Title

James-Stein Type Estimators in Large Samples with Application to the Least Absolute Deviations Estimator

Authors

Kim, Tae-Hwan
White, Halbert

Publication Date

1999-02-01

99-04

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

JAMES-STEIN TYPE ESTIMATORS IN LARGE SAMPLES WITH APPLICATION TO THE LEAST ABSOLUTE DEVIATION ESTIMATOR

BY

TAE-HWAN KIM

AND

HALBERT WHITE

DISCUSSION PAPER 99-04
FEBRUARY 1999

James-Stein Type Estimators in Large Samples with Application to The Least Absolute Deviation Estimator

Tae-Hwan Kim[*]
School of Economics
The University of Nottingham
University Park
Nottingham NG7 2RD
United Kingdom
(Phone) +44-0115-951-5466 (Fax) +44-0115-951-4159
leztk@len1.nottingham.ac.uk

Halbert White
Department of Economics, 0508
The University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
(Phone) 619-534-3502 (Fax) 619-534-7040
hwhite@albert.ucsd.edu

[*] Tae-Hwan Kim is Lecturer, School of Economics, The University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom (E-mail: leztk@len1.nottingham.ac.uk); and Halbert White is Professor, Department of Economics, The University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508 (E-mail: hwhite@albert.ucsd.edu). We thank Clive Granger, James Hamilton, Patrick Fitzsimmons and Bruce Lehmann for helpful comments.

ABSTRACT

We explore the extension of James-Stein type estimators in a direction that enables them to preserve their superiority when the sample size goes to infinity. Instead of shrinking a base estimator towards a fixed point, we shrink it towards a data-dependent point, which makes it possible that the "prior" becomes more accurate as the sample size grows. We provide an analytic expression for the asymptotic risk of James-Stein type estimators shrunk towards a data-dependent point and prove that they have smaller asymptotic risk than the base estimator. Shrinking an estimator toward a data-dependent point turns out to be equivalent to combining two random variables using the James-Stein rule. We propose a general combination scheme which includes random combination (the James-Stein combination) and the usual nonrandom combination as special cases. As an example, we apply our method to combine the Least Absolute Deviations estimator and the Least Squares estimator. Our simulation study indicates that the resulting combination estimators have desirable finite sample properties when errors are drawn from symmetric distributions. Finally, using stock return data we present some empirical evidence that the combination estimators have the potential to improve out-of-sample prediction in terms of both mean square error and mean absolute error.

KEY WORDS: Shrinkage; Asymptotic risk; Combination estimator

1. INTRODUCTION

Shrinkage techniques for the linear regression model have been studied extensively since the seminal works by Stein (1955) and James and Stein (1960), who proved that the usual estimator for the mean of multivariate normal distribution is inadmissible and there exists an improved estimator with smaller risk when the dimension of the multivariate normal vector is greater than two.

Even though this discovery was surprising, its usage has been restricted to small sample situations because the advantage of smaller risk tends to disappear as the sample size grows (Saleh and Sen 1985; Sen and Saleh 1987). Schmoyer and Arnold (1989) proposed a James-Stein type estimator that can achieve risk improvement in large samples, at a cost of imposing a very restrictive assumption on the prior information: the mean of the prior distribution is assumed to converge to the parameter of interest at the rate $O(n^{-1/2})$. We follow an approach taken by Green and Strawderman (1991) for fixed n to shrink a given base estimator towards a data-dependent point; here, however, unlike Green and Strawderman, our data dependent point can be either asymptotically biased or correlated with the base estimator, and we consider what happens as $n \to \infty$. The resulting shrinkage estimator in its general form asymptotically dominates both the base estimator and the data-dependent point in terms of quadratic loss. The data-dependent point can be another estimator under some mild restrictions.

To illustrate our results we choose the Least Absolute Deviations estimator as the base estimator and the Least Squares estimator as the data-dependent point. Our estimator in this case is an optimal mix of the information contained in the two estimators. This example can be viewed

as a multivariate extension of what Laplace (1818) did when he combined the sample median and the sample mean by minimizing the asymptotic variance.

