

1983

Sampling Distribution of Shrinkage Estimators and Their F-Ratios in the Regression Model

Aman Ullah

RAL Carter

Virendra K. Srivastava

Follow this and additional works at: <https://ir.lib.uwo.ca/economicsresrpt>

 Part of the [Economics Commons](#)

Citation of this paper:

Ullah, Aman, RAL Carter, Virendra K. Srivastava. "Sampling Distribution of Shrinkage Estimators and Their F-Ratios in the Regression Model." Department of Economics Research Reports, 8324. London, ON: Department of Economics, University of Western Ontario (1983).

11724
ISBN: 0-7714-0469-7
ISSN: 0318-725X

RESEARCH REPORT 8324

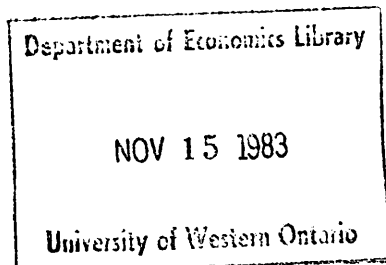
SAMPLING DISTRIBUTION OF SHRINKAGE
ESTIMATORS AND THEIR F-RATIOS IN THE
REGRESSION MODEL

by

Aman Ullah
Richard A. L. Carter
Virendra K. Srivastava¹

University of Western Ontario
and
Lucknow University

September, 1983



1. INTRODUCTION

Several families of adaptive shrinkage estimators have been developed for the estimation of coefficients in a linear regression model, and conditions for their dominance over least squares according to the total mean squared error (risk) criterion have been obtained; see, for example, Judge et al (1980, Ch. 12), Judge and Bock (1978) and Vinod and Ullah (1981, Ch. 5). In a special case of the Stein-type shrinkage estimator, Ullah (1974, 1982) has attempted to analyze its approximate marginal sampling distribution analytically. However, no attempts have been made to evaluate the sampling distribution of other adaptive shrinkage estimators and especially of the t and F ratios which are of significant interest for applied researchers. This paper is an attempt in these directions. We have considered a general family of adaptive shrinkage estimators and have derived Edgeworth-type asymptotic expansions in order to approximate the exact multivariate as well as marginal sampling distributions. The asymptotic expansions of the F ratio are also derived. Although all these expansions are derived by assuming σ (standard error of the disturbance) to be small, the asymptotic results for large T are also given. It is shown that the small- σ approximation provides high order of accuracy to the exact sampling distribution of the F ratio compared to the large- T approximation. This is a new result and it is useful for getting the correct size of the test. Further, it suggests the use of the small- σ technique to derive the approximate distributions of various test statistics in econometrics whose exact sampling distributions are not known. Another interesting result in the paper is the examination

of the performance of shrinkage estimators with respect to the least squares estimator according to a new criterion; viz., concentration of the distribution around the true value. We find that the dominance condition under the mean squared error criterion is robust against the new criterion.

The plan of the paper is as follows. We present the family of shrinkage estimators in Section 2 and its multivariate as well as marginal distributions in 2.1. In 2.1 we also compare the performance of the shrinkage estimators with the least squares estimator. Then in Section 2.2 we present the sampling distribution of the F ratio. In Section 3 a numerical experiment is conducted to assess the accuracy of the small- σ approximation compared to the large-T approximation. Finally in Section 4 we present the proofs of the theorems given in Section 2.

2. THE MAIN RESULTS

Consider a linear regression model:

$$(2.1) \quad y = X\beta + u$$

where y is a $T \times 1$ vector of T observations on the variables to be explained, X is a $T \times p$ full column rank matrix of T observations on p explanatory variables, β is a $p \times 1$ vector of regression coefficients and u is a $T \times 1$ vector of disturbances assumed to follow a multivariate normal distribution $N(0, \sigma^2 I_T)$.

The least squares estimator of β is

$$(2.2) \quad b = (X'X)^{-1}X'y$$

which is unbiased and has a multivariate normal distribution $N(\beta, \sigma^2 (X'X)^{-1})$.

Now consider a general class of shrinkage estimators

$$(2.3) \quad \hat{\beta} = [I + hA]^{-1}b,$$

where A is any known $p \times p$ positive definite matrix and the stochastic scalar h is

$$(2.4) \quad h = \frac{k \hat{u}' \hat{u}}{b' B b}; \quad \hat{u} = y - Xb = Mu \text{ and } M = I - X(X'X)^{-1}X'.$$

In (2.4), k is a positive constant and B is a known positive definite matrix.

When $A = I_p$ and $B = X'X$, $\hat{\beta}$ becomes the Stein-type Ullah and Ullah (1978) estimator

$$(2.5) \quad \hat{b} = \left[1 - \frac{k \hat{u}' \hat{u}}{b' X' X b + k \hat{u}' \hat{u}}\right] b.$$

Also for $A = (X'X)^{-1}$, $\hat{\beta}$ is an adaptive ridge-type estimator. Various other Bayesian and non-Bayesian adaptive shrinkage estimators can also be shown as special cases of (2.3); for details, see Judge and Bock (1978), Judge et al (1980) and Vinod and Ullah (1981, Ch. 13). The mean squared error properties of $\hat{\beta}$ under normal and non-normal disturbances have been analyzed in Ullah et al (1983).

We analyze here the small- σ approximation for the distribution of $\hat{\beta}$ and the F ratio given in (2.26). These results are asymptotic in nature and require σ to be small; see Kadane (1971). Nagar (1959) type large-T asymptotic approximations are easily derived from the small- σ asymptotic expressions. The results have been derived here using the small- σ approach for three reasons. First, it is simpler to develop the asymptotic expansion of a general class of estimators, like $\hat{\beta}$, using this approach: large-T expansions would require assumptions about the nature of k and the matrices

A and B as T grows large. Secondly, Srivastava et al (1980) and Ullah (1980) have shown, for different estimators, that the small- σ approach gives better approximations to the exact moments than does the large-T approach. Section 3 shows that this is also true for the small- σ approximation of the exact distribution of a test statistic. Thirdly, we note that while in small- σ approximations the sample size T is fixed, the large-T approximations make sense only for sufficiently large T. It may be more typical for researchers to have small samples of data from populations with small error variances, in which case small- σ results are more useful.

2.1 The Approximate Distribution Functions of $\hat{\beta}$ and $\hat{\beta}_i$

Before presenting the results we define the estimation error of the vector $\hat{\beta}$ and its i^{th} element $\hat{\beta}_i$, respectively, as

$$(2.6) \quad x = \frac{1}{\sigma}(\hat{\beta} - \beta) \text{ and } \nu' x = \frac{1}{\sigma} \nu' (\hat{\beta} - \beta) = \frac{1}{\sigma}(\hat{\beta}_i - \beta_i)$$

where ν is a $p \times 1$ vector with 1 at the i^{th} place and 0 elsewhere. We also introduce the following notations:

$$(2.7) \quad \theta = \beta' B \beta, \quad q^2 = \nu' Q \nu, \quad Q = (X' X)^{-1}$$

$$n = T - p, \quad m_* = Q \frac{1}{2} m_0$$

where m_0 is a $p \times 1$ vector of fixed constants.

THEOREM 1: The small- σ asymptotic expansion (for fixed T) of the multivariate distribution function of $|x|$, up to order $O(\sigma^2)$, is

$$(2.8) \quad P[|x| \leq m_0] = G(m_*) - G(-m_*) + \frac{\sigma^2 kn}{\theta^2} [G(m_*) - G(-m_*)] \text{tr}(C Q^{\frac{1}{2}} \Gamma Q^{\frac{1}{2}})$$

where $G(m_*)$ represents the standard multivariate normal distribution function, Γ is a $p \times p$ diagonal matrix with j^{th} ($j=1, \dots, p$) diagonal element as $2m_{j*} g(m_{j*}) / (G(m_{j*}) - G(-m_{j*}))$; $g(m_{j*})$ represents the standard univariate normal density function, and

$$(2.9) \quad C = \theta Q^{-1}A - 2Q^{-1}A\beta\beta'B - \frac{1}{2}(n+2)k Q^{-1}A\beta\beta'AQ^{-1}$$

Proof: The proof is given in Section 4.

Remark: The $|x| \leq m_0$ in (2.8) is equivalent to $-m_0 < x \leq m_0$. Thus $P[|x| \leq m_0] = P[-m_0 < x \leq m_0]$. The statement of the theorem in this form is useful for deriving the result in (2.13).² Also, the term $(G(m_*) - G(-m_*))$ on the right-hand side of (2.8) is $P[-m_* < (b-\beta)/\sigma \leq m_*]$; that is, it is the probability for the least squares case. The second term is, to order $O(\sigma^2)$, the extent to which the probability of $\hat{\beta}$ is pulled away from that of b .

