FINITE SAMPLE MOMENTS OF A BOOTSTRAP ESTIMATOR OF THE JAMES-STEIN RULE

Lee C. Adkins

Department of Economics
Oklahoma State University
Stillwater, OK 74078

Key Words and Phrases: *Stein-rules; bootstrap; shrinkage estimators*

ABSTRACT

The finite sample moments of the bootstrap estimator of the James-Stein rule are derived and shown to be biased. Analytical results shed some light upon the source of bias and suggest that the bootstrap will be biased in other settings where the moments of the statistic of interest depends on nonlinear functions of the parameters of its distribution.

INTRODUCTION

The bootstrap is used in many settings where the exact sampling distributions of estimators are either unknown or intractable. Freedman and Peters (1984) use the bootstrap to improve upon the usual asymptotic results in a feasible generalized least squares setting. Runkle (1987) estimates standard errors for variance decompositions and impulse response functions associated with estimation of vector autoregressive models. In the literature on improved estimation of the parameters of the linear regression model Chi and Judge (1985) and

Brownstone (1990) use the bootstrap nonparametrically to estimate standard errors for the Stein-rule estimators.

One of the benefits of the bootstrap for these and other problems is that it can be used more or less automatically to approximate the sampling distribution of a statistic whenever exact results are unavailable. Under fairly general circumstances it is well-known that the bootstrap yields consistent estimates of an estimator's sampling moments as sample size and the number of bootstrap replications gets large [see Bickel and Freedman (1981), Beran (1982)], but few have examined its behavior in finite samples. One exception is Sim (1989) who derives small sample results for estimators of the regression model. Also, based on Monte Carlo evidence Adkins (1990) and Adkins and Hill (1990) report that near the origin, percentile bootstrap confidence intervals and ellipsoids for the James-Stein estimator [see James and Stein (1961)], tend to be larger than necessary to cover at nominal levels. Their results suggest that the bootstrap estimators of the James-Stein's covariance and standard error are biased upward near the origin. The source of the problem is demonstrated analytically below as the finite sample moments of parametric bootstrap estimators of the JS mean and covariance are derived.

The basic finding is that bootstrap estimators of the mean and standard errors of the James-Stein rule are biased in finite samples because the moments of the Stein-rule contain nonlinear functions of the parameters associated with its distribution. In addition, the results suggest that the bootstrap is least accurate in Stein estimation when the location parameters lie at or near the origin. In most cases the researcher will have more specific nonsample information than that embodied by the James-Stein rule and shrinkage will be directed toward points other than the origin. In any event, the bootstrap estimator of standard error is superior to that derived from estimates of the asymptotic covariance matrix (which coincides with that of least squares). The results obtained here should be useful to those involved in

Stein estimation and shed some light on the properties of the bootstrap in similar settings (e.g., ridge regression).

Before proceeding it is worth mentioning that there are alternatives to bootstrapping for estimating the moments of the Stein-rule estimators [e.g., empirical Bayes approach described in Judge and Yancy (1986), unbiased estimates of the MSE matrix, Carter et al. (1990), and asymptotic approximation, Ullah et al. (1984)], but these are not considered here.

THE MODEL AND ITS ESTIMATORS

The classical normal linear regression model (CNLRM) is represented by

$$y = X\beta + e \qquad e \sim N(0,\sigma^2 I_T) \tag{1}$$

where y is a Tx1 vector of observable random variables, X is a nonstochastic TxK matrix of rank K, β is a Kx1 vector of unknown parameters, and e is a Tx1 vector of unobservable normally and independently distributed random variables having zero mean and finite variance. The ordinary least squares (LS) and maximum likelihood estimator of β is $b=(X'X)^{-1}X'y \sim N(\beta,\sigma^2(X'X)^{-1})$ and the minimum variance unbiased estimator of σ^2 is

$$\hat{\sigma}^2=(y-Xb)'(y-Xb)/(T-K),$$

with $(T-K)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{T-K}$ and independent of b.

The James-Stein estimator (JS) dominates the MLE of β in the CNLRM under weighted quadratic loss with weight matrix W. The JS estimator is

$$\delta(b) = (1-as/b'Sb)b \tag{2}$$

where $S=X'X$, and $s=(y-Xb)'(y-Xb)$. The estimator is minimax if the scalar 'a' is chosen to lie within the interval $(0,a_{max})$, where

$$a_{max}=[2/(T-K+2)](\lambda_L^{-1}\ tr[WX'X]-2),$$

and λ_L is the largest characteristic root of (WX'X) [Judge and Bock (1978) p. 235]. The value of the constant 'a' which minimizes quadratic risk is the interval's midpoint.

The James-Stein estimator has mean

$$E[\delta(b)] = \beta - a(T-K)E[1/\chi^2_{K+2,\lambda}]\beta \tag{3}$$

and covariance matrix

$$E[(\delta - E[\delta])(\delta - E[\delta])'] = \sigma^2 S^{-1} -$$
$$\sigma^2 [2a(T-K)\ E(l_1) - a^2(T-K)(T-K+2)\ E(l_1)^2]\ S^{-1} +$$
$$\beta\beta'\ \{2a(T-K)[E(l_1)-E(l_2)] +$$
$$a^2(T-K)[(T-K+2)E(l_2)^2 - (T-K)(E[l_1])^2]\} \qquad (4)$$

where $l_1 = \chi^2_{T-K}/\chi^2_{K+2,\lambda}$, $l_2 = \chi^2_{T-K}/\chi^2_{K+4,\lambda}$, $\lambda = \beta' S\beta/2\sigma^2$ and $\chi^2_{n,\lambda}$ is a chi-square random variable with n degrees of freedom and noncentrality parameter λ [Judge and Bock (1978)]. The estimator is biased in small samples when $\beta \neq 0$ and its covariance matrix depends on the unknown location and scale parameters. The James-Stein estimator is not an MLE and is neither linear nor normally distributed; consequently, exact hypothesis tests and confidence intervals of a given size cannot be formed in the usual way, i.e., based on Wald or likelihood ratio principles [see Engle (1984)]. Asymptotically, the mean and covariance of the James-Stein rule converge to those of the least squares and any risk advantage associated with its use vanishes.

