1982

Asymptotic Expansion of the Distribution of Stein-Rule Estimators when Disturbances Are Small

V. K. Srivastava

Aman Ullah

ISSN: 0318-725X
ISBN: 0-7714-0352-6

RESEARCH REPORT 8213

ASYMPTOTIC EXPANSION OF THE DISTRIBUTION OF STEIN-RULE ESTIMATORS WHEN DISTURBANCES ARE SMALL

by

V. K. Srivastava and A. Ullah*

Department of Statistics
Lucknow University

and

Department of Economics
University of Western Ontario

May, 1982

*The research support from NSERC is gratefully acknowledged.

1. Introduction

For the estimation of coefficients in a classical linear regression model, several families of improved estimators crafting the Stein-rule technology have been developed and conditions for their dominance over least squares estimator according to risk criterion under a general quadratic loss function have been obtained; see, e.g., Alam and Hawkes [1], Casella [3] and Strawderman [8] for a few recent ones. In some cases, the exact expressions for bias vector and mean squared error matrix have been worked out; see, e.g., Srivastava and Chaturvedi [6] and Ullah and Ullah [9]. No efforts have been made to evaluate the sampling distributions of improved estimators probably because they will be intricate enough and meaningful conclusions will be difficult to deduce. However, such is not the case with their asymptotic expansions which may be fruitful for studying the performance of estimators. This paper is an attempt in this direction. We have considered a simple family of Stein-rule estimators and have derived Edgeworth-type asymptotic expansion, assuming disturbances to be small, in order to approximate the exact sampling distribution. An interesting feature of our result is that the approximate distribution function can be easily evaluated on a desk calculator. We have also examined the performance of improved estimators with respect to least squares estimator according to the concentration of the distribution around the true value. Similar results can be obtained for other families of shrinkage estimators without much difficulty.

2. The Main Results:

Consider a classical linear regression model:

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

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

The least squares estimator of β is

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

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

A simple family of Stein-rule estimators for β is given by

$$\hat{\beta} = [1 - k\,\frac{(y - Xb)'(y - Xb)}{b'X'Xb}]b \qquad (2.3)$$

where k is any positive scalar characterizing the estimator.

The estimator $\hat{\beta}$ dominates b, for instance, according to total mean squared error criterion when

$$0 < k < \frac{2}{(n+2)}\,[d - 2]; \qquad d > 2 \qquad (2.4)$$

where

$$n = (T - p)$$
$$d = \lambda p \sum_{i=1}^{p} \frac{1}{\lambda_i} \qquad (2.5)$$

$\lambda_1 \geq \lambda_2 \geq \ldots. \geq \lambda_p$ being the characteristic roots of $Q^{-1} = (X'X)$.

Theorem: Defining the estimation error of the standardized estimator of β_i, the i^{th} component of β, as

$$\xi_i = \frac{1}{\sigma}(\hat{\beta}_i - \beta_i)$$

the small-disturbance asymptotic expansion of the density function of ξ_i is given by

$$(2.6) \qquad g(\xi_i) = [1 - \sigma k n \frac{\beta_i}{\beta' X' X\beta} \cdot \frac{\xi_i}{q_{ii}} + \sigma^2 \frac{kn}{\beta' X' X\beta} \{\frac{k(n+2)+4}{2\beta' X' X\beta} \cdot \frac{\beta_i^2}{q_{ii}} - 1\} (\frac{\xi_i^2}{q_{ii}} - 1)] \cdot \frac{1}{\sqrt{q_{ii}}} f(\frac{\xi_i}{\sqrt{q_{ii}}})$$

where $f(\cdot)$ denotes the univariate standard normal density function and q_{ii} is the i^{th} diagonal element of $Q = (X'X)^{-1}$.

Using the approximate density function to order $O(\sigma^2)$, we can evaluate moments of ξ_i. For instance, the first moment is

$$(2.7) \qquad E(\xi_i) = \frac{1}{\sigma} E(\hat{\beta}_i - \beta_i) = -\sigma\, kn \frac{\beta_i}{\beta' X' X\beta}$$

whence the bias vector, to order $O(\sigma^2)$, is

$$(2.8) \qquad E(\hat{\beta} - \beta) = -\sigma^2 \frac{kn}{\beta' X' X\beta}\beta$$

which tallies with Srivastava and Upadhyaya [7, (2.1) on page 8] to the order of our approximation. Further, it matches with the large concentration parameter approximation of Ullah and Ullah [9, (3.15) on page 712] remembering that large concentration parameter approximations with sample size fixed are equivalent to small-disturbance approximations; see Anderson [2].