2. ASYMPTOTIC RISK IMPROVEMENT

Suppose data are generated according to $y_t = X_t'b^0 + e_t \quad t = 1, 2, ..., n$, where $b^0 \in R^k$ and e_t is assumed to be identically distributed and independent. Let b_n be an estimator for b^0. The quadratic loss is $L(b_n, b^0) \equiv (b_n - b^0)'Q_n(b_n - b^0)$ where Q_n is a symmetric and positive definite matrix. The expectation of the loss function $E(L(b_n, b^0))$ is called the risk and denoted by $R(b_n, b^0)$. Let $\{b_n\}$ be a sequence of estimators of b^0 and let $\{L(b_n, b^0)\}$ be a sequence of loss values. Suppose $L(b_n, b^0)$ converges in distribution to an integrable random variable Ψ. The asymptotic risk of $\{b_n\}$ for $\{L(b_n, b^0)\}$ is then defined to be $AR(\{b_n\}, b^0) \equiv E(\Psi)$. We denote by g_n the mean of the prior distribution towards which a base estimator is shrunk. Classical James-Stein type estimators are obtained by setting g_n to a fixed number. In this paper, we allow g_n to be data-dependent. We now provide formal conditions for our analysis.

Assumption 1. $n^{-1}Q_n \xrightarrow{p} Q$ where Q is a nonstochastic symmetric and positive definite matrix.

Assumption 2. $\begin{bmatrix} n^{1/2}(b_n - b^0) \\ n^{1/2}(g_n - b^0) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0_{k\times 1} \\ q_{k\times 1} \end{bmatrix}, \begin{bmatrix} A_{k\times k} & \Delta_{k\times k} \\ \Delta'_{k\times k} & B_{k\times k} \end{bmatrix} \right)$ where q, A, B and Δ are nonstochastic and finite, and A and B are symmetric and positive definite matrices.

Assumption 3. $P[b_n = g_n] = 0$ for all n and $P[U_1 = U_2] = 0$.

Applying the approach of Green and Strawderman (1991), the natural James-Stein type shrinkage is

$$d^{JS}_{c_1}(b_n, g_n) \equiv \{1 - c_1 / \| b_n - g_n \|^2_n\}(b_n - g_n) + g_n \tag{1}$$

where c_1 is a constant and $\| b_n - g_n \|^2_n \equiv (b_n - g_n)' Q_n (b_n - g_n)$. When the sample size is fixed and the prior is independent of the base estimator, this is identical to the estimator in Green and Strawderman (1991). We are mainly interested in what will happen as the sample size grows. The following lemma describes the limiting distribution and the asymptotic risk.

Lemma 1. Suppose that Assumptions 1, 2 and 3 hold. Then

(i) $n^{1/2}(d^{JS}_{c_1}(b_n, g_n) - b^0) \xrightarrow{d} d^{JS}_{c_1}(U_1, U_2)$.

(ii) $AR(\{ d^{JS}_{c_1}(b_n, g_n)\}, b^0) = R(d^{JS}_{c_1}(U_1, U_2), 0)$, provided the expectation exists.

In writing the limiting random variable $d_{c_1}^{JS}(U_1,U_2)$ we abuse notation somewhat to have $d_{c_1}^{JS}(U_1,U_2) \equiv \{1-c_1/(U_1-U_2)'Q(U_1-U_2)\}(U_1-U_2)+U_2$. We use the notation similarly in writing $d_{c_2}^{NR}(U_1,U_2)$ and $d_l^{OW}(U_1,U_2)$ in what follows. The limiting random variable is not the usual normal but a nonlinear function of (correlated) normal random variables. The asymptotic risk is the risk of the limiting random variable with the parameter of interest being zero. We now provide conditions ensuring that the shrinkage estimator (1) dominates the base estimator and specify the optimal value for c_1.

Theorem 1. Suppose that Assumptions 1, 2 and 3 hold. Then

(i) $AR(\{d_{c_1}^{JS}(b_n,g_n)\},b^0)$ is strictly convex in c_1.

(ii) Let $c_1^* \in$ argmin $AR(\{d_{c_1}^{JS}(b_n,g_n)\},b^0)$. Then $c_1^* = n/w$ where

$n \equiv E[U_1'Q(U_1-U_2)/(U_1-U_2)'Q(U_1-U_2)]$ and $w \equiv E[1/(U_1-U_2)'Q(U_1-U_2)]$.

