho(\sigma^2,\tilde{\sigma}^2_M)=2\sigma^4/(v+2)$. The maximum likelihood estimator of σ^2 is $\tilde{\sigma}^2_{ML}=(y-Xb)'(y-Xb)/T$ and $\rho(\sigma^2,\tilde{\sigma}^2_{ML})=(2v+k^2)\sigma^4/T^2$.

Imposing the restrictions in (3), we estimate β by $b^*=b+S^{-1}R'[RS^{-1}R']^{-1}(r-Rb)$, which, with $D=S^{-1}R'[RS^{-1}R']^{-1}RS^{-1}$, has a risk under squared error loss of $\rho(\beta,b^*)= \sigma^2 tr\left(S^{-1}-D\right) + tr\left(S^{-1}R'[RS^{-1}R']^{-1}\delta\delta'[RS^{-1}R']^{-1}RS^{-1}\right)$. b^* is unbiased if and only if the restrictions are correct ($\delta=0$). Further, as D is at least positive semi-definite, $var(b^*_i)\leq var(b_i)$, $i=1,2,\ldots,k$, and within the class of linear estimators of β, b^* is BLUE in this case.

The corresponding restricted estimators of σ^2 are $\sigma^{*2}_L=(y-Xb^*)'(y-Xb^*)/(v+m)$, $\sigma^{*2}_M=(y-Xb^*)'(y-Xb^*)/(v+m+2)$, and $\sigma^{*2}_{ML}=(y-Xb^*)'(y-Xb^*)/T$. Now $\left((y-Xb^*)'(y-Xb^*)/\sigma^2\right)\sim\chi^{2'}_{v+m;\lambda}$ where the non-centrality parameter, λ, defined by $\lambda=\delta'[RS^{-1}R']^{-1}\delta/2\sigma^2$, is a measure of the validity of the linear restrictions. If the restrictions are true, $\delta=0$ and so $\lambda=0$. It is straightforward to show that $\rho(\sigma^2,\sigma^{*2}_L) = 2(2\lambda^2+4\lambda+v+m)\sigma^4/(v+m)^2$, $\rho(\sigma^2,\sigma^{*2}_M) = 2(2\lambda^2+v+m+2)\sigma^4/(v+m+2)^2$, and $\rho(\sigma^2,\sigma^{*2}_{ML}) = \left(2(m+v+4\lambda)+(m-k+2\lambda)^2\right)\sigma^4/T^2$.

Several studies[2] have considered the conditions under which the risk of b dominates that of b^* and *vice versa*. These conditions are generally data specific, as their risks depend on X. This limits the generality of any comparisons based on quadratic risk and so to avoid this complication we will, as others have, concentrate on the conditional forecast of y rather than on β itself. This is equivalent to assuming orthonormal regressors (*i.e.* $X'X = I_k$) in the β space. So, though similar conclusions are drawn from comparing the risk functions, the mapping from the conditional mean (or orthonormal regressors) case to that of considering the unweighted risk of estimators of β (i.e. nonorthonormal regressors) is *not* direct and is significantly more complicated[3]. The risk of Xb, the unrestricted estimator of E(y), is

---

[2] For instance, see Toro-Vizcarrondo and Wallace (1968), Wallace and Toro-Vizcarrondo (1969), Wallace (1972), Goodnight and Wallace (1972), Yancey *et al.* (1973), Bock *et al.* (1973). See also Judge and Bock (1978) for a summary and a discussion.

[3] See, for instance, Wallace (1972), Brook (1972, 1976), Bock *et al.* (1973), Yancey *et al.* (1973), and Judge and Bock (1978). Brook (1972, 1976), Bock *et al.* (1973), and Judge and Bock (1978) also consider the unweighted risk function of the pre-test estimator of β itself.

$$\rho\left(E(y),Xb\right) = E\left((Xb-X\beta)'(Xb-X\beta)\right) = E\left((b-\beta)'X'X(b-\beta)\right)$$
$$= \sigma^2 k, \tag{4}$$

while that of the restricted estimator, Xb*, is

$$\rho\left(E(y),Xb^*\right) = \sigma^2(k-m+2\lambda). \tag{5}$$

Comparing (4) and (5) we see that the risk of Xb* is less than or equal to that of Xb if $\lambda \leq m/2$.

Similarly, there is a λ-range over which the risk of the restricted estimator of σ^2 is less than or equal to that of the unrestricted estimator. The values of λ, λ_j^*, (j=L,M,ML) for which the risks are equal depend on the estimation method, but it is readily shown that $\lambda_j^* \neq m/2$ and so, if the researcher desired the minimum risk estimators of E(y) and σ^2, there will be some λ-range over which his strategy should be to use the restricted estimator of σ^2 but the unrestricted estimator of E(y). This suggests considering a joint risk function for E(y) and σ^2, something which has not been pursued in the literature.

We now consider the situation where the researcher undertakes a pre-test of the validity of the restrictions. Traditionally,

$$H_0 : \delta=0 \quad \text{vs} \quad H_1 : \delta\neq 0$$

is tested using the Wald (and Lagrange Multiplier) statistic

$$\ell = \left((Rb-r)'[RS^{-1}R']^{-1}(Rb-r)v\right) / \left(m(y-Xb)'(y-Xb)\right).$$

If H_0 is correct, the test statistic ℓ has a central F distribution with m and v degrees of freedom, $F_{(m,v)}$. If one or more of the restrictions are invalid, ℓ has a non-central F distribution with m and v degrees of freedom and non-centrality parameter λ. We reject H_0 if $\ell > F^{\alpha}_{(m,v)} = c$, where the critical value, c, is determined for a given significance level of the test α, by $\int_0^c dF_{(m,v)} = \Pr.\left(F_{(m,v)} \leq c\right) = (1-\alpha)$. This is a UMPI size-$\alpha$ test of the validity of the restrictions. If H_0 is rejected we use the unrestricted estimators of E(y) and σ^2. If $\ell \leq c$, we assume the restrictions are correct and use the restricted estimators of E(y) and σ^2. So, the estimators of E(y) and σ^2 actually reported are the PTE's

$$\hat{Xb} = \begin{cases} Xb & \text{if } \ell > c \\ Xb^* & \text{if } \ell \leq c \end{cases}, \tag{6}$$

and

$$\hat{\sigma}_j^2 = \begin{cases} \sigma_j & \text{if } \ell > c \\ \sigma_j^{*2} & \text{if } \ell \leq c \end{cases} , \tag{7}$$

j=(L, M, ML). It is useful to rewrite (6) and (7) as

$$X\hat{b} = I_{[0,c]}(\ell)Xb^* + I_{(c,\infty)}(\ell)Xb \tag{8}$$

and

$$\hat{\sigma}_j^2 = I_{[0,c]}(\ell)\sigma_j^{*2} + I_{(c,\infty)}(\ell)\tilde{\sigma}_j^2 \tag{9}$$

where $I_{(.,.)}(\ell)$ is an indicator function which takes the value unity if ℓ falls within the subscripted range and zero otherwise. From (8) and (9), it is clear that the PTE's are functions of the data, the hypothesis, and the significance level of the test. Representing a PTE in this way highlights the difficulty of deriving its sampling properties; it is the sum of two parts, both of which are composed of products of non-independent random variables.

Bancrofts' second problem is a special case of the above - he considers a single "zero" restriction when k=2, and derives the bias of $\hat{b}_1$. Toro-Vizcarrondo (1968) derives the MSE of $\hat{b}_1$ for this estimator, and Brook (1972, 1976) generalizes these results, by deriving the unweighted risks of the pre-test estimators of β and E(y) for the general multiple restrictions problem, as outlined here. Sclove *et al.* (1972) also derive the risk of the PTE of β in the orthonormal regressor model and Bock *et al.* (1973) extend their analysis to the non-orthonormal case[4].

The risk of $X\hat{b}$, under squared error loss, is

$$\rho\left(E(y),X\hat{b}\right) = \sigma^2\left(k+(4\lambda-m)P_{20}-2\lambda P_{40}\right) , \tag{10}$$

where[5]

$$P_{ij} = \Pr.\left[F'_{(m+i,v+j;\lambda)} \leq \left(cm(v+j)\right)/\left(v(m+i)\right)\right] \; ; \; i,j=0,1, \ldots . \tag{11}$$

Figure 1 illustrates typical risk functions of Xb, Xb*, and $X\hat{b}$ (for $c\epsilon(0,\infty)$). Some features are:

(a) If the restrictions are valid the pre-test risk is less than that of the unrestricted estimator but higher than that of the restricted estimator.

---

[4]These results are discussed, for instance by Wallace and Asher (1972), Wallace (1977), Judge and Bock (1978), who also further generalise this research, and Judge *et al.* (1985).

[5]$F'_{(m+i,v+j;\lambda)}$ is a non-central F-statistic with m+i and v+j degrees of freedom, and non-centrality parameter λ, defined above.

Intuitively, if $\lambda=0$, the PTE will lead us to use the restricted estimator $100(1-\alpha)\%$ of occasions but $100\alpha\%$ of the time we will erroneously ignore the prior information.

(b) $\rho\left(E(y),Xb\right)=\rho\left(E(y),X\hat{b}\ |c\in(0,\infty)\right)$ occurs for a value of λ, $\lambda_1\in[m/4,m/2]$. So, for $\lambda\in[0,\lambda_1)$ the risk of the PTE, $X\hat{b}$, is less than that of the unrestricted estimator, Xb, but higher than that of the restricted estimator, Xb*, while for $\lambda\in(\lambda_2,\infty)$, $X\hat{b}$ has smaller risk that that of Xb* but is dominated by Xb, where λ_2 is the value of λ such that $\rho\left(E(y),Xb^*\right)=\rho\left(E(y),X\hat{b}\right)$. For $\lambda\in(\lambda_1,\lambda_2)$, $X\hat{b}$ has higher risk than that of *both* Xb and Xb*. Thus, pre-testing is *never* the preferable strategy.

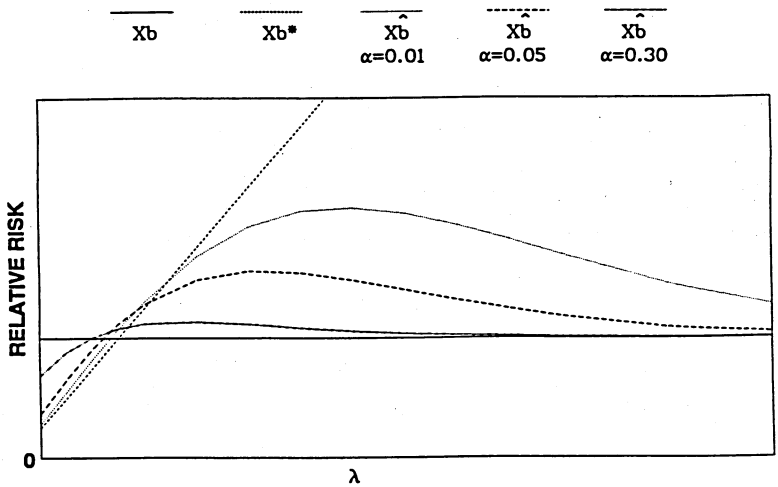

Figure 1. Relative risk functions for Xb, Xb*, and $X\hat{b}$.

(c) For finite c, as δ increases, the risk of $X\hat{b}$ monotonically increases to a maximum, which occurs at a value of $\lambda>\lambda_2$, it then monotonically decreases and as $\lambda\to\infty$, $\rho\left(E(y),X\hat{b}\right)\to\rho\left(E(y),Xb\right)$. Intuitively, when the prior information is so wrong that λ is very large, then pre-testing will lead us to do the right thing: to ignore the restrictions.

(d) The smaller α is (the larger c is), the closer $\rho\left(E(y),X\hat{b}\right)$ is to $\rho\left(E(y),Xb^*\right)$ as a smaller test size increases the probability of accepting

the null hypothesis. This results in a risk gain in the region to the left of λ_1 but at the cost of a (possibly) much higher risk for relatively large λ. An analogous argument can be made for large α. Clearly, from (8), if c is chosen to be zero (infinity), we always reject (accept) the hypothesis, and the PTE degenerates to the unrestricted (restricted) estimator.

(e) Of the estimators considered, no one strictly dominates any of the others. Cohen (1965) proves, under certain assumptions and a squared error loss function, that the PTE is inadmissible. Basically, this is because the estimator is a discontinuous function of the test statistic, ℓ, with a single jump at ℓ=c. Nevertheless, practitioners continue to report the conventional PTE and so, given the lack of dominance of either $X\hat{b}$, Xb, or Xb* and the fact that λ is rarely known, the next obvious question to ask is "is there an 'optimal' pre-test estimator?". The answer will certainly depend on the definition of 'optimal' but, more importantly, it will be linked to the choice of test size. This question is taken up below[6].

We now consider the PTE of σ^2, which has received less attention in the literature than have the PTE's of β and E(y). As σ^2 is often regarded as a nuisance parameter, this is perhaps not surprising. However, an estimator of σ^2 is often used as a measure of the model's "goodness of fit" and if one is interested in forming standard errors, prediction or confidence intervals or undertaking certain hypothesis tests, after pre-testing, then the PTE of σ^2 needs to be investigated. The risk functions of $\hat{\sigma}_j^2$, are derived by Clarke *et al.* (1987a,b). They depend on the data only through T, k, m and λ. Figure 2 depicts typical risk functions for $\tilde{\sigma}_{ML}^2$, σ_{ML}^{*2}, and $\hat{\sigma}_{ML}^2$. The following points may be noted:

(a) As the pre-test size increases, we reject the hypothesis more frequently, and so the risk of the PTE approaches that of the unrestricted estimator. This has the effect of decreasing the maximum risk of $\hat{\sigma}_j^2$ but at the expense of increasing its minimal risk value. A converse argument can be given for a decrease in the test size.

(b) When using the ML components pre-testing is never the preferred strategy, and it can be the worst alternative.

---

[6]Though we do not discuss the Stein-rule family of estimators, it is worth noting that the above analysis has also been considered, by, for example, Judge, Yancey and Bock (1983), using Stein-rule estimators as the component estimators. They show that if (k-m)≥3 then their Stein PTE dominates, under squared error loss, the traditional PTE.

(c) Among the component estimators of σ^2 considered, under a minimax criterion with respect to risk, those based on the principle of minimum mean squared error are preferable when constructing the PTE of σ^2. Clarke *et al.* (1987b) show that the PTE $\hat{\sigma}^2_M$, though composed of the minimum MSE unrestricted and restricted (when H_0 is true) estimators of σ^2, is not itself best invariant.

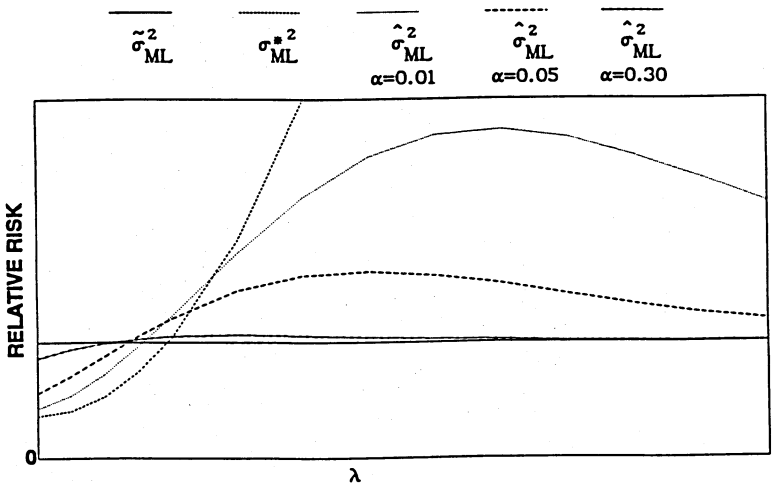

Figure 2. Relative risk functions for $\tilde{\sigma}^2_{ML}$, σ^{*2}_{ML}, and $\hat{\sigma}^2_{ML}$.

(d) The risk of the restricted estimator is smaller than that of the unrestricted estimator and of the PTE when the restrictions are true. The restricted estimator continues to dominate the pre-test and unrestricted estimators for $\lambda \in [0,\lambda^*_j)$. However, as the hypothesis error grows and approaches infinity, the risk of the restricted estimator is unbounded, while the pre-test risk approaches that of the unrestricted estimator.

For certain values of c the risks of $\hat{\sigma}^2_L$ and $\hat{\sigma}^2_M$ approach that of the unrestricted estimator from below. That is, the PTE can strictly dominate the unrestricted estimator. This feature, which is noted by Ohtani (1988a), contrasts with the results found when estimating E(y) (or β) after a

pre-test for linear restrictions. It does, however, also occur when estimating the error variance after a pre-test for homogeneity, in the two sample model, as we shall see below. Ohtani (1988a), extends the work of Clarke *et al.* (1987b), by deriving the improved estimator of the variance proposed by Stein (1964), which dominates the unrestricted estimator, $\tilde{\sigma}^2_M$. He shows that this estimator, say $\hat{\sigma}^2_S$, is in fact a PTE with a critical value equal to v/(v+2). Using numerical evaluations Ohtani proposes that $\hat{\sigma}^2_S$ has the minimum risk among the pre-test estimators which dominate the unrestricted estimator. This result is proved by Gelfand and Dey (1988a)[7].