A comparison of b and $\hat{\beta}$ can be made with respect to the concentration of probability around β . For this purpose we consider the standardized multivariate distribution functions of b and $\hat{\beta}$ and assume for the sake of simplicity that all the elements of m_0 are identical to a scalar m .

It is easy to see that

$$(2.10) \quad P\left[Q^{-\frac{1}{2}} \frac{(b-\beta)}{\sigma} \leq m\right] = G(m)$$

and so $P[-m < Q^{-\frac{1}{2}} (b-\beta)/\sigma \leq m] = G(m) - G(-m)$, where $G(m)$ is the standard multivariate normal distribution function. Similarly, from (2.8) and (2.9)

$$(2.11) \quad P\left[-m < Q^{-\frac{1}{2}} \frac{(\hat{\beta}-\beta)}{\sigma} \leq m\right] = G(m) - G(-m) + \frac{\sigma^2 knv}{\theta^2} [\theta \text{tr}A - 2\beta'BA\beta - \frac{1}{2}(n+2)k\beta'AQ^{-1}A\beta] \\ \times (G(m) - G(-m)),$$

where

$$(2.12) \quad v = \frac{2mg(m)}{G(m) - G(-m)}.$$

We can now present the following result.

COROLLARY 1: Under the criterion of concentration of probability around β the shrinkage estimators $\hat{\beta}$ dominates the least squares estimator b , that is,

$$(2.13) \quad P[-m < \frac{Q}{\sigma} \frac{-1}{2} (\hat{\beta} - \beta) \leq m] - P[-m < \frac{Q}{\sigma} \frac{-1}{2} (b - \beta) \leq m] > 0,$$

when

$$(2.14) \quad 0 < k < \frac{2}{n+2} \lambda_{\min} \mu_{\max} (d-2), \quad d = \frac{1}{\mu_{\max}} \text{tr } A > 2,$$

where λ_{\min} and μ_{\max} are the minimum and maximum characteristic roots of
 $B(A'X'XA)^{-1}$ and A , respectively.

Proof: Using (2.10) and (2.11), and noting from Rao (1973, p. 74) that for any two matrices A and B , $\min(\beta' A \beta / \beta' B \beta)$ is the minimum root of $|A - \lambda B| = 0$, the result in Corollary 1 follows. Q.E.D.

The result in (2.14) provides the dominance condition for the shrinkage estimators $\hat{\beta}$ under a new criterion. It is interesting to observe that this condition of dominance is the same as the dominance condition in Ullah et al (1983, p. 394) under the weighted mean squared error criterion $E(\hat{\beta} - \beta)' Q^{-1} (\hat{\beta} - \beta) - E(b - \beta)' Q^{-1} (b - \beta)$ where $Q^{-1} = X'X$. Note that the condition (2.14) is independent of the parameters of the model.

In the special case of the Stein-type estimator \hat{b} in (2.5) for which $A = I_p$ and $B = X'X$ the condition in (2.14) reduces to

$$(2.15) \quad 0 < k < \frac{2}{n+2}(p-2), \quad p > 2.$$

The results for other special cases of shrinkage estimators can be written similarly with appropriate substitutions of A and B .

Now we consider the distribution of individual elements of $\hat{\beta}$.

THEOREM 2: The small- σ asymptotic expansion (for fixed T) of the distribution function of $z'x$, up to order $O(\sigma^2)$, is

$$(2.16) \quad P[z'x \leq m] = G\left(\frac{m}{q}\right) + \left(\xi_0 + \frac{m}{q} \xi_1\right)g\left(\frac{m}{q}\right)$$

where $g(m/q)$ and $G(m/q)$ represent the standard univariate normal density function and distribution function, respectively, m is a scalar constant,

$$(2.17) \quad \xi_0 = \frac{\sigma kn}{\theta q} z' A \beta, \quad \xi_1 = -\frac{\sigma^2 kn}{\theta} \frac{q_1}{q^2}$$

and

$$(2.18) \quad q_1 = z' \left[\frac{1}{2\theta} \{k(n+2)A\beta\beta' A + 4A\beta\beta' BQ\} - AQ \right] z.$$

Proof: The proof is given in Section 4.

Remark: An alternative representation of the distribution function of $z'x = z'(\hat{\beta} - \beta)/\sigma$ which is correct up to the order of approximation considered above is

$$(2.19) \quad P[z'x \leq m] = G\left(\xi_0 + \frac{m}{q}(\xi_1 + 1)\right).$$

This expression can be evaluated by using the standard normal table.³ For the special cases of $\hat{\beta}$, the distribution of $z'x$ can be written immediately from (2.16). We also note that the $P[-m < z'x \leq m] = P[z'x \leq m] - P[z'x < -m]$ coincides with the result in (2.8) when $p = 1$.

COROLLARY 2: Under the criterion of concentration of probability around $z'\beta$ the shrinkage estimators $z'\hat{\beta}$ dominates the least squares estimator $z'b$, that is,

$$(2.20) \quad P[-m < \frac{z'(\hat{\beta} - \beta)}{\sigma} < m] - P[-m < \frac{z'(b - \beta)}{\sigma} < m] > 0,$$

when

$$(2.21) \quad 0 < k < \frac{2}{n+2} \varphi_1 (\varphi_2 - 2); \quad \varphi_1 = \frac{z' A \beta \beta' B (X' X)^{-1} z}{z' A \beta \beta' A z}, \quad \varphi_2 = \frac{(z' A (X' X)^{-1} z) \theta}{z' A \beta \beta' B (X' X)^{-1} z} > 2.$$

Proof: From (2.16) we note that

$$P[-m < \frac{z'(\hat{\beta} - \beta)}{\sigma} < m] - P[-m < \frac{z'(b - \beta)}{\sigma} < m] = 2 \frac{m}{q} \xi_1 g\left(\frac{m}{q}\right),$$

where ξ_1 is given in (2.17). The result in (2.21) then follows by observing the condition under which ξ_1 is positive. Q.E.D.

The condition in (2.21) implies that, under the criterion considered, the single element of the shrinkage estimator $\hat{\beta}$ cannot dominate the single element of the least squares estimator b in the entire parameter space.

When the matrices A and B are such that $A = (X'X)^{-1}B = A'$, the condition in (2.21) reduces to

$$(2.22) \quad 0 < k < \frac{2}{n+2}(\varphi_2 - 2); \quad \varphi_2 = \frac{(z'_0 B z_0)(\beta' B \beta)}{(z'_0 B \beta)^2} > 2, \quad z_0 = (X'X)^{-1}z.$$

We note, using the Cauchy-Schwartz inequality [Rao (1973), p. 54], that $\varphi_2 \geq 1$. This implies that for some part of the parameter space β the condition $\varphi_2 > 2$ can be satisfied.

For the special cases of $\hat{\beta}$ the conditions of dominance follow by appropriate substitution of A and B . For example, in the case of the Stein-type estimator \hat{b} in (2.5) we substitute $B = X'X$ in (2.22).

To obtain the large- T approximation we first consider k to be any arbitrary constant such that the $\lim_{T \rightarrow \infty}(T k)$ is a constant; thus k is of order $O(T^{-1})$. Next we observe that for the different cases of $\hat{\beta}$ the orders of the matrices A and B are different. For example, for the Lawless and Wang (1976) adaptive ridge estimator $A = (X'X)^{-1}$ and $B = X'X$; that is, the orders of A and B are $O(T^{-1})$ and $O(T)$, respectively. Similarly, for the Stein-type estimator \hat{b} in (2.5) the orders of $A = I_p$ and $B = X'X$ are $O(T^0)$ and $O(T)$, respectively. Below we present the result for \hat{b} . The distribution of various other shrinkage estimators can be written similarly by appropriate substitutes of the matrices A and B .

COROLLARY 3: The large-T asymptotic expansion of the distribution function of $\nu'(\hat{b}-\beta)/\sigma$, up to order $O(T^{-1})$, is given by

$$(2.23) \quad P\left[\frac{\sqrt{T} \nu'(\hat{b}-\beta)}{\sigma} \leq m\right] = G\left(\frac{m}{\sqrt{Tq}}\right) + \left(\xi_0^* + \frac{m}{\sqrt{Tq}} \xi_1^*\right) g\left(\frac{m}{\sqrt{Tq}}\right)$$

where $\xi_0^* = \nu' \beta (\sigma k T / \theta q)$ is of order $O(T^{-\frac{1}{2}})$ and

$$(2.24) \quad \xi_1^* = -\frac{\sigma^2 k T (k T + 4) (\nu' \beta)^2}{\theta q^2} - \nu' (X' X)^{-1} \nu$$

is of order $O(T^{-1})$.

Proof: Noting that k is of order $O(T^{-1})$, and $A = I$ and $B = X' X$ are of orders $O(T^0)$ and $O(T)$, respectively, the result in (2.23) follows.⁴ Q.E.D.