(iii) $AR(\{d_{c_1^*}^{JS}(b_n,g_n)\},b^0) = -n^2/w + k$ where $k \equiv E[U_1'QU_1]$.

(iv) $AR(\{d_{c_1^*}^{JS}(b_n,g_n)\},b^0) \leq AR(\{b_n\},b^0)$ where the equality holds only when $n = 0$.

(v) $AR(\{d_{c_1}^{JS}(b_n,g_n)\},b^0) \leq AR(\{b_n\},b^0)$ if $c_1 \in [\min\{0,2n/w\},\max\{0,2n/w\}]$

where the equality holds only when $n = 0$.

We call the shrinkage estimator (1) with the optimal c_1^* the James-Stein Mix (JSM). As long as $\nu \neq 0$, which we call the Relative Efficiency Condition (REC), we can achieve an asymptotic risk improvement with respect to the base estimator. The relative efficiency condition does not

allow us to choose an asymptotically efficient estimator as the base estimator b_n unless either we select a super-efficient estimator as our "prior" g_n or the "prior" has an asymptotic bias. Suppose we choose an asymptotically efficient estimator as the base estimator. Then any data-dependent point which is not super-efficient and has no bias can be expressed as the sum of the asymptotically efficient estimator and a random noise, which is asymptotically uncorrelated with the asymptotically efficient estimator and converges to zero as the sample size goes to infinity. It follows that $Cov(U_1, U_1 - U_2) = 0$ which, because of normality, is equivalent to U_1 being independent of $U_1 - U_2$. This implies that $n = 0$.

While the sign of w is positive by Assumption 3, the sign of n is not determined. Therefore the sign of c_1^* depends on the sign of n. It is interesting to note that the ratio n/w is equal to $(k-2)s^2$ when U_1 and U_2 are independent; i.e. $\Delta = 0$. In this case, the optimal combination weight is exactly the James-Stein optimal weight. The deviation of the ratio n/w from $(k-2)s^2$ depends on the degree of the asymptotic correlation between the base estimator and the data-dependent point. To illustrate the James-Stein Mix, we give some simple examples. For comparison, note that for all of our examples $AR(\{b_n\}, b^0) = ks^2$.

Example 1. Suppose that $A = s^2 I$, $B = t^2 I$, $\Delta = 0$, $Q = I$, $q \neq 0$ and $k \geq 3$. Then

(i) $c_1^* = (k-2)s^2$.

(ii) $AR(\{d^{JS}_{c_1^*}(b_n, g_n)\}, b^0) = ks^2 - (k-2)^2 s^4/(s^2+t^2)E[1/(k-2+2P)]$

where P has a Poisson distribution with mean $q'q/2(s^2+t^2)$.

Example 2. Suppose that $A = s^2 I$, $B = t^2 I$, $\Delta = 0$, $Q = I$, $q = 0$ and $k \geq 3$. Then

(i) $c_1^* = (k-2)s^2$.

(ii) $AR(\{d^{JS}_{c_1^*}(b_n, g_n)\}, b^0) = ks^2 - (k-2)s^4/(s^2 + t^2)$.

While the James-Stein Mix is a combination of two random variables using a random weight, conventional combination estimators use a nonrandom weight. This nonrandom combination has been studied mainly for independent estimators. See Cohen (1976) and Green and Strawderman (1991). Laplace (1818) combined the sample median and the sample mean by minimizing asymptotic variance. Here we consider combining multi-dimensional correlated estimators by minimizing asymptotic risk. The usual combination gives

$$d^{NR}_{c_2}(b_n, g_n) \equiv \{1 - c_2\}(b_n - g_n) + g_n \tag{2}$$

where c_2 is a constant. Using the same arguments as in Lemma 1 and Theorem 1, we obtain the following results.

Lemma 2. Suppose that Assumptions 1, 2 and 3 hold. Then

(i) $n^{1/2}(d^{NR}_{c_2}(b_n, g_n) - b^0) \xrightarrow{d} d^{NR}_{c_2}(U_1, U_2)$.

(ii) $AR(\{d^{NR}_{c_2}(b_n, g_n)\}, b^0) = R(d^{NR}_{c_2}(U_1, U_2), 0)$, provided the expectation exists.

Theorem 2. Suppose that Assumptions 1, 2 and 3 hold. Then

(i) $AR(\{d^{NR}_{c_2}(b_n, g_n)\}, b^0)$ is strictly convex in c_2.

