EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 3, No. 3, 2010, 371-381
ISSN 1307-5543 – www.ejpam.com

SPECIAL ISSUE ON
GRANGER ECONOMETRICS AND STATISTICAL MODELING
DEDICATED TO THE MEMORY OF PROF. SIR CLIVE W.J. GRANGER

Bootstrapping the Shrinkage Least Absolute Deviations Estimator

Tae-Hwan Kim[1*] and Halbert White[2]

[1] *School of Economics, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul, 120-749, Korea*
[2] *Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508*

Abstract. Kim and White [13] studied a James-Stein type estimator that shrinks towards a data-dependent point rather than a fixed point. This was subsequently extended and applied to combining the OLS and 2SLS estimators by [12, 14]. This approach can be used to combine any two estimators in an optimal way. While the risk dominance properties of the new shrinkage estimator have been well established, a clear prescription for how to conduct inference and hypothesis testing has been missing. In this paper, we close this gap using a bootstrap approach.

2000 Mathematics Subject Classifications: 62

Key Words and Phrases: Shrinkage, James-Stein type estimators

---

1. Introduction

In commemorating Clive W.J. Granger's life and work, it is fitting to note his delight in tackling interesting issues from unconventional angles, frequently opening up new areas of research and new ways of thinking about important topics in econometrics. As Jim Stock noted in his discussion of Granger's work in a memorial session at the 2010 American Economic Association meetings, perhaps one of the most interesting developments in modern econometrics has been the very different asymptotic distribution theory required to treat the estimators emerging from Granger's Nobel Prize-winning work on cointegration [6, 5], compared to the standard asymptotic normality results that previously prevailed.

*Corresponding author.

Email addresses: `tae-hwan.kim@yonsei.ac.kr` (T. Kim), `hwhite@ucsd.edu` (H. White)

Here we offer analysis that also involves unusual asymptotics, growing in a perhaps unanticipated way that we hope Clive would find gratifying, from distribution theory developed by [18], designed to better describe *finite* sample estimator distributions. The estimators we study are shrinkage estimators that have themselves been taken in a novel direction. Indeed, the seminal work of [17] and [11] transformed statistical thinking by showing the inadmissibility of the ordinary least squares (OLS) estimator and initiating an entire field devoted to the study of shrinkage techniques designed to gain the "optimal trade-off" between bias and variance. At the outset, most research in this area focused exclusively on shrinking a given base estimator towards a fixed point, such as zero. As is well understood, this shrinkage vanishes asymptotically. Subsequently, researchers' understanding of the issues involved in shrinkage deepened, and a different shrinkage approach emerged, one that shrinks estimators towards a data-dependent point. See, for example, [15, 16]. In contrast to fixed point shrinkage, data-dependent shrinkage does not vanish asymptotically.

This approach has been extended to a fairly general context by [13], where the data-dependent point can be any other potentially biased estimator for the same parameter of interest and is allowed to be correlated with the base estimator itself. The method proposed in [13] can thus be viewed as a general way to combine two different estimators to improve estimation risk and prediction precision. A finite sample counterpart of the Kim and White estimator has been developed by [12]; in [14], this approach was used to combine the biased OLS estimator and the unbiased 2SLS estimator in the presence of endogeneity.

Nevertheless, a drawback of the shrinkage estimators developed in [13] has been the lack of a measure of precision. The usual Gaussian asymptotics do not apply, as the limiting distribution of these shrinkage estimators is, as alluded to above, not normal. Instead, it is a nonlinear function of a normal random vector. Here we propose bootstrap methods to fill this gap.

Using the bootstrap to obtain the sampling distribution of shrinkage estimators is a familiar approach in the literature. Vinod and Raj [20] apply a bootstrap method to a ridge-type shrinkage estimator to investigate economic issues in the Bell System divestiture. Brownstone [3] bootstraps two shrinkage estimators: Mundlak's restricted principal-components estimator and a Stein-rule estimator that shrinks the OLS estimator toward the Mundlak estimator. Brownstone uses non-pivotal statistics, i.e., the percentile method, to get the sampling distributions. He shows that the non-parametric bootstrap provides a good estimate of the estimator's risk and standard errors. Vinod [19] provides a solution to the non-pivotal problem for ridge regression by applying Beran's double bootstrap. This method involves a bootstrap within a bootstrap, which is computationally intensive.