Similarly, the second moment is

$$(2.9) \qquad E(\xi_i^2) = \frac{1}{\sigma^2} E(\hat{\beta}_i - \beta_i)^2$$

$$= q_{ii} + \sigma^2 \frac{kn}{\beta' X' X\beta} [(\frac{k(n+2)+4}{\beta' X' X\beta})\beta_i^2 - 2q_{ii}]$$

which agrees with the expression for the i^{th} diagonal element obtained by Srivastava and Upadhyaya [7, (2.12) on page 8] and Ullah and Ullah [9, (3.16) on page 713] to the order of our approximation.

From (2.9), it is observed that

$$(2.10) \qquad \frac{1}{\sigma^2} E(b-\beta)'(b-\beta) - \frac{1}{\sigma^2} E(\hat{\beta}-\beta)'(\hat{\beta}-\beta)$$

$$= \sigma^2 \frac{kn}{\beta' X' X\beta} [2\,\mathrm{tr}\,Q - (k(n+2)+4) \frac{\beta'\beta}{\beta' X' X\beta}]$$

which is positive if

$$(2.11) \qquad 0 < k < \frac{2}{(n+2)} [\frac{\beta' X' X\beta}{\beta'\beta} \mathrm{tr}\,Q - 2]; \quad \frac{\beta' X' X\beta}{\beta'\beta} \mathrm{tr}\,Q > 2.$$

The above inequality holds as long as (2.4) is true.

2.1 Comparison of Distribution Functions

A further comparison of b and $\hat{\beta}$ can be made with respect to the concentration of probability around β. For this purpose, we consider the distribution functions of b_i and $\hat{\beta}_i$. It is easy to see that

$$(2.12) \qquad \overset{*}{G}(m) = P[\frac{1}{\sigma}(b_i - \beta_i) \leq m]$$

$$= F\left(\frac{m}{\sqrt{q_{ii}}}\right)$$

where $F(\cdot)$ denotes the distribution function of a standard normal variate.

From (2.6), we can obtain the distribution function of ξ_i to order $0(\sigma^2)$:

$$(2.13)\qquad G(m) = P[\xi_i \le m]$$

$$= P[\frac{1}{\sigma}(\hat{\beta}_i - \beta_i) \le m]$$

$$= F\left(\frac{m}{\sqrt{q_{ii}}}\right) + \sigma kn \frac{\beta_i}{\beta' X' X\beta} \circ \frac{1}{\sqrt{q_{ii}}} f\left(\frac{m}{\sqrt{q_{ii}}}\right)$$

$$- \sigma^2 \frac{kn}{\beta' X' X\beta} [\frac{k(n+2)+4}{2\,\beta' X' X\beta} \cdot \frac{\beta_i^2}{q_{ii}} - 1] \frac{m}{q_{ii}^{3/2}} f\left(\frac{m}{\sqrt{q_{ii}}}\right)$$

Now consider the difference between b_i and $\hat{\beta}_i$ in terms of concentration in an interval symmetric around the true value β_i. From (2.12) and (2.13), we have

$$(2.14)\qquad P[|\frac{1}{\sigma}(b_i - \beta_i)| \le m] - P[|\frac{1}{\sigma}(\hat{\beta}_i - \beta_i)| \le m]$$

$$= [\overset{*}{G}(m) - \overset{*}{G}(-m)] - [G(m) - G(-m)]$$

$$= 2\,\sigma^2 \frac{kn}{\beta' X' X\beta} [\frac{k(n+2)+4}{2\,\beta' X' X\beta} \circ \frac{\beta_i^2}{q_{ii}} - 1] \frac{m}{q_{ii}^{3/2}} f\left(\frac{m}{\sqrt{q_{ii}}}\right)$$

which is positive if

$$(2.15)\qquad 0 < k < \frac{2}{(n+2)} [\frac{q_{ii}}{\beta_i^2} \beta' X' X\beta - 2] \; ; \; \frac{q_{ii}}{\beta_i^2} \beta' X' X\beta > 2.$$