(ii) Let $c_2^* \in \text{argmin } AR(\{d^{NR}_{c_2}(b_n, g_n)\}, b^0)$. Then $c_2^* = r/a$ where

$a \equiv E[(U_1 - U_2)'Q(U_1 - U_2)]$ and $r \equiv E[U_1'Q(U_1 - U_2)]$.

(iii) $AR(\{d^{NR}_{c_2^*}(b_n, g_n)\}, b^0) = -r^2/a + k$.

(iv) $AR(\{d^{NR}_{c_2^*}(b_n, g_n)\}, b^0) \leq AR(\{b_n\}, b^0)$ where the equality holds only when $r = 0$.

(v) $AR(\{d^{NR}_{c_2}(b_n, g_n)\}, b^0) \leq AR(\{b_n\}, b^0)$ if $c_2 \in [\min\{0, 2r/a\}, \max\{0, 2r/a\}]$

where the equality holds only when $r = 0$.

(vi) $AR(\{d^{NR}_{c_2^*}(b_n, g_n)\}, b^0) \leq AR(\{g_n\}, b^0)$ where the equality holds only when

$\gamma = 0$ with $\gamma = E[(U_1 - U_2)'QU_2]$.

We call the combination in (2) with the optimal c_2^* the Nonrandom Mix (NRM). If both $r \neq 0$ and $\gamma \neq 0$, the analog of the relative efficiency condition in the present context, then the asymptotic risk of the Nonrandom Mix is strictly smaller than that of both the base estimator and the data-dependent point. We give some examples for illustration. As before, $AR(\{b_n\}, b^0) = ks^2$.

Example 3. Suppose that $A = s^2 I$, $B = t^2 I$, $\Delta = 0$ and $Q = I$. Then

(i) $c_2^* = ks^2/\{q'q + k(s^2 + t^2)\}$.

(ii) $AR(\{d^{NR}_{c_2^*}(b_n, g_n)\}, b^0) = ks^2 - k^2 s^4/\{q'q + k(s^2 + t^2)\}$.

Example 4. Suppose that $A = s^2 I$, $B = t^2 I$, $\Delta = 0$, $Q = I$ and $q = 0$. Then

(i) $c_2^* = s^2/(s^2 + t^2)$.

(ii) $AR(\{d^{NR}_{c_2^*}(b_n, g_n)\}, b^0) = ks^2 - ks^4/(s^2+t^2)$.

Comparing Examples 2 and 4, we see that here the James-Stein Mix is dominated by the Nonrandom Mix when there is no asymptotic bias in the prior. The following example provides a condition under which the opposite is true in the presence of a bias.

Example 5. Suppose that $A = s^2 I$, $B = t^2 I$, $\Delta = 0$, $Q = I$, $q \neq 0$ and $k \geq 3$. Then

$$AR(\{d^{JS}_{c_1^*}(b_n, g_n)\}, b^0) < AR(\{d^{NR}_{c_2^*}(b_n, g_n)\}, b^0)$$

$$\Leftrightarrow \; E\left[\frac{1}{k-2+2P}\right] > \frac{k^2(s^2+t^2)}{(k-2)^2[q'q + k(s^2+t^2)]}.$$

So far, we have investigated the James-Stein Mix and the Nonrandom Mix separately. A general combination scheme is

$$d^{OW}_l(b_n, g_n) \equiv \{1 - l_1 - l_2 / \| b_n - g_n \|^2_n\}(b_n - g_n) + g_n \tag{3}$$

where $l = (l_1, l_2)'$ is a constant vector. The limiting distribution and the asymptotic risk are as follows.

Lemma 3. Suppose that Assumptions 1, 2 and 3 hold. Then

(i) $n^{1/2}(d^{OW}_l(b_n, g_n) - b^0) \xrightarrow{d} d^{OW}_l(U_1, U_2)$.

(ii) $AR(\{d^{OW}_l(b_n, g_n)\}, b^0) = R(d^{OW}_l(U_1, U_2), 0)$, provided the expectation exists.

In order to study the domination and optimality properties of this general mix, we add the following assumption.

Assumption 4. $(U_1 - U_2)'Q(U_1 - U_2)$ is nondegenerate.

