Preliminary-Test Estimation of the Error Variance in Linear Regression
Author(s): Judith A. Clarke, David E. A. Giles, T. Dudley Wallace
Source: *Econometric Theory*, Vol. 3, No. 2 (Aug., 1987), pp. 299-304
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/3532467
Accessed: 19/07/2009 21:08

Econometric Theory, **3**, 1987, 299–304. Printed in the United States of America.

MISCELLANEA

Preliminary-Test Estimation of the Error Variance in Linear Regression

JUDITH A. CLARKE
University of Canterbury

DAVID E. A. GILES
University of Canterbury

T. DUDLEY WALLACE
Duke University

We derive exact finite-sample expressions for the biases and risks of several common pretest estimators of the scale parameter in the linear regression model. These estimators are associated with least squares, maximum likelihood and minimum mean squared error component estimators. Of these three criteria, the last is found to be superior (in terms of risk under quadratic loss) when pretesting in typical situations.

1. INTRODUCTION

This paper generalizes the results of Clarke et al. [2] for a pretest estimator of σ^2 in the model

$$y = X\beta + e; \qquad e \sim N(o, \sigma^2 I)$$

where y and e are $(T \times 1)$, β is $(k \times 1)$ and X is a $(T \times k)$ non-stochastic matrix of full rank. Consider

$$H_0 : R\beta = r \qquad \text{vs} \qquad H_1 : R\beta \neq r,$$

where R and r are non-stochastic, with $R(m \times k)$ and of rank m, so that $\tilde{\beta} = (X'X)^{-1}X'y$, and $\beta^* = \tilde{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\tilde{\beta})$, are the unrestricted and restricted maximum likelihood (ML) estimators of β, respectively.

A uniformly most powerful invariant size$-\alpha$ test of H_0 may be based on $u = [\nu(e^{*\prime}e^* - \tilde{e}'\tilde{e})]/[m(\tilde{e}'\tilde{e})]$, where $\tilde{e}$ and e^* are residual vectors corresponding to $\tilde{\beta}$ and β^*,

$$\nu = (T - k), \qquad u \sim F'_{(m,\nu;\lambda)},$$

and

$$\lambda = (R\beta - r)'[R(X'X)^{-1}R']^{-1}(R\beta - r)/2\sigma^2.$$

 0266-4666/87 $5.00

The pretest estimator of β is

$$\hat{\beta} = \begin{cases} \tilde{\beta}, & \text{if } u > c, \\ \beta^*, & \text{if } u \leqq c, \end{cases}$$

where $\int_0^c dF_{(m,\nu)} = (1 - \alpha)$, and this suggests a pretest estimator of σ^2:

$$\hat{\sigma}^2 = \begin{cases} \tilde{\sigma}^2, & \text{if } u > c, \\ \sigma^{*2}, & \text{if } u \leqq c, \end{cases} \quad \text{(1)}$$

where $\tilde{\sigma}^2 = (\tilde{e}'\tilde{e})/(T + \delta)$ and $\sigma^{*2} = (e^{*\prime}e^*)/(T + \gamma)$.

The unrestricted and restricted ML estimators of σ^2 correspond to $\delta = \gamma = 0$, and the risk of $\hat{\sigma}^2$ *in this one case* is discussed[1] in [2]. The least squares estimators correspond to $\delta = -k$ and $\gamma = (m - k)$, while the best invariant (minimum mean squared error (MSE)) estimators correspond to $\delta = (2 - k)$ and $\gamma = (m + 2 - k)$ (when H_0 is true), respectively. In general, the pretest estimator has properties which differ from those of its components, $\tilde{\sigma}^2$ and σ^{*2}. For example, $\hat{\sigma}^2$ constructed from the best invariant components is *not* itself the best invariant in the family (1).

2. RISKS AND RELATIVE BIASES

The relative bias of $\hat{\sigma}^2$ is $B(\hat{\sigma}^2) = (E(\hat{\sigma}^2) - \sigma^2)/\sigma^2$ and its risk is $\rho(\hat{\sigma}^2) = E(L(\hat{\sigma}^2))$, which is its relative MSE if $L(\hat{\sigma}^2) = (\hat{\sigma}^2 - \sigma^2)^2/\sigma^4$. Define $z_j \sim \chi^2_{(\nu+j)}$ and $w_i \sim \chi^{2\prime}_{(m+i;\lambda)}$ for $i,j = 0,1,\ldots$, so that

$$((T + \gamma)\sigma^{*2}/\sigma^2) = w_\nu, \qquad ((T + \delta)\tilde{\sigma}^2/\sigma^2) = z_0,$$