2.2 Approximate Null Distributions of F Ratio

Consider the problem of testing $H_0: \beta = \beta_0$ against $H_1: \beta \neq \beta_0$ in model (2.1). The least squares test statistic for this hypothesis is⁵

$$(2.25) \quad F = \frac{n}{p} \frac{(b-\beta)' X' X (b-\beta)}{\hat{u}' \hat{u}} = F_{p, n}$$

The F in (2.25) follows the central F distribution with p and n degrees of freedom under the null hypothesis. Note that when $p=1$ the F becomes the square of a t -ratio.

The improved F is obtained by using $\hat{\beta}$ in place of b in (2.25); that is⁶

$$(2.26) \quad F_{\hat{\beta}} = \frac{n}{p} \frac{(\hat{\beta}-\beta)' X' X (\hat{\beta}-\beta)}{\hat{u}' \hat{u}}$$

In what follows we present the small- σ and large- T asymptotic expansions of $F_{\hat{\beta}}$.

THEOREM 3: The small- σ asymptotic expansion (for fixed T) of the distribution of $F_{\hat{\beta}}$, up to order $O(\sigma^2)$, is given by

$$(2.27) \quad P[p F_{\hat{\beta}} \leq m] = G(p F_{p, n}) + \sum_{l=1}^1 \sum_{l'=1}^4 \eta_{l, l'} G\left(\frac{n(p+2l)}{n+2l'}\right) F_{p+2l, n+2l'}$$

where $G\left(\frac{nr}{w} F_{r,w}\right) = P\left[\frac{nr}{w} F_{r,w} \leq m\right]$, for any r and w, represents the distribution of central F with the r and w degrees of freedom, and $\eta_{-1,1} = \eta_{0,1} = \eta_{-1,2} = \eta_{1,3} = \eta_{0,4} = \eta_{1,4} = 0$,

$$(2.28) \quad \eta_{1,1} = -\frac{\sigma^2 kn(n+2)}{\theta(n+p+2)} [\text{tr} A - \frac{2}{\theta} \beta' B \beta], \quad \eta_{0,2} = \eta - \eta_{1,1}; \quad \eta = \frac{\sigma^2 k^2}{2\theta^2} \frac{n(n+2)(n+4)}{(n+p+2)} \beta' A X' X A \beta$$

$$\eta_{1,2} = -\frac{\eta(n+2)}{(n+p+4)}, \quad \eta_{-1,3} = -\eta, \quad \eta_{0,3} = \frac{2\eta(n+4)}{(n+p+4)}, \quad \eta_{-1,4} = -\frac{\eta(n+6)}{(n+p+4)}$$

such that $\sum_{\ell=1}^1 \sum_{\ell'=1}^4 \eta_{\ell, \ell'} = 0$.

Proof: The proof is given in Section 4.

The results for the special cases of shrinkage estimators $\hat{\beta}$ can be written from (2.27) by appropriate substitutions of the matrices A and B. Also, the large-T asymptotic expansions can be obtained easily. We present below the results for Stein-type estimator \hat{b} in (2.5) for which $A = I_p$ and $B = X'X$. These results have been numerically evaluated in Section 3.

COROLLARY 4: The small- σ asymptotic expansion, up to order $O(\sigma^2)$, and the large-T asymptotic expansion, up to order $O(T^{-1})$, of the distribution of $F_b^{\hat{}}$, are respectively given by

$$(2.29) \quad P[pF_b^{\hat{}} \leq m] = G(pF_{p,n}) + \sum_{\ell=1}^1 \sum_{\ell'=1}^4 \eta_{\ell, \ell'}^* G\left(\frac{n(p+2\ell)}{n+2\ell'} F_{p+2\ell, n+2\ell'}\right)$$

and

$$(2.30) \quad P[pF_b^{\hat{}} \leq m] = G(\chi_p^2) + \sum_{\ell=1}^1 \mu_{\ell} G(\chi_{p+2\ell}^2)$$

where $\eta_{\ell, \ell'}^*$ is $\eta_{\ell, \ell'}$ in (2.28) with $A = I$ and $B = X'X$, $G(\chi_r^2) = P[\chi_r^2 \leq m]$ is the distribution of central χ^2 with r degrees of freedom, and

$$(2.31) \quad \mu_0 = \frac{\sigma^2 k T}{\theta} [p - 2 + \frac{3}{2} k T], \quad \mu_{-1} = -\frac{\sigma^2 k^2 T^2}{\theta}, \quad \mu_1 = -\frac{\sigma^2 k T}{\theta} [p - 2 + \frac{1}{2} k T]$$

such that $\sum_{\ell=1}^1 \mu_{\ell} = 0$.

Proof: The result in (2.29) follows by direct substitution of $A = I$ and $B = X'X$ in (2.27) and (2.28). Further, the result in (2.30) is obtained by observing that $A = O(T^0)$ and $B = O(T)$, and noting the fact that the distribution of $(p + 2l)F_{p+2l, n+2l'}$ tends to χ_{p+2l}^2 as $T \rightarrow \infty$.

Remark: We observe from (2.29) (also (2.27)) that while the small- σ expansion of the distribution of $F_{\hat{\beta}}$ is the weighted sum of the central F distributions, from (2.30) the large-T expansion is the weighted sum of central χ^2 distributions.⁷ The sum of the weights in each case is unity. The numerical experiment in the following section shows that the small- σ expansion provides high order accuracy compared to large-T expansion for approximating the exact distribution of $F_{\hat{\beta}}$. This result suggests the use of small- σ approximation in terms of central F distributions for the various other test statistics used in econometrics whose exact sampling distributions are not known.

3. NUMERICAL RESULTS

Of all the expansions derived above the most useful for applied work are those dealing with the F ratio; equations (2.27), (2.29) and (2.30), since they are what would be used to build confidence intervals for, or test hypotheses about, linear combinations of regression coefficients. The aim of this section is to use numerical methods to compare the accuracy of the small- σ and large-T approximations to the exact distribution function (df) for these improved F ratios. Since this exact df is unknown, the approximate df's are numerically compared to an empirical df obtained by Monte Carlo simulation.

For the model given in (2.1), assume that the focus of interest is $H_0: \beta = \beta_0$. The least square test statistic for this hypothesis is as given in (2.25). The improved F ratio is obtained by using $\hat{\beta}$ in place of b and it is as given in (2.26).

In this section the version of $\hat{\beta}$ considered is \hat{b} , given by (2.5). This is a special case of the double k -class estimator of Ullah and Ullah (1978).

$$\hat{b}(k_1, k_2) = \left[1 - \frac{k_1 \hat{u}' \hat{u}}{y' y - k_2 \hat{u}' \hat{u}} \right] b,$$

obtained by setting $k_1 = k$ and $k_2 = 1 - k_1$. Another member of this class is the James-Stein (1961) estimator for which $k_2 = 1.0$. The disadvantage of the James-Stein estimator is that it can have the opposite sign to the least squares estimator. This change of sign can be prevented by choosing k_2

so that the shrinkage factor $1 - \frac{k_1 \hat{u}' \hat{u}}{y' y - k_2 \hat{u}' \hat{u}}$ is positive. It is easily shown

that a sufficient condition for this is $k_2 \leq 1 - k_1$. Note that \hat{b} in (2.5) satisfies this sufficient condition but the James-Stein estimator does not. We prefer this solution to the sign change problem, over those suggested by Baranchik (1964) and Stein (1966), on the grounds that its constant form over repeated samples makes its small sample properties easier to discover.

Since the ratio $F_{\hat{\beta}} / \beta$ will be used to form a confidence region for β , we wish to set the value of $k_1 = k$ so as to minimize the expectation of $(\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta)$ and hence minimize the radius of the confidence region. The results of Ullah and Ullah can be used to show that, to order σ^4 , the value which does this is $k_1 = k = (p-2)/(n+2)$: See also Corollary 1 to Theorem 1 above.

In order to numerically evaluate the small- σ and large- T approximations, as well as to do the simulation, the value of T was set to 36 and p was

set to 10, implying $n = 26$ and $k = k_1 = .2857$. The approximations also involve the ratio of two population parameters

$$(3.1) \quad \gamma_0 = \frac{\sigma^2}{\theta} = \frac{\sigma^2}{\beta' X' X \beta}$$

Now if we define the population goodness-of-fit to be

$$(3.2) \quad \rho^2 = \frac{\beta' X' X \beta}{\beta' X' X \beta + T \sigma^2},$$

then we can write this ratio in terms of ρ^2 as

$$(3.3) \quad \gamma_0 = \frac{1 - \rho^2}{T}.$$

This is more convenient than the original form because one typically has a more intuitive understanding of what are "large" or "small" values of ρ^2 than of σ^2 and because this form involves only one unknown parameter instead of two. For the computations discussed here ρ^2 was set to .70 so that $\gamma_0 = .0083$.