The James-Stein rule has not met with widespread acceptance among applied researchers. One reason is that in many econometric applications it gives results which coincide with least squares [see for example Aigner and Judge (1977)]. This may occur because the implicit nonsample information contained in the James-Stein estimator (i.e., that $\beta = 0$) is not supported by the data. General versions of the Stein-rules which shrink the MLE toward points other than the origin are available and often yield more interesting results [e.g., see Hill, Ziemer, and White (1981)]. In fact, the quadratic risk performance of the Stein-rule improves as the nonsample information it embodies becomes more accurate. Numerous other improvements have been made to the original James-Stein estimator and the reader is referred to Judge and Bock (1978), Brownstone (1990), Judge, Hill, and Bock (1990) and the references contained therein for details.

Even though the James-Stein estimator and its various "improved" versions offer significant quadratic risk gains over the MLE in certain regions of the parameter space in small samples, it is difficult to derive and use their exact sampling distributions [see Phillips (1984)]. Consequently, exact confidence sets and hypothesis tests are not readily available at this time.

Approximate confidence intervals and ellipsoids of given size centered at the James-Stein rule have been constructed using the bootstrap by Chi and Judge (1985), and Adkins and Hill (1990). Both studies indicate that bootstrap confidence intervals and ellipsoids centered at the James-Stein rule tend to be larger than necessary to cover at nominal rates near the origin; coverage frequency improves as $|\beta|$ increases. Evidence in Brownstone's (1990) nonparametric application is consistent with these findings. This suggests that bootstrap standard errors for the JS estimator may be more accurate estimators of the actual standard errors in some parts of the parameter space than in others. This conjecture is verified below.

BOOTSTRAP ESTIMATORS OF MEAN AND COVARIANCE

The bootstrap can be used in a variety of ways [see Efron (1982)]. In order to avoid confusion over the way the technique is used, each method employed below is discussed briefly.

The most common form of the bootstrap is nonparametric and uses the least squares estimates b to obtain the set of residuals $\hat{e}=y-Xb$; these serve as the estimates of the true disturbances of the model and are thought to capture its underlying structure. After rescaling the least squares residuals using $\hat{e}^*_t=(T/(T-K))^{1/2}\hat{e}_t$, t=1,...,T, [see Wu (1986) or Sim (1989) for discussion] a bootstrap sample of size T is drawn randomly and with replacement from $\hat{e}^*=[\hat{e}^*_1,...,\hat{e}^*_T]'$ and denoted e^*. Then the bootstrap sample $y^*=Xb+e^*$ is obtained and the bootstrap estimate $b^*=(X'X)^{-1}X'y^*$ is computed. A large number, N, of size T random samples are drawn from the empirical distribution $\hat{e}^*$ and the sequences $(y^*)_1^N$ and $(b^*)_1^N$ are computed.

Using the assumption that the errors in this model are independently, identically distributed normal random variables the sequences $(y^*)_1^N$ and $(b^*)_1^N$ can also be generated using a parametric bootstrap. Parametric bootstrapping of the James-Stein estimator consists of obtaining the least squares estimates of β and σ^2 and drawing a bootstrap sample y^* of size T from the $N(Xb,\hat{\sigma}^2 I_T)$ distribution. Drawing N random samples of size T, $\{y^*\}_1^N$, we compute the sequence of bootstrap estimates, $\{b^*\}_1^N$. Given $\{b^*\}_1^N$ the bootstrap estimator of the LS covariance is

$$Cov(b^*) = (B^*-\overline{B}^*)' (B^*-\overline{B}^*)/N-1 \tag{5}$$

where $B^* = [b_1^*, b_2^*, \ldots , b_N^*]'$ is the NxK matrix of bootstrap estimates with columns $\{b_i^*\}_1^N$, $i=1,\ldots,K$, $\overline{B}^*=j_N \otimes \overline{b}^{*\prime}$, j_N is an Nx1 vector of ones, $\overline{b}^*$ is the Kx1 vector whose elements are $\sum_{n=1}^{N} b_{in}^*/N$, and N is the number of bootstrap samples.[1] Efron (1982) and Wu (1986) show that $E(b^*)=b$ and $E[Cov(b^*)]=\hat{\sigma}^2(X'X)^{-1}$. If b^* is obtained using the parametric bootstrap, then $(b^*-b)\sim N(0,\hat{\sigma}^2(X'X)^{-1})$ and $Cov(b^*)\sim W_K(N-1,\hat{\sigma}^2(X'X)^{-1})$ where W_K is the Wishart distribution with N-1 degrees of freedom and covariance $\hat{\sigma}^2(X'X)^{-1}$ [see Anderson (1984)].

The nonparametric bootstrap is essentially biased for the standard errors of b in finite samples while the parametric bootstrap is not. Intuitively, the relationship between parametric and nonparametric bootstrapping for LS depends on the sample size, T, as well as the number of bootstrap samples taken, N. In general, as $N\to\infty$ the bootstrap approximation to the empirical distribution function (d.f.), which is obtained by random resampling the least squares residuals, converges to the exact d.f. of the LS residuals, $\hat{e}$. The exact d.f. of $\hat{e}$ converges to that of e as $T\to\infty$ [see Beran (1982)]; hence, convergence of the

---

[1]For notational simplicity, the symbols $\{y^*\}_1^N$ and $\{b^*\}_1^N$ are used to denote the sequences of bootstrap samples and LS estimates for both the parametric and nonparametric bootstraps. Unless otherwise noted these symbols will henceforth refer to sequences obtained using the parametric
bootstrap.

nonparametric bootstrap to the parametric bootstrap estimator requires T as well as N to be large. When T is small we know only that increasing the number of pseudo-samples, N, *improves* the approximation of the empirical d.f. to that of $\hat{e}$; it will not eliminate the bias associated with the nonparametric bootstrap which arises because the exact empirical distribution of the least squares residuals differs from that of the model's errors (for finite T).

Having obtained b^* a bootstrap estimate of the James-Stein estimator

$$d^*(b^*) = [1-as^*/b^{*\prime}Sb^*]b^* \tag{6}$$

can be computed using $s^*=(y^*-Xb^*)'(y^*-Xb^*)$ and the sequence $\{d^*\}_1^N$ formed. The bootstrap covariance matrix is

$$\text{Cov}(d^*) = (D^*-\bar{D}^*)' \ (D^*-\bar{D}^*)/N-1 \tag{7}$$

where $D^* = [d_1^*, d_2^*, \ \ldots \ , d_N^*]'$ is the NxK matrix of bootstrap estimates with columns $\{d_i^*\}_1^N$, $i=1,\ldots,K$, $\bar{D}^*=j_N \otimes \bar{d}^{*\prime}$, and $\bar{d}^*$ is the Kx1 vector whose elements are $\sum_{n=1}^{N} d_{in}^*/N$.

