COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 682

THE EXACT DISTRIBUTION OF THE STEIN-RULE ESTIMATOR

Peter C. B. Phillips

November 15, 1983

THE EXACT DISTRIBUTION OF THE STEIN-RULE ESTIMATOR

by

P. C. B. Phillips*

Cowles Foundation for Research in Economics
Yale University

"We attach too much importance to unbiased estimates. Sometimes biased estimates are better than the best unbiased ones. An example was given by Halmos" Abraham Wald (1946).

0. ABSTRACT

This paper derives the exact density of the Stein-rule estimator in the setting of the general linear regression. The derivation is facilitated by the author's development of new algebraic methods that involve an extension of the Weyl calculus. General formulae for the moments of the estimator are also provided.

---

*My thanks go to Glena Ames for her skill and effort in typing the manuscript and to the NSF for support under Grant No. SES 8218792.

1. INTRODUCTION

The idea that biased estimators may dominate the best classical procedures seems to have occurred to many early researchers in mathematical statistics. The quotation that heads this article was unearthed by the author in the unpublished research archives of the Cowles Commission and gives convincing evidence that at least two prominent scholars in mathematics and mathematical statistics had thought seriously about this idea long before 1950. Nevertheless, it was the article by James and Stein (1961) that really excited professional interest in this subject and which laid the foundation for subsequent work on the class of biased estimators which now come under the generic name of the Stein-rule family.

In view of the dominance of the Stein-rule estimator over classical methods in multivariate settings, knowledge of the sampling properties of this estimator is very important to us if we are to properly understand its behavior. Unfortunately, and, in spite of the attention that the Stein-rule estimator has received in the theoretical literature, its sampling properties are still very imperfectly understood. Some scalar summary characteristics are known and these are useful in determining the dominance property; but it would seem to be of greater importance to study the whole sampling distribution of the Stein-rule estimator, particularly when it has been advanced as a serious contender for estimation problems.

Undoubtedly the nonlinearity of the Stein-rule estimator has been the major obstacle in the development of a complete distribution theory. The simplest way to proceed is, of course, by the use of Edgeworth expansions or by approximating distributions that utilize known moment formulae. The first approach has been facilitated by recent advances in the theory

of asymptotic expansions and is currently being pursued in other work (see, for instance, Ullah (1982)).

The object of the present paper is to show that an exact theory is also well within reach. Specifically, this paper provides a mathematical derivation of the exact probability density function (p.d.f.) of the Stein-rule estimator in the setting of the general linear regression. Moment formulae are then deduced directly from our general result.

These derivations are made possible by the deployment of new algebraic methods that are developed in the article. These methods involve the use of fractional calculus and have many exciting potential applications in statistical distribution theory. They seem especially appropriate in the case of statistics like the Stein-rule family that embody nonlinearities which cannot be treated by more traditional algebraic methods. Matrix variate extensions of the techniques given here are also being developed by the author in the context of other work.

2. THE MODEL AND NOTATION

We will work with the linear regression model

$$y = X\beta + u \tag{1}$$

where y is a vector of T observations on a dependent variable, X is a $T \times m$ observation matrix of full rank $m < T$ of nonrandom independent variables and u is a vector of disturbances that is assumed to be distributed as $N(0, \sigma^2 I)$. To simplify the formulae that follow without loss of generality we assume that orthonormalizing transformations have already been performed which make the reduction to canonical form $T^{-1}X'X = I$. Results that apply directly to the untransformed system are

of asymptotic expansions and is currently being pursued in other work (see, for instance, Ullah (1982)).

The object of the present paper is to show that an exact theory is also well within reach. Specifically, this paper provides a mathematical derivation of the exact probability density function (p.d.f.) of the Stein-rule estimator in the setting of the general linear regression. Moment formulae are then deduced directly from our general result.