The X matrix used in the Monte Carlo simulation was derived from a set of 36 observations on 10 regressors which was originally published by Daniel and Wood (1971) and has been used in simulations by Hoerl, Kennard and Baldwin (1975) and Lawless and Wang (1976). Each regressor was first standardized to have zero mean and unit variance so that the 10 slope coefficients could be estimated independently of the intercept. Finally the set of regressors was transformed so that $X'X$ was a diagonal matrix with the following eigenvalues on the main diagonal: .07402, .1479, .2318, .3605, .6571, .9446, 1.051, 1.300, 1.550, 3.686. This transformation simplifies computation but has no effect on F or $F_{\hat{\beta}}$ or any aspect of their sampling distributions. The ratio of the largest eigenvalue to the smallest is 49.8 which indicates a nontrivial degree of multicollinearity among the regressors.

The ten items in the vector β , after the diagonalizing transformation, were each set equal to 1.0. This implies that $\sigma^2 = \gamma_0 \beta' X' X \beta = .119$.

For these parameter values the expectation of the numerator of F is

$$(3.4) \quad E[(b-\beta)' X' X (b-\beta)] = \sigma^2 p = 1.19.$$

The results of Carter (1981) can be used to find the expectation of the numerator of $F_{\hat{\beta}}$, to order σ^8 ,

$$(3.5) \quad E[(\hat{\beta}-\beta)' X' X (\hat{\beta}-\beta)] = 1.11.$$

The generation of random disturbances proceeded in several stages.

First the IMSL (1979) subroutine GGJW was used to generate a pair of pseudo-random numbers, ϵ_1 and ϵ_2 , from a uniform (0,1) distribution. If ϵ_1 and ϵ_2 are such that $(2\epsilon_1 - 1)^2 + (2\epsilon_2 - 1)^2 \geq 1.0$ they are rejected and another pair drawn. If ϵ_1 and ϵ_2 pass this test they are used to construct

$$(3.6) \quad \zeta = \left[\frac{-2 \ln[(2\epsilon_1 - 1)^2 + (2\epsilon_2 - 1)^2]}{(2\epsilon_1 - 1)^2 + (2\epsilon_2 - 1)^2} \right]^{\frac{1}{2}}$$

which is then used to find

$$(3.7) \quad \tau_t = \zeta (2\epsilon_1 - 1)^2$$

$$(3.8) \quad \tau_{t+1} = \zeta (2\epsilon_2 - 1)^2 .$$

The pair τ_t and τ_{t+1} are pseudo-random drawings from $N(0,1)$. This was repeated for $t=1,3,5,\dots,35$ to obtain a vector of τ of 36 pseudo-random drawings from $N(0,1)$. Then $u = \sigma\tau$ and a sample y , of length 36, is obtained from $y = X\beta + u$. The procedure was repeated, using the same X , to obtain 2500 samples, each of size 36, which led to 2500 values of F and $F_{\hat{\beta}}$.

The cumulative frequency distribution was obtained for both F and $F_{\hat{\beta}}$ at intervals of 1%. A few selected points from these cumulative distributions are given in Table 3.1 together with points from the df for the F distribution with 10 and 26 degrees of freedom.

TABLE 3.1Selected Points for Cumulative Distributions

Cumulative Frequency or Probability	$F_{10,26}$	F	Estimated Sampling Error	Raw $F_{\hat{\beta}}$	Adjusted $F_{\hat{\beta}}$
.50	.958721	.935374	-.023347	.874836	.898183
.75	1.365645	1.355979	-.009666	1.264360	1.274026
.90	1.855028	1.863848	-.008820	1.736350	1.727530
.95	2.219718	2.258913	-.039195	2.070771	2.031576
.99	3.094108	3.180216	.086108	2.987363	2.901255

TABLE 3.2Right Tail Points: Approximate and Simulated Distributions

F-value	Cumulative Probability			
	Simulated	Small- σ	Large-T	$F_{10,26}$
1.7275	.9000	.9024	.9753	.8726
1.7863	.9100	.9138	.9823	.8861
1.8462	.9200	.9239	.9879	.8983
1.9155	.9300	.9341	.9927	.9109
1.9664	.9400	.9407	.9953	.9191
2.0316	.9500	.9482	.9978	.9285
2.1753	.9600	.9614	1.0001	.9456
2.3418	.9700	.9724	1.002	.9603
2.5521	.9800	.9818	1.002	.9732
2.9013	.9900	.9907	1.001	.9859

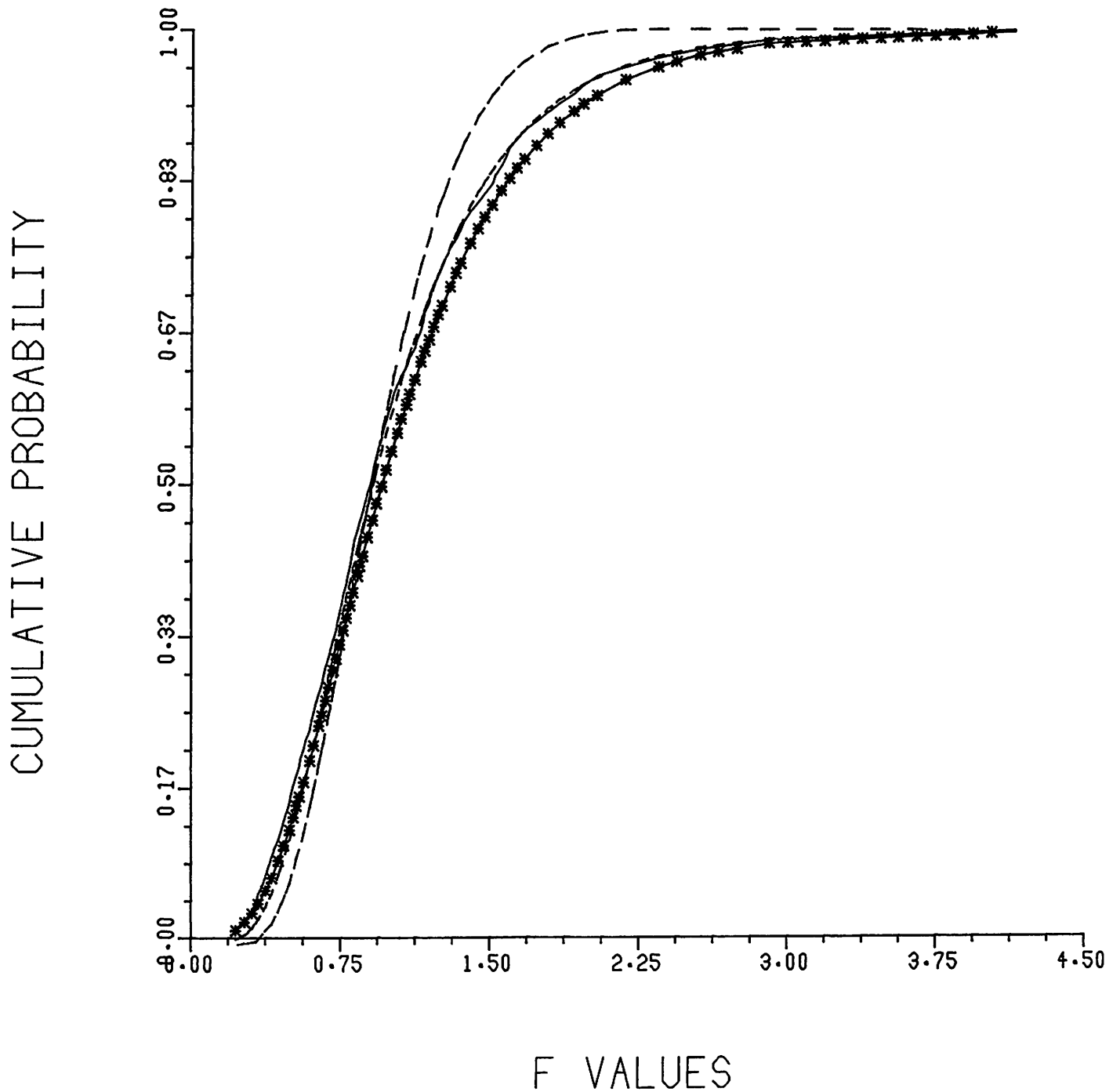
The set of F ratios is a pseudo-random sample of size 2500 from $F_{10,26}$. Therefore, the difference between an $F_{10,26}$ point and an F point in Table 3.1, or the more complete set from which Table 3.1 was drawn, is an estimate of the sampling error inherent in the Monte Carlo simulation. That estimated sampling error was subtracted from each raw $F_{\hat{\beta}}$ point (column 5 of Table 3.1) to obtain adjusted $F_{\hat{\beta}}$ points (column 6 of Table 3.1) which we believe are very nearly equal to the exact $F_{\hat{\beta}}$ points. Both the small- σ and large- T approximations were evaluated at each of the 100 adjusted $F_{\hat{\beta}}$ points.