Giles (1991a) shows that a similar result holds when using the least squares component estimators. Then there exists a family of PTE's with $c \in (0,1]$ which strictly dominate the unrestricted estimator, and it is optimal to use c=1. She also suggests, from her numerical evaluations, that when $m \leq 2$ the PTE which uses c=1 strictly dominates *both* of its component estimators. Giles (1990) shows that there is no corresponding case when using the maximum likelihood estimators. Then the smallest risk results from either using the unrestricted or the restricted estimator.

Clarke (1986, 1990) derives and analyses the PTE of the "standard error of estimate", σ, after a preliminary test of linear restrictions on the coefficients; this PTE, say $\hat{\sigma}$, is not equal to $(\hat{\sigma}^2)^{1/2}$. We will not discuss this research here; it suffices to say that the results are found to be *qualitatively* similar whether one is estimating σ^2 or σ.

From a practical viewpoint, applied researchers should be aware that pre-test estimation of the regression coefficients, after a test of restrictions on the coefficient vector, is *never* better than estimating the model without a prior test. Indeed, from a MSE viewpoint it may be the *worst* of the three basic strategies that can be adopted. The same is true if the maximium likelihood estimator of the error variance is obtained after the pre-test, and there is some advantage in using the minimum MSE estimator instead. Finally, when using the least squares PTE of the error variance, it is best to set the pre-test critical value to unity, especially if only one or two restrictions are being tested.

---

[7]See also Gelfand and Dey (1988b).

2.2 Homoscedasticity Pre-test Regression Estimators

Frequently we wish to estimate models for which we suspect that the assumption of a scalar error covariance matrix is invalid. For example, the errors may be autocorrelated or the observations may be drawn from different populations, which may result in different error variances. Then, least squares is generally an unbiased but inefficient estimator of the coefficient vector; the generalized least squares estimator (GLS) is minimum variance unbiased. So, we might test for the presence of a non-spherical error covariance matrix prior to estimating the model. Here we consider one such case, that of pre-testing for homogeneity of the error variances in the two-sample linear model. This is a natural extension of Bancroft's first problem. The other obvious case of pre-testing for autocorrelated errors is discussed in Section 7.1. So, we assume that model (2) involves two samples, with T_1 and T_2 observations ($T_1+T_2=T$) respectively:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}, \quad \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N\left[0, \begin{bmatrix} \sigma_1^2 I_{T_1} & 0 \\ 0 & \sigma_2^2 I_{T_2} \end{bmatrix} \right] \qquad (12)$$

where y_i and e_i are $(T_i \times 1)$, X_i is $(T_i \times k_i)$, β_i is $(k_i \times 1)$ and $k_i < T_i$, $i = 1,2$. In Section 2.2.1 we consider the estimation of σ_1^2, given the uncertainty about whether the second sample comes from the same population as the first. We examine the estimation of the coefficient vector, assuming that $\beta_1=\beta_2=\beta$, in Section 2.2.2.

2.2.1 Estimation of the Scale Parameter

If the variances are equal then the two samples may be pooled and an unbiased estimator of σ_1^2 is $s_A^2=(v_1 s_1^2+v_2 s_2^2)/(v_1+v_2)$ where $v_i=T_i-k_i$, $s_i^2=(y_i-X_i b_i)'(y_i-X_i b_i)/v_i$, $b_i=S_i^{-1}X_i' y_i$, $S_i=(X_i' X_i)$, $i=1,2$. We call s_A^2 the always-pool estimator of σ_1^2. Conversely, if the variances are unequal, an unbiased estimator of σ_1^2 is $s_N^2=s_1^2$. We call s_N^2 the never-pool estimator.

The usual procedure, to decide which estimator of σ_1^2 to use, is to undertake a preliminary test of the hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad vs \quad H_1 : H_0 \text{ not true.}$$

The alternative hypothesis can be one- or two-sided depending on the researcher's prior beliefs. A test statistic for homoscedasticity is $J=s_2^2/s_1^2$ (or $J^*=s_1^2/s_2^2$, depending on H_1), with $f(J)=\phi^{-1}f\left(F_{(v_2,v_1)}\right)$, where

$F_{(v_2,v_1)}$ is a central F variate with v_2 and v_1 degrees of freedom. $\phi=(\sigma_1^2/\sigma_2^2)$ is a measure of the hypothesis error. Assuming for simplicity the one-sided alternative $H_1:\sigma_1^2<\sigma_2^2$, we accept H_0 if $J\le F^{\alpha}_{(v_2,v_1)}=c$ where the critical value of the test, c, satisfies $\int_0^c dF_{(v_2,v_1)}=\Pr.\left(F_{(v_2,v_1)}\le c\right)=(1-\alpha)$, for a size-$\alpha$ test.

If the outcome of the pre-test suggests that the variances are equal ($J\le c$) then we estimate σ_1^2 using the always-pool estimator s_A^2, while we employ the never-pool estimator s_N^2 if $J>c$; that is, when we reject H_0. After such a (pre-)testing procedure, the estimator of σ_1^2 actually reported is the PTE

$$s_P^2 = \begin{cases} s_N^2 & \text{if} \quad J > c \\ s_A^2 & \text{if} \quad J \le c, \end{cases} \quad , \tag{13}$$

The risks of s_N^2, s_A^2, and s_P^2 are[8]

$$\rho(\sigma_1^2,s_N^2) = 2\sigma_2^4\phi^2/v_1. \tag{14}$$

$$\rho(\sigma_1^2,s_A^2) = \sigma_2^4\left(\phi^2(2v_1+v_2^2)-2v_2^2\phi+v_2(v_2+2)\right)/(v_1+v_2)^2. \tag{15}$$

$$\rho(\sigma_1^2,s_P^2) = \rho(\sigma_1^2,s_N^2)+\sigma_2^4\left(\phi^2v_2\left[2v_1(v_1+v_2)Q_{02}-(v_1+2)(v_2+2v_1)Q_{04}\right]+2v_1v_2\phi\right.$$
$$\left.\cdot\left[v_1Q_{22}-(v_1+v_2)Q_{20}\right]+v_1v_2(v_2+2)Q_{40}\right)/\left(v_1(v_1+v_2)^2\right) \quad , \tag{16}$$

where, for i,j=0,1,2,...,

$$Q_{ij} = \Pr.\left[F_{(v_2+i,v_1+j)} < \left(v_2(v_1+j)c\right)/\left(v_1(v_2+i)\right)\right]. \tag{17}$$

Bancroft's first problem is equivalent to the one described here with an appropriate re-definition of the degrees of freedom. He derives the bias and variance of s_P^2 and finds that the bias of s_P^2 is smaller than that of s_A^2 when ϕ is close to zero: that range of ϕ where the bias of s_A^2 is highest. We recall that the always-pool estimator is only unbiased when the variances

---

[8]See, for instance, Bancroft (1944) or Toyoda and Wallace (1975).

are equal. Intuitively, the pre-test is leading us to follow the correct path when ϕ is small - reject the null. From the MSE comparisons Bancroft finds that the pre-test which uses c=1 results in a MSE equal to or smaller than that of s_N^2 for all possible values of ϕ - that is, this PTE *strictly dominates* the never-pool estimator. These estimators are considered further by Toyoda and Wallace (1975).

Figure 3 illustrates typical risk functions for s_N^2, s_A^2, and s_P^2, for various values of $c\in(0,\infty)$. Note that when c=0 we always reject the hypothesis and so, $\rho(\sigma_1^2,s_P^2)=\rho(\sigma_1^2,s_N^2)$. Conversely, $\rho(\sigma_1^2,s_P^2)=\rho(\sigma_1^2,s_A^2)$ when $c=\infty$, so that we always accept the hypothesis. This figure highlights the following points:

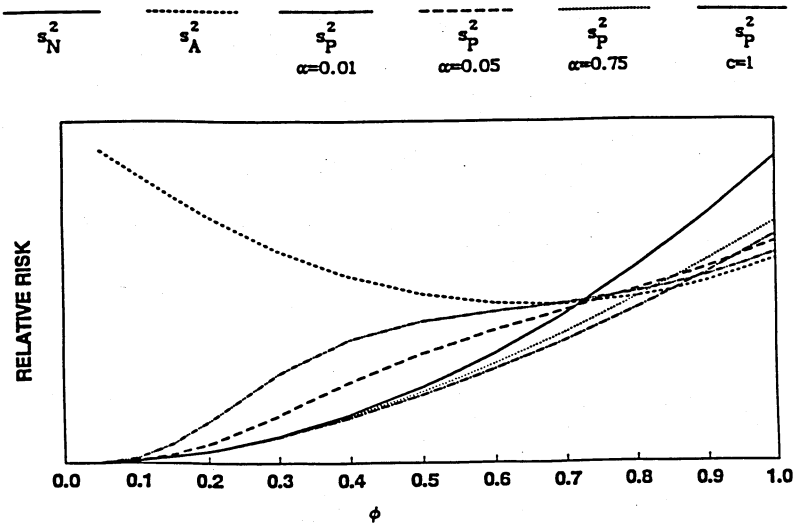

Figure 3. Relative risk functions for s_N^2, s_A^2, and s_P^2 .

(a) Comparing equations (14) and (15), there are two possible values of ϕ, ϕ_1 and ϕ_2, for which $\rho(\sigma_1^2,s_A^2)$ and $\rho(\sigma_1^2,s_N^2)$ intersect, provided that $v_1v_2-4v_1-2v_2\neq 0$ (see Toyoda and Wallace (1975)). In any particular case only one of these values, say ϕ_1, will lie in the interval (0,1]. If $0<\phi<\phi_1$ then s_N^2 dominates s_A^2. Intuitively, the variances are so different that the gain in sampling error from the extra degrees of freedom is outweighed by the bias from pooling the (unequal) variances. Alternatively, s_A^2 has smaller risk than s_N^2 when $\phi_1<\phi<1$.

(b) There exist values of $c\in(0,2)$ such that s_P^2 strictly dominates s_N^2 for

all possible values of ϕ, $0<\phi\leq 1$. Though these particular PTE's do not dominate s_A^2 for all ϕ, they do so over a wide range of ϕ. It is only within the neighbourhood of $\phi=1$ that the risk of s_A^2 is smaller. Ohtani and Toyoda (1978) prove, for a given value of ϕ and $c\in[0,1]$, that the minimum pre-test risk occurs when c=1; so s_N^2 is inadmissible and specifically, is dominated (at least) by the PTE with c=1. These features raise the question of an optimal pre-test critical value - we return to this issue in Section 2.3.

2.2.2 Estimation of the Coefficient Vector

Assuming that $\beta_1=\beta_2=\beta$, model (12) is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\beta + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}, \quad \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N\left[0, \begin{bmatrix} \sigma_1^2 I_{T_1} & 0 \\ 0 & \sigma_2^2 I_{T_2} \end{bmatrix}\right] \tag{18}$$

or

$$y = X\beta + e, \qquad e\sim N(0,\Sigma) . \tag{19}$$

If the variances are equal we estimate β from the T_1+T_2 observations and $b_A = S^{-1}X'y$, which is the usual least squares (and maximum likelihood estimator) of β, is BLUE. b_A is the always-pool estimator of β. However, if the variances are unequal, a feasible GLS estimator of β is the "two-step" Aitken estimator (2SAE) $b_N=[S_1/s_1^2+S_2/s_2^2]^{-1}[X_1'y_1/s_1^2+X_2'y_2/s_2^2]$. b_N is the never-pool estimator of β. The PTE of β is

$$b_P = \begin{cases} b_N & \text{if } J > c \\ b_A & \text{if } J \leq c, \end{cases} \tag{20}$$

The research on this particular pre-test problem has either worked within the framework of the orthonormal model[9] or a reparameterised version of the model[10] given by:

$$y = X^*\beta^* + e, \tag{21}$$

where $X^*=XP$, $\beta^*=P^{-1}\beta$, and $P=T\times\text{diag}\left((1+\mu_i)^{-1/2}\right)$ is a non-singular matrix, and μ_i are the roots of the polynomial $|X_2'X_2/\sigma_2^2-\mu X_1'X_1/\sigma_1^2|=0$ (i=1,...,k). The matrix T is chosen so as to diagonalise $X_1'X_1$ and $X_2'X_2$ simultaneously.

---

[9]See, for example, Ohtani and Toyoda (1980), Yancey *et al.* (1984) and Judge and Yancey (1986).

[10]For example, Taylor (1977, 1978), Greenberg (1980) and Mandy (1984).

Taylor (1978) establishes the finite sample moments of the ith element of the 2SAE within the context of (21). Let this estimator be β^*_{Ni}, i=1,2,...,k. He shows that β^*_{Ni} is an unbiased estimator of β^*_i and that under appropriate conditions the 2SAE is consistent and asymptotically efficient. The least squares estimator of β^* is also unbiased, and Taylor shows that neither estimator dominates, in terms of risk, though he concludes that substantial gains can result from using the 2SAE, depending on the values of v_1, v_2, ϕ, and μ_i[11].

Greenberg (1980) follows Taylor's approach and derives the risk of the two-sided pre-test estimator, $\beta^*_{Pi}{}'$, corresponding to the ith element of b'_P, for the reparameterised model and where the test statistic is J* rather than J. He shows that $\beta^*_{Pi}{}'$ is an unbiased estimator of β^*_i, and that no one estimator, of those evaluated, strictly dominates the others. Nevertheless, the results would seem to favour the 2SAE, unless one had a very strong belief that the variances were equal.

Ohtani and Toyoda (1980) derive the risk of the PTE, for the orthonormal model, when the alternative hypothesis is $H_1{:}\sigma_1^2>\sigma_2^2$. They show that in this situation the 2SAE is inadmissible, as it is dominated by the PTE when the critical value is chosen appropriately. In particular, if one adopts the criterion of minimizing average risk, then the optimal critical value is unity. Mandy (1984) generalises Ohtani and Toyoda's analysis to the non-orthonormal case. He shows that if the direction of the alternative hypothesis is correct then the (inequality) PTE that takes this directional information into account is superior, in terms of risk, to the two-sided (equality) PTE analysed by Greenberg (1980). However, of course, if the alternative hypothesis should be $H_1{:}\sigma_1^2<\sigma_2^2$ then the inequality PTE is risk inferior to the equality PTE.

Finally, Adjibolosoo (1989, 1990a) suggests that this traditional pre-test procedure may lead the researcher to use the 2SAE when in fact the degree of heteroscedasticity may be such that it is still preferable to use OLS. Consequently, he considers a PTE (the "probabilistic heteroscedasticity PTE") which chooses between b_A and b_N according to a measure of the degree of severity of the heteroscedasticity rather than according to the Goldfeld-Quandt J test. Using a Monte Carlo experiment,

---

[11] See Özcam (1987) for some extensions to this work.

Adjibolosoo shows first, that this new test is generally more powerful than the J test and secondly, he shows that the probabilistic heteroscedasticity PTE is typically preferable, in terms of MSE, to its traditional Goldfeld-Quandt counterpart.

Some of the most important practical implications for the applied researcher who pre-tests for error variance homogeneity in the two-sample case are the following. First, if estimation of the error variance is of direct interest, then there are advantages in pre-testing with a critical value of unity, rather than using simply the "never-pool" or "always-pool" estimators. Second, as far as estimation of the coefficient vector is concerned, the preferred strategy may depend on the form of the alternative hypothesis for the pre-test itself. If this alternative is that the sub-sample variances are unequal, then the use of the 2SAE (without pre-testing) seems advisable. On the other hand, if the alternative hypothesis is one-sided then pre-testing with a critical value of unity is again a good strategy.

2.3 Another Homoscedasticity Pre-Test Estimator of the Error Variance

Several studies examine the problem of estimating the error variance in the classical linear regression model, $y=X\beta+e$, $e\sim N(0,\sigma^2 I_T)$, after a preliminary test of $H_0:\sigma^2=\sigma_0^2$, where σ_0^2 is some known value available from previous experience. The alternative hypothesis, H_A, can be one- or two-sided. If we accept H_0 then we use σ_0^2 as our "estimator" of σ^2, while we use $\hat{\sigma}^2=(y-Xb)'y-Xb)/v$, the usual least squares estimator of σ^2, if we reject H_0. The PTE is then:

$$\sigma_P^2 = \begin{cases} \sigma_0^2 \text{ , if accept } H_0 \\ \hat{\sigma}^2 \text{ , if reject } H_0 \end{cases} .$$

Assuming orthonormal regressors, Yancey *et al.* (1983) (see also Srivastava (1976)) derive the risk, under squared error loss, of σ_P^2 assuming $H_A:\sigma^2\geq\sigma_0^2$; and also the risk of the PTE, say $\overset{*}{\sigma}{}^2$, which would arise after testing $H_0:\sigma^2\geq\sigma_0^2$ *vs.* $H_A:\sigma^2<\sigma_0^2$. They numerically evaluate their exact risk expressions for a 5% significance level and find, of the estimators investigated, that there is no strictly dominating estimator, though when the direction of the hypothesis is correct, the risk of $\overset{*}{\sigma}{}^2$ is always equal to or less than the risk of $\hat{\sigma}^2$. Comparing the risks of $\overset{*}{\sigma}{}^2$ and σ_P^2 their results suggest that if $(\sigma^2/\sigma_0^2) \in [0.75,1.25]$ then $\overset{*}{\sigma}{}^2$ has smaller risk than σ_P^2, while the converse is typically the case for other values of (σ^2/σ_0^2).