These derivations are made possible by the deployment of new algebraic methods that are developed in the article. These methods involve the use of fractional calculus and have many exciting potential applications in statistical distribution theory. They seem especially appropriate in the case of statistics like the Stein-rule family that embody nonlinearities which cannot be treated by more traditional algebraic methods. Matrix variate extensions of the techniques given here are also being developed by the author in the context of other work.

2. THE MODEL AND NOTATION

We will work with the linear regression model

$$y = X\beta + u \tag{1}$$

where y is a vector of T observations on a dependent variable, X is a $T \times m$ observation matrix of full rank $m < T$ of nonrandom independent variables and u is a vector of disturbances that is assumed to be distributed as $N(0, \sigma^2 I)$. To simplify the formulae that follow without loss of generality we assume that orthonormalizing transformations have already been performed which make the reduction to canonical form $T^{-1}X'X = I$. Results that apply directly to the untransformed system are

of asymptotic expansions and is currently being pursued in other work (see, for instance, Ullah (1982)).

The object of the present paper is to show that an exact theory is also well within reach. Specifically, this paper provides a mathematical derivation of the exact probability density function (p.d.f.) of the Stein-rule estimator in the setting of the general linear regression. Moment formulae are then deduced directly from our general result.

These derivations are made possible by the deployment of new algebraic methods that are developed in the article. These methods involve the use of fractional calculus and have many exciting potential applications in statistical distribution theory. They seem especially appropriate in the case of statistics like the Stein-rule family that embody nonlinearities which cannot be treated by more traditional algebraic methods. Matrix variate extensions of the techniques given here are also being developed by the author in the context of other work.

2. THE MODEL AND NOTATION

We will work with the linear regression model

(1) $$y = X\beta + u$$

where y is a vector of T observations on a dependent variable, X is a $T \times m$ observation matrix of full rank $m < T$ of nonrandom independent variables and u is a vector of disturbances that is assumed to be distributed as $N(0, \sigma^2 I)$. To simplify the formulae that follow without loss of generality we assume that orthonormalizing transformations have already been performed which make the reduction to canonical form $T^{-1}X'X = I$. Results that apply directly to the untransformed system are

then obtained by transforming $\beta \to (X'X/T)^{-1/2}\beta$.

The Stein-rule estimator of β in (1) is given by:

$$(2) \qquad r = \left[1 - \frac{a}{T}\left(\frac{s}{b'b}\right)\right]b$$

where $b = T^{-1}X'y$ is the least squares estimator of β , $s = y'(I - P_X)y$, $P_X = X(X'X)^{-1}X'$ and a is a scalar constant. The dominance criterion of r over b under quadratic loss requires $0 < a \leq 2(m-2)/(T-m+2)$ and $m \geq 3$ (see, for instance, Judge and Bock (1978)).

3. THE EXACT DENSITY OF THE STEIN-RULE ESTIMATOR

We start by considering the characteristic function (c.f.) of r . Since b is $N(\beta, (\sigma^2/T)I)$ and s/σ^2 is independent and χ^2_{T-m} we have:

$$(3) \quad cf(t) = E(e^{it'r}) = \int \exp\{it'b - i(as/Tb'b)t'b\}pdf(b)pdf(s)dbds \ .$$

To reduce the integral in (3) we introduce a new operator calculus. Specifically, we now write:

$$(4) \quad \exp\{it'b - i(as/Tb'b)t'b\}$$

$$= \exp\{it'\partial x - i(as/T\Delta_x)t'\partial x\}e^{x'b}\Big|_{x=0}$$

where ∂x denotes the vector operator $\partial/\partial x$ and Δ_x is the Laplacian operator $\partial x'\partial x$. Negative powers of Δ_x are interpreted in (4) by appealing to the following definition of the (possibly fractional) operator $\Delta_x^{-\alpha}$ $(\alpha > 0)$:

$$(5) \quad \Delta_x^{-\alpha}f(x) = \frac{1}{\Gamma(\alpha)}\int_0^\infty [\exp\{-\Delta_x w\}f(x)]w^{\alpha-1}dw \ , \quad \alpha > 0$$

provided the integral converges. This definition is inspired by the form of the gamma integral and, if Δ_x were the simple differential operator d/dx for a scalar variable x, then (5) would correspond to one form of the Weyl fractional integral. As it stands (5) extends the Weyl calculus by the use of fractional powers of polynomials such as Δ_x in the operator ∂x. Further generalizations along these lines are possible and some of these have been developed in other work by the author (1983). Readers unfamiliar with fractional operators are referred to Ross (1974) for an introduction to the subject.