Using theorems given in Judge and Bock (1978), the mean and covariance of d^* (taken over the distribution of y^*) are

$$E_*[d^*(b^*)] = b - a(T-K)E[1/\chi^2_{K+2,\hat{\lambda}}]b \tag{8}$$

and

$$\begin{aligned} E_*[(d^*-E_*[d^*])(d^*-E_*[d^*])'] &= \hat{\sigma}^2 S^{-1} - \\ \hat{\sigma}^2 [2a(T-K)E(l_1) &- a^2(T-K)(T-K+2) E(l_1)^2] S^{-1} + \\ bb'\{2a(T-K)&[E(l_1)-E(l_2)] + \\ a^2(T-K)[(T-K+2)E(l_2)^2 &- (T-K)(E[l_1])^2]\} \end{aligned} \tag{9}$$

where $l_1=\chi^2_{T-K}/\chi^2_{K+2,\hat{\lambda}}$ and $l_2=\chi^2_{T-K}/\chi^2_{K+4,\hat{\lambda}}$, and $\hat{\lambda}=b'Sb/2\hat{\sigma}^2$ (proofs in Appendix A). This amounts to replacing β and σ in (3) and (4) with b and $\hat{\sigma}$, respectively. Comparing (8) and (9) with (3) and (4), respectively, gives us some idea of why the bootstrap estimates moments of the James-Stein rule with bias. Intuitively, the problem occurs because in general $E[\phi(bb')]\neq\phi(\beta\beta')$ and $E[\phi(b'b)]\neq\phi(\beta'\beta)$, where ϕ is any Borel measurable function.

Now, using the fact that $E(g(b^*))=E_y(E_*[g(b^*|b)])$ we take the expectations with respect to the distribution of y and obtain the following results

$$E_y(E_*[d^*(b^*)]) = \beta - a(T-K)E[1/\chi^2_{K+2,\eta_0}]\beta \tag{10}$$

and

$$E_y(E_*[(d^*-E[d^*])(d^*-E[d^*])']) =$$
$$\sigma^2\left\{1-2a(T-K)E_*(l_{22})+a^2(T-K)(T-K+2)E_*[(l_{22})^2]\right\} S^{-1}+$$
$$\sigma^2\left\{2a(T-K)(E_*(l_{20}) - E_*(l_{40})) +\right.$$
$$\left.a^2(T-K)((T-K+2)E_*[(l_{40})^2] - (T-K)\ E_*[(l_{20})^2])\right\}S^{-1} +$$
$$\left\{2a(T-K)(E_*(l_{21}) - E_*(l_{41})) +\right.$$
$$\left.a^2(T-K)((T-K+2)E_*[(l_{41})^2] - (T-K)\ E_*[(l_{21})^2])\right\}\beta\beta' \tag{11}$$

where $\eta_0 = \frac{1}{2}\frac{(T-K)}{(T-K-2)}E[\chi^2_{K+2,\lambda}] = \frac{1}{2}\frac{(T-K)}{(T-K-2)}(K+2+2\lambda)$, $\eta_1 = \frac{1}{2}\frac{(T-K)}{(T-K-2)}E[\chi^2_{K+4,\lambda}] = \frac{1}{2}\frac{(T-K)}{(T-K-2)}(K+4+2\lambda)$, $\eta_2 = \frac{1}{2}E[\chi^2_{K,\lambda}] = \frac{1}{2}(K+2\lambda)$,
$E_*(l_{22})=E_*[1/\chi^2_{K+2,\eta_2}]$, $E_*(l_{42})=E_*[1/\chi^2_{K+4,\eta_2}]$,
$E_*(l_{20})=E_*[1/\chi^2_{K+2,\eta_0}]$, $E_*(l_{40})=E_*[1/\chi^2_{K+4,\eta_0}]$,
$E_*(l_{21})=E_*[1/\chi^2_{K+2,\eta_1}]$, and $E_*(l_{41})=E_*[1/\chi^2_{K+4,\eta_1}]$. Proofs of (8)-(11) are given in Appendix A.

The results demonstrate that in general, the sampling moments of the bootstrap estimator d^* are not equal to the moments of the James-Stein rule. Since $\lambda<\eta_0$ for all β, $E[1/\chi^2_{K+2,\eta_0}] > E[1/\chi^2_{K+2,\lambda}]$ and $E[d^*]>E[\delta]$. Determining the relationship between $Cov(d^*)$ and $Cov(\delta)$ is more complicated but the same general story applies. Rather than seek an analytical solution to this problem, the results of a small simulation documenting the behavior of $Cov(d^*)$ and $Cov(\delta)$ for various values of λ are given below. Once again the bootstrap estimator of the James-Stein's covariance (11) is biased for (4) and in general, $E_y[Cov(d^*)]>Cov(\delta)$. Consequently, standard errors obtained using (11) with the parametric bootstrap will be larger on average than the actual standard error of the James-Stein estimator.

The results also suggest ways to reduce the bias associated with bootstrap estimation of the JS standard errors. Comparing equation (4) to equation (9) it can be seen that the parametric bootstrap estimator of Cov(d) essentially replaces $\beta'\beta$ and $\beta\beta'$ with b'b and bb', respectively. Near the origin these are poor estimates; when β lies farther from the origin the bias is reduced. Replacing b by the JS estimator, δ, should reduce bias near the origin and δ will converge to b as $|\beta|$ increases. Hence, in the nonparametric setting we could resample randomly from the JS residuals $\hat{e}_\delta = y - X\delta$ and generate bootstrap samples using $y^* = X\delta + e^*_\delta$ where e^*_δ represents a random resample from the JS residuals $\hat{e}_\delta$. This approach is similar in spirit to that taken by Brownstone (1990). Smaller standard errors can also be attained from Cov(d^*) by simply resampling randomly from the scaled least squares residuals e^* and using $y^* = X\delta + e^*$. Either of these modifications can be justified on theoretical grounds since δ is consistent for β.