and

$$((T + \gamma)\sigma^{*2} - (T + \delta)\tilde{\sigma}^2)/\sigma^2 = w_0,$$

where z_0 and w_0 are independent. Then,

$$B(\tilde{\sigma}^2) = -((k + \delta)/(T + \delta)),$$

$$B(\sigma^{*2}) = (m - k - \gamma + 2\lambda)/(T + \gamma),$$

$$\rho(\tilde{\sigma}^2) = (2\nu + (k + \delta)^2)/(T + \delta)^2,$$

$$\rho(\sigma^{*2}) = (2(m + \nu + 4\lambda) + (m - k - \gamma + 2\lambda)^2)/(T + \gamma)^2.$$

These expressions and those in Theorem 1 depend on X only through λ and the values of T and k.

THEOREM 1.

$$B(\hat{\sigma}^2) = [(T+\delta)(2\lambda P_{40} + mP_{20}) - (T+\gamma)(k+\delta) + \nu(\delta-\gamma)P_{02}]/((T+\gamma)(T+\delta)),$$

$$\rho(\hat{\sigma}^2) = 1 + \{4\lambda(T+\delta)^2[\lambda P_{80} + (m+2)P_{60} + \nu P_{42} - (T+\gamma)P_{40}] + \nu(\nu+2)(T+\gamma)^2 - 2(T+\gamma)(T+\delta)[\nu(T+\gamma) + \nu(\delta-\gamma)P_{02} + m(T+\delta)P_{20}] + m(T+\delta)^2[2\nu P_{22} + (m+2)P_{40}] + \nu(\nu+2)(\delta-\gamma)(2T+\delta+\gamma)P_{04}\}/((T+\gamma)(T+\delta))^2,$$

where $P_{ij} = \Pr.[F'_{(m+i,\nu+j;\lambda)} \leqq (cm(\nu+j))/(\nu(m+i))]$.

The proof follows by noting that

$$\hat{\sigma}^2 = \tilde{\sigma}^2 + (\sigma^{*2} - \tilde{\sigma}^2)I_{[0,c]}(u),$$
$$= \sigma^2\{z_0/(T+\delta) + [w_0/(T+\gamma) + z_0(\delta-\gamma)/((T+\delta)(T+\gamma))] \times I_{[0,c]}(\nu w_0/(mz_0))\},$$

where $I_{[0,c]}(u)$ is an indicator function with value unity if $u \in [0,c]$, zero otherwise, and repeatedly applying the results in the appendix of [2]. ■

Note that $\hat{\sigma}^2 \to \tilde{\sigma}^2$ as $\lambda \to \infty$ or $\alpha \to 1$ ($P_{ij} \to 0$); and $\hat{\sigma}^2 \to \sigma^{*2}$ as $\alpha \to 0$ ($P_{ij} \to 1$). Both $B(\hat{\sigma}^2)$ and $\rho(\hat{\sigma}^2)$ depend on α, m, k, T, and λ (and hence R, r, β, σ^2, and X), as well as the choice of γ and δ. We have evaluated the risk and bias of $\hat{\sigma}^2$ for various choices of α, m, k, T, and the three choices of γ and δ noted earlier. Some representative results appear[2] in Figures 1 and 2, these relating to typically moderate values of ν and m and the commonly used least squares component estimators of σ^2.

Overall, our results suggest that if a pretest strategy is adopted and if a mini-max criterion is used with respect to the absolute value of relative bias when estimating σ^2, then among the three choices of γ and δ considered it is preferable to use the best invariant component estimators when $\alpha = 0.01$, but least squares components[3] when $\alpha \geqq 0.05$. Four basic features of the risk results emerge: there are always λ-ranges for which $\rho(\tilde{\sigma}^2)$ is less than both $\rho(\sigma^{*2})$ and $\rho(\hat{\sigma}^2)$; for which $\rho(\sigma^{*2})$ is less than both $\rho(\tilde{\sigma}^2)$ and $\rho(\hat{\sigma}^2)$; for which $\rho(\hat{\sigma}^2)$ exceeds[4] both $\rho(\tilde{\sigma}^2)$ and $\rho(\sigma^{*2})$; but there is *no* λ-range for which $\rho(\hat{\sigma}^2)$ is less than both $\rho(\tilde{\sigma}^2)$ and $\rho(\sigma^{*2})$, simultaneously. At least for the values of γ and δ considered, these results are analogous to those for the pretest estimation of β ([3]). Our results suggest that if a pretest estimator of σ^2 is used and one adopts a mini-max rule with respect to risk under quadratic loss, then of the three component estimators we have considered, it may be advisable to use those based on the minimum MSE principle.