If h is a $p \times 1$ vector with all elements 0 except 1 at the i^{th} component, we have from Rao [4, page 74]

$$(2.16)\qquad \frac{\beta'X'X\beta}{\beta_i^2} = \frac{\beta'X'X\beta}{\beta'hh'\beta} \geq \lambda p$$

so that (2.15) is satisfied so long as

$$(2.17)\qquad 0 < k < \frac{2}{(n+2)}[q_{ii}\lambda p - 2]\ ;\ q_{ii}\ \lambda_p > 2.$$

If q denotes the smallest diagonal element of Q, it follows from (2.17) that the estimator $\hat{\beta}$ dominates b according to concentration around β at least as long as

$$(2.18)\qquad 0 < k < \frac{2}{(n+2)}\ [q\lambda p - 2]\ ;\ q\lambda p > 2.$$

It may be pointed out that the quantity $\frac{q_{ii}}{\beta_i^2}\beta'X'X\beta$ assumes values larger than 1 which essentially follows from a result in Rao [4, page 60].

3. Proof of Theorem

In order to derive the small-disturbance approximation for the exact sampling distribution of the estimator $\hat{\beta}$, we write the model (2.1) as

$$(3.1)\qquad y = X\beta + \sigma v \qquad [u = \sigma v]$$

so that v follows a multivariate normal distribution $N(0, I_T)$.

Using (2.2) and (3.1), we can express

$$(3.2)\qquad (\hat{\beta} - \beta) = \sigma(X'X)^{-1}X'v - \sigma^2 k\, \frac{v'[I_T - X(X'X)^{-1}X']v}{\beta'X'X\beta + 2\sigma\beta'X'v + \sigma^2 v'X(X'X)^{-1}X'v} \cdot [\beta + \sigma(X'X)^{-1}X'v]$$

$$= \sigma\, Q^{\frac{1}{2}}Z - \sigma^2 k\, \frac{(w+n)}{\theta}[1 + 2\sigma\, \frac{\beta'Q^{-\frac{1}{2}}Z}{\theta} + \sigma^2\, \frac{Z'Z}{\theta}]^{-1} \cdot [\beta + \sigma Q^{\frac{1}{2}}Z]$$

where

$$(3.3)\quad Z = (X'X)^{-\frac{1}{2}}X'v \ , \ \theta = \beta'X'X\beta = \beta'Q^{-1}\beta,$$
$$w = v'[I_T - X(X'X)^{-1}X']v - n.$$

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

Expanding the expression in first square brackets on the right-hand side of (3.2), we can express the estimation error as

$$(3.4)\quad (\hat{\beta} - \beta) = \sigma e_1 + \sigma^2(e_2 - \frac{kn}{\theta}\beta) + \sigma^3 e_3 + \sigma^4 e_4 + 0(\sigma^j) \qquad j \geq 5$$

where

$$e_1 = Q^{\frac{1}{2}} Z$$
$$e_2 = -\frac{kw}{\theta}\beta$$
$$e_3 = -\frac{k(w+n)}{\theta} Q^{\frac{1}{2}}(I_p - \frac{2}{\theta} Q^{-\frac{1}{2}}\beta\beta' Q^{-\frac{1}{2}})Z$$
$$e_4 = \frac{k(w+n)}{\theta^2} [Z'(I_p - \frac{4}{\theta} Q^{-\frac{1}{2}}\beta\beta' Q^{-\frac{1}{2}}) Z \cdot \beta + 2\beta' Q^{-\frac{1}{2}} Z \cdot Q^{\frac{1}{2}} Z]$$

Writing

$$(3.5)\quad e(Z,w) = \sigma e_1 + \sigma^2 e_2 + \sigma^3 e_3 + \sigma^4 e_4$$

we observe that $e(0,0) = 0$ whence it follows from approximation theorem of Sargan [5] that $\frac{1}{\sigma}\, e(Z,w)$ has a valid Edgeworth expansion.