More formal procedures which can be used to reduce bias include smoothing the empirical distribution of the residuals before resampling (Efron 1982) or, in the case of confidence intervals and sets, prepivoting (Beran 1988) and double bootstrapping (Beran 1990).

SIMULATION RESULTS

In this section the results from a small simulation are reported in order to document the relationship between bootstrap estimates of standard error and the actual values associated with the JS estimator. Results from both parametric and nonparametric bootstraps are given in an attempt to measure the magnitude of the bias associated with each.

For the Monte Carlo each of 8 regressors in the X matrix has been standardized to have zero mean, variance 1, and to be mutually orthogonal. This design is referred to as the orthonormal regression model (i.e., $X'X = I$) and corresponds to the model of the mean of a multivariate population which is widely studied in statistics. The orthonormal model is computationally convenient and the results generalize to nonorthogonal models for which it is a canonical form. The weight matrix W which appears in the scalar 'a' is chosen to be $W=I_K$ while 'a' is the midpoint of the interval $(0,a_{max})$.

A total of 500 pseudo-random samples of size T=30 was drawn from the N(0,1) density. The same set of deviates were used for each of the ten parameter points generated using

$$\ell=(R^2\ T\ \sigma^2/\ [(1-R^2)j'j])^{1/2}$$

where $\beta=\ell j$, j is 8x1 vector of ones, $\sigma^2=1$, and population goodness-of-fit $R^2=[.00001,\ 0.01,\ 0.025,\ 0.05,\ 0.075\ 0.10,\ 0.25,\ 0.50,\ 0.75,\ 0.90]$. Thus, the bias of bootstrap moments for the James-Stein estimator is studied for several points in the parameter space which lie along a ray extending from the origin. As R^2 increases, the degree of noncentrality λ increases as well.

The parametric bootstrap results from the simulation are reported in Table I along with the expected values of Stein-rule means and standard deviations obtained using equations (3), (4), (10) and (11). Let d_{kmn} be the k(th) element of d computed from the n(th) bootstrap sample from the m(th) Monte Carlo iteration. The elements of tables I and II are computed as follows: $\bar{\delta}_i = \sum_K \sum_M \delta_{km}/(KM)$, $\bar{d}_i^* = \sum_K \sum_M \sum_N d^*_{kmn}/(KMN)$, $\bar{\sigma}_\delta = [\sum_K \sum_M (\delta_{km} - \bar{\delta}_{k.})^2/(KM)]^{1/2}$ where $\bar{\delta}_{k.} = \sum_M \delta_{km}/M$, and $\bar{\sigma}^*_d = \sum_K \sum_M \sigma^*_{km}/KM$ where σ^*_{km} is the square root of the k(th) diagonal element of $Cov(d^*)$ from the m(th) Monte Carlo iteration. In the first column the values of the noncentrality parameter and the corresponding R^2 are given. Each value of λ (or R^2) represents a different point in the parameter space. In the next two columns the average value of the elements of δ and its standard error are reported for the 500 Monte Carlo samples. The corresponding theoretical values, $E(\delta_i)$ and $E(\sigma_\delta)$, were obtained using equations (3) and (4) and appear in columns 4 and 5. Similar statistics and their theoretical values for the parametric bootstrap appear in the remaining columns. Because the elements

TABLE I**

R^2	Monte Carlo: $y \sim N(X\beta, \sigma^2 I)$ $\bar{\delta}_i$	$\bar{\sigma}_\delta$	$E(\delta_i)$	$E(\sigma_\delta)$	Parametric Bootstrap: $y^* \sim N(Xb, \hat{\sigma}^2 I)$ $\bar{d}_i^*$	$\bar{\sigma}_d^*$	$E(d_i^*)$	$E(\sigma_d^*)$
.00001 [.00015]	.0018 (.009)	.572 (.041)	.0019	.559	.0041 (.010)	.730 (.002)	.0041	.727
.0100 [.1515]	.061 (.009)	.578 (.032)	.064	.566	.131 (.010)	.734 (.002)	.132	.731
.0250 [.3846]	.106 (.009)	.586 (.024)	.112	.578	.211 (.011)	.740 (.002)	.213	.738
.0500 [.7894]	.173 (.010)	.603 (.019)	.181	.596	.307 (.011)	.751 (.002)	.311	.749
.0750 [1.216]	.239 (.010)	.621 (.020)	.248	.622	.388 (.011)	.760 (.002)	.393	.759
.1000 [1.667]	.306 (.012)	.641 (.020)	.316	.645	.460 (.011)	.771 (.002)	.468	.770
.2500 [5.000]	.737 (.012)	.762 (.023)	.753	.766	.869 (.014)	.826 (.002)	.886	.828
.5000 [15.00]	1.63 (.014)	.890 (.025)	1.64	.890	1.68 (.015)	.909 (.002)	1.71	.906
.7500 [45.00]	3.15 (.015)	.958 (.029)	3.17	.959	3.17 (.015)	.969 (.002)	3.18	.961
.9000 [135.0]	5.69 (.015)	.983 (.045)	5.81	1.00	5.69 (.016)	.996 (.002)	5.81	1.00
Least Squares (b_i)								
Bias	0.000				0.000			
S.E.	0.997				1.010			

**$T=30$, $K=8$, $X'X=I_K$, $W=I_K$, $a=a_{max}/2$. Monte Carlo standard errors appear in parentheses. Noncentrality parameter appears in brackets, $[\lambda]$.

of β were chosen to be equal to one another and $X'X=I_K$, we can average over the additional dimension K. Monte Carlo standard errors appear in parentheses.

Near the origin, $R^2=.00001$ and the average value of δ_i over the Monte Carlo samples is .0018 and has standard error in the experiment of .572. This compares to expected values of .0019 and .559, which are derived from equations (3) and (4), respectively. Using the parametric bootstrap the average values of the i(th) element of the James-Stein estimator is .0041, while its average estimated standard error is .730. At the 5% level, these are not significantly different than the theoretical values $E(d_i^*)$ and $E(\sigma_d^*)$ which are derived from equations (10) and (11), respectively. Comparing $E(\sigma_d^*)$ to $E(\sigma_\delta)$ indicates that the parametric bootstrap estimate of standard error is expected to be nearly 30% larger than its theoretical value at the origin. Notice that the degree of bias falls as λ increases in size. As the sample size, T, increases the James-Stein estimator converges toward least squares and the bias associated with parametric bootstrapping will diminish. Increasing the number of bootstrap samples, N, cannot be expected to eliminate the bias; this is seen in equations (10) and (11) where the expectations have been taken over y^* and y and depend on T, not N.