The four cumulative distributions obtained in this way are plotted in Figure 3.1. The first conclusion to be drawn from this figure is that the small- σ approximate df is nearly identical to the simulated df while the large- T approximate df is noticeably different from the simulated df. This is true even in the right tails of the distributions which are of most interest for inference. Table 3.2 shows this in more detail. This comparison is strictly true for only the simulated distribution. However, we believe that the use of F as a control variate together with the large number of replications makes it true in general.

These results imply that inference should be based on the small- σ approximation. Using the large- T approximation would give a nominal level of confidence which was larger than the true level; that is, the nominal level of significance would be smaller than the true level. Using $F_{10,26}$ as an approximation (i.e. ignoring terms in the approximations beyond the first) would give a nominal level of confidence lower than the true level; that is, the nominal level of significance would be higher than the true level.

Figure 3.1

CUMULATIVE DISTRIBUTIONS
 RHO-SQUARED=.7 T=36 P=10 K=.2857



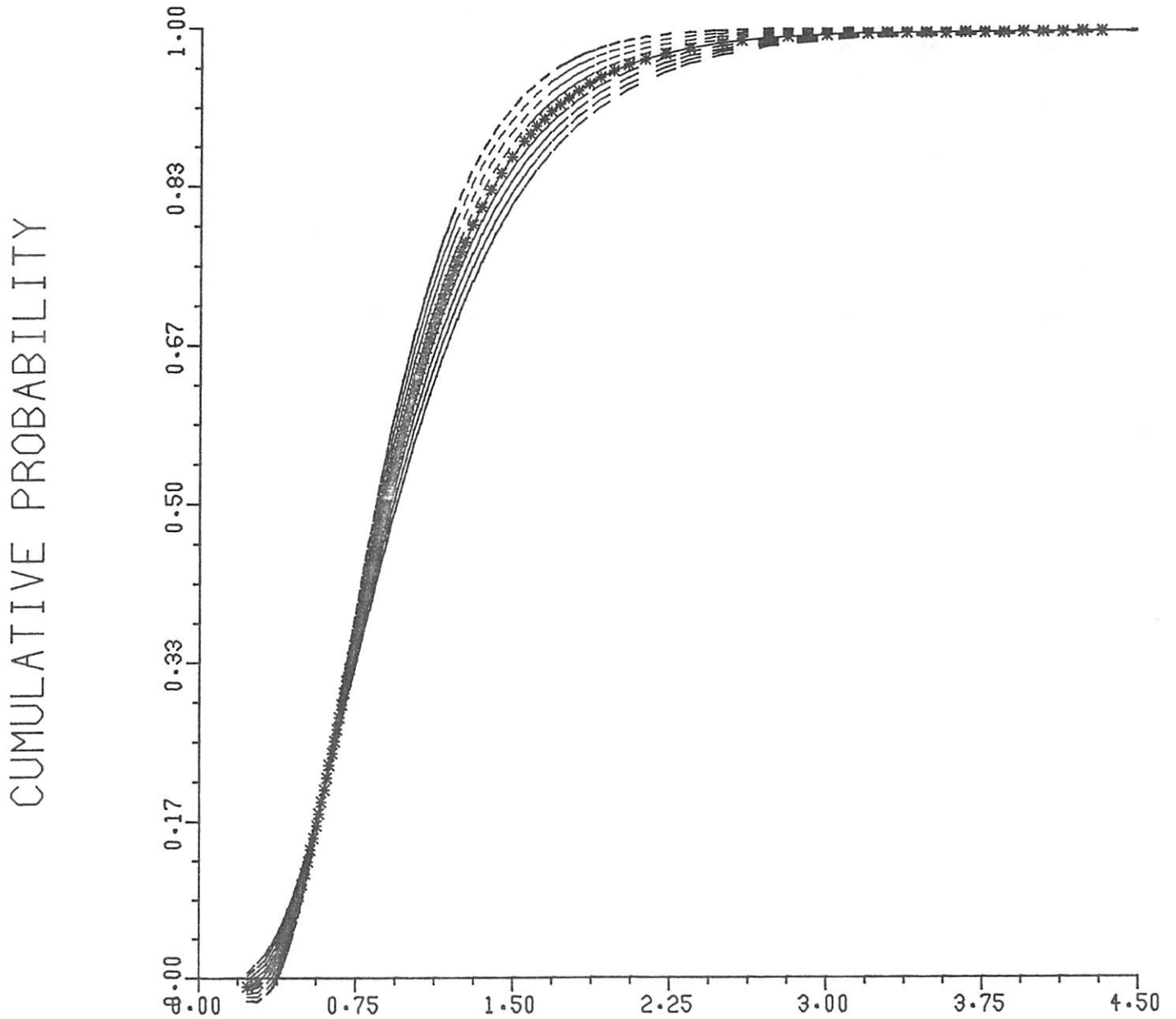
_____ SIMULATED F
 - - - - - APPROXIMATION TO ORDER SIGMA-SQUARED
 - . - . - APPROXIMATION TO ORDER 1/T
 * * * * * THEORETICAL F(10,26)

Although the small- σ approximate df is extremely close to the simulated (and, we believe, the exact) df, it is not an operational basis for inference because it depends upon the unknown parameter $\gamma_0 = (1-\rho^2)/T$. Fortunately, the range of γ_0 is known because the range of ρ^2 is known, that is, $0 \leq \rho^2 \leq 1$ so that $0 \leq \gamma_0 \leq 1/T$. At one extreme, as the population goodness-of-fit becomes closer to 1.0, $\hat{\beta}$ grows closer to b so the df of $F_{\hat{\beta}}$ gets closer to $F_{p,n}$. For example, the df of $F_{10,26}$ in Figure 3.1 is as far down and right as the approximate (and exact) df can shift. As ρ^2 grows smaller, making γ_0 larger, the approximate (and exact) df shifts up and left. To illustrate this behavior the small- σ approximate df is plotted in Figure 3.2 for values of ρ^2 of .1 to .9 in increments of .1. The furthest left curve corresponds to $\rho^2 = .1$, the furthest right to $\rho^2 = .9$. The curves in Figure 3.2 show rather small variation as ρ^2 changes, given T and p. Some more detail from the right tails is shown in Table 3.3.

Empirical researchers may have precise prior beliefs about the value of ρ^2 . In the absence of such priors a good rule of thumb seems to be to set γ_0 to the middle of its range, $1/2T$; that is, set $\rho^2 = .5$. The results shown in Figure 3.2 and Table 3.3 suggest that employing such a rule would yield nominal confidence (or significance) levels close to the true levels so long as the nominal confidence level was high.

Figure 3.2

SMALL-SIGMA CUMULATIVE DISTRIBUTIONS
T=36 P=10 K=.2857 RHO-SQUARED=.1 TO .9



F VALUES

- RHO-SQUARED LESS THAN .5
- ***** RHO-SQUARED EQUAL TO .5
- RHO-SQUARED MORE THAN .5

TABLE 3.3Approximate Right Tail Cumulative Probabilities for Various ρ^2 Values

F-Values	ρ^2 Values								
	.1	.2	.3	.4	.5	.6	.7	.8	.9
1.61	.943	.932	.920	.909	.897	.886	.874	.863	.851
1.91	.980	.972	.965	.957	.949	.941	.933	.926	.918
2.04	.988	.982	.975	.967	.962	.956	.949	.943	.936
2.12	.991	.986	.980	.974	.968	.963	.957	.951	.946
2.34	.997	.993	.988	.984	.980	.976	.972	.968	.964
2.71	1.000	.998	.995	.993	.991	.989	.987	.984	.982

4. DERIVATION OF RESULTS

In this section we provide proofs of the results in Section 2.⁸

4.1 Proof of Theorem 1

Let us rewrite the model (2.1) as

$$(4.1) \quad y = X\beta + \sigma v \quad [u = \sigma v]$$

so that v follows a multivariate normal distribution $N(0, I_T)$ and as σ approaches 0, the disturbance term tends to be small.