Using (4) we may now write (3) in the form

$$(6)\quad cf(t) = \int_s [\exp\{it'\partial x - i(as/T\Delta_x)t'\partial x\}\int_b e^{x'b}pdf(b)db]_{x=0}\,pdf(s)ds$$

since the integral over b-space converges uniformly. Upon evaluation we find:

$$cf(t) = \int_s\left[\exp\{it'\partial x - i(as/T\Delta_x)t'\partial x\}e^{x'\beta+\sigma^2x'x/2T}\right]_{x=0} pdf(s)ds$$

$$(7)\quad = \left[\exp(it'\partial x)\int_0^\infty \exp\{-i(as/T\Delta_x)t'\partial x\}2^{-(T-m)/2}\Gamma((T-m)/2)^{-1}\sigma^{-(T-m)}e^{-s/2\sigma^2}s^{(T-m)/2-1}\right.$$
$$\left. e^{x'\beta+\sigma^2x'x/2T}\right]_{x=0}$$

$$(8)\quad = \left[\exp(it'\partial x)\{1 + 2i(a\sigma^2t'\partial x/T\Delta_x)\}^{-(T-m)/2}e^{x'\beta+\sigma^2x'x/2T}\right]_{x=0}.$$

The order of differentiation and integration may be interchanged once again in view of the uniform convergence of the gamma integral in (7). If we set $t = sh$ for an arbitrary real scalar s and m-vector h we deduce from (8) the characteristic function of the linear form $y = h'r$:

$$(9) \quad cf(s) = \left[\exp(ish'\partial x)\{1 + 2is(a\sigma^2 h'\partial x/T\Delta_x)\}^{-(T-m)/2} e^{x'\beta+\sigma^2 x'x/2T}\right]_{x=0}$$

We observe that, since $\exp(x'\beta + \sigma^2 x'x/2T)$ is analytic,

$$\exp(ish'\partial x)\exp(x'\beta+\sigma^2 x'x/2T) = \exp\{\beta'(x+ish) + \sigma^2(x+ish)'(x+ish)/2T\}$$

and (9) becomes:

$$(10) \quad cf(s) = \exp(is\beta'h - \sigma^2 s^2 h'h/2T)\left[(1+2is\zeta_x)^{-(T-m)/2}\exp\{x'(\beta+is\sigma^2 h/T)+\sigma^2 x'x/2T\}\right]_{x=0}$$

where $\zeta_x = a\sigma^2 h'\partial x/T\Delta_x$.

Inversion of (9) now yields the p.d.f. of y in the integral form:

$$(11) \quad pdf(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isy}\exp(is\beta'h - \sigma^2 s^2 h'h/2T)\Big[(1+2is\zeta_x)^{-(T-m)/2}$$

$$\cdot\exp\{x'(\beta+is\sigma^2 h/T) + \sigma^2 x'x/2T\}\Big]_{x=0} ds \; .$$

Using the integral representation of $(1+2is\zeta_x)^{-(T-m)/2}$ in (11), and interchanging the order of the operators, which is permissable because of the uniform convergence of the integrals and the series, we obtain:

$$\text{pdf}(y) = \frac{1}{\Gamma\left(\frac{T-m}{2}\right)} \int_0^\infty e^{-w} w^{(T-m)/2-1} \frac{1}{2\pi} \int_{-\infty}^\infty e^{-2is\zeta_x w}$$