The importance of using pivotal statistics to get a better bootstrap confidence interval is well known. See, for example, [2, 8, 9]. In this paper, we focus on bootstrapping James-Stein-type estimators that shrink towards a data-dependent point, working with pivotal statistics obtained by using lemmas 1 and 2 of [18]. We then use these pivotal statistics for the bootstrapping approximation to obtain standard errors and confidence intervals for our James-Stein-type shrinkage estimators.

2. The Shrinkage Estimator and its Asymptotic Moments

Suppose that one is interested in estimating a k-dimensional parameter vector β^0 from data generated as $y_t = X_t'\beta^0 + \varepsilon_t$, $t = 1, 2, \ldots, n$, where we assume that (i) (X_t', ε_t) is independent and identically distributed (IID); and (ii) $E(\varepsilon_t) = 0$, $0 < \sigma^2 \equiv \text{var}(\varepsilon_t) < \infty$. Suppose that there is an estimator b_n of β^0 based on n observations of y_t and X_t. Traditional James-Stein estimators shrink towards a fixed point, usually zero when there is no prior information. Extending earlier proposals, [13] propose shrinking b_n towards a data-dependent point, g_n, where the base estimator b_n and the data-dependent point g_n are assumed to satisfy the following mild condition:

$$\begin{bmatrix} n^{1/2}(b_n - \beta^0) \\ n^{1/2}(g_n - \beta^0) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim N(\xi, \Sigma), \tag{1}$$

where $\xi \equiv \begin{bmatrix} 0 \\ \theta \end{bmatrix}$, $\Sigma \equiv \begin{bmatrix} A & \Delta \\ \Delta' & B \end{bmatrix}$, and A, B, and Σ are symmetric positive definite matrices. In their application, Kim and White used the least absolute deviations (LAD) estimator for b_n and the OLS estimator for g_n. For concreteness, we will use the same estimators here. The "Optimal Weighting Scheme" (OWS) estimator is

$$\delta^{OW}(b_n, g_n) = \left(1 - \lambda_1 - \frac{\lambda_2}{(b_n - g_n)'Q_n(b_n - g_n)}\right)(b_n - g_n) + g_n, \tag{2}$$

where λ_1 and λ_2 are weights to be optimally chosen [see 13] and Q_n is a random symmetric positive definite matrix defining the quadratic loss of estimation, $(b_n - \beta^0)'Q_n(b_n - \beta^0)$. We assume that $n^{-1}Q_n$ converges in probability to a nonstochastic symmetric positive definite matrix Q.

The limiting distribution of the OWS estimator is not a member of a location-scale family. If it were, then the usual studentization would be enough to yield a pivotal statistic. Nevertheless, the result of [1] shows that the bootstrap can still estimate the true sampling distribution up to second-order terms, justifying studentization for non-location-scale families.

With the same assumptions and notations used in [13] and using lemmas 1 and 2 of [18], it is straightforward to show that

$$n^{1/2}(\delta^{OW}(b_n, g_n) - \beta^0) \xrightarrow{d} \left(1 - \lambda_1 - \frac{\lambda_2}{(U_1 - U_2)'Q(U_1 - U_2)}\right)(U_1 - U_2) + U_2. \tag{3}$$

We let $h(U)$ denote the limiting distribution in (3). Note that there is no asymptotic bias in this case since the data-dependent point (the OLS estimator) is consistent. Since Σ is positive definite, there is a matrix P, such that $\Sigma = PP'$. Let $Z = P^{-1}U$. Then Z is normally distributed with mean vector $\mu = \begin{bmatrix} 0_{k\times 1} \\ 0_{k\times 1} \end{bmatrix}$ and identity covariance matrix. The limiting random variable, $h(U)$, can be rewritten as

$$h(U) = \left(1 - \lambda_1 - \frac{\lambda_2}{Z'P'J_1'QJ_1PZ}\right)J_1PZ + J_2PZ \equiv H(Z), \tag{4}$$

where $J_1 \equiv [I_k; -I_k]$ and $J_2 \equiv [0_k; I_k]$. Let $M_1 \equiv P'J_1'QJ_1P$, $M_2 \equiv J_1P$, and $M_3 \equiv J_2P$. Then we can further simplify $H(Z)$ to

$$H(Z) = MZ - \begin{bmatrix} \frac{\lambda_2 M_2 Z}{Z'M_1 Z} \end{bmatrix}, \tag{5}$$

where $M \equiv (1-\lambda_1)M_2 + M_3$. Using lemmas 1 and 2 of [18], the following new results provide the first and second moments of the limiting distribution of the OWS estimator.