Inada (1989) considers the problem of estimating the variance of a normal variate after a pre-test that the variance lies in the neighbourhood of a known value; that is, $H_0:[\sigma_0^2/c_0 \le \sigma^2 \le \sigma_0^2 c_0]$, where c_0 is a known positive constant. He considers two PTE's, say PT_1 and PT_2, where

$$PT_1 = \begin{cases} \sigma_0^2 & \text{if } c_1^{-1} < u/\sigma_0^2 < c_1 \\ wu & \text{if } u/\sigma_0^2 \ge c_1 \\ w^{-1}u & \text{if } u/\sigma_0^2 \le c_1^{-1} \end{cases},$$

and

$$PT_2 = \begin{cases} wu & \text{if } u/\sigma_0^2 > 1 \\ w^{-1}u & \text{if } u/\sigma_0^2 \le 1 \end{cases},$$

where $u = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$, the weight w is a constant such that $0<w\le 1$, and c_1 is an appropriate critical value. Inada derives the risks of PT_1 and PT_2, and solves for the values of w such that PT_1 and PT_2 are minimax estimators, given the value of c_0. He compares these PTE's with the traditional PTE σ_P^2, and shows that it is preferable to use PT_1, in small or moderate samples, when H_0 is in the neighbourhood of being true, but in large samples, the risks of PT_1, PT_2, and u are virtually indistinguishable.

Ohtani (1991b) (see also Ohtani (1991a)) considers the PTE, $\overset{+}{\sigma}{}^2$, which arises after testing $H_0:\sigma^2=\sigma_0^2$ *vs.* $H_A:\sigma^2>\sigma_0^2$, when we use the Stein (1964) estimator of σ^2, say σ_S^2, rather than $\hat{\sigma}^2$ if we reject H_0. Recall from the discussion in Section 2.1 that σ_S^2 is itself a PTE, so $\overset{+}{\sigma}{}^2$ is a special type of multi-stage PTE. We discuss this further in Section 7.2. Ohtani shows that if the direction of the prior information is valid, and the size of the pre-test on σ^2 is chosen appropriately, then $\overset{+}{\sigma}{}^2$ strictly dominates σ_S^2.

2.4 The Choice of Significance Level

One feature of these pre-test risk functions considered so far is their dependence on the choice of significance level. If the test size is varied, the pre-test risk function changes, and so too do the differences between the risk of the PTE and the risks of its component estimators. A second feature is that for any particular problem, there exists no dominating estimator; in general, the risks of the PTE and its component estimators cross somewhere in the hypothesis error space.

As the extent to which the non-sample information is true or false is unknown, these features raise the question: "Is there an optimal choice of

test size such that the pre-test risk is as close as possible to the smallest that could be achieved?". Several studies have addressed this issue. Among other things, the answer depends on the pre-test under investigation and the chosen optimality criterion.

First, we review those studies which have considered the optimal choice of test size after a pre-test for linear restrictions. From Figure 1, the minimum risk that could conceivably be achieved, for all λ, is given by the boundary traced out by the risk of the restricted estimator for $\lambda \in [0,m/2]$, and for $\lambda \in [m/2,\infty)$ by the risk of the unrestricted estimator[12]. So we desire a choice of test size which results in the risk of the PTE being as close as possible to this boundary. As α increases the risk of the PTE moves down (up) toward the risk of the unrestricted estimator to the right (left) of $\lambda=m/2$, and there is a trade-off between the proximities of the pre-test risk and the minimum risk boundary. There are various ways of measuring this distance.

One possibility is the criterion of minimax regret. For a given test size, we determine the maximum regret of $\rho\left(E(y),X\hat{b}\right)$ from the boundary for all λ, then solve for the value of the critical value, c, which minimizes the maximum regret. This value of c is the optimal critical value. For the case of a single hypothesis involving a t test, Sawa and Hiromatsu (1973) use this criterion and find an optimum value of c of about 1.8. (See also Farebrother (1975).) For the situation of multiple restrictions, Brook (1972, 1976) chooses values of c, say c^*, that minimise the maximum regret on either side of $\lambda=m/2$. This is a slight modification of the Sawa and Hiromatsu criterion.

For the conditional mean forecast problem (or, when the regressors are orthonormal), Brook finds that c^* is generally very close to two, regardless of the degrees of freedom. This result gives *some* comfort to researchers who traditionally use the 5% significance level: two is an *approximate* critical value when the degrees of freedom are moderate to high, say greater than 25, and $m>4$. The robustness of this result to model mis-specification is considered in Section 3.

Another way of defining the optimal critical value is as follows. Instead of searching for the maximum regret for each level of α, we could

---

[12] Note that James-Stein type estimators exist which have smaller risk than this boundary. See, for instance, Sclove *et al.* (1972) and Judge and Bock (1978).

take into account the regret for each value of λ and search for the value of α which minimises their sum or average. That is, minimise the area between the pre-test risk and the minimum possible risk boundary. This criterion is considered by Toyoda and Wallace (1976), who find that it leads to a critical value of zero (*i.e.* use the least squares estimator) if the number of restictions is less than five. For $m \leq 5 < 60$, they find that the (non-constant) optimal critical value is smaller than that observed by Brook (1976), and approximately equal to Brook's values for $m \geq 60$.

Brook (1976) and Toyoda and Wallace (1976) effectively assume a diffuse (non-informative) prior for λ. This may be giving too little weight to small λ, as the investigator must believe λ is in the neighbourhood of zero to be pre-testing at all[13]. Wallace (1977) postulates that with a strong prior on λ weighted towards zero, the minimum average risk critical value would be increased. Toyoda and Ohtani (1978) extend the analysis of Toyoda and Wallace (1976) to include prior knowledge about λ, by assuming a gamma prior density on λ which allows one to weight the likely values of the hypothesis error. They find that if more weight is given to values of λ around the null hypothesis then the optimal critical values do increase from those proposed by Toyoda and Wallace. Nevertheless, typically, their results do not support the use of the common test sizes of 1% and 5%.

Brook and Fletcher (1981) extend the analyses of Toyoda and Wallace (1976) and Brook (1976) to the case of multicollinear (non-orthonormal) regressors. Then the optimal critical value depends on the level of multicollinearity. They consider testing $H_0:\beta_2=0$ in $y=X_1\beta_1+X_2\beta_2+e$, under the usual classical assumptions, where β_1 is $((k-m)\times 1)$ and β_2 is $(m\times 1)$, and show that the optimal critical value of the pre-test according to the Toyoda and Wallace average risk criterion can be well approximated by $c^*_{TW}=v(m+t-4)/(m(v+2))$, where $v=T-k$, $t=\text{trace}(C_{22})$ and C_{22} is the $(m\times m)$ sub-matrix of

$$(X'X)^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} .$$

When the regressors are orthonormal $C_{22}=I_m$ and $t=m$, but as the columns of X exhibit higher degrees of collinearity then t increases. Brook and Fletcher find c^*_{TW} to be very accurate, especially for large m and v values, and they

---

[13] This is not always the case. For instance, stepwise regression is an obvious counter-example. See, for example, Wallace (1977).

show that the optimal critical value of the prior F-test for $t \leq 4$ is 0; that is, it is preferable to ignore the prior information. This is analagous to the result found by Toyoda and Wallace (1976). For $t \geq 4$, c^*_{TW} increases with m and v, and it increases as t increases for a given m and v, implying a higher probabilty of choosing the restricted estimator.

Under a minimax regret criterion Brook and Fletcher show that the optimal critical value, for multicollinear X, is well approximated by $c^*_B=(1+t/m)$. Recall that for orthonormal regressors t=m and so $c^*_B \cong 2$, as found by Brook (1976). c^*_B depends only on t/m, and not on v, and increases as the relative degree of multicollinearity (t/m) increases. Typically these optimal critical values are still substantially higher than those implied by the traditional 1% and 5% significance levels, and c^*_B is close to c^*_{TW} for reasonably large m and v.

Until recently there has been no research into the choice of an optimal critical value when estimating the error variance after a pre-test for exact linear restrictions. Then, when using the least squares component estimators, the PTE which uses c=1 strictly dominates the unrestricted estimator and can also strictly dominate the restricted estimator for $m \leq 2$. For the latter case there is then no optimal size problem - it is always optimal to pre-test using c=1 *even if the restrictions are valid.* When using the minimum mean squared error component estimators Ohtani (1988a) shows that the PTE using c=v/(v+2) strictly dominates the unrestricted estimator but that there is still a range in the neighbourhood of the null hypothesis where the restricted estimator has smaller risk. Finally, Giles (1990) shows that it is never better to pre-test when using the maximum likelihood components.

Giles and Lieberman (1991b) consider the choice of optimal critical value for a pre-test of exact linear restrictions when estimating the regression error variance. They calculate the critical value, c*, according to a minimax regret criterion and show that regardless of which component estimators are used c* is not constant. This contrasts with Brook's general finding. However, for a given m, k and estimation procedure, c* is relatively constant as v varies. Giles and Lieberman also compare the risk functions of the PTE which uses c* and that which uses the critical value which minimises the pre-test risk function (c=1 for the L estimators, c=v/(v+2) for the M estimators and c=0 for the ML estimators). They find that generally the risk of the PTE which uses these latter (easier to apply) critical values is smaller than that which uses the critical value

derived from the minimax regret criterion.

We now consider the question of the optimal size for a pre-test for homogeneity. Toyoda and Wallace (1975), Hirano (1978), Ohtani and Toyoda (1978) and Bancroft and Han (1983) each investigate this problem when the parameter being estimated is the error variance, σ_1^2, while Ohtani and Toyoda (1980) seek an optimal critical value for the PTE of the location vector in the orthonormal model.

Toyoda and Wallace base their choice of optimal critical value on the minimum average risk criterion, with a diffuse prior. They prove that the necessary condition for the minimum is attained when c=1 and they numerically check the sufficiency and the uniqueness of this minimum. They show that this optimal critical value typically implies a type one error ranging from 40 to 60 percent. Relatively high optimal levels of significance are also reported by Hirano (1978). He considers the choice of significance level one should adopt for the pre-test on the basis of minimising Akaike's information criterion.

A minimax regret criterion is employed by Ohtani and Toyoda (1978). When the alternative hypothesis is one-sided, they find that the optimal critical value depends on the degrees of freedom and varies from about 1.7 to 2.8. This contrasts with the results of Toyoda and Wallace (1975).

Bancroft and Han (1983) investigate yet another criterion: relative efficiency of the PTE to the never-pool estimator. For given values of v_1, v_2, and α, and a one-sided alternative hypothesis, they numerically solve for the maximum and minimum values of this efficiency. For certain values of α, the PTE strictly dominates the never-pool estimator; and so, they suggest selecting a test size such that maximum efficiency is the largest and minimum efficiency is no less than unity. This procedure should ensure the largest gain in efficiency. Bancroft and Han find that this criterion results in optimal significance levels in the region of 30% to 50%.

Ohtani and Toyoda (1980) adopt the criterion of minimising average relative risk when they seek the optimal critical value of the pre-test for homogeneity, prior to estimating the location vector in the orthonormal model. They consider a one-sided alternative hypothesis and show that the 2SAE is inadmissible, as it is strictly dominated by the PTE with a critical value of unity. Ohtani and Toyoda derive the extrema of the average relative risk function and conclude that the optimal critical value for the pre-test is $c^*=1$.

From these studies we see the influence of the chosen criterion on

the proposed optimal test size. Nevertheless, these results suggest optimal values of α that are substantially different from those traditionally used in practice. Further, depending on the criterion adopted, the optimal critical values may vary with the degrees of freedom.

There are some clear prescriptions here for the applied economist who adopts pre-test estimation strategies. If the pre-test relates to linear restrictions on the coefficients then one should apply the F-test with a critical value of two if low predictive risk is desired. If attention focuses on estimation of the coefficient vector itself, then the c^*_{TW} and c^*_B critical value formulae of Brook and Fletcher provide clear guidelines. On the other hand, when estimating the error variance after this same pre-test, it is generally advisable to use a critical value of unity, $v/(v+2)$, or zero, depending on whether one uses the OLS, minimum MSE, or ML variants of the scale parameter estimator. Finally, if the pre-test is one for variance homogeneity, a crtitical value of unity seems advisable when estimating the coefficient vector, assuming a one-sided alternative hypothesis.

3. ROBUSTNESS OF PRE-TEST ESTIMATORS

In any econometric application there is some chance of mis-specifying the model. The errors may not obey the usual "ideal" assumptions; some irrelevant regressors may be included in the model, or relevant ones excluded; the error term may be non-Normal, serially correlated, or heteroscedastic; or the functional form of the model may be mis-represented.

The traditional pre-testing literature in econometrics is based on the premise that there are no such mis-specifications. No other "complications" are allowed for. Recently, this situation has been rectified, and several studies have considered some of the consequences of pre-testing in the context of models that are already mis-specified in some way. Specifically, models which are incorrectly specified in terms of the regressors or with respect to the error term assumptions have now been analysed, but pre-testing in the context of a model whose functional form is mis-specified has yet to be researched.

3.1 Mis-Specification of the Regressors

Mis-specification of the regressor matrix in a linear regression model is a common situation. Extraneous regressors may be included in the model, but it is more likely that relevant regressors will be omitted. The latter situation may arise either because of the researcher's lack of understanding

of the underlying theory, or because certain data are unavailable. In the latter case, another type of mis-specification may also arise - a proxy variable may be substituted for the "real" regressor.

With this in mind, several authors have reappraised some of the standard pre-test estimation strategies, allowing for such model mis-specification. The inclusion of extraneous regressors is easily dealt with. Giles (1986) shows that in this case the risks of the OLS, RLS and pre-test estimators of the regression coefficient vector, after a test of exact linear restrictions on this vector, are the same as in the properly specified model except for a simple scaling of the results. Accordingly, there are the usual regions in the parameter space over which the relative dominance of one of these estimators over the others arises, as in Figure 1. In particular, such pre-testing is never the best of these three strategies, and can be worst. Moreover, the results relating to the optimal choice of pre-test size are unaffected by such a mis-specification.

This situation changes fundamentally if relevant regressors are excluded from the regression. Effectively, this possibility and that of including extraneous regressors was first studied by Ohtani (1983). He considered a pre-test for exact restrictions on the regression coefficients when the model includes proxy variables - that is, effectively, relevant regressors are omitted and irrelevant ones are also included in the model. Unaware of this work, Mittelhammer (1984) dealt with the more extreme case of pre-testing in the context of omitted regressors. Measuring performance in terms of squared error predictive risk, imposing *valid* restrictions no longer guarantees dominance of RLS over OLS, or of the PTE over OLS! This should be contrasted with result (a) noted in connection with Figure 1. Further, referring to result (b) associated with that diagram, the region in which the pre-test and OLS predictive risks must cross is unaltered if the model is mis-specified in this way. Finally, as the degree of model mis-specification increases, the OLS, RLS, and pre-test predictive risks are all unbounded, for a given level of hypothesis error.

When the model is mis-specified in this way, it is also natural to ask whether or not the optimal choice of pre-test size is affected. Intuitively, one would expect that the omission of relevant regressors would generally affect this choice, given the preceeding comments about the effects on the risk functions themselves. Giles, Lieberman and Giles (1992) re-consider Brook's (1976) result relating to a preliminary test of linear restrictions on the coefficient vector when the regressors are orthonormal.

They find that Brook's mini-max regret criterion no longer leads to an optimal critical F-value of approximately two when the model is mis-specified. In fact, the optimal critical value is then sensitive to the degrees of freedom in the problem, and can differ substantially from Brook's value. Further, for a given number of restrictions and regression degrees of freedom, the opimal choice of pre-test critical value declines, and the optimal pre-test size increases, monotonically as the model becomes increasingly mis-specified. This has the effect of accentuating the other strong result from Brooks' analysis - the optimal choice of pre-test size in this problem is often much greater than the commonly assigned values such as 5% or 1% - when relevant regressors are omitted from the model.

Giles and Clarke (1989) study the estimation of the regression scale parameter after the same pre-test in the same mis-specified model. Qualitatively, they come to the same conclusions as Mittelhammer in the case of predictive risk. In particular, imposing valid restrictions need not lead to lower risk than if the prior information is ignored or if a pre-test is undertaken. Clearly, there can be serious costs in omitting relevant regressors. Giles (1991b) extends the analyses of Mittelhammer (1984) and Giles and Clarke (1989) when the disturbances are incorrectly assumed to be normal and we have simultaneously omitted relevant regressors. We discuss this study forther in the next section.

Estimation of the scale parameter in the context of omitted regressors is also considered by Ohtani (1987a), but for a different preliminary test, namely $H_0:\sigma^2=\sigma_0^2$ *vs.* $H_A:\sigma^2\neq\sigma_0^2$ or $H_A:\sigma^2>\sigma_0^2$. He finds that under the one-sided alternative (but not under the two-sided one), there exists a family of PTE's for σ^2 which strictly dominate the unrestricted estimator. This dominance is robust to mis-specification through the omission of regressors. He considers a numerical example with v=20 degrees of freedom, and conjectures that the PTE based on a size of 45% has minimum risk in this dominating family. It is straightforward to show, using the approach of Giles (1991a,b, 1992b), that the optimal such critical value is c=v (regardless of model mis-specification). This implies a pre-test size of 45.8% if v=20.