Now for the derivation of the result in Theorem 1 we first substitute (4.1) and (2.2) in (2.4) and note that $h = \sigma^2 k v' M v / b' B b$ is at least of order σ^2 . Thus, for sufficiently small σ in Kadane's (1971) sense, using (2.2) we can write (2.3) as

$$(4.2) \quad \hat{\beta} = (I + hA)^{-1} b = (I - hA + h^2 A^2) b + \dots$$

or

$$(4.3) \quad \hat{\beta} - \beta = \sigma (X'X)^{-1} X'v - hA[\beta + \sigma (X'X)^{-1} X'v] + h^2 A^2 [\beta + \sigma (X'X)^{-1} X'v] + \dots$$

where

$$(4.4) \quad h = \sigma^2 k \frac{v' M v}{\beta' B \beta} \left[1 + 2\sigma \frac{\beta' B (X'X)^{-1} X'v}{\beta' B \beta} + \sigma^2 \frac{v' X (X'X)^{-1} B (X'X)^{-1} X'v}{\beta' B \beta} \right]^{-1}.$$

Define

$$(4.5) \quad s = v' M v, \quad z = (X'X)^{-1/2} X'v = Q^{1/2} X'v.$$

Notice that z follows a multivariate normal distribution $N(0, I_p)$ and s has a central χ^2 -distribution with n degrees of freedom. Further, they are stochastically independent.

Expanding the expression in square brackets on the right-hand side of (4.4), we can express the estimation error in (4.3) up to order $O(\sigma^3)$, as

$$(4.6) \quad \hat{\beta} - \beta = \sigma e_1 + \sigma^2 e_2 + \sigma^3 e_3$$

where

$$(4.7) \quad e_1 = Q^{1/2} z, \quad e_2 = -\frac{ks}{\theta} A\beta, \quad e_3 = -\frac{ks}{\theta} D.z; \quad D = (A - \frac{2}{\theta} A\beta\beta' B)Q^{1/2}$$

and $\theta = \beta' B \beta$ is as given in (2.7).

Thus, up to order $O(\sigma^2)$,

$$(4.8) \quad x = \frac{1}{\sigma}(\hat{\beta} - \beta) = e_1 + \sigma e_2 + \sigma^2 e_3$$

and its characteristic function is

$$(4.9) \quad \begin{aligned} \Psi(t) &= E[e^{it'x}] \\ &= E_s E_z [e^{it'(e_1 + \sigma e_2 + \sigma^2 e_3)}] \\ &= E_s [e^{-\sigma i \frac{ks}{\theta} t' A \beta} E_z (e^{i\delta' z} | s)] \end{aligned}$$

where

$$(4.10) \quad \delta' = t'[Q^{1/2} - \frac{\sigma^2 ks}{\theta} D].$$

It is easy to see that

$$(4.11) \quad \begin{aligned} E_z (e^{i\delta' z} | s) &= e^{-1/2 \delta' \delta} \\ &= e^{-\frac{1}{2}(t' Q t - \frac{2\sigma^2 ks}{\theta} t' D Q^{1/2} t)} \\ &= e^{-\frac{1}{2} t' Q t [1 + \frac{\sigma^2 ks}{\theta} t' D Q^{1/2} t]} \end{aligned}$$

up to order $O(\sigma^2)$. Substituting (4.11) in (4.9) we get, up to order $O(\sigma^2)$,

$$(4.12) \quad \Psi(t) = E_s \left[\left(1 - \sigma i \frac{k_s}{\theta} t' A \beta - \frac{\sigma^2}{2} \frac{k_s^2}{\theta^2} (t' A \beta)^2 \right) \left(1 + \sigma^2 \frac{k_s}{\theta} t' D Q^2 t \right) e^{-\frac{1}{2} t' Q t} \right]$$

$$= e^{-\frac{1}{2} t' Q t} \left[1 - \sigma i \frac{k_n}{\theta} t' A \beta + \sigma^2 \frac{k_n}{\theta^2} t' \left\{ \theta D Q^2 - \frac{1}{2} (n+2) k A \beta \beta' A \right\} t \right].$$

Next, using the Inversion Theorem, the multivariate density function of x to order $O(\sigma^2)$ is

$$(4.13) \quad g_o(x) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-it'x} \Psi(t) dt$$

$$= g(x) \left[1 - \sigma \frac{k_n}{\theta} x' Q^{-1} A \beta + \sigma^2 \frac{k_n}{\theta^2} \left\{ \theta \text{tr} A - 2\beta' B A \beta \right. \right.$$

$$\left. - \frac{1}{2} (n+2) k \beta' A Q^{-1} A \beta - x' (\theta Q^{-1} A - 2Q^{-1} A \beta \beta' B - \frac{1}{2} (n+2) k Q^{-1} A \beta \beta' A Q^{-1}) x \right]$$

where we have utilized the following results:

$$\frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-it'x - \frac{1}{2} t' Q t} dt = g(x) = \frac{e^{-\frac{1}{2} x' Q^{-1} x}}{(2\pi)^{\frac{p}{2}} |Q|^{\frac{1}{2}}}$$

$$(4.14) \quad \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \delta_o' t e^{-it'x - \frac{1}{2} t' Q t} dt = -i \delta_o' Q^{-1} x g(x)$$

$$\frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} t' D_o t e^{-it'x - \frac{1}{2} t' Q t} dt = [\text{tr} D_o Q^{-1} - x' Q^{-1} D_o Q^{-1} x] g(x)$$

for any fixed vector δ_o and fixed matrix D_o .

Finally, noting that for any fixed vector δ_1 and the matrix D_1

$$(4.15) \quad \int_{-m_1}^{*} \dots \int_{-m_p}^{*} x' \delta_1 g(x) dx = 0$$

and

$$(4.16) \quad \int_{-m_1}^{*} \dots \int_{-m_p}^{*} x' D_1 x g(x) dx = (G(m^*) - G(-m^*)) \left[\text{tr} D_1 Q - \text{tr} Q^{\frac{1}{2}} D_1 Q^{\frac{1}{2}} \Gamma \right]$$

we get $P[-m_0 < x < m_0]$ as stated in Theorem 1 ($m_x = Q^{-\frac{1}{2}} m_0$).

4.2 Proof of Theorem 2

From (4.7) and (4.8)

$$(4.17) \quad z'x = \frac{1}{\sigma} z'(\hat{\beta} - \beta) = \delta_2' z - \sigma \frac{ks}{\theta} z'AB$$

where

$$(4.18) \quad \delta_2' = z'Q^{\frac{1}{2}} [I_p - \sigma^2 \frac{ks}{\theta} Q^{-\frac{1}{2}} A Q^{\frac{1}{2}} (I_p - \frac{2}{\theta} Q^{-\frac{1}{2}} \beta \beta' B Q^{\frac{1}{2}})].$$

Now the characteristic function of $z'x$ is

$$(4.19) \quad \begin{aligned} \Psi(t) &= E[e^{it(z'x)}] \\ &= E_s E_z [e^{it(\delta_2' z - \sigma \frac{ks z' AB}{\theta})} | s] \\ &= E_s [e^{-it(\sigma \frac{ks z' AB}{\theta})} E_z \{e^{it(\delta_2' z)} | s\}] \end{aligned}$$

It is easy to see that

$$(4.20) \quad E_z \{e^{it(\delta_2' z)} | s\} = e^{-\frac{1}{2} \delta_2' \delta_2 t^2}.$$

But, up to order $O(\sigma^2)$

$$(4.21) \quad \delta_2' \delta_2 = z'Qz - 2\sigma^2 \frac{ks}{\theta} \{z'AQz - \frac{2}{\theta} z'AB\beta' BQz\}.$$

Thus putting (4.21) in (4.20), and then substituting (4.20) in (4.19) along with

$$(4.22) \quad e^{-it(\frac{\sigma ks z' AB}{\theta})} = 1 - \sigma it \frac{ks z' AB}{\theta} - \sigma^2 \frac{t^2}{2} (\frac{ks z' AB}{\theta})^2$$

we find, up to order $O(\sigma^2)$

$$\begin{aligned}
(4.23) \quad \Psi(t) &= e^{-\frac{t^2 q^2}{2}} E_s \left[1 - \sigma it \frac{ks z'_{AB}}{\theta} - \sigma^2 \frac{t^2 k^2 s^2 (z'_{AB})^2}{2\theta^2} \right. \\
&\quad \left. + \sigma^2 \frac{t^2 ks}{\theta} \left\{ z'_{AQ} z - \frac{2}{\theta} z'_{AB} z'_{BQ} z \right\} \right] \\
&= e^{-\frac{t^2 q^2}{2}} \left[1 - \sigma \frac{itknz'_{AB}}{\theta} - \sigma^2 \frac{t^2 kn}{\theta} \left\{ \frac{k(n+2)}{2\theta} (z'_{AB})^2 \right. \right. \\
&\quad \left. \left. + \frac{2}{\theta} z'_{AB} z'_{BQ} z - z'_{AQ} z \right\} \right]
\end{aligned}$$

where $q^2 = z'_{Q} z$ as given in (2.7).