In Table II the results from nonparametric bootstrapping are given. The Monte Carlo averages are repeated from Table I for convenience. In columns 4 and 5 nonparametric bootstrap samples were generated from randomly resampled and rescaled LS residuals from mean Xb. The next two columns use the LS residuals from mean $X\delta$. The final two columns contain results generated using unscaled and centered James-Stein residuals from mean $X\delta$.

Near the origin, finite sample standard errors are estimated most accurately using nonparametric bootstrap samples generated using $y^*=X\delta+e_\delta^*$. Underestimation of the finite sample standard errors of the James-Stein estimator occurs when $R^2\geq.05$ using $y^*=X\delta+e_\delta^*$ and when $R^2\geq.1$ using $y^*=X\delta+e^*$. Of the nonparametric bootstraps considered, that based on the LS residuals and mean

II

TABLE**

R^2	Monte Carlo $y \sim N(X\beta, \sigma^2 I)$ $\bar{\delta}_i$	$\bar{\sigma}_\delta$	$y^* = Xb + e^*$ $\bar{d}^*_i$	$\bar{\sigma}^*_d$	Nonparametric Bootstrap $y^* = X\delta + e^*$ $\bar{d}^*_i$	$\bar{\sigma}^*_d$	$y^* = X\delta + e^*_\delta$ $\bar{d}^*_i$	$\bar{\sigma}^*_d$
.00001	.0018	.572	.0040	.701	.001	.605	.0012	.590
[.00015]	(.009)	(.041)	(.010)	(.002)	(.006)	(.002)	(.005)	(.002)
.0100	.061	.578	.128	.706	.035	.607	.004	.589
[.1515]	(.009)	(.032)	(.010)	(.002)	(.006)	(.002)	(.006)	(.002)
.0250	.106	.586	.206	.711	.061	.610	.068	.590
[.3846]	(.009)	(.024)	(.010)	(.002)	(.006)	(.002)	(.006)	(.002)
.0500	.173	.603	.300	.721	.101	.617	.112	.592
[.7894]	(.010)	(.019)	(.011)	(.002)	(.006)	(.002)	(.006)	(.002)
.0750	.239	.621	.379	.731	.143	.624	.157	.595
[1.216]	(.010)	(.020)	(.011)	(.002)	(.006)	(.002)	(.006)	(.002)
.1000	.306	.641	.451	.740	.186	.632	.203	.600
[1.667]	(.012)	(.020)	(.011)	(.002)	(.007)	(.002)	(.006)	(.002)
.2500	.737	.762	.858	.797	.500	.858	.539	.652
[5.000]	(.012)	(.023)	(.012)	(.002)	(.009)	(.002)	(.007)	(.002)
.5000	1.63	.890	1.68	.879	1.34	.873	1.41	.759
[15.00]	(.014)	(.025)	(.014)	(.002)	(.014)	(.002)	(.014)	(.002)
.7500	3.15	.958	3.16	.939	2.95	.933	3.00	.814
[45.00]	(.015)	(.028)	(.015)	(.002)	(.014)	(.002)	(.014)	(.002)
.9000	5.69	.983	5.69	.967	5.57	.966	5.60	.833
[135.0]	(.015)	(.028)	(.015)	(.002)	(.015)	(.002)	(.015)	(.002)
Least Squares (b_i)								
Bias		.000	0.000					
S.E.		.997	0.998					

**$T=30$, $K=8$, $X'X=I_K$, $W=I_K$, $a=a_{max}/2$. Monte Carlo standard errors appear in parentheses. Noncentrality parameter appears in brackets, $[\lambda]$.

(i.e., using $y^*=Xb+e^*$) is generally a more conservative estimator of Cov(δ) since it underestimates the true standard error over a smaller portion of the parameter space than its competitors.

Increasing the number of bootstrap samples to 1000 and 2000 had virtually no effect on the estimates obtained. In fact, estimated standard errors reported in Tables I and II changed by less than .002 at the origin when setting N=2000. The similarity between the parametric and nonparametric approaches for given N is expected to increase as T increases, i.e., as the empirical distribution of $\hat{e}$ converges to that of e. However, as T increases the Stein estimator itself converges to LS and its use in large samples offers little if any risk advantage over least squares.

CONCLUSION

The parametric and nonparametric bootstrapping are reasonably accurate ways to estimate finite sample standard errors for the James-Stein rule. The bootstrap provides a closer approximation of the actual standard error than estimates based on the asymptotic covariance matrix which for the JS rule coincides with that of least squares. The parametric bootstrap estimator of the James-Stein's standard error tends to be the most conservative of the versions considered here, but is the least accurate when β lies close to the origin. At this point the percentage bias is nearly 30%. The farther β lies from 0, the more accurate the bootstrap becomes. The degree of bias near the origin can be reduced in a number of practical ways which include creating pseudo-samples generated using $y^*=X\delta+e^*_\delta$ or $y^*=X\delta+e^*$. However, these methods tend to cause underestimation of standard error over much of the parameter space. The parametric bootstrap for which finite sample results are obtained is less prone to this problem. In fact the theoretical results reported in this paper indicate that it is biased upward no matter where β lies.

A general implication of these results is that the finite sample accuracy of the bootstrap approximation depends on the values of the model's parameters. That is, its accuracy depends on the very quantity that we are trying to estimate. This suggests that it may be important to study the bootstrap's properties before conclusions are drawn based on its use in a given application. For the James-Stein rule the situation is simpler than it will usually be because we know its finite sample moments and are able to derive straightforward results for the bootstrap moments using normal distribution theory. For other estimators these results are often unavailable (e.g., impulse response functions in the VAR application) and our best recourse is to compare moments obtained from a Monte Carlo study with the average value of the bootstrap moments obtained over the same simulation. Such computer intensive exercise tends to be expensive and the results often are difficult to generalize. Nevertheless, the finite sample bias of the bootstrap even for a fairly simple estimator like the James-Stein rule can be significant and depends on the way the unknown parameters enter the actual moments of the estimator. In a similar way we would also expect bootstrapping to be biased for the covariance of certain ridge estimators where the unknown parameters appear as nonlinear functions in the moments of the estimator.