Assuming a one-sided alternative, Giles (1991c) extends Ohtani's (1987a) and Giles' (1992b) analyses to the testing of homogeneity in the two-sample linear heteroscedasticity model when relevant regressors are omitted from the model's for each sample (possibly different regressors) and the disturbances are spherically symmetric. Then the J test for homogeneity

is invalid under the null, as its distribution depends on all aspects of the problem, including the degree of mis-specification and the variance mixing distribution. She also shows that the critical values, identified by Giles (1992b) (see the next section) which minimise the pre-test risk in the correctly specified model also hold this property for the mis-specified model. Analagous to Ohtani's results, there is a family of PTE's which strictly dominate the never-pool estimator, and also in some cases the always-pool estimator. It is never preferable to always-pool the samples without testing the validity of the null hypothesis, nor is it optimal to ignore the prior information.

Ohtani's (1983) contribution focusses on predictive risk in the context of proxy variables, when the pre-test involves coefficient restrictions. Implicitly, it subsumes the essential pre-test results of Mittelhammer (1984) and Giles (1986). One of Ohtani's most important results is that the pre-test strategy can have lower risk than *both* of its component estimators. This is contrary to the situation in the properly specified regression model, as depicted in Figure 1, and it again underscores the point that once we move away from the make-believe world of a properly specified model to the real-life situation of invalid models, our standard textbook results need to be re-assessed. In this context, perhaps the most important lesson for applied econometricians is that extreme care must be taken over the model's specification. With a mis-specified model it is difficult to offer many helpful prescriptions.

3.2 Non-Normal Regression Errors

Our discussion so far has assumed that the regression disturbances are normally distributed, but there is a large literature which suggests that this assumption is sometimes unrealistic. In particular, many economic data series exhibit more kurtosis (and hence fatter tails) than the normal distribution[14]. This has obvious implications for the distribution of the regression disturbance term, and accordingly there has been increasing interest in the sampling properties of estimators and test statistics for non-normal disturbances. Many studies have considered this issue and various distributions have been investigated (see, for example, Judge *et al.*

---

[14]Such studies include those of Mandelbröt (1963), Fama (1965), Blattberg and Gonedes (1974), Rainbow and Praetz (1986), and Lau *et al.* (1990) in respect of returns analyses in the stock, financial and commodity markets.

(1985)). Two general forms of non-normality are usually analysed. The first assumes that the errors are dependent but are uncorrelated (for example, multivariate Student-t errors), while the second assumes that the non-normal errors are identically and independently distributed (for example, univariate Student-t).

Little work has been undertaken on the investigation of the properties of PTE's with non-normal disturbances. Assuming particular non-normal distributions, Mehta (1972) and Giles (1992b) consider the risk, under squared error loss, of estimators of the error variance after a pre-test for homogeneity of the variances in the two-sample linear regression model, while Giles (1991a, b) derives the risk of PTE's of the prediction vector and of the error variance after a pre-test for exact linear restrictions.

Mehta (1972) considers a family of symmetric distributions given by $f(x|\theta_1,\sigma_1^2,\gamma) = \left\{\Gamma\left[\left(\beta+3\right)/2\right]2^{(\beta+3)/2}\sigma_1^2\right\}^{-1}\exp\left\{-\frac{1}{2}|(x-\theta_1)/\sigma_1^2|^{2/(1+\beta)}\right\}$, which includes the normal, double exponential and rectangular distributions as special cases. Mehta considers the problem of estimating the scale parameter from a random sample which follows this distribution when we also have a second, independent, random sample which follows the same distribution but with σ_1^2 different from σ_2^2. The interest in this problem to economists was outlined in Section 1.2.

Mehta derives the MSE of two PTE's of σ_1^2. The first is analagous to the PTE for this problem that was discussed in Section 2.2 - this PTE is a dis-continuous function of the test statistic. He also derives the MSE of a PTE which is a continuous function of the test statistic, and he compares the MSEs of the estimators. For the cases investigated, the qualitative results are the same for all values of β, the non-normality parameter. He suggests that a test size of between 25%-50% be used.

The remaining pre-test literature in this area considers that the departure from normality is to the spherically symmetric family of distributions, which includes the multivariate Student-t (Mt) and normal as special cases. Aside from the normal distribution, this family results in dependent uncorrelated disturbances. One particularly strong motivation for considering this family is that a particular subclass, the so-called compound normal family, can be expressed as a variance mixture of normals. That is, $f(e)=\int_0^\infty f_N(e)f(\tau)d\tau$, where $f(e)$ is the probability density function (pdf) of e, $f_N(e)$ is the pdf of e when $e\sim N(0,\tau^2 I)$ and $f(\tau)$ is the pdf of τ supported on $[0,\infty)$. Non-normal regression disturbances can arise, even if

each e_i (i=1,...,T) is normally distributed, when the variance of e_i is itself a random variable[15]. For example, the Mt distribution arises if τ is an inverted gamma variate.

Many studies have investigated linear regression models with spherically symmetric disturbances[16]. Of particular relevance to this paper, Box (1952) shows that the null distribution of ℓ, the test statistic for exact linear restrictions, is the same for all members of the spherically symmetric family[17]. Thomas (1970) derives the non-null distribution of ℓ and shows that it depends on the specific form of the variance mixing distribution. King (1979) extends many of Thomas' results. In particular, he shows that if a test has an optimal power property for normal disturbances over all possible values of τ^2 then it maintains this property when the errors are compound normal. Consequently, ℓ is a UMPI size-α test for compound normal disturbances. King also proves that if any function of y (be it a test statistic or an estimator) is invariant to the values taken by τ^2 when $e \sim N(0,\tau^2 I_T)$ then the function has the same distribution for the wider class of spherically symmetric distributions (in fact, elliptically symmetric). So, assuming a correctly specified design matrix, the test statistic for homoscedasticity, J, has the same null and non-null distributions under the wider error term assumption (see also Chmielewski (1981b)).

Giles (1992b) considers the same pre-test problem as Mehta (1972) (and for instance, Bancroft (1944) and Toyoda and Wallace (1975) under normal errors) when the disturbances follow the compound normal family of elliptically symmetric distributions, but are wrongly assumed to be normal. She derives the risk of the PTE and also broadens the standard assumption that the never-pool variance estimators are based on the least squares principle. Two families of variance mixing distributions are considered for specific illustrations - the inverted gamma density and the gamma density. The former mixture results in the Mt family of densities, while Teichroew

---

[15]Further discussion of the family of distributions is beyond the scope of this paper. See, for example, Kelker (1970), Chmielewski (1981a), and Muirhead (1982).

[16]For example, Box (1952), Thomas (1970), King (1979), Chmielewski (1981a, b), Sutradhar and Ali (1986), Sutradhar (1988), Brandwein and Strawderman (1990).

[17]In fact, this result holds for all members of the wider elliptically symmetric family. See King (1979).

(1957) derives the density and distribution functions of a random variable generated from the latter member of the spherically symmetric family. Giles shows that the results are qualitatively invariant to which of these mixing distributions is used and to the choice of estimation method used to form the never-pool estimator.

The key results from the Giles (1992b) study are first, that the PTE can strictly dominate *both* of its component estimators for sufficiently non-normal disturbances. Secondly, it may be preferable to use the maximum likelihood principle to form the never-pool estimators rather than the least squares principle for non-normal disturbances. Finally, she shows that the risk function of the pre-test estimator has *a* minimum when $c^*=1$ for the least squares component estimators, $c^*=v_1T_2/(v_2T_1)$ for the (usual) maximum likelihood component estimators, and $c^*=\left(v_1(v_2+2)\right)/\left(v_2(v_1+2)\right)$ for the (usual) minimum MSE component estimators[18]. Giles (1991c) extends this analysis to the simultaneous possibility of omitted regressors, as discussed in the previous section.

Assuming a correctly specified design matrix, Giles (1991a) derives the risk of PTE's of the prediction vector and of the error variance after a pre-test for linear restrictions when the disturbances are compound normal. Her study suggests that the risk properties of the PTE of the prediction vector are qualitatively the same for all members of the compound normal family as presented in Section 2.1 for normal disturbances. In particular, pre-testing is never the preferable strategy. This is incorrect. Wong and Giles (1991) show that it is possible for the PTE to dominate both of its component estimators over some of the λ-range. The investigations of Wong and Giles for Mt disturbances, suggest first that the existence and magnitude of the dominating region for the PTE depends on the values of m and ν, the degrees of freedom parameter of the Mt distribution[19]. Secondly, their results show that there is no strictly dominating PTE.

Giles (1991a) also shows that the wider error distribution assumption can have a substantial impact on the risk function of the estimators of the error variance. She considers the least squares estimators of the error

---

[18] That is, the ML amd M estimators under a normality assumption.

[19] $\nu=\infty$ corresponds to normal errors.

variance[20] and she shows that the pre-test risk function has an extremum when $c=0$, ∞, and $c=1$ - the PTE can dominate both of its component estimators. In fact, using the Mt distribution to illustrate, there exists a family of PTE's with $c\in(0,1]$ which strictly dominate the unrestricted estimator for all λ, and the PTE using a critical value of unity has the smallest risk of those PTE's with $c\in[0,1]$ for all λ.

This family of PTE's also dominates the restricted estimator over part of the λ-range, and the numerical evaluations suggest that this will be strict dominance for small values of ν, say $\nu<15$. The results also suggest that this may occur for normal disturbances if m is small, say, equal to one. Thus, when estimating the error variance, using the least squares estimators, it is never preferable to ignore any linear restrictions on the coefficients. Pre-testing is always preferable, and the optimal pre-test critical value is unity. Further, it is better to pre-test using $c=1$ than to impose the restrictions without testing, unless there is a strong belief that the restrictions are valid. Then pre-testing is better only if ν is small (that is, the tails of the marginal distribution of the disturbances are "fat" in relation to normality) or m is small.

Giles (1991b) extends the Giles (1991a) study to the omitted variables model. She finds that the results of Mittelhammer (1984) and Giles and Clarke (1989) assuming normal errors carry over to the wider error term assumption. In particular, imposing valid restrictions does not guarantee a reduction in risk if we have omitted relevant regressors.

The question of the optimal size of a pre-test for linear restrictions with non-normal disturbances has received little attention. The evaluations of Giles (1991a) show that Brook's optimal critical value of two does not extend to all members of the compound normal family, though she offers no alternative critical value. Wong and Giles (1991) consider the extension of the Brook minimax regret criterion to Mt disturbances. They show first, that the optimal critical values are not constant for all values of ν. Secondly, for a given value of ν, the optimal critical values are relatively invariant to the degrees of freedom and the number of restrictions. For instance, for $\nu=5$ the optimal critical value is approximately 2.4, approximately 2.1 when $\nu=10$, and the optimal critical value of 2 suggested

---

[20] Giles (1990) extends these results to the maximum likelihood and the minimum mean squared error component estimators.

by Brook for normal errors holds reasonably well for the case of Mt errors when $\nu \geq 20$. Wong and Giles also suggest, if ν is unknown, that a researcher could be (practically) content to continue to use Brook's optimal critical value prescribed for normal errors.

Further research is obviously required on deriving the properties of PTE's under non-normal disturbances. In particular, it would be of interest to know whether the observed results extend to the situation of non-normal but identically, independently distributed disturbances. However, from a practical viewpoint it is clear that in the likely event of non-normal disturbances, the prescriptions offered so far may have to be re-examined.

3.3. Other Forms of Model Mis-specification

Recent studies of the effects of model mis-specification on the properties of standard pre-test strategies have proved to be most enlightening, in the sense of overturning a number of apparently strong results which in fact rely on a correct model specification for their validity. A final form of mis-specification that has been considered in this context is that of a nonscalar covariance matrix for the regression errors. Given the likelihood of autocorrelated or heteroscedastic errors in practice, it is natural to ask what effects these may have on some of the standard pre-testing results.

The only two contributions to date which respond to this question are those of Albertson (1991) and Giles, Giles and Wong (1992). Albertson considers the estimation of the regression coefficient vector after a pre-test of linear restrictions on the coefficients, and where the researcher fails to take account of the fact that the errors have an arbitrary nonscalar covariance matrix. Exact analytic results are derived for the OLS, RLS and PTE risks under quadratic loss, and these are evaluated for various data sets and for the cases of AR(1), AR(4) and MA(1) errors.

The form of the regressor variables appears to have some bearing on the results, and several interesting points emerge. First, in the case of trended data and positive AR(1) or MA(1) errors, the usual PTE can be strictly dominated by OLS. Second, in the case of nontrended data, or negative autocorrelation, it is generally preferable to pre-test than to use OLS estimation. Accordingly, prior information about the error process is helpful in prescribing an overall strategy, though it should be noted that autocorrelation pre-test testing raises other considerations, as is discussed in Section 6.4. Third, heteroscedastic errors affect the

properties of the PTE in a less systematic way, though again it is possible for this strategy to be strictly dominated by OLS, something which cannot occur if the model is properly specified.

Other work in progress in this area considers the consequences of this type of mis-specification for optimal pre-test size, pre-test estimation of the regression scale parameter, and investigates the implications of simultaneously mis-specifying the error term properties and omitting relevant regressors.

Giles, Giles and Wong (1992a) consider the robustness of the exact restrictions PTE for the prediction vector in the multiple regression model, to the presence of ARCH or GARCH errors. As such an error distribution is typically more leptokurtic than under normality, it is not surprising that the results are qualitatively somewhat similar to those reported by Giles (1992b) in the case of errors which follow multivariate Student-t and certain other compound Normal distributions. In particular, Giles, Giles and Wong find that when the conventional pre-test is applied (based on the assumption of Normal errors), but the disturbances actually follow a sufficiently strong GARCH process, it is possible for the PTE to strictly dominate *both* of the OLS and RLS estimators in terms of quadratic risk.

Again, the intention of both this latter study and that undertaken by Albertson is to base the analysis of pre-test strategies in a more realistic environment, thus making the results more useful to applied econometricians.

4. PRE-TESTING WITH INEQUALITY RESTRICTIONS

Frequently, we may wish to test the validity of inequality restrictions on the coefficients of a regression model, as opposed to testing exact equality restrictions, as we have discussed so far. For example, after the estimation of a consumption equation we may test whether the marginal propensity to consume is less than unity. Suppose in the classical linear regression model, $y=X\beta+e$, that we have prior information on the coefficient vector which we express as a *single* inequality constraint, $C'\beta \geq r$, where C' is a $(1\times k)$ known vector and r is a known scalar. The estimator of β which ignores the prior constraint is simply the OLS estimtor b, while the estimator which includes the non-sample information is the so-called inequality restricted estimator b^{**}. Rather like a PTE, b^{**} comprises two components: if b satisfies the inequality constraint then $b^{**}=b$, but if $C'b<r$ then $b^{**}=b^{*}$, the equality restricted estimator,

$b^* = b - (X'X)^{-1}C[C'(X'X)^{-1}C]^{-1}(C'b - r)$.

The sampling properties of b^{**} are well known. Zellner (1961) showsthat b^{**} is biased and that it has a truncated normal distribution (see also Judge and Takayama (1966)). This implies, for instance, that a standard t-test based on b^{**} can be misleading (Lovell and Prescott (1970)). The superiority of b^{**} relative to b is considered by, for example, Lovell and Prescott (1970), Liew (1976), Judge *et al.* (1980), Judge and Yancey (1981, 1986), Thomson (1982), and Thomson and Schmidt (1982). They find that when the direction of the inequality constraint is correct it is preferable, in terms of quadratic risk, to use b^{**} rather than b. Further, if the direction of the constraint is in fact incorrect then it is still preferable to use b^{**} rather than b in the neighbourhood of H_0, but b has smaller risk than b^{**} for a sufficiently large hypothesis error. These features are evident in Figure 4, which illustrates typical risk functions for this problem.

The sampling properties of the estimator of the model's parameters which results after a pre-test for the validity of $H_0 : C'\beta \geq r$ *vs.* $H_A : C'\beta < r$ have not received as much attention. Assuming that σ^2 is known, Judge *et al.* (1980) and Judge and Yancey (1981, 1986) derive the exact risk of the PTE defined by:

$$\hat{\beta} = \begin{cases} b \text{ if we reject } H_0 \\ b^{**} \text{ if we cannot reject } H_0; \end{cases} \quad b^{**} = \begin{cases} b \text{ if } C'b \geq r \\ b^* \text{ if } C'b < r \end{cases}.$$

So, $\hat{\beta}$ is the unrestricted OLS estimator of β, b, if we reject the validity of the constraint while it is the inequality restricted estimator of β, b^{**}, if we cannot reject H_0. Figure 4 depicts a typical risk result (under quadratic loss) and shows, in particular, that it is never preferable to pre-test. In fact, pre-testing is sometimes the worst strategy. These results are qualitatively the same as those that we discussed in Section 2.1 with reference to the pre-test for exact linear restrictions. Hasegawa (1989) considers the unknown σ^2 case and shows that qualitatively there is no change in the results. He also considers some Bayesian estimators and shows that these can be preferable to the classical estimators, in a risk sense.

Yancey *et al.* (1989) and Judge *et al.* (1990) extend this literature to the multi-parameter hypothesis case (see also Judge and Yancey (1986)). They consider the case of two inequality constraints and investigate a number of potential PTE's. Yancey *et al.* (1989) examine two multivariate

inequality PTE's; the first after a pre-test of H_0 : $R\beta=r$ *vs.* H_A : $R\beta>r$ and the second after a pre-test of H_0' : $R\beta\geq r$ *vs.* H_A' : not H_0'. They show that neither of the inequality PTE's strictly dominates the other.