Theorem 1. *Let $H_i(Z)$ be the ith component of $H(Z)$. Then, under regularity conditions allowing the exchange of limit and integral, $E(H_i(Z)) = 0$.*

The next theorem shows the asymptotic second moment of the OWS estimator with the following definitions; let M_{ij} and M_{2ij} be the $(i,j)^{th}$ element of M and M_2 respectively, and let the indicator matrix I_{ij} be the zero matrix except with element (i,j) equal to one.

Theorem 2. *Let v_{ij} be the (i,j)th element of $E(H(Z)H(Z)')$. Then, under the same conditions as in Theorem 3 in [13], we have*

$$v_{ij} = a_{ij} - b_{ij} - c_{ij} + d_{ij},$$

where

1. *a_{ij} is the (i,j)th element of MM';*
2. *$b_{ij} = \lambda_2 \sum_{a=1}^{2k} \sum_{b=1}^{2k} M_{ia} \mathscr{E}_{ab} M_{2jb}$, with $\mathscr{E}_{ab} = \Gamma(1)^{-1} \int_0^\infty \left| N_{0t} \right|^{-1/2} tr(I_{ab} N_{0t}^{-1})\, dt$ and $N_{0t} = I + 2tM_1$;*
3. *$c_{ij} = b_{ji}$;*
4. *$d_{ij} = \lambda_2^2 \sum_{a=1}^{2k} \sum_{b=1}^{2k} M_{2ia} \mathscr{F}_{ab} M_{2jb}$, with $\mathscr{F}_{ab} = \Gamma(2)^{-1} \int_0^\infty t \left| N_{0t} \right|^{-1/2} tr(I_{ab} N_{0t}^{-1}) dt$.*

The diagonal terms v_{ii} are the asymptotic variances of the OWS estimates. We use these to construct a studentized statistic for bootstrapping. The only input needed to estimate v_{ii} is a consistent estimator of Σ. The diagonal sub-matrices A and B of Σ are easily estimated, since these are the covariance matrices of the LAD and the LS estimators. A consistent estimator of the off-diagonal matrix Δ is also provided in [13] as follows:

$$\widehat{\Delta} = \left[\widehat{f}(0) n^{-1} \sum_{t=1}^{n} X_t X_t' \right]^{-1} \left[n^{-1} \sum_{t=1}^{n} S_{1t}, S_{2t}' \right] \left[n^{-1} \sum_{t=1}^{n} X_t X_t' \right]^{-1}, \tag{6}$$

where $S_{1t} = -X_t(1_{[\varepsilon_t \leq 0]} - 0.5)$, $S_{2t} = X_t \varepsilon_t$, and $\widehat{f}(0)$ is a kernel estimator of the density of ε_t evaluated at zero.

3. Bootstrap Confidence Intervals

In this section, we show how the results of the previous section can be used in bootstrapping the OWS estimator. Bootstrap methods for the LAD estimator (the base estimator in our case) are not new. [7] bootstraps the quantile estimator and shows that the bootstrap distribution converges weakly to the limiting distribution of the quantile estimator. However, bootstrapping a shrinkage LAD estimator (the OWS estimator in our case) has not previously been studied.

We generate artificial data as $y = X\beta^0 + \varepsilon$, where $\varepsilon \in \mathbb{R}^n$ and $\beta^0 \in \mathbb{R}^k$, with $n = 80$ and $k = 3$. We set $\beta^0 = 0$ without loss of generality. We draw errors from the standard normal distribution. Each row of X is drawn from the joint normal distribution[†], $N(1, \Omega)$, where the covariances are each 0.8 and the variances are one. Once the first set of data is generated, we take it to be our original data and pretend not to know the true value of β^0. The OWS estimator, δ_{ni}, and its asymptotic standard deviation, s_i, are computed using the original data and the method described in the previous section.