ACKNOWLEDGEMENTS

I'd like to thank the two anonymous referees for their careful reading of the manuscript and for their many suggestions on ways to improve the paper.

REFERENCES

Adkins, L. C., (1990). "Small Sample Performance of Jackknife Confidence Intervals for the James-Stein Estimator," Communications in Statistics, B19, 401-418.

Adkins, L. C. and Hill, R. C., (1990). "An Improved Confidence Ellipsoid for the Linear Regression Model," Journal of Statistical Computation and Simulation, 36, 9-18.

Aigner, D. J. and Judge, G. G., (1977). "Application of Pre-Test and Stein Estimators to Economic Data," Econometrica, 45, 1279-1288.

Anderson, T. W., (1984). An Introduction to Multivariate Statistical Analysis, Second Edition. New York: John Wiley and Sons.

Beran, R., (1982). "Estimated Sampling Distributions: The Bootstrap and Competitors," Annals of Statistics, 10, 212-225.

Beran, R., (1988). "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," Journal of the American Statistical Association, 83, 687-697.

Beran, R., (1990). "Refining Bootstrap Simultaneous Confidence Sets," Journal of the American Statistical Association, 85, 417-426.

Bickel, P. and Freedman, D. A., (1981). "Some Asymptotic Theory for the Bootstrap." Annals of Statistics, 9, 1196-1217.

Brownstone, D., (1990). "Bootstrapping Improved Estimators for Linear Regression Models." Journal of Econometrics, 44, 171-188.

Carter, R. A. L., Srivastava, M. S., Srivastava, V. K., and Ullah, A., (1990). "Unbiased Estimation of the MSE Matrix of Stein-Rule Estimators, Confidence Ellipsoids, and Hypothesis Testing." Econometric Theory, 6, 63-74.

Chi, X. W. and Judge, G., (1985). "On Assessing the Precision of Stein's Estimator." Economics Letters, 18, 143-148.

Efron, B., (1982). The Jackknife, Bootstrap, and Other Resampling Plans. Philadelphia: Society for Industrial and Applied Mathematics.

Engle, R. F., (1984). "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics." In Handbook of Econometrics, Volume II. Edited by Z. Griliches and M. D.

Intriligator. Amsterdam: North-Holland.

Freedman, D. A. and Peters, S. C., (1981). "Bootstrapping a Regression Equation: Some Empirical Results." Journal of the American Statistical Society, 79, 97-106.

Hill, R. C., Ziemer, R. F., and White, Fred C., (1981). "Mitigating the Effects of Multicollinearity Using Exact and Stochastic Restrictions: The Case of an Aggregate Agricultural Production Function in Thailand," The American of Agricultural Economics, 63, 298-300.

James, W. and Stein, C., (1961). "Estimation with Quadratic Loss." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1. Berkeley: University of California Press, 361-379.

Judge, G. G., Hill, R. C., and Bock, M. E., (1990). "An Adaptive Empirical Bayes Estimator of the Multivariate Normal Mean Under Quadratic Loss." Journal of Econometrics, 44, 189-213.

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

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

Phillips, P. C. B., (1984). "The Exact Distribution of the Stein-Rule Estimator." Journal of Econometrics, 25, 123-132.

Runkle, D. E., (1987). "Vector Autoregressions and Reality." Journal of Business and Economic Statistics, 5, 437-442.

Sim, A. B., (1989). "Bootstrapping Single Equation Regression Models: Some Finite Sample Results." Ph.D. Thesis, Concordia University, Montreal, Canada.

Ullah, A., Carter, R. A. L., and Srivastava, V. K., (1984). "Sampling Distribution of Shrinkage Estimators and their F-Ratios in the Regression Model." Journal of Econometrics, 25, 109-122.

Wu, C. J., (1986). "Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis." Annals of Statistics, 14, 1261-1295.

APPENDIX A

In this appendix, theorems used to prove the results in the paper are given and a sketch of how the results were derived are presented. The symbol $\phi(.)$ denotes a Borel measurable function.

Theorem 1: (Judge and Bock, p. 321)

If the Jx1 vector $w \sim N(\omega, I_J)$, then

$$E[\phi(w'w)w] = \omega E[\phi(\chi^2_{J+2,\omega'\omega/2})]. \tag{A1}$$

Theorem 2: (Judge and Bock, p.323)

If the Jx1 vector $w \sim N(\omega, I_J)$, then

$$E[\phi(w'w)ww'] = E[\phi(\chi^2_{J+2,\omega'\omega/2})]I_J + E[\phi(\chi^2_{J+4,\omega'\omega/2})]\omega\omega' . \tag{A2}$$

Using theorems 1 and 2 we will derive $E(d^*)$ and $Cov(d^*)$ for the orthonormal model. Let

$$y = Z\theta + e \qquad e \sim N(0,\sigma^2 I_T),\ Z'Z = I_K \tag{A3}$$

The least squares estimator of θ is $\hat{\theta} = Z'y \sim N(\theta, \sigma^2 I)$ and the bootstrap estimator $\hat{\theta}^* | \hat{\theta} = Z'y^* \sim N(\hat{\theta}, \hat{\sigma}^2 I)$ where $\hat{\sigma}^2 = (y - Z\hat{\theta})'(y - Z\hat{\theta})/(T-K)$ with $(T-K)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{T-K}$ and independent of $\hat{\theta}$.

Theorem 3: If $\hat{\theta}^* | \hat{\theta} \sim N(\hat{\theta}, \hat{\sigma}^2 I)$, $\hat{\theta} = Z'y \sim N(\theta, \sigma^2 I)$, and $d^* = (1 - as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)\hat{\theta}^*$ where $s^* = (y^* - Z\hat{\theta}^*)'(y^* - Z\hat{\theta}^*)$, then

$$E_y[d^*] = [1 - a(T-K)E_*(1/\chi^2_{K+2,\eta_0})]\theta \tag{A4}$$

where $\eta_0 = \frac{1}{2}\frac{(T-K)}{(T-K-2)}E[\chi^2_{K+2,\lambda}] = \frac{1}{2}\frac{(T-K)}{(T-K-2)}(K+2+2\lambda)$ and $\lambda = \theta'\theta/2\sigma^2$.