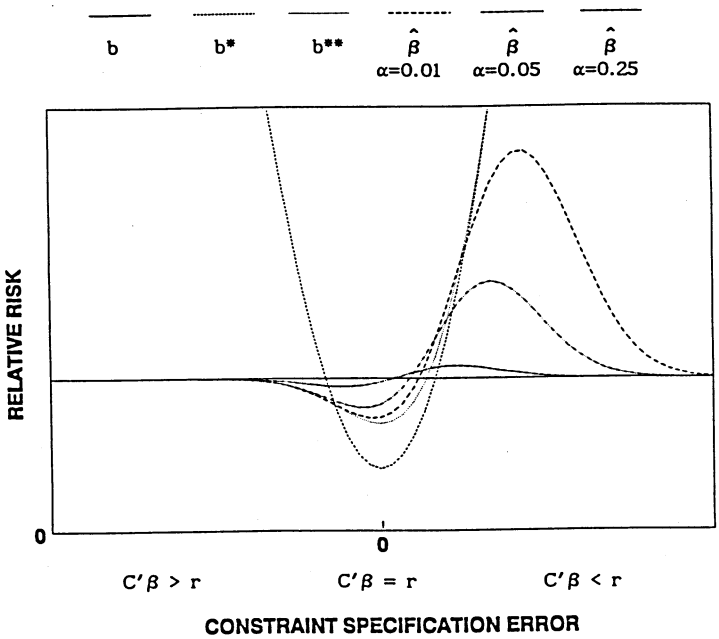

Figure 4. Relative risk functions for b, b*, b**, and $\hat{\beta}$.

The pre-tests examined by Judge *et al.* (1990) are similar though, unlike Yancey *et al.* (1989), they consider the same test statistic for each of the hypothesis tests. Judge *et al.* (1990) find that no one PTE strictly dominates any other and over some parts of the hypothesis error space the equality restricted PTE has smaller risk than the inequality PTE's.

The current literature on inequality pre-testing has considered only the unrealistic situation of a properly specified model. The effects of model mis-specifications on the above results have only recently begun to receive attention. Wan (1992) extends the analyses of Judge *et al.* (1980), Judge and Yancey (1981, 1986), and Hasegawa (1989) to the case of a researcher who

unwittingly omits relevant regressors from the design matrix[21]. Wan shows that the use of valid prior information in an underfitted model does not necessarily guarantee a reduction in risk. This is consistent with the results found for the exact linear restrictions pre-test which we discussed in Section 3.1. He also shows that many of the results of Judge and Yancey carry over qualitatively to the mis-specified case. An exception is the dominance of b by $\hat{\beta}$ when the direction of the constraint is valid. This need not occur when we have omitted relevant regressors.

Given these results there is an obvious question of the choice of an optimal critical value for the pre-test of inequality constraints. Wan (1992) investigates this issue using the Toyoda and Wallace (1976) criterion of minimising the average relative risk. His results suggest that for the case of testing *one* inquality restriction we should simply ignore the prior information and use b. This result is analagous to that obtained by Toyoda and Wallace (1976) for the pre-test of exact linear restrictions.

We can also define a corresponding inequality PTE of the scale parameter. Wan (1992) derives the exact risk of this estimator for both the correctly specified and omitted variables models. He finds that qualitatively many of the results noted in Section 2.1 and Section 3.1 for the estimation of σ^2 after a pre-test for exact linear restrictions carry over to the pre-test of inequality constraints on the coefficient vector. In particular, he shows that the choices of c which result in stationary points of the risk function of the PTE are identical to those reported by Giles (1990, 1991a,b).

Hasegawa (1991) considers the PTE of the coefficient vector when the pre-test relates to the validity of an interval constraint, $H_0: r_1 \leq C'\beta \leq r_2$, where r_1 and r_2 are known scalars. He assumes that the testing procedure is undertaken in two steps. First, we test $H_{01}: C'\beta \geq r_1$ *vs.* $H_{A1}: C'\beta < r_1$ using the usual standard normal test statistic (assuming σ^2 is known). If H_{01} is rejected we use the OLS estimator b as our estimator of β.

If, on the other hand, we cannot reject H_{01} we proceed to the second test, H_{02}: $C'\beta \leq r_2$ *vs.* H_{A2}: $C'\beta > r_2$. If H_{02} is rejected then b is used as the estimator of β, while we use the so-called interval constrained least squares estimator b^+ if we cannot reject H_{02}. This latter estimator is

---

[21] See Ohtani (1991c) for related work. He investigates the properties of the inequality restricted estimator when there is a proxy variable in the model, but he does not examine the inequality pre-test estimator.

given by:

$$b^+ = \begin{cases} r_1 & \text{if } b<r_1 \\ b & \text{if } r_1 \le b \le r_2 \\ r_2 & \text{if } b>r_2 \end{cases} .$$

The properties of this estimator are examined by, for example, Escobar and Skarpness (1986, 1987), and Ohtani (1987d). So, the PTE is

$$\hat{\beta}^+ = \begin{cases} b \text{ if we reject } H_{01} \text{ or we accept } H_{01} \text{ and reject } H_{02} \\ b^+ \text{ if we accept } H_{01} \text{ and accept } H_{02} \end{cases} .$$

Hasegawa derives the risk, under quadratic loss, of $\hat{\beta}^+$ and he also solves for the critical value of the second stage test, given that this test depends on the outcome of the first stage test. He compares the risks of b^+ and $\hat{\beta}^+$, finding first that neither strictly dominates the other, and secondly that the preference for one estimator over the other depends on the relative width of the interval constraint. In particular, when the distance of the interval constraint increases, b^+ dominates $\hat{\beta}^+$ over a wider range.

5. EXACT DISTRIBUTIONS OF PRE-TEST ESTIMATORS

As will be apparent from the discussion so far, the main emphasis in the pre-test literature has been on the first two moments of PTE's. In particular, the literature emphasises the use of risk under quadratic loss as a measure of estimator performance, and so it focusses on Mean Squared Error and the associated trade-off between estimator bias and precision. Accordingly, the emphasis is on the consequences of pre-testing for *point estimation*, rather than *interval estimation*.

To deal with the latter important topic we need more information. For example, to determine the effects of pre-testing on the probability content of a confidence interval we need knowledge of the full sampling distribution of the PTE. Given the additional demands that this places on the analysis, it is not surprising that the econometrics literature was virtually silent on this point until quite recently.

To date, there appear to be only two exact results and one simulation experiment relating to the full sampling distribution of PTE's which are of direct interest to econometricians. Fittingly, the exact results relate to the two problems first studied by Bancroft (1944), as discussed in Section

1.2. Giles (1992a) considers the sampling distribution of the estimator of a variance parameter after a preliminary test of variance homogeneity across two Normal populations; and Giles and Srivastava (1990) derive the sampling distribution of the OLS estimator of a coefficient in a two-regressor model after a preliminary t-test of the significance of the other regressor.

The first of these problems has an econometric interpretation in terms of the estimation of the error term's scale after a pre-test for homoscedasticity in a regression model which may be subject to structural change. So, it relates directly to the earlier discussion in Section 2.2.1. Giles (1992a) considers two random samples, $\{x_{ij}\}_{i=1}^{N_j} \sim N(\mu_j, \sigma_j^2)$; $j = 1,2$. The usual unbiased estimator of σ_j^2 is $s_j^2 = \frac{1}{n_j} \sum_{i=1}^{N_j} (x_{ij} - \bar{x}_j)^2$, where $\bar{x}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_{ij}$, and $n_j = N_j - 1$; $j = 1,2$.

The hypothesis under test is $H_0 : \sigma_1^2 = \sigma_2^2$ *vs.* $H_A : \sigma_1^2 > \sigma_2^2$. As is well known, the statistic (s_1^2/s_2^2) is F-distributed with n_1 and n_2 degrees of freedom if H_0 is true. If H_0 is accepted there is an incentive to pool the samples and estimate σ_1^2 by $s^2 = (n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2)$, which leads to the following PTE of σ_1^2:

$$\hat{\sigma}_1^2 = \begin{cases} s_1^2 & ; \text{ if } (s_1^2/s_2^2) > F_c \\ s^2 & ; \text{ if } (s_1^2/s_2^2) \leq F_c \end{cases}$$

where $F_c = F_c(\alpha)$ is the critical F-value for a significance level of α.

The sampling properties of $\hat{\sigma}_1^2$ differ from those of the "never pool" estimator, s_1^2, and of the "always pool" estimator, s^2. In particular, $\hat{\sigma}_1^2$ is *biased* in finite samples. Clearly, misleading inferences may be drawn if one constructs confidence intervals centered on $\hat{\sigma}_1^2$, but with limits chosen as if no pre-testing had occurred. To analyse this situation fully, Giles (1992a) derives the full c.d.f. of $\hat{\sigma}_1^2$, which is shown to be a rather complicated function of the various parameters of the problem, *but it does not depend on the sample values* and is easily evaluated numerically. Given such evaluations, the pdf for the pre-test estimator is readily obtained by numerical differentiation, and is found to be uni-modal.

Extending earlier related work by Bennett (1956), Giles (1992a) uses these results to examine the extent to which confidence intervals for σ_1^2 are distorted when they are based on $\hat{\sigma}_1^2$, but with the confidence limits (wrongly) determined from the $\chi^2_{n_1}$ distribution of s_1^2 or the $\chi^2_{n_1+n_2}$

distribution of s^2. It transpires that as long as H_0 is not too false, confidence intervals based on pre-testing have higher probability content than those based on s_1^2, while they have lower probability content than those based on s^2. The converse applies for large departures from the null hypothesis. Substantial departures from the assumed confidence level can arise in practice, so extreme care must be taken in applied work.

As with point estimation, the choice of critical value, F_c, can crucially affect the results. Interestingly, when the optimal values suggested by Toyoda and Wallace (1975) and Bancroft and Han (1983) are chosen, there are regions of the parameter space for which the pre-test confidence interval for σ_1^2 has higher probability content than do *either* of the intervals based on s_1^2 or s^2.

Broadly speaking, similar conclusions emerge with the problem analysed by Giles and Srivastava (1990). They consider the estimation of β_1 in the model

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \quad ; \quad t = 1,\ldots,T$$

where the u_t's are iid $N(0,\sigma^2)$, after a pre-test of $H_0 : \beta_2 = 0$ *vs.* $H_A : \beta_2 \neq 0$. Their result extends earlier related work for Normal means by Bennett (1952). The cdf of the PTE $\hat{\beta}_1$ is readily evaluated, and its (uni-modal) pdf is again obtained numerically. Using these results to assist in the evaluation of confidence limits, results corresponding to those of Giles (1992a) above emerge. It is clear that while pre-testing may affect the true confidence level of an interval estimate either adversely or favourably, only by applying the unrestricted estimator without any prior testing, can we be sure that the true and nominal levels coincide.

Monte Carlo simulation results reported by Basmann and Hwang (1988) focus on the effect of certain pre-tests on the location and shape of the sampling distributions of estimators of a simultaneous structural model. More specifically, the pre-testing situation that is considered relates to the application of a likelihood ratio test of the over-identifying restrictions on the parameters prior to estimation of the latter by Two Stage Least Squares (2SLS) or Three Stage Least Squares (3SLS).

The Monte Carlo experiment is based on Klein's Model I of the U.S. economy. Using the Kolmogorov-Smirnov goodness-of-fit test, the authors conclude that the sampling distributions of the PTE's under consideration are not particularly sensitive to the size of the pre-test itself. The same goodness-of-fit test, and the first four moments are used to compare the

conditional and marginal distributions of the 2SLS and 3SLS estimators.

On the basis of such a limited study it is difficult to offer strong practical recommendations, but one conclusion that emerges is that the pre-test 2SLS estimator is generally superior to the pre-test 3SLS estimator in terms of bias and MSE, but the latter exhibits less skewness than the former. Also, as it seems that the identifiability test statistic is *not* distributed independently of the restricted structural coefficient estimators, the usual diagnostic statistics such as "t-ratios" and R^2 may be extremely misleading, and one should be aware of this in applied work.

6. PRE-TEST TESTING

6.1 Introduction

The econometrics pre-testing literature has emphasized pre-test *estimation*, much more than pre-test *testing*, resulting in something of an imbalance. However, work in progress may soon alter this situation. To some degree, this imbalance has arisen because the properties of many of the PTE's that we have considered can be established analytically and without resorting to asymptotic approximations, while the properties of their pre-test test counterparts are generally somewhat less tractable. So, there has been a tendency to focus attention on those problems for which exact results are more readily forthcoming.

The two pre-test testing examples cited in Section 1.1 exposed the importance of the independence of successive tests if a controllable pre-test situation is to be attained. For instance, the researcher's ability to control the overall test size in the context of "nested" hypotheses relies on the independence of such tests, if appropriately formulated, and can be established by appealing to Basu's (1955) Independence Theorem. Moreover, Anderson (1971, pp. 34-43, 116-134, 270-276) shows, for the case of normal linear models, that such a nested testing procedure is UMP in the class of procedures which fix the probabilities of accepting a less restrictive hypothesis than the true one. Seber (1964) provides some asymptotic justification when nested testing is conducted in the context of nonlinear models. Here, there is no substantive pre-test testing "problem". Parenthetically it is worth noting that this result provides strong justification for testing from the "general" to the "specific" when deciding on a regression specification, and underscores Leamer's (1988) point that most "stepwise regression" routines are probably

better described as "stepUNWISE".

The second pre-test testing example given in Section 1.1 also involves an independence issue, but one of a different sort. Extensions of that independence result to a range of other interesting testing situations involving statistics based on quadratic forms in normal random vectors are considered by Phillips and McCabe (1989). However, even with this independence there remains a pre-testing issue. The Chow test would be used only if the null for the pre-test could not be rejected. Otherwise an alternative test, such as a Wald test, for structural stability would be used. So, even with the advantage of statistical independence there remains a pre-test effect of substance in this case - the true size of the two-part test for structural change will differ from the nominal sizes assigned to the Chow test and its alternative, because of the prior test for variance homogeneity.

Furthermore, pre-test testing generally involves an element of non-independence between the "component" test statistics. In such cases, there are immediate effects on the true (as opposed to nominal) sizes of the second and subsequent tests, as well as on their powers. These issues have been considered primarily in the context of three testing situations in the econometrics literature, each of which is considered in turn below. The properties of certain estimators after more than one preliminary test are discussed briefly in Section 7 under the title of "multi-stage pre-testing". If several tests are undertaken simultaneously, rather than sequentially, then one can consider the global or multiple significance level for the collection of tests. A good discussion of this situation is given by Krämer and Sonnberger (1986, pp.147-154), and we do not pursue this topic here.

6.2 Testing the Coefficient Vector After a Pre-Test of the Error Variance

Several studies have used exact analytic methods to consider the properties of the test for linear restrictions on the regression model's coefficient vector, after a pre-test of some hypothesis relating in one way or another to the variance of the error term. In this class of problems, the form of the test for linear restrictions depends on the outcome of the pre-test. For instance, Ohtani and Toyoda (1985) consider a pre-test of the form $H_0:\sigma^2=\sigma_0^2$ vs. $H_A:\sigma^2\neq\sigma_0^2$, where σ^2 is the regression error variance, and σ_0^2 is a fixed value, supposedly known on the basis of previous experience or analysis. In this case, the test of primary interest (that of linear restrictions on the coefficient vector) is based on the usual F-statistic if

the pre-test null is rejected, and otherwise on a Chi-square statistic. Ohtani (1987b) extends this analysis to the case where σ_0^2 is unknown, and it is estimated from an "auxiliary" regression; and Ohtani (1987c) considers a closely related problem, one where the model includes a "proxy" regressor, and the main test of linear restrictions is simply one of the significance of the unobservable regressor which is being proxied. In each of these studies, similar results emerge. The true size of the two stage test can differ substantially from the nominal size of the main test, especially for small degrees of freedom, but this can be corrected by assigning a 25%-30% significance level to the pre-test. Further, the two-stage test can often be more powerful than the single main test, once size correction has been taken into account. Both of these results have obvious implications for applied researchers.

Toyoda and Ohtani (1986) and Ohtani and Toyoda (1988) generalise earlier work by Gurland and McCullough (1962) by considering a more familiar testing problem of the type discussed in the second example cited in the Introduction to this section. Specifically, the pre-test is for variance homogeneity (against either a one-sided or two-sided alternative) in a two sample linear regression situation, followed by a Chow test or an alternative test for the structural stability of the coefficient vector across the two samples, depending on whether the first null is accepted or rejected. The first of these two studies employs a modified (approximate) F-test as the alternative to the Chow test, while the second study uses the Wald test. Consistent with the results noted above, it is found that a pre-test size of up to 80% may be advisable in order to avoid size distortion in the test for structural change. Then, the two-stage test is generally more powerful than the single stage test suggested by Jayatissa (1977) and the Wald test, though typically less powerful than the size-corrected variant of the latter suggested by Rothenberg (1984). Again, there are some clear implications here for applied econometricians.

6.3 Testing the Error Variance After a Pre-test on the Coefficient Vector

A pre-test testing strategy which is essentially the reverse of the ones just discussed is considered by Ohtani (1988b). Specifically, his pre-test relates to exact linear restrictions on the regression coefficient vector, while the main test of interest is of $H_0:\sigma^2=\sigma_0^2$ against $H_A:\sigma^2\neq\sigma_0^2$, where σ_0^2 is known.