We consider (i) the equal tail percentile$-t$ method (studentized); (ii) the equal tail percentile method (unstudentized); (iii) the naïve percentile method; and (iv) the "normal approximation" method for constructing confidence intervals for β^0. For the percentile$-t$ method[‡], the population equation is given by

$$\Pr[t_L^p < n^{1/2}(\delta_{ni} - \beta_i^0)/s_i < t_U^p] = 1 - \alpha. \tag{7}$$

The ideal $(1-\alpha)\%$ confidence interval is (t_L^p, t_U^p), but we cannot obtain this interval because the exact distribution of δ_{ni} is not known. The population equation in (7) can be approximated by the following sample equation:

$$\Pr[t_L^s < n^{1/2}(\delta_{ni}^* - \delta_{ni})/s_i^* < t_U^s] = 1 - \alpha \tag{8}$$

where δ_{ni}^*, s_i^* are bootstrap estimates.

Even though the sample equation in (8) can be solved in principle, in most cases this is intractable because the empirical distribution from which the bootstrap re-sampling is taken is not continuous. Hence, we approximate the solution to the sample equation using bootstrap re-sampling. By bootstrapping (y_t, X_t) pairs[§], we generate $\{\gamma_i : i = 1, \ldots, m\}$, where $\gamma_i = n^{1/2}(\delta_{ni}^* - \delta_{ni})/s_i^*$ and m is the number of bootstrap re-samples. We take the $\alpha/2$ percentile $(\widehat{t_L^s})$ and $(1-\alpha/2)$ percentile $(\widehat{t_U^s})$ to be approximations for t_L^s and t_U^s, respectively. The $(1-\alpha)\%$ bootstrap confidence interval is given by

$$[\delta_{ni} - \widehat{t_U^s} s_i/n^{1/2}, \delta_{ni} - \widehat{t_L^s} s_i/n^{1/2}]. \tag{9}$$

[†]We use the multivariate normal random vector generator, DNRVG, which is a FORTRAN subroutine in NSWC (Naval Surface Warfare Center) Library. We set the seed to be 3833981.

[‡]For the percentile method the population equation is Pr $[t_L^p < \delta_{ni} - \beta_i^0 < t_U^p] = 1 - \alpha$ and the sample equation is $\Pr[t_L^s < \delta_{ni}^* - \delta_{ni} < t_U^s] = 1 - \alpha$. We compute the naïve percentile confidence intervals by taking the $\alpha/2$ and $(1-\alpha/2)$ percentiles of $\{\delta_{ni}^*; i = 1, 2, \ldots, m\}$. The normal approximation interval is given by $[\delta_{ni} - 1.96 s_i/n^{1/2}$, $\delta_{ni} + 1.96 s_i/n^{1/2}]$.

[§][3, 19] bootstrap residuals. We prefer bootstrapping pairs because this is more robust to assumptions on the error term. See [4].

Table 1: Standard Errors and t-values.

Estimator	Variable	Coefficient	Std. Error	t-value
LS	x_1	0.077618	0.172397	0.450227
	x_2	−0.005561	0.169351	−0.032840
	x_3	0.015524	0.190734	0.081391
LAD	x_1	−0.001885	0.221519	−0.008510
	x_2	0.050000	0.217604	0.229775
	x_3	0.045658	0.245080	0.186299
OWS	x_1	0.072559	0.171099	0.424076
	x_2	−0.002030	0.168950	−0.011990
	x_3	0.017442	0.189935	0.091832

Table 1 summarizes the regression results using the original data. The LAD estimates have uniformly higher standard errors than the OLS estimates as we should expect, since the OLS estimator is the maximum likelihood estimator here. The standard errors of the OWS estimator are quite comparable to those of the OLS estimator.

Table 2[¶] shows the bootstrap 95% confidence intervals and some descriptive statistics such as the length and shape of the intervals[‖]. The percentile$-t$ bootstrap confidence intervals are [-.243, .434], [-.336, .296], and [-.349, .406]. The confidence intervals cover the true β^0 correctly. Note that these are not symmetric intervals, according to the reported shape statistics. The non-symmetry can be visualized using the histograms of the standardized bootstrap estimates given in Figures 1-3. In many cases, enforcing symmetry will cause size distortions. Allowing for asymmetry can be considered an advantage of using the bootstrap method. The percentile confidence intervals are fairly comparable to the percentile$-t$ confidence intervals. They also are not symmetric and are a little bit shorter. As expected, the naïve percentile intervals have the exact same length as the percentile intervals, because the naïve percentile intervals are a shifted version of the percentile intervals.