Next, using the Inversion Theorem, the density function of z'_{x} to order $O(\sigma^2)$ is

$$\begin{aligned}
(4.24) \quad g_0(z'_{x}) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it z'_{x}} \Psi(t) dt \\
&= \frac{1}{q} g\left(\frac{z'_{x}}{q}\right) \left[1 - \sigma \frac{knz'_{AB}}{\theta} \frac{z'_{x}}{q^2} + \sigma^2 \frac{kn}{\theta q^2} \left\{ \frac{k(n+2)}{2\theta} (z'_{AB})^2 \right. \right. \\
&\quad \left. \left. + \frac{2}{\theta} z'_{AB} z'_{BQ} z - z'_{AQ} z \right\} \left(\frac{(z'_{x})^2}{q^2} - 1 \right) \right]
\end{aligned}$$

where we have utilized the following results:

$$\begin{aligned}
(4.25) \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{t^2 q^2}{2} - it z'_{x}} dt &= \frac{1}{q} g\left(\frac{z'_{x}}{q}\right) \\
\frac{1}{2\pi} \int_{-\infty}^{\infty} it e^{-\frac{t^2 q^2}{2} - it z'_{x}} dt &= \frac{z'_{x}}{3} g\left(\frac{z'_{x}}{q}\right) \\
\frac{1}{2\pi} \int_{-\infty}^{\infty} (it)^2 e^{-\frac{t^2 q^2}{2} - it z'_{x}} dt &= \left(\frac{(z'_{x})^2}{q^2} - 1 \right) \frac{1}{3} g\left(\frac{z'_{x}}{q}\right).
\end{aligned}$$

Finally, noting that the distribution function $P[z'_{x} \leq m]$ $= \int_{-\infty}^m g_0(z'_{x}) dz'_{x}$, we get the result in Theorem 2.

4.3 Proof of Theorem 3

Let us write by using (4.6) and (4.7)

$$(4.26) \quad (\hat{\beta} - \beta)' W (\hat{\beta} - \beta) = \sigma^2 e_1' W e_1 + 2\sigma^3 e_1' W e_2 + \sigma^4 (e_2' W e_2 + 2e_1' W e_3),$$

up to order $O(\sigma^4)$, where W is any positive definite matrix and

$$(4.27) \quad \begin{aligned} e_1' W e_1 &= z' Q^{\frac{1}{2}} W Q^{\frac{1}{2}} z, \quad e_1' W e_2 = -\frac{ks}{\theta} \beta' A W Q^{\frac{1}{2}} z \\ e_2' W e_2 &= \frac{k^2 s^2}{\theta^2} \beta' A W A \beta, \quad e_1' W e_3 = -\frac{ks}{\theta} z' Q^{\frac{1}{2}} W A Q^{\frac{1}{2}} \left(I_p - \frac{2}{\theta} Q^{-\frac{1}{2}} \beta \beta' B Q^{\frac{1}{2}} \right) z. \end{aligned}$$

Substituting (4.27) in (4.26) and dividing both sides of (4.26) by $\sigma^2 s$ we can write the improved F ratio in (2.26) as

$$(4.28) \quad \frac{P}{n} F_{\hat{\beta}} = F_{\hat{\beta}}^0 = \frac{(\hat{\beta} - \beta)' W (\hat{\beta} - \beta)}{\sigma^2 s} = z' P_1 z + \sigma \delta_3' z + \sigma^2 s \beta' P_2 \beta$$

where

$$(4.29) \quad \begin{aligned} P_1 &= H_1 + \sigma^2 H_2; \quad H_1 = \frac{Q^{\frac{1}{2}} W Q^{\frac{1}{2}}}{s}, \quad H_2 = -\frac{2k}{\theta} Q^{\frac{1}{2}} W A Q^{\frac{1}{2}} \left[I_p - \frac{2}{\theta} Q^{-\frac{1}{2}} \beta \beta' B Q^{\frac{1}{2}} \right] \\ P_2 &= \frac{k^2}{\theta^2} A W A, \quad \delta_3' = -\frac{2k}{\theta} \beta' A W Q^{\frac{1}{2}}. \end{aligned}$$

Now the characteristic function of $F_{\hat{\beta}}^0$ is

$$(4.30) \quad \begin{aligned} \phi(t) &= E e^{it F_{\hat{\beta}}^0} = E_s E_z \left[e^{it(z' P_1 z + \sigma \delta_3' z + \sigma^2 s \beta' P_2 \beta)} \Big|_s \right] \\ &= E_s \left[e^{it \sigma^2 s \beta' P_2 \beta} E_z \left\{ e^{it(z' P_1 z + \sigma \delta_3' z)} \Big|_s \right\} \right]. \end{aligned}$$

Using (4.29), it is easy to see that up to order $O(\sigma^2)$

$$\begin{aligned}
(4.31) \quad E_z \{ e^{it(z'P_1z + \sigma\delta_3'z)} \mid s \} &= |I_p - 2itP_1|^{-\frac{1}{2}} e^{-\frac{t^2}{2} \sigma^2 \delta_3'(I_p - 2itP_1)^{-1} \delta_3} \\
&= |J_1|^{-\frac{1}{2}} [1 + it\sigma^2 \text{tr } J_1^{-1} H_2 - \frac{t^2}{2} \sigma^2 \delta_3' J_1^{-1} \delta_3],
\end{aligned}$$

where

$$(4.32) \quad J_1 = I_p - 2it H_1,$$

and we use the result that for any square matrix A , $|I-A|^{-\frac{1}{2}} \approx 1 + \frac{1}{2} \text{tr } A$. Also,

up to order $O(\sigma^2)$

$$(4.33) \quad e^{it\sigma^2 s\beta' P_2 \beta} = 1 + it\sigma^2 s\beta' P_2 \beta.$$

Thus, substituting (4.31) and (4.32) in (4.30) we get, up to order $O(\sigma^2)$,

$$\begin{aligned}
(4.34) \quad \phi(t) &= E_s [(1 + it\sigma^2 s\beta' P_2 \beta) |J_1|^{-\frac{1}{2}} (1 + it\sigma^2 \text{tr } J_1^{-1} H_2 - \frac{t^2}{2} \sigma^2 \delta_3' J_1^{-1} \delta_3)] \\
&= E_s (|J_1|^{-\frac{1}{2}}) + it\sigma^2 \text{tr } H_2 E_s (|J_1|^{-\frac{1}{2}} J_1^{-1}) - \frac{t^2}{2} \sigma^2 \delta_3' E_s (|J_1|^{-\frac{1}{2}} J_1^{-1}) \delta_3 \\
&\quad + it\sigma^2 \beta' P_2 \beta E_s (s |J_1|^{-\frac{1}{2}})
\end{aligned}$$

Considering $W = X'X$ in (4.28), we note that $E_s |J_1|^{-\frac{1}{2}} = E_s (1 - \frac{2it}{s})^{-\frac{p}{2}}$
 $= \phi(F_{p,n}^0, t)$; that is the characteristic function of a central $\frac{p}{n} F = F^0$ ratio
with the p and n degrees of freedom, see Phillips (1982, p. 262). Similarly,
it can easily be verified that $E_s (|J_1|^{-\frac{1}{2}} J_1^{-1}) = E_s [(1 - \frac{2it}{s})^{-\frac{(p+2)}{2}}] I_p$
where $E_s (1 - \frac{2it}{s})^{-\frac{(p+2)}{2}} = \phi(F_{p+2,n}^0, t)$. Also, $E_s \frac{s}{n} |J_1|^{-\frac{1}{2}} = E_s \frac{s}{n} (1 - \frac{2it}{s})^{-\frac{p}{2}}$
 $= \phi(F_{p,n+2}^0, t)$. Using these results in (4.34) we can write

$$(4.35) \quad \phi(t) = \phi(F_{p,n}^0, t) + it\sigma^2(\text{tr } H_2) \phi(F_{p+2,n}^0, t) - \left(\frac{t^2}{2}\right) \sigma^2 \delta_3' \delta_3 \phi(F_{p+2,n}^0, t) \\ + it\sigma^2 n(\beta' P_2 \beta) \phi(F_{p,n+2}^0, t).$$

Now, using the Inversion Theorem, the density function of the F ratio in (4.28), with $W = X'X$, can be written after simplification as

$$(4.36) \quad g_o(F_{\beta}^0) = g(F_{p,n}^0) + \eta[g(F_{p,n+4}^0) - g(F_{p-2,n+6}^0)] + \eta_{1,1}[g(F_{p+2,n+2}^0) \\ - g(F_{p,n+4}^0)] + \frac{2\eta_{1,2}}{(n+2)} \left[\left(\frac{n}{2}+1\right)g(F_{p+2,n+4}^0) + \left(\frac{n}{2}+3\right)g(F_{p-2,n+8}^0) \right. \\ \left. - 2\left(\frac{n}{2}+2\right)g(F_{p,n+6}^0) \right]$$

where η , $\eta_{1,1}$ and $\eta_{1,2}$ are as given in (2.28). In obtaining the result in (4.36) we used the following results (for any positive scalar F^0):