$$\cdot \exp\{-is(y-\beta'h-\sigma^2 x'h/T) - \sigma^2 s^2 h'h/2T\} ds dw \; e^{x'\beta+\sigma^2 x'x/2T}\Big|_{x=0}$$

$$= \frac{1}{\Gamma\left(\frac{T-m}{2}\right)} \sum_{k=0}^\infty \frac{1}{k!} \int_0^\infty e^{-w} w^{(T-m)/2+k-1} dw (-2\zeta_x)^k \frac{1}{2\pi} \int_{-\infty}^\infty (is)^k$$

$$\cdot \exp\{-is(y-\beta'h-\sigma^2 x'h/T) - \sigma^2 s^2 h'h/2T\} ds \; e^{x'\beta+\sigma^2 x'x/2T}\Big|_{x=0}$$

$$= \sum_{k=0}^\infty \frac{\left(\frac{T-m}{2}\right)_k}{k!} \Bigg[(-2\zeta_x)^k \Bigg[(\partial z)^k \frac{1}{2\pi} \int_{-\infty}^\infty$$

$$\cdot \exp\{-is(y-\beta'h-\sigma^2 x'h/T-z) - \sigma^2 s^2 h'h/2T\} ds \Bigg]_{z=0} e^{x'\beta+\sigma^2 x'x/2T} \Bigg]_{x=0}$$

$$= \sum_{k=0}^\infty \frac{\left(\frac{T-m}{2}\right)_k}{k!} \Bigg[-(2\zeta_x)^k \Bigg[(\partial z)^k (2\pi\sigma^2 h'h/T)^{-1/2}$$

$$\cdot \exp\{-T(y-\beta'h-\sigma^2 x'h/T-z)^2/2\sigma^2 h'h\} \Bigg]_{z=0} e^{x'\beta+\sigma^2 x'x/2T} \Bigg]_{x=0} .$$

Thus, the p.d.f. of $y = h'r$ is given by:

$$\text{(12)} \quad \text{pdf}(y) = \left(\frac{T}{2\pi\sigma^2 h'h}\right)^{1/2} \sum_{k=0}^\infty \frac{\left(\frac{T-m}{2}\right)_k}{k!} \Bigg[(-2\zeta_x)^k$$

$$\cdot \Bigg[(\partial z)^k \exp\{-T(y-\beta'h-\sigma^2 x'h/T-z)^2/2\sigma^2 h'h\} \Bigg]_{z=0} e^{x'\beta+\sigma^2 x'x/2T} \Bigg]_{x=0}$$

where $\zeta_x = a\sigma^2 h'\partial x/T\Delta_x$.

Note that when $a = 0$ only the first term of the series in (12) is nonvanishing and we are left with

$$(13)\quad \left(\frac{T}{2\pi\sigma^2 h'h}\right)^{1/2} e^{-T(y-h'\beta)/2\sigma^2 h'h}$$

corresponding to the density of $h'b$. Note also that the analogue of (12) for the case of regressors that are not orthonormalized is obtained by transforming $h \rightarrow (X'X/T)^{-1/2}h$.

When a is of $O(T^{-1})$ in (2), as the dominance criterion requires, (13) gives the crude asymptotic approximation to (12). Higher order approximations may be obtained from (12) by selecting more terms of the series according to their order of magnitude after application of the operator ζ_x . The algebra quickly becomes quite heavy in this process.

4. MOMENT FORMULAE

Exact formulae for the moments of $h'r$ can be deduced quite simply from (12). We note that

$$\left(\frac{T}{2\pi\sigma^2 h'h}\right)^{1/2} \int_{-\infty}^{\infty} y^p \exp\{-T(y-\beta'h-\sigma^2 x'h/T-z)^2/2\sigma^2 h'h\}dy$$

$$= \sum_{j=0}^{[p/2]} \binom{p}{2j} (\beta'h-\sigma^2 x'h/T-z)^{p-2j} \left(\frac{\sigma^2 h'h}{T}\right)^j$$