Theorem 3. Suppose that Assumptions 1, 2, 3, and 4 hold. Then

(i) $AR(\{d_l^{OW}(b_n, g_n)\}, b^0)$ is strictly convex in l.

(ii) Let $l^* \in \operatorname{argmin} AR(\{d_l^{OW}(b_n, g_n)\}, b^0)$. Then

$l_1^* = (aw - 1)^{-1}(rw - n)$ and $l_2^* = (aw - 1)^{-1}(an - r)$.

(iii) $AR(\{d_{l^*}^{OW}(b_n, g_n)\}, b^0)$

$= (aw - 1)^{-2}\{-ar^2w^2 - (2arn - a^2n^2 + r^2)w + (an^2 - 2rn)\} + k.$

(iv) $AR(\{d_{l^*}^{OW}(b_n, g_n)\}, b^0) \leq AR(\{b_n\}, b^0)$ where the equality holds only when

$r = 0$ and $n = 0$.

We call the general mix in (3) with the optimal weight l^* the Optimal Weighting Mix (OWM) and $(a, r, n, w)'$ the Combination Control Parameters (CCP).

We now discuss some properties of the optimal weight and combination control parameters. First, note that the denominator in the optimal weight $aw - 1$ is positive, which can be shown using Assumption 4 and the strong form of Jensen's Inequality. This observation leads us to the necessary and sufficient condition for determining the sign of the optimal weight:

$$Sign\{\ l_1^*\} = Sign\{Cov[U_1'Q(U_1 - U_2), 1/\|U_1 - U_2\|]\} \tag{4}$$

$$Sign\{\ l_2^*\} = Sign\{Cov[\|U_1 - U_2\|, U_1'Q(U_1 - U_2)/\|U_1 - U_2\|]\}, \tag{5}$$

where $\|U_1 - U_2\|^2 \equiv (U_1 - U_2)Q(U_1 - U_2)$. We again see that the relative efficiency condition governs the optimal weights. We now prove that the Optimal Weighting Mix asymptotically dominates both the James-Stein Mix and the Nonrandom Mix.

Corollary 1. Suppose that Assumptions 1, 2, 3, and 4 hold. Then

(i) $AR(\{\ d_{l^*}^{OW}(b_n, g_n)\}, b^0) \leq AR(\{\ d_{c_1^*}^{JS}(b_n, g_n)\}, b^0)$ where the strict inequality holds if l_1^* is not equal to zero.

(ii) $AR(\{\ d_{l^*}^{OW}(b_n, g_n)\}, b^0) \leq AR(\{\ d_{c_2^*}^{NR}(b_n, g_n)\}, b^0)$ where the strict inequality holds if l_2^* is not equal to zero.

The optimal weights l_1^*, l_2^* can be viewed as the contribution of the nonrandom mix and the random mix respectively. In the special case where $A = s^2 I$, $B = t^2 I$, $\Delta = 0$, $Q = I$, $k \geq 3$ and $q = 0$, we can show that $l_1^* = s^2/(s^2 + t^2)$ and $l_2^* = 0$. The random mix makes no contribution. On the other hand, if $A = s^2 I$, $B = t^2 I$, $\Delta = 0$, $Q = I$, $q \neq 0$, and $k \geq 3$, then $l_2^* \neq 0$ if and only if $(k-2)E[1/(k-2+2P)] \neq k(s^2 + t^2)/[q'q + k(s^2 + t^2)]$ where P has a Poisson distribution with mean $q'q/2(s^2 + t^2)$.

3. ESTIMATION

Even though the Optimal Weighting Mix has some nice properties, it contains the four unknown parameters $(a, r, n, w)'$. This is why we use the term "mix" instead of "estimator". We now discuss how to estimate the combination control parameters consistently. For simplicity, we consider only the case where there is no asymptotic bias. Before we proceed, we define some random variables used in our results. Let $U \equiv (U_1, U_1 - U_2)'$. Then $U \sim N(0_{2k\times 1}, \Sigma_{2k\times 2k})$ where $\Sigma \equiv \begin{bmatrix} A & A-\Delta \\ A-\Delta' & A-\Delta-\Delta'+B \end{bmatrix}$. There exists a matrix P such that $\Sigma = PP'$. Let $Z \equiv P^{-1}U$. Then $Z \sim N(0_{2k\times 1}, I_{2k\times 2k})$. Define $M_1 \equiv P'N_1P$ where $N_1 \equiv 1/2\begin{bmatrix} 0_{k\times k} & Q \\ Q & 0_{k\times k} \end{bmatrix}$ and $M \equiv P'NP$ where $N \equiv \begin{bmatrix} 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & Q \end{bmatrix}$. It can be shown with some simple algebra that $U_1'Q(U_1 - U_2) = Z'M_1Z$ and $(U_1 - U_2)'Q(U_1 - U_2) = Z'MZ$.