For h to be any $p \times 1$ fixed vector, let us define

$$\mathcal{G}_h = \frac{1}{\sigma} h'e(Z,w)$$

(3.6) $$d' = h'Q^{\frac{1}{2}}[I_p - \sigma^2 \frac{k(w+n)}{\theta} (I_p - \frac{2}{\theta} Q^{-\frac{1}{2}} \beta\beta' Q^{-\frac{1}{2}})]$$

$$C = \frac{k(w+n)}{\theta^2} [h'\beta (I_p - \frac{4}{\theta} Q^{-\frac{1}{2}} \beta\beta' Q^{-\frac{1}{2}}) + 2Q^{-\frac{1}{2}} \beta h' Q^{\frac{1}{2}}]$$

so that $\mathcal{G}_h$ contains terms up to order $0(\sigma^3)$ and

(3.7) $$\mathcal{G}_h = d'Z - \sigma^2 \frac{kwh'\beta}{\theta} + \sigma^3 Z'CZ.$$

Now the characteristic function of $\mathcal{G}_n$ is

(3.8) $$\Psi(t) = E[e^{it\mathcal{G}_h}]$$

$$= E_w E_Z [e^{it(d'Z - \sigma \frac{kwh'\beta}{\theta} + \sigma^3 Z'CZ)} | w]$$

$$= E_w [e^{-\sigma it \frac{kwh'\beta}{\theta}} E_Z \{e^{it(d'Z + \sigma^3 Z'CZ)} | w\}]$$

It is easy to see that

(3.9) $$E_Z \{e^{it(d'Z + \sigma^3 Z'CZ)} | w\}$$

$$= \frac{1}{(2\pi)} \frac{p}{2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{it(d'Z + \sigma^3 Z'CZ)} e^{-\frac{1}{2} Z'Z} dZ$$

$$= |I_p - 2\sigma^3 it\, C|^{-\frac{1}{2}} e^{-\frac{t^2}{2} d'(I_p - 2\sigma^3 it\, C)^{-1} d}$$

But

(3.10) $$|I_p - 2\sigma^3 it\, C|^{-\frac{1}{2}} = 1 + 0(\sigma^j) \qquad j \geq 3$$

$$d'(I_p - 2\sigma^3 it\, C) d = d'(I_p + 2\sigma^3 it\, C \ldots\ldots) d$$

$$= h'Qh - 2\sigma^2 \frac{k(w+n)}{\theta} \{h'Qh - \frac{2}{\theta}(h'\beta)^2\} + 0(\sigma^j) \qquad j \geq 3$$

$$(3.11)\quad e^{-\frac{t^2}{2} d'(I_p - 2\sigma^3 itC)^{-1} d} = e^{-\frac{t^2}{2} h'Qh}\left[1 + \sigma^2 t^2 \frac{k(w+n)}{\theta}\left\{h'Qh - \frac{2}{\theta}(h'\beta)^2\right\} + 0(\sigma^j)\right] \qquad j \geq 3$$

Thus putting (3.10) and (3.11) in (3.9), and then substituting (3.9) in (3.8) along with

$$(3.12)\quad e^{-\sigma it \frac{kwh'\beta}{\theta}} = 1 - \sigma it \frac{kwh'\beta}{\theta} - \sigma^2 \frac{t^2}{2}\left(\frac{kwh'\beta}{\theta}\right)^2 + 0(\sigma^j) \qquad j \geq 3$$

we find

$$(3.13)\quad \Psi(t) = e^{-\frac{t^2}{2} h'Qh} E_w\left[1 - \sigma it \frac{kwh'\beta}{\theta} - \sigma^2 \frac{t^2 k^2 w^2 (h'\beta)^2}{2\theta^2}\right.$$

$$\left. + \sigma^2 \frac{t^2 k(w+n)}{\theta}\left\{h'Qh - \frac{2}{\theta}(h'\beta)^2\right\} + 0(\sigma^j)\right] \qquad j \geq 3$$

$$= e^{-\frac{t^2}{2} h'Qh}\left[1 - \sigma^2 \frac{t^2 kn}{\theta}\left\{\left(\frac{k+2}{\theta}\right)(h'\beta)^2 - h'Qh\right\}\right] + 0(\sigma^j) \qquad j \geq 3$$

whence taking $\Psi(t)$ to order $0(\sigma^2)$ and using Inversion Theorem, the density function of G_h to order $0(\sigma^2)$ is