where [ ] denotes the integer part of its argument. Then, in view of the uniform convergence of the series which defines the moment $E(y^p)$, we integrate termwise and deduce the following general expression:

$$(14)\quad E(y^p) = \sum_{k=0}^{\infty} \frac{\left(\frac{T-m}{2}\right)_k}{k!}\Bigg[(-2\zeta_x)^k(\partial z)^k \sum_{j=0}^{[p/2]}\binom{p}{2j}$$

$$\cdot(\beta'h-\sigma^2x'h/T-z)^{p-2j}\left(\frac{\sigma^2h'h}{T}\right)^j e^{x'\beta+\sigma^2x'x/2T}\Bigg]_{\substack{x=0\\z=0}} .$$

This general formula may be written out more explicitly by using rules for composite function differentiation. But for low order moments these are not needed. Thus, when $p = 1$ we find directly that

$$(15)\quad E(y) = \beta'h - (T-m)\left[\zeta_x e^{x'\beta+\sigma^2x'x/2T}\right]_{x=0} .$$

Consider

$$\left[\zeta_x e^{x'\beta+\sigma^2x'x/2T}\right]_{x=0} = \frac{a\sigma^2}{T}e^{-T\beta'\beta/2\sigma^2}\left[h'\partial x\Delta_x^{-1}\exp\left\{\frac{\sigma^2}{2T}\left(x+\frac{T\beta}{\sigma^2}\right)'\left(x+\frac{T\beta}{\sigma^2}\right)\right\}\right]_{x=0}$$

$$(16)\qquad = \frac{a\sigma^2}{T}e^{-\theta}\left[h'\partial w\Delta_w^{-1}e^{\sigma^2w'w/2T}\right]_{w=T\beta/\sigma^2}$$

where $\theta = T\beta'\beta/2\sigma^2$ and $w = x+t\beta/\sigma^2$.

Now

$$(17)\quad \Delta_w^{-1}e^{\sigma^2w'w/2T} = \int_0^{\infty} e^{-s\Delta_w}e^{\sigma^2w'w/2T}ds$$

$$= \int_0^{\infty}\sum_{\ell=0}^{\infty}\frac{(-s)^{\ell}}{\ell!}\Delta_w^{\ell}e^{\sigma^2w'w/2T}ds$$

and

$$\Delta_w^{\ell} e^{\sigma^2 w'w/2T} = \sum_{q=0}^{\infty} \frac{(\sigma^2/2T)^q}{q!} \Delta_w^{\ell}(w'w)^q$$

$$= \sum_{u=0}^{\infty} \frac{(\sigma^2/2T)^{\ell+u}}{(\ell+u)!} 2^{2\ell}(w'w)^u (u+1)_{\ell}\left(\frac{n}{2}+u\right)_{\ell}$$

where we use the fact that:

$$(18) \quad \Delta_w^{\ell}(w'w)^q = \begin{cases} 0 & , \quad q < \ell \\ 2^{2\ell}(w'w)^u(u+1)_{\ell}\left(\frac{n}{2}+u\right)_{\ell} & , \quad q = \ell+u \geq \ell \end{cases}$$

as can be verified by direct differentiation. Next, we apply the operator $h'\partial w$ that appears in (16). We have

$$h'\partial w e^{-s\Delta_w} e^{\sigma^2 w'w/2T}$$

$$= (h'\partial w) \sum_{\ell=0}^{\infty} \frac{(-s)^{\ell}}{\ell!} \sum_{u=0}^{\infty} \frac{(\sigma^2/2T)^{\ell+u}}{(\ell+u)!} 2^{2\ell}(w'w)^u(u+1)_{\ell}\left(\frac{n}{2}+u\right)_{\ell}$$

$$= (2h'w) \sum_{\ell=0}^{\infty} \sum_{u=0}^{\infty} \frac{(-2\sigma^2 s/T)^{\ell}}{\ell!(\ell+u)!} u(\sigma^2/2T)^u (w'w)^{u-1}(u+1)_{\ell}\left(\frac{n}{2}+u\right)_{\ell}$$

$$= (2h'w) \sum_{u=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{(-2\sigma^2 s/T)^{\ell}\left(\frac{n}{2}+u\right)_{\ell}}{\ell!u!} u(\sigma^2/2T)^u (w'w)^{u-1}$$

$$(19) \quad = \left(\frac{\sigma^2 h'w}{T}\right) \sum_{u=1}^{\infty} \frac{(\sigma^2 w'w/T)^{u-1}}{(u-1)!} (1+2\sigma^2 s/T)^{-n/2-u} .$$