This transformation permits us to use Ullah (1990)'s results on moments of the ratio of quadratic forms of normal random variables. Suppose that $\hat{A}_n, \hat{B}_n, \hat{\Delta}_n, \hat{Q}_n$ are consistent estimators for A, B, Δ, Q respectively. We consider the following estimators for the combination control parameters.

$$\hat{a}_n \equiv tr[(\hat{A}_n - \hat{\Delta}_n - \hat{\Delta}_n' + \hat{B}_n)\hat{Q}_n] \tag{6}$$

$$\hat{r}_n \equiv tr[(\hat{A}_n - \hat{\Delta}_n')\hat{Q}_n] \tag{7}$$

$$\hat{w}_n \equiv \int_0^\infty |\hat{N}_{1tn}|^{-1/2}\, dt \tag{8}$$

$$\hat{n}_n \equiv \int_0^\infty |\hat{N}_{0tn}|^{-1/2}\, tr\left[\hat{M}_{1n}\hat{N}_{0tn}^{-1}\right] dt \tag{9}$$

where $\hat{N}_{1tn} \equiv I + 2t\hat{\Sigma}_{22n}\hat{Q}_n$, $\hat{N}_{0tn} \equiv I + 2t\hat{M}_n$ and $\hat{\Sigma}_{22n} \equiv \hat{A}_n - \hat{\Delta}_n - \hat{\Delta}_n' + \hat{B}_n$. Before we show the consistency results, we first establish that the control parameters are finite.

Assumption 5. $\Sigma_{22} \equiv A - \Delta - \Delta' + B$ is positive definite.

Lemma 4. Suppose that Assumptions 2 and 5 hold. Then

(i) $|a| < \infty$.

(ii) $|r| < \infty$.

(iii) $|w| < \infty$ if $k > 2$.

(iv) $|n| < \infty$ if $k \neq 2$ and $k \neq 4$.

We now prove that the estimators defined in (6) – (9) converge to the combination control parameters in probability.

Theorem 4. Suppose that Assumptions 1, 2, 3, 4, and 5 hold. Suppose that $k > 2$ and $k \neq 4$. Then

(i) $\hat{a}_n \xrightarrow{p} a$.

(ii) $\hat{b}_n \xrightarrow{p} b$.

(iii) $\hat{w}_n \xrightarrow{p} w$.

(iv) $\hat{n}_n \xrightarrow{p} n$.

Once we obtain consistent estimators for the control parameters, a natural way to approximate the Optimal Weighting Mix is given by

$$d^{OW}_{\hat{l}_n}(b_n, g_n) \equiv \{1 - \hat{l}_{1n} - \hat{l}_{2n} / \| b_n - g_n \|^2_n\}(b_n - g_n) + g_n \tag{10}$$

where $\hat{l}_{1n} \equiv (\hat{a}_n \hat{w}_n - 1)^{-1}(\hat{r}_n \hat{w}_n - \hat{n}_n)$ and $\hat{l}_{2n} \equiv (\hat{a}_n \hat{w}_n - 1)^{-1}(\hat{a}_n \hat{n}_n - \hat{r}_n)$. We call the estimator in (10) the Optimal Weighting Scheme (OWS) Estimator. An interesting question is whether we can still achieve optimality (minimum asymptotic risk) for this estimator. The following corollary answers this question.

Corollary 2. Suppose that Assumptions 1, 2, 3, 4, and 5 hold. Suppose that $k > 2$ and $k \neq 4$. Then

(i) $\hat{l}_{1n} \xrightarrow{p} l^*_1$ and $\hat{l}_{2n} \xrightarrow{p} l^*_2$.