[¶]Let 'cl' be the lower bound and 'cu' be the upper bound of a confidence interval. Then 'Length' and 'Shape' are defined as follows. (1) Length = cu - cl. (2) Shape = (cu - b)/(b - cl) where b is the JSLAD estimate computed from the original data set. Shape measures how asymmetric the bootstrap confidence interval is around its center (b). If Shape > 1, then (cu - b) > (b - cl). If Shape < 1, then (cu - b) < (b - cl).

[‖]We use the DQAGI subroutine in NSWC Library which allows us to compute the required integrals.

Table 2: Bootstrap 95% Confidence Intervals.

Method		Confidence Interval Lower Bound	Upper Bound	Length	Shape
Percentile-t (Studentized)	x_1	−0.24321	0.434032	0.677245	1.144725
	x_2	−0.33606	0.296353	0.632408	0.893269
	x_3	−0.34916	0.406528	0.755685	1.061344
Percentile-t (Unstudentized)	x_1	−0.25978	0.395289	0.655070	0.971079
	x_2	−0.33165	0.301504	0.633151	0.920841
	x_3	−0.32338	0.383159	0.706539	1.073044
Naïve (Percentile)	x_1	−0.25017	0.404900	0.655070	0
	x_2	−0.30556	0.327596	0.633151	0
	x_3	−0.34828	0.358264	0.706539	0
Normal (Approximation)	x_1	−0.26280	0.407913	0.670709	1
	x_2	−0.33317	0.329116	0.662283	1
	x_3	−0.35483	0.389714	0.744544	1

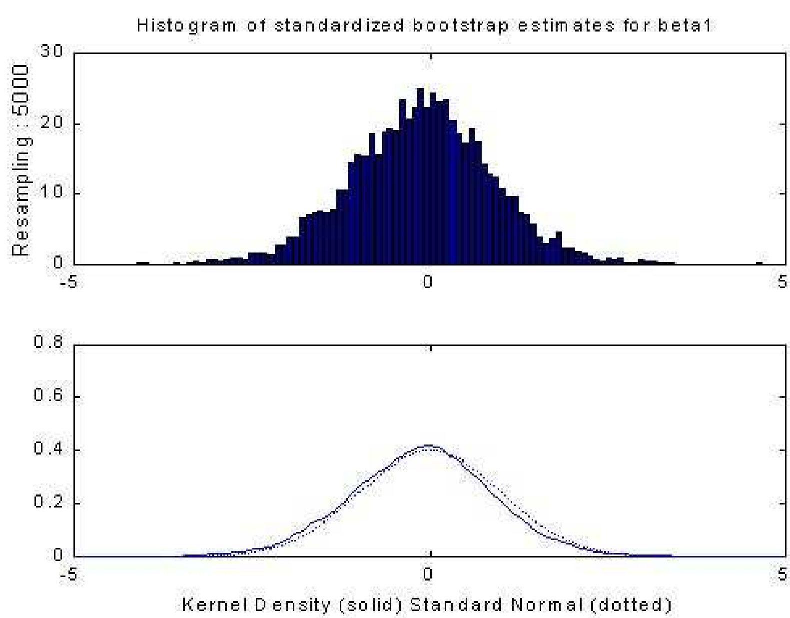

Figure 1: Histogram of Standardized Bootstrap OWS Estimates for β_1.