Proof: Let $w = \frac{\hat{\theta}^*}{\hat{\sigma}} \sim N(\frac{\hat{\theta}}{\hat{\sigma}}, I_K)$ and note $s^*/\hat{\sigma}^2 \sim \chi^2_{T-K}$ which is statistically independent of w.

$$E_y(d^*) = E_y[E_*(d^*|\hat{\theta})]$$

$$E_*(d^*|\hat{\theta}) = E\left[\hat{\sigma}[1 - a(s^*/\hat{\sigma}^2)/(\hat{\theta}^{*\prime}\hat{\theta}^*/\hat{\sigma}^2)]\ \hat{\theta}^*/\hat{\sigma}\right]$$

$$= E\left[\hat{\sigma}\ [1 - a\ \chi^2_{T-K} / w'w\ ]\ w\right] \tag{A5}$$

Apply theorem 1 to (A5). This yields

$$E_*(d^*|\hat{\theta}) = \left[1 - a\ (T-K)\ E(1/\chi^2_{K+2,\hat{\lambda}})\right]\hat{\theta} \tag{A6}$$

where $\hat{\lambda} = (\hat{\theta}'\hat{\theta})/(2\hat{\sigma}^2)$. Redefine $w = \frac{\hat{\theta}}{\sigma} \sim N(\frac{\theta}{\sigma}, I_K)$ and note $\hat{\lambda} = (\hat{\theta}'\hat{\theta}/\sigma^2)/(2\hat{\sigma}^2/\sigma^2) = \frac{(T-K)\ w'w}{2\ \chi^2_{T-K}}$

Rewrite $E_*(d^*|\hat{\theta}) = \sigma \left[1-a\ (T-K)\ E(1/\chi^2_{K+2,\hat{\lambda}}) \right] w$ and take the expectation with respect to y by applying theorem 1, noting that w and $\hat{\sigma}^2/\sigma^2 \sim \chi^2_{T-K}/(T-K)$ are independent.

$$E_y(d^*) = E_y[E_*(d^*|\hat{\theta})] = \left[1-a\ (T-K)\ E(1/\chi^2_{K+2,\eta_0}) \right] \theta \quad \blacksquare \qquad \text{(A7)}$$

Theorem 4.

$$E_y(E_*[(d^*-E[d^*])(d^*-E[d^*])']) =$$
$$\sigma^2(1-2a(T-K)E_*(l_{22})+a^2(T-K)(T-K+2)E_*[(l_{22})^2]\ I_K$$
$$+ \sigma^2 \Big\{ 2a(T-K)(E_*(l_{20}) - E_*(l_{40})) +$$
$$a^2(T-K)((T-K+2)E_*[(l_{40})^2] - (T-K)\ E_*[(l_{20})^2]) \Big\} I_K$$
$$+ \Big\{ 2a(T-K)(E_*(l_{21}) - E_*(l_{41})) +$$
$$a^2(T-K)((T-K+2)E_*[(l_{41})^2] - (T-K)\ E_*[(l_{21})^2]) \Big\}\ \theta\theta'$$

where $\eta_0 = \frac{1}{2}\frac{(T-K)}{(T-K-2)} E[\chi^2_{K+2,\lambda}] = \frac{1}{2}\frac{(T-K)}{(T-K-2)}(K+2+2\lambda)$, $\eta_1 = \frac{1}{2}\frac{(T-K)}{(T-K-2)} E[\chi^2_{K+4,\lambda}] = \frac{1}{2}\frac{(T-K)}{(T-K-2)}(K+4+2\lambda)$, $\eta_2 = \frac{1}{2}E[\chi^2_{K,\lambda}] = \frac{1}{2}(K+2\lambda)$,
$E_*(l_{22}) = E_*[1/\chi^2_{K+2,\eta_2}]$, $E_*(l_{42}) = E_*[1/\chi^2_{K+4,\eta_2}]$,
$E_*(l_{20}) = E_*[1/\chi^2_{K+2,\eta_0}]$, $E_*(l_{40}) = E_*[1/\chi^2_{K+4,\eta_0}]$,
$E_*(l_{21}) = E_*[1/\chi^2_{K+2,\eta_1}]$, $E_*(l_{41}) = E_*[1/\chi^2_{K+4,\eta_1}]$ and $\lambda = \theta'\theta/2\sigma^2$.

Proof:

$$E_*((d^*-E[d^*])(d^*-E[d^*])') =$$
$$E_*\Big[(1-as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)^2\ \hat{\theta}^*\hat{\theta}^{*\prime} - (1-as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)\hat{\theta}^*\hat{\theta}' -$$
$$(1-as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)\hat{\theta}\hat{\theta}^{*\prime} + [1-a\ (T-K)\ E(1/\chi^2_{K+2,\hat{\lambda}})]^2\ \hat{\theta}\hat{\theta}' \Big] =$$

$$E_*\Big[(1-as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)^2\ \hat{\theta}^*\hat{\theta}^{*\prime} - [1-a\ (T-K)\ E(1/\chi^2_{K+2,\hat{\lambda}})]^2\ \hat{\theta}\hat{\theta}' \Big] \qquad \text{(A8)}$$

Let $w = \frac{\hat{\theta}^*}{\hat{\sigma}} \sim N(\frac{\hat{\theta}}{\hat{\sigma}}, I_K)$ and note $s^*/\hat{\sigma}^2 \sim \chi^2_{T-K}$.

Consider the term $E_* \left[(1-as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)^2 \; \hat{\theta}^*\hat{\theta}^{*\prime} \right]$ from (A8).

$$E_* \left[(1-as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)^2 \; \hat{\theta}^*\hat{\theta}^{*\prime} \right] = E_* \left[\hat{\sigma}^2 \; (1 - a\, \chi^2_{T-K} \;/\; w'w)^2 \; ww' \right] \quad \text{(A9)}$$

Apply theorem 2 to the r.h.s. of (A9).

$$E_* \left[(1-as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)^2 \; \hat{\theta}^*\hat{\theta}^{*\prime} \right] =$$

$$\hat{\sigma}^2 \left[1 - 2a(T-K)\; E(1/\chi^2_{K+2,\hat{\lambda}}) + a^2\; (T-K)(T-K+2)\; E\{(1/\chi^2_{K+2,\hat{\lambda}})^2\} \right] I_K$$

$$+ \left[1 - 2a(T-K)\; E(1/\chi^2_{K+4,\hat{\lambda}}) + \right.$$

$$\left. a^2\; (T-K)(T-K+2)\; E\{(1/\chi^2_{K+4,\hat{\lambda}})^2\} \right] \hat{\theta}\hat{\theta}' \quad \text{(A10)}$$