(ii) $n^{1/2}(d^{OW}_{\hat{l}_n}(b_n, g_n) - b^0) \xrightarrow{d} d^{OW}_{l^*}(U_1, U_2)$.

(iii) $AR(\{d^{OW}_{\hat{l}_n}(b_n, g_n)\}, b^0) = AR(\{d^{OW}_{l^*}(b_n, g_n)\}, b^0)$.

The Optimal Weighting Scheme estimator has the same limiting distribution as the Optimal Weighting Mix, and therefore achieves the minimum bound.

Remark 1. The same analysis as for the Optimal Weighting Scheme estimator applies to the James-Stein Mix and the Nonrandom Mix. We call the resulting estimators the James-Stein Combination (JSC) Estimator and the Nonrandom Combination (NRC) Estimator respectively.

4. APPLICATION

In this section we discuss heuristically how our method can be utilized in combining two possibly correlated estimators. We choose the Least Absolute Deviations (LAD) estimator as the base estimator and the Ordinary Least Squares (OLS) estimator as the data-dependent point. The resulting estimator is an optimal combination of the two estimators. There has been some interesting research on this issue. As previously mentioned, Laplace (1818) combined the sample median and the sample mean by minimizing the asymptotic variance. Taylor (1974) suggested a two step procedure; first apply the LAD estimator to identify outliers to be trimmed and then apply the OLS estimator. Arthanari and Dodge (1981) combined the objective functions of the LAD estimator and the LS estimator.

As shown in Bates and White (1993), both the LAD estimator and the OLS estimator are members of a RCASOI (Regular Consistent Asymptotically Second Order Indexed) class under some regularity conditions. For any member b_n in a RCASOI class, there is a score representation (s_n^0) and Hessian representation (H_n^0) such that $b_n - b^0 = H_n^{0^{-1}} s_n^0 + o_p(n^{-1/2})$.

Accordingly, we have the following representation for the two estimators; $s_n^{LS} = 2\sum_{t=1}^{n} X_t e_t$, $H_n^{LS} = 2\sum_{t=1}^{n} E(X_t X_t')$, $s_n^{LAD} = -2\sum_{t=1}^{n} X_t(1_{[e_t \le 0]} - 1/2)$ and $H_n^{LAD} = 2f(0)\sum_{t=1}^{n} E(X_t X_t')$ where $f(0)$ is the value of the density of e_t at zero. Given these representations it is not difficult to show the joint asymptotic normality which, follows from

$$n^{1/2}\begin{bmatrix}(b_n^{LAD} - b^0)\\ (b_n^{LS} - b^0)\end{bmatrix} = \begin{bmatrix} n^{-1}H_n^{LAD} & 0_{k\times k} \\ 0_{k\times k} & n^{-1}H_n^{LS}\end{bmatrix}^{-1} n^{-1/2}\begin{pmatrix} s_n^{LAD} \\ s_n^{LS}\end{pmatrix} + o_p(1). \tag{11}$$

The identical and independent distribution assumption is sufficient (though not necessary) to deliver the desired result. The asymptotic covariance between the two estimators is given by $\Delta = \{4f(0)\}^{-1} E(X_t X_t')^{-1} E(S_{1t} S_{2t}') E(X_t X_t')^{-1}$ where $S_{1t} \equiv -2X_t(1[e_t \le 0] - 1/2)$ and $S_{2t} \equiv 2X_t e_t$. We estimate the asymptotic covariance by the plug-in principle: $\hat{\Delta}_n \equiv \{4\hat{f}_n(0)\}^{-1}[n^{-1}\sum_{t=1}^{n} X_t X_t']^{-1} n^{-1}\sum_{t=1}^{n} \hat{S}_{1t}\hat{S}_{2t}' [n^{-1}\sum_{t=1}^{n} X_t X_t']^{-1}$ where $\hat{f}_n(0)$ is an estimate for the density at zero and $\hat{S}_{1t}, \hat{S}_{2t}$ are estimates for S_{1t}, S_{2t} using $\hat{e}_t \equiv y_t - X_t' b_n$. We study the behavior of various combinations of LAD and OLS in the following sections.