The above proof applies when s lies in the interval $0 < s < T/2\sigma^2$; but the formula holds by analytic continuation over the entire interval $(0,\infty)$. We may therefore deduce from (17) and (19) that:

$$(h'\partial w)\Delta_w^{-1} e^{\sigma^2 w'w/2T} = \left(\frac{\sigma^2 h'w}{T}\right)\sum_{v=0}^{\infty} \frac{(\sigma^2 w'w/T)^v}{v!\,(n/2+v}\left(\frac{T}{2\sigma^2}\right)$$

$$= \left(\frac{h'w}{2}\right)\sum_{v=0}^{\infty} \frac{(\sigma^2 w'w/T)^v (n/2)_v \Gamma(n/2)}{v!\,(n/2+1)_v \Gamma(n/2+1)}$$

$$(20) \quad = \left(\frac{h'w}{2}\right)\frac{\Gamma(n/2)}{\Gamma(n/2+1)}\ {}_1F_1\left(\frac{n}{2}, \frac{n}{2}+1;\ \frac{\sigma^2 w'w}{T}\right).$$

Term by term integration is justified by the uniform convergence of the series. Evaluating (20) at $w = T\beta/\sigma^2$ and substituting in (16) we obtain:

$$(21) \quad \tfrac{1}{2}ah'\beta e^{-\theta}\ \frac{\Gamma(n/2)}{\Gamma(n/2+1)}\ {}_1F_1\left(\frac{n}{2}, \frac{n}{2}+1;\ \theta\right).$$

Hence, from (15) and (21) we deduce that:

$$(22) \quad E(y) = h'\beta - \frac{T-m}{2}ah'\beta e^{-\theta}\frac{\Gamma(n/2)}{\Gamma(n/2+1)}\ {}_1F_1\left(\frac{n}{2}, \frac{n}{2}+1;\ \theta\right)$$

where $\theta = T\beta'\beta/2\sigma^2$. (22) yields as a special case the expression first found by Ullah (1974) for the bias of the Stein-rule estimator of an individual element of β .

5. CONCLUSION

Formula (12) is very convenient for mathematical work with the exact density, as the moment derivations of Section 4 demonstrate. But the generalized operator form of (12) is not the most convenient for numerical exercises. More explicit reductions of the formula to a form that is amenable to computational work can be achieved by direct use of the extended Weyl calculus developed here. These reductions will be reported in subsequent work.

6. REFERENCES

James, W. and C. Stein (1961), "Estimation with Quadratic Loss," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 361-379 (University of California Press, Berkeley).

Judge, G. G. and M. E. Bock (1978), The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics (North-Holland, Amsterdam).

Phillips, P. C. B. (1983), "The Exact Distribution of the SUR Estimator," mimeographed, Yale University, March.

Ross, B. (1974), Fractional Calculus and its Applications (Springer-Verlag, Berlin).

Ullah, A. (1974), "On the Sampling Distribution of Improved Estimators for Coefficients in Linear Regression," Journal of Econometrics, 2, 143-150.

_________ (1982), "The Approximate Distribution of the Stein-Rule Estimator," Economics Letters, 10, 305-308.

Wald, A. (1946), Comments recorded in the Minutes of Discussion of Papers on Multivariate Analysis for Non-Experimental Data and Related Problems of Matrix Computation by Cowles Commission Staff Members on August 23-24, 1946 (Ithaca, New York).