As before, the size of the two-stage test can be controlled to that of

the main test if the nominal pre-test size is much larger than would commonly be chosen. Specifically, a size of the order of 30% is recommended. Again, with such a choice, pre-test testing can be more powerful than applying the main test by itself.

6.4 Autocorrelation Pre-Test Testing

In the next section we consider the properties of various estimators involving a preliminary test for the serial independence of the regression errors. However, autocorrelation pre-testing also gives rise to some important pre-test testing situations. Primarily, attention has focussed here on the following problem - what are the properties of the usual t-test for the significance of a regression coefficient after a prior Durbin-Watson (or similar) test for autocorrelation? In this case, the form of the "t-test" varies according to the outcome of the pre-test - if the null of independent errors is accepted then the t-test is based on OLS estimates; but if the errors are found to be autocorrelated then the second test is based on Cochrane-Orcutt or full maximum likelihood estimates, for example. Attention has centred on tests of single linear restrictions at the second stage, and the nature of the problem has necessitated Monte Carlo simulation rather than exact analytical treatment.

Nakamura and Nakamura (1978) and King and Giles (1984) have studied this problem. The former study is the more limited, being based on a model with a single trended regressor, and considering only the Durbin-Watson and Geary autocorrelation tests. A much wider range of tests and design matrices are examined by King and Giles, and (following the findings of Fomby and Guilkey (1978)) both 5% and 50% sizes are considered for the pre-tests. Nakamura and Nakamura find that the size of the pre-test "t-test" exceeds its nominal size, increasingly so as the degree of positive first-order autocorrelation in the errors increases. They also suggest that simply adjusting for autocorrelation and then testing is preferable to either pre-test testing, or t-testing on the basis of OLS estimation in this context. This preference is based on the degree of size distortion.

King and Giles find that their results are somewhat sensitive to the regressor data, and that trended regressors can produce extreme results, so this raises some questions as to the strength of the results just cited. Generally, they find that the size of the pre-test "t-test" is distorted less than those of the two component "t-tests", especially if a 50% pre-test size is used. Further, once size distortion is taken into account, the

powers of the pre-test "t-test" and the "t-test" conducted after an automatic allowance for AR(1) errors are found to be similar, and greater than that of the simple OLS-based t-test. *These results are quite supportive of autocorrelation pre-testing, especially if a larger than usual pre-test size is adopted*, and they are broadly consistent with the principal results discussed in the last two subsections. The converse autocorrelation pre-testing problem is discussed by Giles and Lieberman (1991a). That is, they consider the size and power of the (exact) Durbin-Watson test after a preliminary t-test. Their results are based on a mixture of analytical and Monte Carlo results. When there is no pre-testing the exact power properties of the Durbin-Watson test are easily computed for any regressor data, but a simulation experiment is needed to analyse the pre-test strategy.

Working with a range of data sets, different sample sizes, and pre-test sizes of 5% and 50%, Giles and Lieberman find that pre-testing distorts the true size of the Durbin-Watson test above its nominal value, unless the restriction under test is true. This distortion can be reduced in percentage terms by choosing a large nominal size (say, 50%) for the Durbin-Watson test. The results with respect to power are less clear, but there are situations in which pre-testing can result in increased power of the Durbin-Watson test. These results suggest that autocorrelation pre-testing, defined in either of the above ways, is not necessarily a "bad" thing. Moreover, they lend considerable support to the case made by Fomby and Guilkey (1978) for assigning much larger sizes than usual to tests for autocorrelation, and this should be borne in mind by applied econometricians.

7. OTHER DEVELOPMENTS

There are several other important pre-test problems which have attracted attention in the more recent econometrics literature. Some of these are discussed briefly in this section.

7.1 Autocorrelation Pre-Testing Estimation

The second motivating example given in Section 1.1 serves as an introduction to an interesting pre-test situation that arises frequently in applied econometrics. That example relates to the use of the Durbin-Watson test for serial independence of the regression errors, as a basis for choosing between least squares estimation and Cochrane-Orcutt estimation of

the regression coefficients. Pre-test strategies of this general type have been analysed to some degree.

Judge and Bock (1978, pp. 143-164) consider the empirical risks of such autocorrelation PTE's based on both the Durbin-Watson and Berenblutt and Webb tests, with various component estimators. Specifically, the Durbin, Cochrane-Orcutt and Prais-Winsten estimators are considered when a positive AR(1) error process is detected by one of the tests. A limited Monte Carlo experiment, based on a model with four orthonormal regressors, forms the basis of their analysis, and estimator performance is measured in terms of risk under squared error loss. Only positive autocorrelation is considered. The risks of the component estimators and of the various PTE's increase with the degree of autocorrelation.

This study suggests, among other things, that it is the choice of estimators, rather than the choice of test, which determines the risk consequences of autocorrelation pre-testing. Strategies which incorporate the Durbin estimator when serial correlation is detected are especially favoured. It also indicates that autocorrelation pre-testing doesn't incur particularly large risk losses (relative to no pre-testing), even when there is slight serial correlation, and for moderate to large degrees of serial correlation there can be substantial gains from pre-testing.

These conclusions are corroborated in recent work by Chen and Saleh (1991). Their experimental design is virtually identical to that of Judge and Bock, but they emphasise the performance of autocorrelation pre-test strategies based on shrinkage estimators of the type proposed by Sen and Saleh (1985) and Saleh and Sen (1985). In particular, they find that autocorrelation PTE's of this type uniformly dominate their conventional autocorrelation pre-test counterparts. This illustrates the (known) inadmissibility of standard autocorrelation pre-test strategies, in the same way that Judge and Bock (1978, pp. 189-195) illustrate the inadmissibility of the conventional linear restrictions PTE for the regression coefficients. (Of course, the autocorrelation PTE's suggested by Chen and Saleh are themselves inadmissible, being discontinuous functions of the random data.)

King and Giles (1984) extend the results of Judge and Bock in several directions by considering additional tests for serial independence as well as the full maximum likelihood estimator to allow for AR(1) errors. They also consider the effects of autocorrelation pre-testing on the power of the test for regressor significance, as was discussed in Section 6 above, and on predictive performance. Their Monte Carlo experiment is more comprehensive

than those discussed above, with an allowance for different (nonorthogonal) design matrices with three to six regressors, different sample sizes, and both the 5% and 50% pre-test significance levels. The latter size reflects results obtained by Fomby and Guilkey (1978), and extended recently by Kennedy and Simons (1991). Fomby and Guilkey showed that, in terms of maximising the (MSE) efficiency of the PTE relative to that of OLS, after a preliminary Durbin-Watson test, the size of the latter test should be of the order of 50%, rather than the conventional 1% or 5%.

One of the most important findings from the study by King and Giles is that the form of the regressor matrix is crucial in determining the risks of autocorrelation PTE's. This implies that other related results based on more limited studies should be interpreted cautiously. This study also finds that PTE's which incorporate the full maximum likelihood estimator to allow for autocorrelation are preferable to those which involve the Durbin estimator, except for problems involving relatively small degrees of freedom. At conventional significance levels, and especially with small sample sizes, the results also suggest a slight preference for the Berenblutt and Webb test, or King's point optimal test as the pre-test. The choice of pre-test is generally less consequential if a 50% significance level is used.

Griffiths and Beesley (1984) extend the earlier Judge-Bock analysis in a different direction. In a Monte Carlo experiment limited to a single regressor model, but considering both trended and stationary data, they analyse the autocorrelation PTE based on the (two-sided) Durbin-Watson test and either OLS or Maximum Likelihood estimation. They also consider what is effectively a multi-step PTE (further examples of which appear in the next subsection). In particular, they analyse the consequences of using the Durbin-Watson test to discriminate between serial independence and autocorrelation of the errors, and Akaike's Information Criterion to select between AR(1) and MA(1) disturbances in the latter case. In terms of reasonably uniform point estimation relative efficiency over the entire parameter space, the multi-step PTE is found to be quite successful. However, in the case of interval estimation, the OLS-AR(1) PTE of the form considered by the other authors is found to be preferable to multistage pre-testing, at least in terms of the tests considered by Griffiths and Beesley. As with the results of King and Giles, the type of regressor data is found to be an important determinant of the estimators' relative performances.

Folmer (1988) provides a more complete summary of the studies discussed above, and provides Monte Carlo and bootstrap approximations to the full distribution of the autocorrelation PTE based on the Durbin-Watson test and the OLS and Durbin estimators. His results accord with those noted already, and he concludes that (wrongly) applying standard results, when autocorrelation pre-testing has taken place, can be very misleading. This is especially so in the case of confidence intervals, a result which accords with the conclusions reached in Section 5 above.

Autocorrelation pre-testing in the context of models which include a lagged dependent variable as a regressor is considered by Giles and Beattie (1987). Again, given the nature of the problem, this study is based on Monte Carlo analysis. The risks of nine PTE's are evaluated, and are found to exhibit similar shapes to those encountered in the fixed-regressor studies discussed already. Consistent with those other results, this study also finds that autocorrelation pre-testing can lead to significant reductions in risk over large parts of the parameter space, and that the choice of component estimators is generally more important than the choice of preliminary test. PTE's which incorporate Wallis's (1967) three-step procedure and Durbin's (1970) "m test" are found to be advantageous from a risk viewpoint.

In summary, autocorrelation pre-testing is generally preferable to simply applying OLS. In terms of both point estimation and interval estimation, pre-testing is also preferable to simply estimating under the assumption of AR(1) errors when the degree of autocorrelation is small, and for high degrees of autocorrelation pre-testing is no worse than this alternative approach. Finally, significance levels much higher than those conventionally assigned in econometric work deserve serious consideration in this context.

7.2 Multi-stage Pre-Testing

Though the majority of the pre-test literature concentrates on the properties of PTE's after a single pre-test, the analysis of multi-stage PTE's has recently received attention. This is, of course, closer to the procedure actually undertaken by applied researchers, who would typically undertake a series of pre-tests prior to deciding the "final" version of the model. For instance, a researcher may test for autocorrelation, and on the basis of this test decide whether to use OLS or some feasible GLS estimator. After this decision, he may then undertake a test for heteroskedasticity,

with the test being used dependent on the outcome of the prior pre-test for autocorrelation. The outcome of the second pre-test - the heteroscedasticity test - will determine whether he keeps the "current" version of the model or whether some modification is undertaken and so on.

Clearly, given that each step in this sequential testing procedure is dependent on the outcome of the previous pre-test, and that typically, the tests are not independent of each other, it is difficult to derive the properties of multi-stage PTE's analytically. To date, as far as we are aware, only four studies have attempted to consider some finite sample properties of various multi-stage PTE's - Shukla (1979), Adjibolosoo (1990b), Özcam and Judge (1991), and Ohtani (1991a).

Shukla (1979) considers the estimation of the slope coefficients in the two-sample simple linear regression model

$$y_{ij} = \gamma_j + \beta_j x_{ij} + e_{ij} \quad ; \quad j=1,2;\ i=1,2,\ldots,T_j;$$

where $e_{ij} \sim N(0,\sigma^2)$, after two preliminary tests of significance. The first pre-test is of $H_{01}:\sigma^2=\sigma_0^2$, where σ_0^2 is a known constant, and this test is undertaken using a χ^2-statistic. The second pre-test is of $H_{02}:\beta_2=\beta_1$, for which a z-test is used if we do not reject H_{01}, while if we reject H_{01} a t-test is undertaken. Shukla derives the bias and MSE of the PTE of β_1. Unfortunately, though he notes that the expressions are too complicated for comparisons to be made, he does not undertake any numerical evaluations of the expressions.

Özcam and Judge (1991) extend Shukla's investigation to the multiple regression case and possibly heteroscedastic error variances. The first pre-test is of $H_{01}:\sigma_1^2=\sigma_2^2$ (*vs.* $H_{A1}:\sigma_1^2>\sigma_2^2$) followed by a second pre-test of $H_{02}:\beta_1=\beta_2$ (*vs.* $H_{A2}:\beta_1\neq\beta_2$) in

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

where $e_i \sim N(0,\sigma_i^2 I_{T_i})$ and the usual assumptions on X_i are satisfied; i=1,2.

The two-stage pre-test estimator (2SPE) of β_1, assuming orthonormal regressors, comprises the least squares estimator $\left(LSE=(X_1'y_1+X_2'y_2)/2\right)$, the Two-Step Aitken estimator $\left(2SAE=(\theta X_1'y_1+(1-\theta)X_2'y_2)\right.$, where $\theta=s_2^2/(s_1^2+s_2^2)$ and s_i^2 is the usual unbiased estimator of $\left.\sigma_i^2\right)$, and the Gauss-Markov estimator $(GME=X_1'y_1)$. H_{01} is tested using the Goldfeld-Quandt test, while H_{02} is tested using the Chow test if we accept that the error variances are

homoscedastic, but using the Wald test if we reject homoscedasticity. Özcam and Judge derive the risk, under quadratic loss, of the 2SPE and using numerical evaluations of these exact expressions, they compare the estimators assuming that each test is undertaken at the (nominal) 5% level. For a given value of σ_1^2/σ_2^2 they find that the 2SPE is preferable to the GME, regardless of whether β_1 and β_2 are equal. Further, their results suggest that the risk advantages of the 2SAE and the LSE over the 2SPE under H_{01} disappears as the difference between β_1 and β_2 disappears.

Adjibolosoo (1990b) also considers a two-step estimator of the coefficients in the two-sample heteroscedastic linear regression model. Following the lines of Adjibolosoo (1989, 1990a) he considers the possibility that one may still prefer to use the LSE even if H_{01} is rejected, if the degree of heteroscedasticity is not severe, and suggests that a second step should be incorporated if we reject H_{01} to compare the relative efficiency (strong MSE) of the LSE and the 2SAE. If the LSE is relatively more efficient we use this estimator even though we have rejected homoscedasticity. Using a Monte Carlo experiment, he shows that the risk of the two-step PTE is typically better than that of the traditional PTE discussed in Section 2.2.2.

Ohtani (1991a) investigates the sampling properties of a PTE for the variance of the classical linear regression model, $\overset{+2}{\sigma}$, after a pre-test for homogeneity H_{02}: $\sigma^2=\sigma_0^2$, where σ_0^2 is known (*vs.* H_{A1}:$\sigma^2>\sigma_0^2$) when the estimator we use if we reject the null hypothesis is Stein's (1964) estimator, σ_S^2. We mentioned Ohtani's study in Section 2.3 and we recall that σ_S^2 is itself a PTE with its components being the restricted (σ_M^{*2}) and unrestricted ($\tilde{\sigma}_M^2$) minimum MSE estimators of the error variance when the test critical value is v/(v+2). The pre-test is for exact linear restrictions, H_{01}:$R\beta=r$ *vs* H_{A1}:$R\beta\neq r$. So, we first test H_{01} using a critical value of v/(v+2); if we reject this hypothesis we use $\tilde{\sigma}_M^2$ while we use σ_M^{*2} if we accept H_{01}. Thus, the estimator actually used is σ_S^2. We then test H_{02}; if we accept H_{02} we use σ_0^2, while we continue to use σ_S^2 if we reject H_{02}. Ohtani derives the MSE of this two-stage PTE and compares it with that of σ_S^2 using numerical evaluations. He shows that if the size of the test for H_{02} is chosen appropriately (say, 25%) then the two-stage PTE strictly dominates the Stein estimator, assuming that the direction of H_{A2} is valid, and that it is only around the neighbourhood of H_{02} being true that other choices of test size may be more appropriate.

Given these studies, it is apparent that further research is required

into the properties of multi-stage PTE's. Nevertheless, the papers discussed in this section suggest that multi-step pre-testing may be a preferable option to naively imposing or ignoring prior information.

7.3 Alternative Loss Functions

All of the above discussion reflects the emphasis in the econometrics pre-testing literature on the use of a quadratic loss function. The appeal of this type of loss, of course, is that it leads to risk measures which are in the form of matrix MSE (or its trace), and so incorporate the familiar bias-variance tradeoff when assessing estimator performance. However, an interesting question that naturally arises is, "to what extent are the various results in the pre-testing literature robust to departures from the quadratic loss assumption?"

There is only scant evidence in answer to this question, though this is an important topic currently under research. One feature of a quadratic loss function is that it is symmetric - underestimation and overestimation are equally penalised. Accordingly, it would be interesting to know whether the standard pre-testing results are sensitive to choices of loss functions which are non-quadratic, but still symmetric, as opposed to ones which are asymmetric. With respect to the former, there is apparently no published evidence. Current work by Giles and Lieberman (1992) reconsiders the first of Bancroft's (1944) problems in terms of risk based on an absolute error loss function. The results to date suggest that, except in one respect, the known results under quadratic loss apply qualitatively under absolute error loss. In particular, when estimating a single regression coefficient after a preliminary t-test of the significance of the second regressor, the risk of the unrestricted estimator is still independent of the test statistic's noncentrality parameter; the pre-test estimator's risk has the same shape properties as in Figure 1; but the risk of the restricted estimator is mildly concave to the origin. The same types of regions of estimator dominance hold as for the quadratic loss function. So, pre-testing can be the worst strategy, it is never best, and the same question of the optimal choice of pre-test size arises.

Any symmetric loss function is unduly restrictive in certain estimation situations. For example, underprediction and overprediction of the future exchange rate are unlikely to be equally costly mistakes. Underestimating

the error variance in a regression model will lead to calculated t-statistics which make the regressors appear to be more "significant" than is warranted. A conservative researcher may prefer to err in the opposite direction.