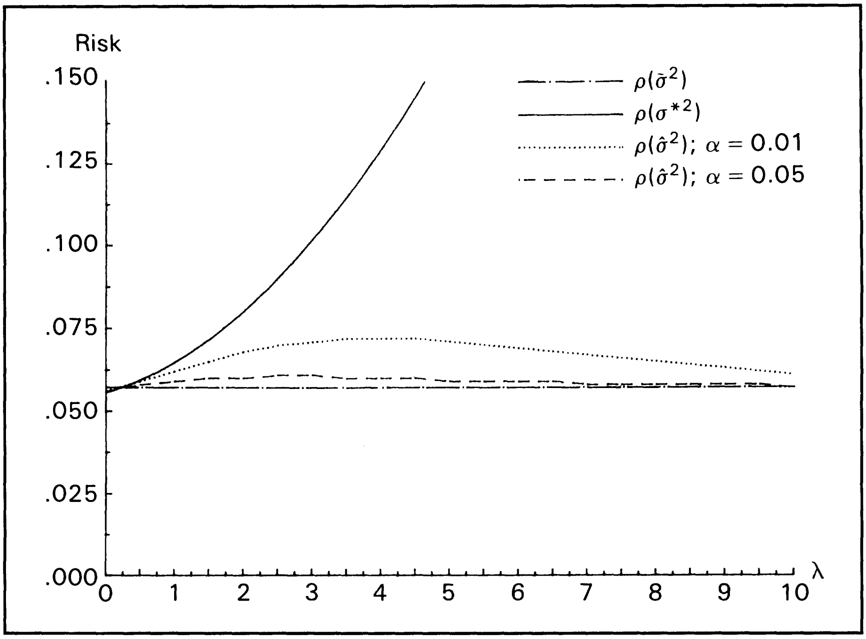

FIGURE 1. Risk (least squares components) at $T = 40$, $k = 5$, and $m = 1$.

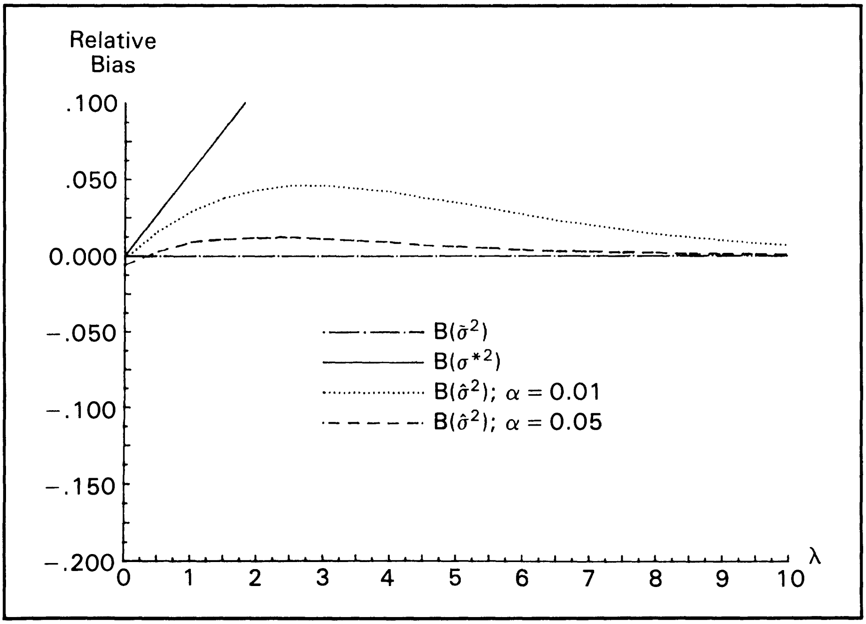

FIGURE 2. Relative bias (least squares components) at $T = 40$, $k = 5$, and $m = 1$.

With regard to the question of whether or not to pretest, we find that for moderate degrees of freedom there is little difference between $\rho(\hat{\sigma}^2)$ and $\rho(\tilde{\sigma}^2)$ over most λ-values, especially when $m = 1$ and $\alpha \geqq 0.05$, and the risks of $\hat{\sigma}^2$, $\tilde{\sigma}^2$, and σ^{*2} are of similar magnitude for small λ. This region of the λ-space is of interest, as H_0 would not be tested unless one held a reasonable prior probability that $\lambda = 0$, in which case a small test size would be chosen. Pretesting emerges as preferable to naively imposing the restrictions in H_0 without testing their validity when estimating σ^2.