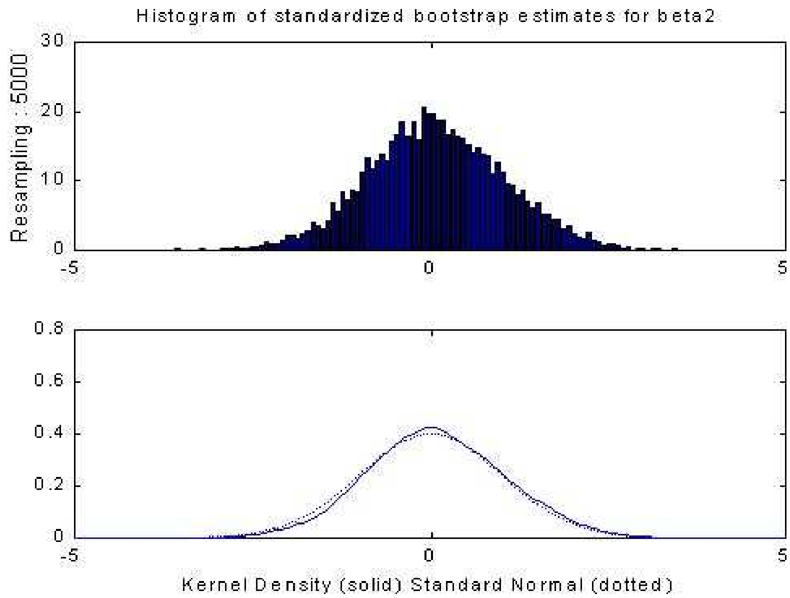

Figure 2: Histogram of Standardized Bootstrap OWS Estimates for β_2.

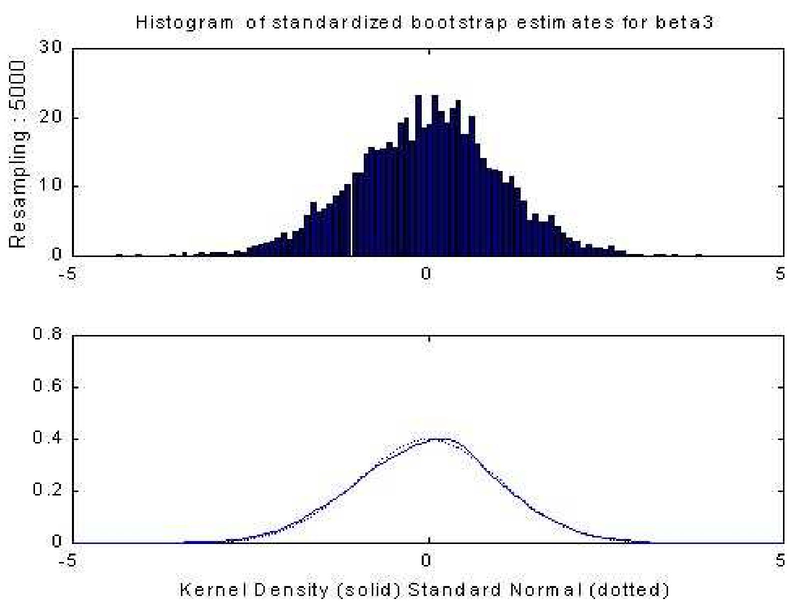

Figure 3: Histogram of Standardized Bootstrap OWS Estimates for β_3.

4. Conclusion

We have used results of [18] to obtain the asymptotic moments of the OWS estimator shrinking to a data-dependent point. This permits us to use a consistent estimator for the asymptotic moments to construct pivotal non-parametric bootstrap statistics. We demonstrate their use by showing how to calculate bootstrap standard errors and confidence intervals for the OWS estimator.

ACKNOWLEDGEMENTS This work was supported by National Research Foundation of Korea - Grant funded by the Korean Government (NRF-2009-327-B00088).

References

[1] Babu, G.J. and Singh, K., "Inference on Means using the Bootstrap," *Annals of Statistics*, 11, 999-1003. 1983.

[2] Beran, R., "Prepivoting to Reduce Level Error of Confidence Sets," *Biometrika*, 74, 457-468. 1987.

[3] Brownstone, D., "Bootstrapping Improved Estimators for Linear Regression Models," *Journal of Econometrics*, 44, 171-187. 1990.

[4] Efron, B. and Tibshirani, R.J., *An Introduction to the Bootstrap*, Chapman & Hall. 1993.

[5] Engle, R. and Granger, C.W.J., "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251-276. 1987.

[6] Granger, C.W.J., "Co-Integrated Variables and Error-Correcting Models," Discussion Paper 83-13, University of California at San Diego. 1983

[7] Hahn, J., "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 11, 105-121. 1995.