Work in progress by Giles, Giles and Wallace (1992) considers the former problem, using the asymmetric LINEX loss function proposed by Varian (1975), and subsequently adopted by Zellner (1986) and Srivastava and Rao (1992). Using this same loss function, of which quadratic loss is a special case, Giles and Giles (1991, 1992) consider the second problem suggested above. They consider the estimation of the scale parameter in a multiple regression model with Normal errors after a preliminary test of exact linear restrictions on the coefficient vector, or after a pre-test of variance homogeneity. They find that the risk functions for the pre-test, unrestricted, and restricted estimators are robust to mild departures from quadratic loss (at least qualitatively). However, as the degree of asymmetry in the loss function increases, rather different results can emerge. On the basis of this limited information, it seems that the PTE properties may be reasonably robust to departures from quadratic loss which are still symmetric, or only mildly asymmetric. Strong asymmetry in the loss function, however, may lead to results at variance with those in the established literature, and this remains a topic for further investigation.

7.4 Other Pre-Test Problems

As we have seen already, many of the tests that are used in econometrics can be reduced to ones of the validity of exact restrictions on the regression coefficients, or to tests that can be linked to homoscedasticity of the regression errors. Accordingly, the detailed attention that we have paid to the related PTE's effectively covers a range of specific situations, such as testing the "significance" of an individual regressor, or testing for structural change. Also, there are other closely related pre-test situations that are of interest to economists.

One recent such example is the problem of estimating the regression coefficient vector after a linear restrictions pre-test, but in a situation where there are two sets of linear constraints, one of which contains only valid information, but the second of which may contain some invalid information. This problem is studied by Hessenium and Trenkler (undated), who derive necessary and sufficient conditions for the dominance of the PTE

in terms of matrix MSE[22].

Another example, considered by Griffiths and Judge (1989), involves estimating regression coefficients after the application of Weerahandi's (1987) test of coefficient stability. This test allows for the possibility that the errors may be heteroscedastic across the two sub-samples. The component estimators in this case are OLS and the restricted 2SAE. Using Monte Carlo analysis, Griffiths and Judge compare this PTE with the corresponding one based on an asymptotic F-test[23] and find that although the former PTE is slightly risk-superior to the latter for small degrees of hypothesis error, generally there is no clear advantage in using Weerahandi's "exact" test, rather than the asymptotic F-test, in this pre-test environment.

There are other interesting pre-test problems of various sorts which deserve brief mention here. For example, Morey (1984) derives the asymptotic risks of specification pre-test estimators based on Wu-Hausman tests in the context of a linear model which may be mis-specified in the sense that certain of the errors may be correlated with the disturbances. The two component parts of this PTE for the coefficients are OLS (if the test suggests independence between the errors and the regressors) and Instrumental Variables (IV) estimation (if the errors and regressors are thought to be correlated). The most interesting feature of Morey's results is that this PTE is strictly dominated by the IV estimator itself. However, to be conservative, and given that the degree of mis-specification will be unknown in practice, Morey recommends against simply using the IV estimator without a pre-test, and instead suggests testing with a much higher significance level than would usually be adopted. In this respect his results accord with those discussed already in the context of autocorrelation pre-testing.

More recently, Özcam *et al.* (1991) consider an important PTE in the context of the Seemingly Unrelated Regressions (SUR) model. The setting for their analysis is a two-equation SUR model with orthonormal regressors, and the PTE is based on OLS and GLS, depending on whether there is evidence of cross-equation (population) correlation between the disturbances. The

---

22 See also Pordzik and Trenkler (undated).

23 This is just a scaled Wald statistic.

latter is tested using the corresponding sample correlation coefficient based on the "restricted" OLS residuals. Although the (squared error) risks of the PTE and its components "cross" in the usual sense, there are regions of the parameter space over which the PTE's risk is the smallest of the three. As we have seen already, contrary to popular mythology (which is based on the standard "exact restrictions" PTE problem) *there are situations in which pre-testing can be advantageous from a MSE viewpoint.*

The pre-test literature has not kept pace fully with certain significant developments in econometrics which are of vital importance to applications in both macroeconomics and financial economics. The modelling of high frequency financial time-series now invariably considers the possibility of leptokurtic errors, specifically by specifying some sort of autoregressive conditionally heteroscedastic (ARCH) process for the disturbances. Very little is known about the properties of pre-test strategies in the context of ARCH or GARCH regression errors. The nature of the problem does not allow an analytical treatment, and finite-sample results must be obtained by Monte Carlo simulation.

Engle *et al.* (1985) consider the following pre-test problem. The model is estimated by OLS and a one-sided version of the LM test for ARCH(1) errors is applied in its nR^2 form. Depending on the outcome of the test, either the OLS results are retained, or the coefficients are re-estimated by MLE. This PTE is found to be unbiased. Moreover, in terms of finite sample efficiency, it is found to be little worse than OLS when the errors are free of an ARCH process, much better than OLS when there is an ARCH effect present, and almost as good as the MLE in the latter case. Giles, Giles and Wong (1992b) consider a related pre-test testing problem. They use Monte Carlo analysis to determine the effects that various pre-tests for ARCH(p) or GARCH(p,q) errors have on the size of a subsequent "t-test" for the significance of a regressor. Their results show that such pre-testing is quite innocuous in samples of 100 or more observations, but in small samples it may be preferable to use a smaller than usual size for the pre-test.

Empirical macroeconomics has been revolutionised in recent years with the recognition that the use of cointegrated time-series has special implications for model formulation, estimation and testing. Several tests for the order of integration of time-series, and for their possible cointegration, are now routinely used. Pre-test strategies also abound in this context in practice, though their implications are only just beginning to be explored. Again, the asymptotic nature of the tests concerned

necessitates the use of Monte Carlo simulation to analyse the finite-sample properties of cointegration pre-test estimators and tests. This remains a fruitful and important area for research, and work underway by the authors[24] considers such issues as the effects that multi-stage sequential pre-testing for order of integration and the presence of cointegration may have on test sizes; and the finite-sample properties of estimators of the parameters of VAR/error correction models after cointegration pre-testing.

8. CONCLUSIONS

In this paper we have attempted to provide an overview of preliminary-test problems as they arise in econometrics, and to indicate some of the recent developments in this field. Inevitably, there will be omissions, but hopefully we will have captured the principal thrust of the associated literature while exposing both the historical and recent connections between the different strands of research into pre-testing problems.

A number of specific practical implications that arise from the econometric pre-testing literature have been presented in the paper. However, by way of conclusion it may be appropriate to offer some general comments. First, pre-test estimation and pre-test testing are perhaps the norm rather than the exception in applied econometric analysis, so the issues and results that we have discussed are of direct relevance to economists who engage in empirical work.

Second, once pre-testing takes place the standard "textbook results" relating to the properties of various estimators and tests generally no longer apply. For instance, estimators which are unbiased under a specified set of assumptions or conditions may be biased if used after one or more preliminary tests relating to the specification of the model. The applied researcher should be aware of this as it may affect the overall strategy that is adopted when specifying, testing and estimating an economic relationship.

Third, and contrary to a commonly encountered viewpoint, this is not to say that pre-testing is necessarily a "bad" thing. On the contrary, we have given examples of situations in which pre-testing may lead to estimators which have uniformly smaller Mean Squared Error than can be achieved by

---

[24] This work is being undertaken with John Small.

applying the estimator without a suitable prior test. The essential point is that the sampling properties (such as bias, efficiency, size and power) of the estimators and tests that economists use in their empirical research are typically altered if pre-testing occurs prior to their application.

Fourth, as with any estimators or tests, the established results are based on various assumptions, including the presumption that the model is "correctly specified". If any of these assumptions are violated then typically the established results are affected. One of the recent developments in the pre-testing literature has been an examination of the robustness of established results to the type of assumption violations (such as the omission of regressors, or non-normal errors) that are likely to occur in practice. This makes the literature more directly relevant to practitioners.

Fifth, there are more established results relating to pre-test *estimation* situations than to pre-test *testing* situations. This is an important imbalance, and is one that is being addressed to some degree in current work in the field. Given the nature of the statistical issues involved, it seems likely that these developments will rely more on simulation analysis than on exact distribution theory. In any event, there is an urgent need for more information about the implications of pre-test testing, especially in the context of ARCH-GARCH tests and tests for integration and cointegration with economic time-series data.

Sixth, most of the available information concerning the implications of pre-testing in econometrics relates to the application of a single preliminary test. In reality, economists engage in multi-stage pre-testing in a regression environment. While we have discussed a few results relating to this situation, the available information is very limited. Certainly, multi-stage pre-testing alters the standard pre-test results, though again it is *not* necessarily the case that things get "worse" (in some sense) as the degree of pre-testing is increased. However, one general point that does emerge in this context (and in many simple pre-test testing situations) is that it is often advisable to conduct our standard diagnostic tests with nominal significance levels that are far different than we typically adopt.

Much remains to be done before the full implications of pre-test strategies of the type that economists actually use in their empirical work are properly understood. More information is needed about pre-test testing, multi-stage pre-testing, the impact of non-normal disturbances, and the full sampling distributions of pre-test estimators. However, the recent work in

this field has provided us with a good deal of information that is of direct practical benefit to econometricians and applied economists alike.

References

Adjibolosoo B-S.K. (1989) Measuring the degree of severity of heteroskedasticity and the choice between the OLS estimator and the 2SAE. *Communications in Statistics: Theory and Methods* A18, 3451-3462.

Adjibolosoo, B-S.K. (1990a) The degree of severity of heteroskedasticity and the traditional Goldfeld and Quandt pretest estimator. *Communications in Statistics: Simulation and Computation* B19, 827-836.

Adjibolosoo, B-S.K. (1990b) A procedure for improving upon the performance of the traditional heteroskedasticity pretest estimator. *Communications in Statistics: Theory and Methods* A19, 1899-1912.

Albertson, K.V. (1991) Pre-test estimation in a regression model with a misspecified error covariance matrix. Discussion Paper No. 9115, Department of Economics, University of Canterbury.

Anderson, T.W. (1971) *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons.

Bancroft, T.A. (1944) On biases in estimation due to the use of preliminary tests of significance. *Annals of Mathematical Statistics* 15, 190-204.

Bancroft, T.A. and Han, C-P. (1977) Inference based on conditional specification: A note and a bibliography. *International Statistical Review* 45, 117-127.

Bancroft, T.A. and Han, C-P. (1983) A note on pooling variances. *Journal of the American Statistical Association* 78, 981-983.

Basmann, R.L. and Hwang, H-S. (1988) A Monte Carlo study of structural estimator distributions after performance of likelihood ratio pre-tests. Mimeo., SUNY-Binghamton.

Basu, D. (1955) On statistics independent of a complete sufficient statistic. *Sankhya* 15, 377-380.

Bennett, B.M. (1952) Estimation of means on the basis of preliminary tests of significance. *Annals of the Institute of Statistical Mathematics* 4, 31-43.

Bennett, B.M. (1956) On the use of preliminary tests in certain statistical procedures. *Annals of the Institute of Statistical Mathematics* 8, 45-52.

Berkson, J. (1942) Tests of significance considered as evidence. *Journal of the American Statistical Association* 37, 325-335.

Blattberg, R.C. and Gonedes, N.J. (1974) A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business* 47, 244-280.

Bock, M.E., Yancey, T.A., and Judge, G.G. (1973) The statistical consequences of the preliminary test estimators in regression. *Journal of the American Statistical Association* 68, 109-116.

Box, G.E.P. (1952) Multi-factor designs of first order. *Biometrika* 39, 49-57.

Brandwein, A.C. and Strawderman, W.E. (1990) Stein estimation: The spherically symmetric case. *Statistical Science* 5, 356-369.

Brook, R.J. (1972) On the use of minimax regret functions to set significance points in prior tests of estimation. Ph.D. thesis, North Carolina State University.

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

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

Chmielewski, M.A. (1981a) Elliptically symmetric distributions: A review and bibliography. *International Statistical Review* 49, 67-74.

Chmielewski, M.A. (1981b) Invariant tests for the equality of K scale parameters under spherical symmetry, *Journal of Statistical Planning and Inference* 5, 341-346.

Chen, E.J. and Saleh A.K.Md.E. (1991) Estimation of regression parameters when the errors are autocorrelated. Mimeo., Department of Mathematics and Statistics, Carleton University.

Clarke, J.A. (1986) Some implications of estimating a regression scale parameter after a preliminary test of restrictions. M.Ec. thesis, Monash University.

Clarke, J.A. (1990) Preliminary-test estimation of the standard error of estimate in linear regression. *Economics Letters* 34, 27-32.

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

Cohen, A. (1965) Estimates of the linear combinations of parameters in the mean vector of a multivariate distribution. *Annals of Mathematical Statistics* 36, 299-304.

Durbin, J. (1970) Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica* 38, 410-421.

Engle, R.F., Hendry, D.F., and Trumble, D. (1985) Small-sample properties of ARCH estimators and tests. *Canadian Journal of Economics* 51, 66-93.

Escobar, L.A. and Skarpness, B. (1986) The bias of the least squares estimator over interval constraints. *Economics Letters* 20, 331-335.

Escobar, L.A. and Skarpness, B. (1987) Mean square error and efficiency of the least squares estimator over interval constraints. *Communications in Statistics: Theory and Methods* A16, 397-406.

Fama, E.F. (1965) The behaviour of stock market prices. *Journal of Business* 38, 34-105.

Farebrother, R.W. (1975) Minimax regret significance points for a preliminary test in regression analysis. *Econometrica* 43, 1005-1006.

Folmer, H. (1988) Autocorrelation pre-testing in linear models with AR(1) errors. In T.K. Dijkstra (ed.), *On Model Uncertainty and its Statistical Implications* (pp. 39-55). Berlin: Springer-Verlag.

Fomby, T.B. and Guilkey, D.K. (1978) On choosing the optimal level of significance for the Durbin-Watson test and the Bayesian alternative. *Journal of Econometrics* 8, 203-213.

Gelfand, A.E. and Dey, D.K. (1988a) Improved estimation of the disturbance variance in a linear regression model. *Journal of Econometrics* 39, 387-395.

Gelfand, A.E. and Dey, D.K. (1988b) On the estimation of variance ratio. *Journal of Statistical Planning and Inference* 19, 121-131.

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

Giles, D.E.A. (1992a) The exact distribution of a simple pre-test estimator. In W.E. Griffiths, Lütkepohl, H. and Bock, M.E. (eds.), *Readings in Econometric Theory and Practice: In Honor of George Judge*. Amsterdam: North-Holland.

Giles, D.E.A. and Beattie, M. (1987) Autocorrelation pre-test estimation in models with a lagged dependent variable. In M.L. King and D.E.A. Giles (eds.), *Specification Analysis in the Linear Model* (pp. 99-116). London: Routledge & Kegan Paul.

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

Giles, D.E.A. and Lieberman, O. (1991a) Some properties of the Durbin-Watson test after a preliminary t-test. *Journal of Statistical Computation and Simulation* 41.

Giles, D.E.A. and Lieberman, O. (1992) Preliminary test estimation under absolute error loss. Mimeo., Department of Economics, University of Canterbury.

Giles, D.E.A., Giles, J.A., and Wallace, T.D. (1992) Pre-test estimation of the regression predictor under a LINEX loss structure. Mimeo., Department of Economics, University of Canterbury.

Giles, D.E.A., Giles, J.A., Wong, J.K. (1992a) Properties of a regression pre-test estimator in the presence of GARCH disturbances. Mimeo., Department of Economics, University of Canterbury.

Giles, D.E.A., Lieberman, O., and Giles, J.A. (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, D.E.A. and Srivastava, V.K. (1990) The exact distribution of a least squares regression coefficient estimator after a preliminary t-test. Discussion Paper No. 9010, Department of Economics, University of Canterbury.

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

Giles, J.A. (1991a) Pre-testing for linear restrictions in a regression model 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.

Giles, J.A. (1991c) Estimation of the scale parameter after a pre-test for homogeneity in a mis-specified regression model. Discussion Paper No. 9116, Department of Economics, University of Canterbury.

Giles, J.A. (1992b) Estimation of the error variance after a preliminary-test of homogeneity in a regression model with spherically symmetric disturbances. *Journal of Econometrics* forthcoming.

Giles, J.A. and Giles, D.E.A. (1991b) Preliminary-test estimation of the regression scale parameter when the loss function is asymmetric. Discussion Paper No. 9104, Department of Economics, University of Canterbury.

Giles, J.A. and Giles, D.E.A. (1992) Estimation of the regression scale after a pre-test for homoscedasticity under LINEX loss. Mimeo., Department of Economics, University of Canterbury.

Giles, J.A., Giles, D.E.A., and Wong, J.K. (1992b) The size of a t-test after a pre-test for ARCH-GARCH regression errors. Mimeo., Department of Economics, University of Canterbury.

Giles, J.A. and Lieberman, O. (1991b) The optimal size of a preliminary test for linear restrictions when estimating the regression scale parameter. *Economics Letters* 37, 25-30.

Goodnight, J. and Wallace, T.D. (1972) Operational techniques and tables for making weak MSE tests for restrictions in regression. *Econometrica* 40, 699-710.

Greenberg, E. (1980) Finite sample moments of a preliminary-test estimator in the case of possible heteroscedasticity. *Econometrica* 48, 1805-1813.