$$(3.14)\quad g(G_h) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itG_h}\Psi(t)\,dt$$

$$= \left[1 + \sigma^2 \frac{kn}{\theta}\left\{\left(\frac{k+2}{\theta}\right) \cdot \frac{(h'\beta)^2}{h'Qh} - 1\right\}\left(\frac{G_h^2}{h'Qh} - 1\right)\right] \frac{1}{\sqrt{h'Qh}}\, f\left(\frac{G_h}{\sqrt{h'Qh}}\right)$$

where we have utilized the following results:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{t^2}{2} h'Qh - itG_h}\,dt = \frac{1}{\sqrt{h'Qh}}\, f\left(\frac{G_h}{\sqrt{h'Qh}}\right)$$

(3.15)

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} (it)^2 e^{-\frac{t^2}{2} h'Qh - itG_h}\,dt = \left(\frac{G_h^2}{h'Qh} - 1\right)\frac{1}{(h'Qh)^{3/2}}\, f\left(\frac{G_h}{\sqrt{h'Qh}}\right)$$

Applying the transformation

$$(3.16) \qquad \alpha_h = a_h - \sigma \frac{knh'\beta}{\theta}$$

and observing that

$$(3.17) \qquad f\left(\frac{\alpha_h + \sigma \frac{knh'\beta}{\theta}}{\sqrt{h'Qh}}\right)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot \frac{(\alpha_h + \sigma \frac{knh'\beta}{\theta})^2}{h'Qh}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot \frac{\alpha_h^2}{h'Qh}} [1 - \sigma \frac{knh'\beta}{\theta} \cdot \frac{\alpha_h}{h'Qh}$$

$$+ \sigma^2 \frac{k^2 n^2 (h'\beta)^2}{2\theta^2 h'Qh} (\frac{\alpha_h^2}{h'Qh} - 1)] + 0(\sigma^j) \qquad j \geq 3$$

we get the density function of α_h to order $0(\sigma^2)$:

$$(3.18) \qquad g(\alpha_h) = \frac{1}{\sqrt{h'Qh}} f\left(\frac{\alpha_h}{\sqrt{h'Qh}}\right) [1 - \sigma \frac{knh'\beta}{\theta} \cdot \frac{\alpha_h}{h'Qh}$$

$$+ \sigma^2 \frac{kn}{\theta h'Qh} \{ (\frac{k(n+2)+4}{2\theta}) (h'\beta)^2 - h'Qh \} (\frac{\alpha_h^2}{h'Qh} - 1) .$$

Setting h equal to a vector with all elements 0 except 1 at the j^{th} place so that $h'\beta = \beta_j$ and $h'Qh = q_{jj}$, we get the result stated in the Theorem.

REFERENCES

[1] Alam, K. and J. S. Hawkes (1978): "Estimation of Regression Coefficients," Scandinavian Journal of Statistics, Vol. 5, pp. 169-172.

[2] Anderson, T. W. (1977): "Asymptotic Expansions of the Distributions of Estimates in Simultaneous Equations for Alternative Parameter Sequences," Econometrica, Vol. 45, pp. 506-518.

[3] Casella, G. (1980): "Minimax Ridge Regression Estimation," Annals of Statistics, Vol. 8, pp. 1036-1056.

[4] Rao, C. R. (1965): Linear Statistical Inference and its Applications, John Wiley and Sons, Inc., New York.

[5] Sargan, J. D. (1976): "Econometric Estimators and the Edgeworth Approximation," Econometrica, Vol. 44, pp.

[6] Srivastava, V. K. and A. Chaturvedi (to appear): "Bias Vector and Mean Squared Error Matrix of Minimax Adaptive Generalized Ridge Regression Estimators," Sankhya.

[7] Srivastava, V. K. and S. Upadhyaya (1977): "Properties of Stein-Like Estimators in the Regression Model When Disturbances are Small," Journal of Statistical Research, Vol. 11, pp. 5-21.

[8] Strawderman, W. E. (1978): "Minimax Adaptive Generalized Ridge Regression Estimators," Journal of the American Statistical Association, Vol. 73, pp. 623-627.

[9] Ullah, A. and S. Ullah (1978): "Double k-Class Estimators Coefficients in Linear Regression," Econometrica, Vol. 46, pp. 705-722.