3. FURTHER DISCUSSION

Our results favor the use of the best invariant component estimators when constructing a pretest estimator of σ^2, but this estimator is not itself the best invariant in the family (1). For the case where $\gamma = (\delta + m)$, the best invariant $\hat{\sigma}^2$ arises when δ is any real root of:

$$\begin{aligned}&(T+\delta)^4(2\lambda P_{40} + \nu + mP_{20}) + (T+\delta)^3[-2\lambda(2\lambda P_{80} + 2(m+2)P_{60}\\
&\qquad + 2\nu P_{42} - mP_{40}) - \nu(\nu+2) + 3m\nu - m(m+2)P_{40}\\
&\qquad - 2m\nu(P_{22} + P_{02}) + m^2P_{20}] + (T+\delta)^2[3m\nu((\nu+2)\\
&(P_{04} - 1) + m(1 - P_{02}))] + (T+\delta)[m^2\nu(3(\nu+2)\\
&(P_{04} - 1) + m(1 - P_{02}))] + m^3\nu(\nu+2)(P_{04} - 1) = 0.\end{aligned}$$

The "optimal" δ is a function of λ, so the best invariant $\hat{\sigma}^2$ is not an operational estimator.[5] However, we have evaluated this "estimator" for several situations[6] and have found its hypothetical risk to be only slightly less than that of $\hat{\sigma}^2$ based on the best invariant component estimators. This reinforces the findings in the last section.

Our results show that the pretest estimator discussed in [2] can be improved upon, in terms of both relative bias and risk under quadratic loss, by adopting a least squares or minimum mean squared error criterion when constructing the component estimators. Pretest estimation of σ (rather than σ^2) is of interest for the construction of "standard errors" and confidence intervals for elements of β. Work in progress by the first author suggests that our findings here also hold (qualitatively) for the estimation of σ.

ACKNOWLEDGMENTS

We are grateful to Ted Anderson, Phillip Edwards, and the referees for their helpful comments.

NOTES

1. Estimators of σ^2 after pretests of other hypotheses are discussed by Yancey et al. [8] and Ohtani and Toyoda [4]. See also Bancroft [1], Paull [5], and Toyoda and Wallace [7].

2. Details of our complete results are available on request. Our calculations were obtained with a double-precision FORTRAN program on a VAX 11-780 computer. Tiku's [6] method was used to evaluate the P_{ij}'s.

3. Recall that $\hat{\sigma}^2 \rightarrow \tilde{\sigma}^2$ (which is unbiased in the least squares case) as $\alpha \rightarrow 1$.

4. Typically, this range is narrow and the risk differences are negligible.

5. For the case covered in Figures 1 and 2, the optimal value of δ ranges from -1 to -3 as λ varies. Estimating λ would produce a suboptimal δ and $\hat{\sigma}^2$ estimator.

6. In all cases, only one real root was plausible, in the sense of implying positive $\tilde{\sigma}^2$ and σ^{*2} for all λ. The FORTRAN subprogram SILJAK was used on a Hewlett Packard 9845B computer.

REFERENCES

1. Bancroft, T. A. On biases in estimation due to the use of preliminary tests of significance. *Annals of Mathematical Statistics* 15 (1944): 190–204.
2. Clarke, J. A., D. E. A. Giles, & T. D. Wallace. Estimating the error variance in regression after a preliminary test of restrictions on the coefficients. *Journal of Econometrics* 34 (1987): 293–304.
3. Judge, G. G. & M. E. Bock. *The statistical implications of pretest and Stein-rule estimators in econometrics.* Amsterdam: North-Holland, 1978.
4. Ohtani, K. & T. Toyoda. Testing linear hypothesis on regression coefficients after a pretest for disturbance variance. *Economics Letters* 17 (1985): 111–114.
5. Paull, A. E. On a preliminary test for pooling mean squares in the analysis of variance. *Annals of Mathematical Statistics* 21 (1950): 539–556.
6. Tiku, M. L. Tables of the power of the *F*-test. *Journal of the American Statistical Association* 62 (1967): 525–539.
7. Toyoda, T. & T. D. Wallace. Estimation of variance after a preliminary test of homogeneity and optimal levels of significance for the pretest. *Journal of Econometrics* 3 (1975): 395–404.
8. Yancey, T. A., G. G. Judge, & D. M. Mandy. The sampling performance of pretest estimators of the scale parameter under squared error loss. *Economics Letters* 12 (1983): 181–186.