Now, take the expectation of (A10) with respect to y. To do this, redefine $w = \frac{\hat{\theta}}{\sigma} \sim N(\frac{\theta}{\sigma}, I_K)$ and note $\hat{\lambda} = (\hat{\theta}'\hat{\theta}/\sigma^2)/\;(2\hat{\sigma}^2/\sigma^2) = \frac{(T-K)\; w'w}{2\; \chi^2_{T-K}}$ with $\hat{\sigma}^2 = \sigma^2\chi^2_{T-K}/(T-K)$. Then

$$E_y \left\{ E_* \left[(1-as^*/\hat{\theta}^{*\prime}\hat{\theta}^*)^2 \; \hat{\theta}^*\hat{\theta}^{*\prime} \right] \right\} =$$

$$E_y \left\{ \frac{\sigma^2\chi^2_{T-K}}{(T-K)} \left[1 - 2a(T-K)\; E(1/\chi^2_{K+2,\hat{\lambda}}) + \right.\right.$$

$$\left. a^2\; (T-K)(T-K+2)\; E\{(1/\chi^2_{K+2,\hat{\lambda}})^2\} \right] I_K$$

$$+ \sigma^2 \left[1 - 2a(T-K)\; E(1/\chi^2_{K+4,\hat{\lambda}}) + \right.$$

$$\left.\left. a^2\; (T-K)(T-K+2)\; E\{(1/\chi^2_{K+4,\hat{\lambda}})^2\} \right] ww' \right\}. \quad \text{(A11)}$$

Note that $E[\phi(\chi^2_n)\chi^2_n] = nE[\phi(\chi^2_{n+2})]$ [see Judge and Bock (1978) p. 320], hence the expectation of the first term in (A11) is

$$E_y \left\{ \frac{\sigma^2\chi^2_{T-K}}{(T-K)} \left[1 - 2a(T-K)\; E(1/\chi^2_{K+2,\hat{\lambda}}) + \right.\right.$$

$$\left.\left. a^2\; (T-K)(T-K+2)\; E\{(1/\chi^2_{K+2,\hat{\lambda}})^2\} \right] I_K \right\}$$

$$= \sigma^2 \left[1 - 2a(T-K)\; E(1/\chi^2_{K+2,\eta_2}) + \right.$$

$$\left. a^2\; (T-K)(T-K+2)\; E\{(1/\chi^2_{K+2,\eta_2})^2\} \right] I_K \quad \text{(A12)}$$

Consider the second term of (A11):

$$E_y \left\{ \sigma^2 \left[1 - 2a(T-K)\ E(1/\chi^2_{K+4,\hat{\lambda}}) + \right.\right.$$

$$\left.\left. a^2\ (T-K)(T-K+2)\ E\{(1/\chi^2_{K+4,\hat{\lambda}})^2\} \right] ww' \right\} .$$

Using Theorem 2 yields

$$\sigma^2 \left[1 - 2a(T-K)\ E(1/\chi^2_{K+4,\eta_0}) + \right.$$

$$\left. a^2\ (T-K)(T-K+2)\ E\{\ (1/\chi^2_{K+4,\eta_0})^2\} \right] I_K +$$

$$\left[1 - 2a(T-K)\ E(1/\chi^2_{K+4,\eta_1}) + \right.$$

$$\left. a^2\ (T-K)(T-K+2)\ E\{(1/\chi^2_{K+4,\eta_1})^2\} \right] \theta\theta' \qquad \text{(A13)}$$

Now, taking the expectation of the second term in (A8) with respect to y using Theorem 2 yields

$$E_y \left[[1-a\ (T-K)\ E(1/\chi^2_{K+2,\hat{\lambda}}\ )]^2\ \hat{\theta}\hat{\theta}' \right] =$$

$$\sigma^2 \left[1 - 2a(T-K)\ E(1/\chi^2_{K+2,\eta_0}) + \right.$$

$$\left. a^2\ (T-K)^2\ \{E(1/\chi^2_{K+2,\eta_0})\}^2 \right] I_K +$$

$$\left[1 - 2a(T-K)\ E(1/\chi^2_{K+2,\eta_1}) + a^2\ (T-K)^2\ \{E(1/\chi^2_{K+2,\eta_1})\}^2 \right] \theta\theta'$$

(A14)

Combining (A14), (A12), and (A13) yields

$$= \sigma^2 \left[1 - 2a(T-K)\ E(1/\chi^2_{K+2,\eta_2}) + \right.$$

$$\left. a^2\ (T-K)(T-K+2)\ E\{\ (1/\chi^2_{K+2,\eta_2})^2\} \right] I_K \Big\} +$$

$$\sigma^2 \left[1-2a(T-K)\ E(1/\chi^2_{K+4,\eta_0}) + a^2(T-K)(T-K+2)\ E\{\ (1/\chi^2_{K+4,\eta_0})^2\} \right] I_K$$

$$+ \left[1 - 2a(T-K)\ E(1/\chi^2_{K+4,\eta_1}) + a^2\ (T-K)(T-K+2)\ E\{\ (1/\chi^2_{K+4,\eta_1})^2\} \right] \theta\theta'$$

$$- \sigma^2 \left[1 - 2a(T-K)\ E(1/\chi^2_{K+2,\eta_0}) + a^2\ (T-K)^2\ \{E(1/\chi^2_{K+2,\eta_0})\}^2 \right] I_K$$

$$- \left[1 - 2a(T-K)\, E(1/\chi^2_{K+2,\eta_1}) + a^2\, (T-K)^2\, (E(1/\chi^2_{K+2,\eta_1}))^2 \right] \theta\theta'$$

Rearranging yields the desired result.■

To prove the results in the text of the paper for the nonorthonormal model we introduce the canonical form.

$$y = X\beta + e = X\, V^{-1}\, V\, \beta + e = Z\theta + e$$

where $Z=XV^{-1}$, $\theta=V\beta$, $Z'Z=I_K$. Replacing θ with $V\beta$, and noting that $Vb=\hat{\theta}$, $Vb^*=\theta^*$, $V'V=X'X$, and $b'X'Xb/2\hat{\sigma}^2= \hat{\theta}'\hat{\theta}/2\hat{\sigma}^2$ enables the above Theorems to be used for the nonorthonormal model.