[8] Hall, P., "On the Bootstrap and Likelihood-based Confidence Regions," *Biometrika*, 74, 481-493. 1987.

[9] Hall, P., "Theoretical Comparison of Bootstrap Confidence Intervals," *Annals of Statistics*, 16, 927-953. 1988.

[10] Hall, P., *The Bootstrap and Edgeworth Expansion*, Springer-Verlag. 1992.

[11] James, W. and Stein C., "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (vol. 1)*, Berkeley, CA: University of California Press, pp. 361-379. 1960.

[12] Judge, G. and Mittelhammer R.C., "A Semi-parametric Stein-like Basis for Combining Estimators under Quadratic Loss," *Journal of the American Statistical Association*, 99, 479-487. 2004.

[13] Kim, T.H. and White H., "James-Stein-Type Estimators in Large Samples with Application to the Least Absolute Deviations Estimator," *Journal of the American Statistical Association*, 96, 697-705. 2001

[14] Mittelhammer, R.C. and Judge G., "Combining Estimators to Improve Structural Model Estimation under Quadratic Loss," *Journal of Econometrics*, 128, 1-30. 2005.

[15] Saleh, A.K.M.E. and Sen P.K., "On Some Shrinkage Estimators of Multivariate Location," *The Annals of Statistics*, 13, 272-281. 1985.

[16] Sen, P. K. and Saleh A.K.M.E., "On Preliminary Test and Shrinkage M-estimation in Linear Models," *The Annals of Statistics*, 15, 1580-1592. 1987

[17] Stein, C., "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability (vol. 1)*, Berkeley, CA: University of California Press, pp. 197-206. 1955.

[18] Ullah, A., "Finite Sample Econometrics: A Unified Approach," in *Contributions to Econometric Theory and Application: Essays in Honour of A.L. Nagar*, eds. R.A.L. Carter, J. Dutta, and A. Ullah. New York: Springer-Verlag, pp. 242-292. 1990.

[19] Vinod, H.D., "Double Bootstrap for Shrinkage Estimators," *Journal of Econometrics*, 68, 287-302. 1995.

[20] Vinod, H.D. and B. Raj, "Economic Issues in Bell System Divesture: A Bootstrap Application," *Applied Statistics*, 37, 251-261. 1988.

Appendix

Proof. (Theorem 1) Let $f(Z) \equiv MZ$ and $g(Z) \equiv \lambda_2 M_2 Z/Z'M_1 Z$. Let $f_i(Z)$ and $g_i(Z)$ be the ith elements of $f(Z)$ and $g(Z)$ respectively. Then $H_i(Z) = f_i(Z) - g_i(Z)$. First, we note that $E(f_i(Z)) = E(M_i Z) = M_i E(Z) = 0$ where M_i is the ith row of M. By defining $g_{1i}(Z) \equiv M_{2i}Z$ and $g_{2i}(Z) \equiv \lambda_2/Z'M_1 Z$, we have that $g_i(Z) = g_{1i}(Z) g_{2i}(Z)$. By definition, $g_{1i}(d) = M_{2i} d = M_{2i}\left[\mu + \frac{\partial}{\partial \mu}\right]$ where M_{2i} is the ith row of M_2; and it can be shown that $E(g_{2i}(Z)) = \lambda_2 \Gamma(1)^{-1} \int_0^\infty \left|N_{0t}\right|^{-1/2} \exp(-1/2\mu' N_{1t}\mu)\, dt$ where $N_{0t} \equiv I + 2tM_1$, $N_{1t} \equiv 2tM_1 N_{0t}^{-1}$, and Γ is the gamma function. We define $W(\mu) \equiv \Gamma(1)^{-1} \int_0^\infty \left|N_{0t}\right|^{-1/2} \exp(-1/2\mu' N_{1t}\mu)\, dt$. Therefore, using lemmas 1 and 2 of [18], we have that $E(g_i(Z)) = E(g_{1i}(Z) g_{2i}(Z)) = g_{1i}\,(d) E(g_{2i}(Z)) =$

$M_{2i}\left[\mu+\frac{\partial}{\partial\mu}\right]\lambda_2 W(\mu)|_{\mu=0} = 0$. Finally, we note that $\frac{\partial}{\partial\mu}W(\mu)|_{\mu=0} = \Gamma(1)^{-1}\int_0^\infty \left|N_{0t}\right|^{-1/2}$ $\exp(-1/2\mu' N_{1t}\mu)(-N_{1t}\mu)\,dt\,|_{\mu=0} = 0$. Hence, we have the desired result: $E(H_i(Z)) = 0$.