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itF^0} \phi(F_{p,n}^0, t) dt = g(F_{p,n}^0) = \frac{\Gamma(\frac{p+n}{2} + 1)}{\Gamma(\frac{p+1}{2})\Gamma(\frac{n}{2})} \frac{F^{(p+2)/2 - 1}}{(1 + F^0)^{(p+2+n)/2}}$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} ite^{-itF^0} \phi(F_{p,n+2}^0, t) dt = \frac{(n+2)(n+4)}{2(p+n+2)} [g(F_{p,n+4}^0) - g(F_{p-2,n+6}^0)]$$

$$(4.37) \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} ite^{-itF^0} \phi(F_{p+2,n}^0, t) dt = \frac{n(n+2)}{2(p+n+2)} [g(F_{p+2,n+2}^0) - g(F_{p,n+4}^0)]$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} t^2 e^{-itF^0} \phi(F_{p+2,n}^0, t) dt = \frac{n(n+2)}{(p+n+2)(p+n+4)} \left[\left(\frac{n+2}{2}\right) \left(\frac{n+1}{2}\right) \{g(F_{p+2,n+4}^0) \right. \right. \\ \left. \left. - g(F_{p,n+6}^0)\right\} - \left(\frac{n+3}{2}\right) \left(\frac{n+2}{2}\right) \{g(F_{p,n+6}^0) \right. \\ \left. \left. - g(F_{p-2,n+8}^0)\right\} \right].$$

Finally, noting that the distribution function of $F_{\hat{\beta}}^0$ is $P[F_{\hat{\beta}}^0 \leq m] = \int_0^m g_0(F_{\hat{\beta}}^0) dF_{\hat{\beta}}^0$, and observing that $P(nF_{\hat{\beta}}^0 \leq m) = P(pF_{\hat{\beta}} \leq m)$ we get the result stated in Theorem 3.

FOOTNOTES

¹This paper is prepared for the special issue of the Journal of Econometrics on Shrinkage and Pre-test Estimators. The research supports to A. Ullah from the NSERC and Air Force Office of Scientific Research at the Center for Multivariate Analysis, University of Pittsburgh, and to R. Carter from SSHRCC, are gratefully acknowledged. The authors are thankful to D. E. A. Giles for helpful discussions and T. Peters for research assistance.

²The multivariate density function of $x = (\hat{\beta} - \beta) / \sigma$ is given in (4.13). Using this, one can easily verify that $E\hat{\beta}$ and $E(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)$ are identical with the corresponding results in Ullah et al (1983) for the normal case. Another point to be noted is that when A in (2.3) is not a symmetric matrix, the results stated in Theorems 1 to 3 remain the same with $\beta'A$ replaced by $\beta'A'$. Since in almost all the shrinkage estimators known in the literature A is a symmetric matrix we have presented here results for this case only.

³Given the distribution of $z'x$ in (2.10) the density function of $z'x$ can be obtained (also see (4.24)). Using this density one can also verify the results indicated in footnote 2.

⁴The result in (2.23) compares with that in Ullah (1982).

⁵The result in this section easily extends to the case of testing linear restrictions $R\beta = r$, where R is an $m \times p$ matrix and r is an $m \times 1$ vector of known constants. This is because we can write $R\beta = r$ as $R\beta = R\beta_0$ or $R(\beta - \beta_0) = 0$, where β_0 is a fixed vector, and hence the F statistic as $(b - \beta_0)'W(b - \beta_0) / \hat{u}'\hat{u}$, where $W = R'[R(X'X)^{-1}R']^{-1}R$. In fact, the proof in Section 4.3 has been

outlined for any positive definite matrix W . We have considered here

$R = I_p (W = X'X)$ for the sake of simplicity in exposition.

⁶Note that $pF_{\hat{\beta}} = (\hat{\beta} - \beta)' [Asy.V(\hat{\beta})]^{-1} (\hat{\beta} - \beta)$ is the well-known Wald (1943) test statistic, where $Asy.V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ is the asymptotic variance of $\hat{\beta}$ and it is to be evaluated at the unrestricted estimator of σ^2 . We have taken $\hat{\sigma}^2 = \hat{u}'\hat{u}/n$ where $\hat{u} = y - X\hat{\beta}$. An alternative to this, that is $\tilde{\sigma}^2 = \tilde{u}'\tilde{u}/n$, where $\tilde{u} = y - X\hat{\beta}$, was also considered. In this case the result in (2.27) of Theorem 3 remained the same with $[tr A - 2\beta'BA\beta/\theta]$ on the right-hand side of $\eta_{1,1}$ (see (2.28)) replaced by $[tr A - 2\beta'BA\beta/\theta + k\beta'AX'XAB/2\theta]$. Further, the numerical findings in Section 3 would be changed only slightly by using $\tilde{\sigma}^2$, or the approximate $V(\hat{\beta})$ given in Ullah and Ullah (1978) instead of $Asy.V(\hat{\beta})$.

⁷A less interesting case of the distribution of the statistic $(\hat{\beta} - \beta)' X'X(\hat{\beta} - \beta)/\sigma^2$ was also examined and it was found that while the small- σ expansion of this statistic was in terms of the central χ^2 distribution and hence different with (2.29), the large- T expansion was the same as in (2.30). This implies that the replacement of σ^2 by $\hat{u}'\hat{u}/n$ did not change the large- T expansion up to the order of approximation considered.

⁸We essentially derive Edgeworth-type expansions. For details about these expansions, see Sargan (1976).

REFERENCES

- Baranchik, A. J., 1964, "Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution," Technical Report 51 (Stanford Department of Statistics).
- Carter, R. A. L., 1981, "Improved Stein-Rule Estimator for Regression Problems," Journal of Econometrics 17, 113-123.
- Daniel, C. and F. Wood, 1971, Fitting Equations to Data (Wiley: New York)
- Hoerl, A. E., R. W. Kennard and K. F. Baldwin, 1975, "Ridge Regression: Some Simulations," Communications in Statistics, 108-115.
- James, W. and C. Stein, 1961, "Estimation with Quadratic Loss," Proceedings of Fourth Berkeley Symposium in Mathematical Statistics and Probability 1, 1020-1034.
- Judge, G. G., and M. E. Bock, 1978, The Statistical Implications of Pretest and Stein-Rule Estimators in Econometrics (North-Holland: New York).
- Judge, G. G., W. E. Griffiths, R. C. Hill and T. C. Lee, 1980, The Theory and Practice of Econometrics (Wiley: New York).
- Kadane, J. B., 1971, "Comparison of k-class Estimators When the Disturbances are Small," Econometrica (Sept.), 723-737.
- Lawless, J. F. and P. Wang, 1976, "A Simulation Study of Ridge and Other Regression Estimators," Communications in Statistics, Ser. A, 307-323.
- Nagar, A. L., 1959, "The Bias and Moment Matrix of the General k-class Estimators of Parameters in Simultaneous Equations," Econometrica (Oct.), 575-595.
- Phillips, P. C. B., 1982, "The True Characteristic Function of the F Distribution," Biometrika 69, 261-264.

- Rao, C. R., 1973, Linear Statistical Inference and its Applications (Wiley: New York).
- Sargan, J. D., 1976, "Econometric Estimators and the Edgeworth Approximation," Econometrica 44, 421-428.
- Srivastava, V. K., T. D. Dwivedi, M. Belinsky and R. Tiwari, 1980, "A Numerical Comparison of Exact, Large-Sample and Small-Disturbance Approximations of Properties of k-class Estimators," International Economic Review (Feb.), 249-252.
- Stein, C., 1966 "An Approach to the Recovery of Inter-Block Information on Balanced Incomplete Block Designs" in F. N. David (ed.), Research Papers in Statistics (Wiley: New York), 351-366.
- Ullah, A., 1974, "On the Sampling Distribution of Improved Estimators for Coefficients in Linear Regression," Journal of Econometrics, 143-150.
- Ullah, A., 1980, "The Exact, Large-Sample and Small-Disturbance Conditions of Dominance of Biased Estimators in Linear Models," Economics Letters, 339-344.
- Ullah, A., 1982, "The Approximate Distribution Function of the Stein-Rule Estimator," Economics Letters, 305-308.
- Ullah, A., and S. Ullah, 1978, "Double k-class Estimators of Coefficients in Linear Regression," Econometrica (May).
- Ullah, A., V. K. Srivastava and R. Chandra, 1983, "Properties of Shrinkage Estimators in Linear Regression When Disturbances Are Not Normal," Journal of Econometrics (Apr.), 389-402.
- Vinod, H. D. and A. Ullah, 1981, Recent Advances in Regression Methods (Marcel Dekkar: New York).
- Wald, A., 1943, "Tests of Hypotheses Concerning Several Parameters When the Number of Observations is Large," Transactions of the American Mathematical Society 54, 426-482.