Griffiths, W.E. and Beesley, P.A.A. (1984) The small sample properties of some preliminary test estimators in a linear model with autocorrelated errors. *Journal of Econometrics* 25, 49-61.

Griffiths, W.E. and Judge, G.G. (1989) Testing and estimating location vectors under heteroscedasticity. Mimeo., University of New England.

Gurland, J. and McCullough (1962) Testing equality of means after a preliminary test of equality of variances *Biometrika* 49, 403-417.

Han, C-P., Rao, C.V., and Ravichandran, J. (1988) Inference based on conditional specification: A second bibliography. *Communications in Statistics : Theory and Methods* A17, 1945-1964.

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.

Hasegawa, H. (1991) The MSE of a pre-test estimator of the linear regression model with an interval constraint on coefficients. *Journal of the Japan Statistical Society* 21, 189-195.

Hessenius, H. and Trenkler, G. (undated) Pre-test estimation in the linear regression model with competing linear constraints. Mimeo., University of Dortmund.

Hirano, K. (1978) A note on the level of significance of the preliminary test in pooling variances. *Journal of the Japan Statistical Society* 8, 71-75.

Inada, K. (1989) Estimation of variance after preliminary conjecture. *Bulletin of Informatics and Cybernetics* 23, 183-198.

Jayatissa, W.A. (1977) Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal. *Econometrica* 45, 1291-1292.

Judge, G.G. and Bock, M.E. (1978) *The Statistical Implications of Pre-test and Stein Rule Estimators in Econometrics*. New York: North-Holland.

Judge, G.G. and Bock, M.E. (1983) Biased estimation. In Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics* (pp. 599-649). Amsterdam: North-Holland.

Judge, G.G., Bohrer, R. and Yancey, T.A. (1990) Some statistical implications of multivariate inequality constrained testing. *Communications in Statistics: Theory and Methods* A19, 413-430.

Judge, G.G., Griffiths, W.E., Hill, and Lee, T-C. (1980) *The Theory and Practice of Econometrics*. New York: John Wiley and Sons.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H. and Lee, T-C. (1985) *The Theory and Practice of Econometrics* (2nd edn.). New York: John Wiley and Sons.

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

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

Judge, G.G. and Yancey, T.A. (1986) *Improved Methods of Inference in Econometrics*. Amsterdam: North-Holland.

Judge, G.G., Yancey, T.A., and Bock, M.E. (1983) Pre-test estimation under squared error loss. *Economics Letters* 11, 347-352.

Kelker, D. (1970) Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya A*32, 419-430.

Kennedy, P. and Simons, D. (1991) Fighting the teflon factor. *Journal of Econometrics* 48, 15-27.

King, M.L. (1979) Some aspects of statistical inference in the linear regression model. Ph.D. thesis, University of Canterbury.

King, M.L. and Giles, D.E.A. (1984) Autocorrelation pre-testing in the linear model: Estimation, testing and prediction. *Journal of Econometrics* 25, 35-48.

Krämer, W. and Sonnberger, H. (1986) *The Linear Regression Model Under Test*. Heidelberg: Physica-Verlag.

Lau, A.H-L., Lau, H-S. and Wingender, J.R. (1990) The distribution of stock returns: New evidence against the stable model. *Journal of Business and Economic Statistics* 8, 217-223.

Leamer, E.E. (1988) Things that bother me. *Economic Record* 64, 331-335.

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

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

Mandelbröt, B.B. (1963) The variation of certain speculative prices. *Journal of Business* 36, 394-419.

Mandy, D.M. (1984) The moments of a pre-test estimator under possible heteroscedasticity. *Journal of Econometrics* 25, 29-33.

Mehta, J.S. (1972) On utilizing information from a second sample in estimating the scale parameter for a family of symmetric distributions. *Journal of the American Statistical Association* 67, 448-452.

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

Mizon, G.E. (1977) Inferential procedures in nonlinear models: An application in a UK cross section study of factor substitution and returns to scale. *Econometrica* 45, 1221-1242.

Morey, M. (1984) The statistical implications of preliminary specification error testing. *Journal of Econometrics* 25, 63-72.

Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*. New York: John Wiley and Sons.

Nakamura, A. and Nakamura, M. (1978) On the impact of the tests for serial correlation upon the test of significance for the regression coefficient. *Journal of Econometrics* 7, 199-210.

Ohtani, K. (1983) Preliminary test predictor in the linear regression model including a proxy variable. *Journal of the Japan Statistical Society* 13, 11-19.

Ohtani, K. (1987a) Some small sample properties of a pre-test estimator of the disturbance variance in a misspecified linear regression. *Journal of the Japan Statistical Society* 17, 81-89.

Ohtani, K. (1987b) On pooling disturbance variances when the goal is testing restrictions on regression coefficients. *Journal of Econometrics* 35, 219-231.

Ohtani, K. (1987c) Some sampling properties of the two-stage test in a linear regression with a proxy variable. *Communications in Statistics: Theory and Methods* A16, 717-729.

Ohtani, K. (1987d) The MSE of the least squares estimator over an interval constraint. *Economics Letters* 25, 351-354.

Ohtani, K. (1988a) Optimal levels of significance of a pre-test in estimating the disturbance variance after the pre-test for a linear hypothesis on coefficients in a linear regression. *Economics Letters* 28, 151-156.

Ohtani, K. (1988b) Testing the disturbance variance after a pre-test for a linear hypothesis on coefficients in a linear regression. *Communications in Statistics: Theory and Methods* A17, 4231-4250.

Ohtani, K. (1988c) Small sample properties of the two-stage test of equality of individual coefficients after a pre-test for homoscedasticity in two linear regressions. *Journal of the Japan Statistical Society* 18, 23-35.

Ohtani, K. (1991a) Some sampling properties of the pre-test estimator for normal variance with the Stein-type estimator. Mimeo., Faculty of Economics, Kobe University.

Ohtani, K. (1991b) 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. (1991c) Some sampling properties of the inequality constrained least squares estimator in a linear regression model with a proxy variable. Mimeo., Faculty of Economics, Kobe University.

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

Ohtani, K. and Toyoda, T. (1980) Estimation of regression coefficients after a preliminary test for homoscedasticity. *Journal of Econometrics* 12, 151-159.

Ohtani, K. and Toyoda, T. (1985) Testing linear hypothesis on regression coefficients after a pre-test for disturbance. *Economics Letters* 17, 111-114.

Ohtani, K. and Toyoda, T. (1988) Small sample properties of the two-stage test of equality of individual coefficients after a pre-test for homoscedasticity in two linear regressions. *Journal of the Japan Statistical Society* 18, 23-35.

Özcam, A. (1987) The two stage pre-test estimator and Stein type estimators for the two-sample heteroscedastic linear model. Ph.D. thesis, University of Illinois.

Özcam, A. and Judge, G. (1991) Some risk results for a two-stage pre-test estimator in the case of possible heteroskedasticity. *Journal of Econometrics* 48, 355-371.

Özcam, A., Judge, G., Bera, A. and Yancey, T. (1991) The risk properties of a pre-test estimator for Zellner's seemingly unrelated regression model. Mimeo., Department of Agricultural and Resource Economics, University of California, Berkeley.

Phillips, G.D.A. and McCabe, B.P. (1983) The independence of tests for structural change in regression models. *Economics Letters* 12, 283-287.

Phillips, G.D.A. and McCabe, B.P. (1989) A sequential approach to testing for structural change in econometric models. In W. Krämer (ed.), *Econometrics of Structural Change* (pp. 87-101). Heidelberg: Physica-Verlag.

Pordzik, P. and Trenkler, G. (undated) Pre-test estimation in the linear regression model based on competing restrictions. Mimeo., University of Dortmund.

Rainbow, K. and Praetz, P.D. (1986) The distribution of returns in Sydney wool futures. In B.A. Goss (ed.), *Futures Markets: Thier Establishment and Performance* (pp. 191-207). London: Croom Helm.

Rothenberg, T.J. (1984) Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica* 45, 1293-1298.

Saleh A.K.Md.E. and Sen P.K. (1985) On shrinkage M-estimators of location parameters. *Communications in Statistics: Theory and Methods* A14, 2313-2329.

Sawa, T. and Hiromatsu, T. (1973) Minimax regret significance points for a preliminary test in regression analysis. *Econometrica* 41, 1093-1106.

Sclove, S.L., Morris, C., and Radhakrishnan, R. (1972) Non-optimality of preliminary-test estimators for the multinormal mean. *Annals of Mathematical Statistics* 43, 1481-1490.

Seber, G.A.F. (1964) Linear hypotheses and induced tests. *Biometrika* 51, 41-48.

Sen, P.K. and Saleh, A.K.Md.E. (1985) On some shrinkage estimators of multivariate location. *Annals of Statistics* 13, 272-281.

Shukla, N.D. (1979) Estimation of a regression coefficient after two preliminary tests of significance. *Metrika* 26, 182-193.

Srivastava, S.R. (1976) A preliminary test estimator for the variance of a normal distribution. *Journal of the Indian Statistical Association* 14, 107-111.

Srivastava, V.K. and Rao, B.B. (1992) Estimation of disturbance variance in linear regression models under asymmetric criterion. *Journal of Quantitative Economics* forthcoming.

Stein, C. (1964) Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Annals of the Institute of Statistical Mathematics* 16, 155-160.

Sutradhar, B.C. (1988) Testing hypothesis with t error variable. *Sankhya B*50, 175-180.

Sutradhar, B.C. and Ali, M.M. (1986) Estimation of the parameters of a regression model with a multivariate t error variable. *Communications in Statistics: Theory and Methods* A15, 429-450.

Taylor, W.E. (1977) Small sample properties of a class of two stage Aitken estimators. *Econometrica* 45, 497-508.

Taylor, W.E. (1978) The heteroscedastic linear model: Exact finite sample results. *Econometrica* 46, 663-675.

Teichroew, D. (1957) The mixture of normal distributions with different variances. *Annals of Mathematical Statistics* 28, 510-512.

Thomas, D.H. (1970) Some contributions to radial probability distributions, statistics, and the operational calculi. Ph.D. thesis, Wayne State University.

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 Schmidt, P. (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.

Toro-Vizcarrondo, C.E. (1968) Multicollinearity and the mean square error criterion in multiple regression: A test and some sequential comparisons. Ph.D. thesis, North Carolina State University.

Toro-Vizcarrondo, C.E. and Wallace, T.D. (1968) A test of the mean square criterion for restrictions in linear regression. *Journal of the American Statistical Association* 63, 558-572.

Toyoda, T. and Ohtani, K. (1978) Optimal pre-testing procedure in regression. *Keizai Kenkyu* 29, 39-43.

Toyoda, T. and Ohtani, K. (1986) Testing equality between sets of coefficients after a preliminary test for equality of disturbance variances in two linear regressions. *Journal of Econometrics* 31, 67-80.

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

Toyoda, T. and Wallace, T.D. (1979) Pre-testing on part of the data. *Journal of Econometrics* 10, 119-123.

Trenkler, G. (1985) Mean square error matrix comparisons of estimators in linear regression. *Communications in Statistics: Theory and Methods* A14, 2495-2509.

Varian, H.R. (1975) A Bayesian approach to real estate assessment. In S.E. Fienberg and A. Zellner (eds.), *Studies in Bayesian Econometrics and Statistics in Honour of Leonard J. Savage, Vol. I* (pp. 195-208). Amsterdam: North-Holland.

Vinod, H.D. and Ullah, A. (1981) *Recent Advances in Regression Methods*. New York: Marcel Dekker.

Wallace, T.D. (1972) Weaker criteria and tests for linear restrictions in regression. *Econometrica* 40, 689-698.

Wallace, T.D. (1977) Pre-test estimation in regression: A survey. *American Journal of Agricultural Economics* 59, 431-443.

Wallace, T.D. and Asher, V.G. (1972) Sequential methods in model construction. *Review of Economics and Statistics* 54, 172-178.

Wallace, T.D. and Toro-Vizcarrondo, C.E. (1969) Tables for the mean square error test for exact linear restrictions in regression. *Journal of the American Statistical Association* 64, 1649-1663.

Wallis, K.F. (1967) Lagged dependent variables and serially correlated errors: A reappraisal of three-pass least squares. *Review of Economics and Statistics* 49, 555-567.

Wan, A.T.K. (1992) The sampling performance of inequality restricted and pre-test estimators in mis-specified linear regressions. Mimeo., Department of Economics, University of Canterbury.

Weerahandi, S. (1987) Testing regression equality with unequal variances. *Econometrica* 55, 1211-1215.

Wong, J.K. and Giles, J.A. (1991) Optimal critical values of a preliminary test for linear restrictions in a regression model with multivariate Student-t disturbances. Discussion Paper No. 9114, Department of Economics, University of Canterbury.

Yancey, T.A., Judge, G.G. and Bock, M.E. (1973) Wallace's weak mean square error criterion for testing linear restrictions in regression: A tighter bound. *Econometrica* 41, 1203-1206.

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

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

Yancey, T.A., Judge, G.G. and Miyazaki, S. (1984) Some improved estimators in the case of possible heteroscedasticity. *Journal of Econometrics* 25, 133-150.

Zellner, A. (1961) Linear regression with inequality constraints on the coefficients. Report No. 6109, International Centre for Management Science, Rotterdam.

Zellner, A. (1986) Bayesian estimation and predistion using asymmetric loss functions. *Journal of the American Statistical Association* 81, 446-451.

LIST OF DISCUSSION PAPERS*

No. 8801 Workers' Compensation Rates and the Demand for Apprentices and Non-Apprentices in Victoria, by Pasquale M. Sgro and David E. A. Giles.

No. 8802 The Adventures of Sherlock Holmes, the 48% Solution, by Michael Carter.

No. 8803 The Exact Distribution of a Simple Pre-Test Estimator, by David E. A. Giles.

No. 8804 Pre-testing for Linear Restrictions in a Regression Model With Student-t Errors, by Judith A. Clarke.

No. 8805 Divisia Monetary Aggregates and the Real User Cost of Money, by Ewen McCann and David Giles.

No. 8806 The Management of New Zealand's Lobster Fishery, by Alan Woodfield and Pim Borren.

No. 8807 Poverty Measurement: A Generalization of Sen's Result, by Prasanta K. Pattanaik and Manimay Sen.

No. 8808 A Note on Sen's Normalization Axiom for a Poverty Measure, by Prasanta K. Pattanaik and Manimay Sen.

No. 8809 Budget Deficits and Asset Sales, by Ewen McCann.

No. 8810 Unorganized Money Markets and 'Unproductive' Assets in the New Structuralist Critique of Financial Liberalization, by P. Dorian Owen and Otton Solis-Fallas.

No. 8901 Testing for Financial Buffer Stocks in Sectoral Portfolio Models, by P. Dorian Owen.

No. 8902 Provisional Data and Unbiased Prediction of Economic Time Series by Karen Browning and David Giles.

No. 8903 Coefficient Sign Changes When Restricting Regression Models Under Instrumental Variables Estimation, by David E. A. Giles.

No. 8904 Economies of Scale in the New Zealand Electricity Distribution Industry, by David E. A. Giles and Nicolas S. Wyatt.

No. 8905 Some Recent Developments in Econometrics: Lessons for Applied Economists, by David E. A. Giles.

No. 8906 Asymptotic Properties of the Ordinary Least Squares Estimator in Simultaneous Equations Models, by V. K. Srivastava and D. E. A. Giles.

No. 8907 Unbiased Estimation of the Mean Squared Error of the Feasible Generalised Ridge Regression Estimator, by V. K. Srivasatva and D. E. A. Giles.

No. 8908 An Unbiased Estimator of the Covariance Matrix of the Mixed Regression Estimator, by D. E. A. Giles and V. K. Srivastava.

No. 8909 Pre-testing for Linear Restrictions in a Regression Model with Spherically Symmetric Disturbances, by Judith A. Giles.

No. 9001 The Durbin-Watson Test for Autocorrelation in Nonlinear Models, by Kenneth J. White.

No. 9002 Determinants of Aggregate Demand for Cigarettes in New Zealand, by Robin Harrison and Jane Chetwyd.

No. 9003 Unemployment Duration and the Measurement of Unemployment, by Manimay Sengupta.

No. 9004 Estimation of the Error Variance After a Preliminary-Test of Homogeneity in a Regression Model with Spherically Symmetric Disturbances, by Judith A. Giles.

No. 9005 An Expository Note on the Composite Commodity Theorem, by Michael Carter.

No. 9006 The Optimal Size of a Preliminary Test of Linear Restrictions in a Mis-specified Regression Model, by David E. A. Giles, Offer Lieberman, and Judith A. Giles.

No. 9007 Inflation, Unemployment and Macroeconomic Policy in New Zealand: A Public Choice Analysis, by David J. Smyth and Alan E. Woodfield.

No. 9008 Inflation — Unemployment Choices in New Zealand and the Median Voter Theorem, by David J. Smyth and Alan E. Woodfield.

No. 9009 The Power of the Durbin-Watson Test when the Errors are Heteroscedastic, by David E. A. Giles and John P. Small.

No. 9010 The Exact Distribution of a Least Squares Regression Coefficient Estimator After a Preliminary t-Test, by David E. A. Giles and Virendra K. Srivastava.

No. 9011 Testing Linear Restrictions on Coefficients in a Linear Regression Model with Proxy variables and Spherically Symmetric Disturbances, by Kazuhiro Ohtani and Judith A. Giles.

(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	Lp-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.

* 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.