Proof. (Theorem 2) We first note the following:

$$H(Z)H(Z)' = f(Z)f(Z)' - f(Z)g(Z)' - g(Z)f(Z)' + g(Z)g(Z)'.$$

The expected value of the first term is $E(f(Z)f(Z)') = E(MZZ'M') = ME(ZZ')M' = MM'$. Hence a_{ij} is the (i,j)th element of MM'. The second term is given by $E(f(Z)g(Z)') = E(MZ\left[\frac{\lambda_2 M_2 Z}{Z'M_1 Z}\right]') = \lambda_2 E\left[\frac{MZZ'M_2'}{Z'M_1 Z}\right] = \lambda_2 ME\left[\frac{ZZ'}{Z'M_1 Z}\right]M_2'$. Let $\mathscr{E}_{ab}$ be the (a,b) element of $E\left[\frac{ZZ'}{Z'M_1 Z}\right]$. Then E_{ab} can be computed using Ullah's lemmas as follows:

$$\begin{aligned}
\mathscr{E}_{ab} &= E\left[\frac{Z'I_{ab}Z}{Z'M_1 Z}\right] \\
&= \Gamma(1)^{-1}\int_0^\infty \left|N_{0t}\right|^{-1/2}\ (\mathrm{tr}(I_{ab}N_{0t}^{-1}) + \mu' N_{2t}\mu)\ \exp(-.5\ \mu' N_{1t}\mu)\,dt\,|_{\mu=0} \\
&= \Gamma(1)^{-1}\int_0^\infty \left|N_{0t}\right|^{-1/2}\ (\mathrm{tr}(I_{ab}N_{0t}^{-1})\,dt,
\end{aligned}$$

where I_{ab} is defined as before and where $N_{2t} = N_{0t}^{-1} I_{ab} N_{0t}^{-1}$. Hence, the (i,j)th element of $E(f(Z)g(Z)')$ is given by $b_{ij} = \lambda_2 \sum_{a=1}^{2k}\sum_{b=1}^{2k} M_{ia}\mathscr{E}_{ba}M_{2jb}$. By symmetry, the (i,j)th element of $f(Z)g(Z)'$ is equal to the (i,j)th element of $E(g(Z)f(Z)')$. Hence, $c_{ij} = g_{ji}$. The expected value of the last term is $E(g(Z)g(Z)') = E\left(\left[\frac{\lambda_2 M_2 Z}{Z'M_1 Z}\right]\left[\frac{\lambda_2 M_2 Z}{Z'M_1 Z}\right]'\right) = \lambda_2^2 E\left[\frac{ZZ'}{(Z'M_1 Z)^2}\right]$. Let $\mathscr{F}_{ab}$ be the (a,b) element of $E\left[\frac{ZZ'}{(Z'M_1 Z)^2}\right]$. Then $\mathscr{F}_{ab}$ can be computed using Ullah's lemmas as follows:

$$\begin{aligned}
\mathscr{F}_{ab} &= E\left[\frac{Z'I_{ab}Z}{(Z'M_1 Z)^2}\right] \\
&= \Gamma(2)^{-1}\int_0^\infty t\ \left|N_{0t}\right|^{-1/2}\ (\mathrm{tr}(I_{ab}N_{0t}^{-1}) + \mu' N_{2t}\mu)\ \exp(-.5\ \mu' N_{1t}\mu)\,dt|_{\mu=0} \\
&= \Gamma(2)^{-1}\int_0^\infty t\ \left|N_{0t}\right|^{-1/2}\ (tr(I_{ab}N_{0t}^{-1})\,dt.
\end{aligned}$$

Hence, the (i,j)th element is $d_{ij} = \lambda_2^2 \sum_{a=1}^{2k}\sum_{b=1}^{2k} M_{2ia}\mathscr{F}_{ab}M_{2jb}$ which completes the proof.