5. SIMULATION

We conduct Monte Carlo experiments designed to investigate the finite sample properties of James-Stein type estimators. For purposes of comparison, we also include stable estimators (the

Ridge estimator, the Garrotte estimator and the Non-Negative Garrotte estimator). Stable estimators have been shown to have good prediction performance (Breiman 1995, 1996). The definitions for stable estimators considered here are given in Table 1. In the simulation we combine the LAD and the OLS estimators described in Section 4 to obtain examples of the NRC, JSC and OWS estimators.

The basic model for the simulation is $y_t = X_t' b^0 + e_t$ where $t = 1, 2, ..., n$, $b^0 \in R^k$, $n = 500$ and $k = 5$. We set $b^0 = (1,1,1,1,1)'$. The number of replications is 1,000. We obtain the LAD estimator using the efficient L_1 algorithm developed by Barrodale and Roberts (1974). The simulation was carried out on a 266MHz PC using MATLAB. The random number generator used in the simulation is that from the MATLAB Statistics Toolbox.

Table 1. Definition of Estimators

Estimator	Definition
Ridge (b^R)	$b^R \in \text{argmin} \parallel y - Xb \parallel^2$ s.t. $b'b < s$
Garrotte (b^G)	$b^G \in \text{argmin} \parallel y - Z g \parallel^2$ s.t. $Z_{ij} = X_{ij} b_j^{LS}$ and $g'g < s$
Non-Negative Garrotte (b^N)	$b^N \in \text{argmin} \parallel y - Z g \parallel^2$ s.t. $Z_{ij} = X_{ij} b_j^{LS}$ and $g'i < s,\ g \geq 0$

NOTE: Values for s are determined by k-fold cross-validation.

We choose four symmetric distributions and two non-symmetric distribution for e_t. Symmetric distributions are the Uniform distribution within [-4,4], the standard normal distribution, the student *t*-distribution with 3 degrees of freedom, and the Cauchy distribution with interquartile range 1. These represent moderate, heavy and very heavy tailed distributions. For

the non-symmetric distributions, we choose the shifted Chi-square distribution centered at zero with 12 degrees of freedom and the shifted Rayleigh distribution centered at zero with parameter 4. Independent variables are generated using the joint normal distribution $N(0,\Sigma)$ where the covariances are set to 0.5 and the variances are one. The first entry of X_t is one. We estimate the required density $\hat{f}_n(0)$ using a kernel method with Gaussian kernel. For each replication we compute the quadratic loss value for each estimator. We approximate the risk by averaging the loss values over all replications. The results are collected in Table 2.

Table 2. Finite Sample Risk Comparison over Different Error Distributions ($n=500$)

	Uniform	Normal	Student-*t*	Cauchy	c^2	Rayleigh
OLS	26.306	5.006	14.789	463971952.194	121.770	35.089
LAD	77.029	7.791	9.345	12.789	398.972	99.337
NRC	20.595	5.035	8.908	12.531	126.610	35.982
JSC	72.765	6.306	9.104	12.707	396.751	96.982
OWS	20.617	5.035	8.910	12.525	126.925	35.830
RIDGE	25.504	4.974	14.471	433311022.768	115.666	33.612
GAR	27.241	5.103	15.216	463971921.290	126.777	36.226
NNGAR	26.311	5.006	14.789	463970894.151	123.116	35.090

It is known that the performance of the median is worse than the sample mean when the error is distributed uniformly. As expected, the risk of the LAD estimator (77.029) is greater than the risk of the OLS estimator (26.306) for the Uniform distribution with [-4,4]. All combination methods give negative weight to the LAD estimator. As a result, both the OWS estimator and the NRC estimator dominate the OLS estimator. When the regression error is normal, the OLS

estimator is asymptotically efficient. The risk of the OLS estimator (5.006) is smaller than that of the LAD estimator (7.791). All combination methods again give negative weights to the LAD estimator and have smaller risk than the LAD estimator, but greater risk than the OLS estimator. However, the deterioration of the OWS estimator and the NRC estimator relative to the OLS estimator is not large (-0.57 %). The Student-*t* distribution with 3 degrees of freedom has a relatively fat tail. As expected, the risk of the LAD estimator (9.345) is smaller than the risk of the OLS estimator (14.789). All combination estimators have smaller risk than both the LAD estimator and the OLS estimator. The improvement of the combination estimators over the LAD estimator and the OLS estimator is about 2 - 5 % and 38 - 40 % respectively. The Cauchy distribution represents a very heavy tailed distribution. The risk performance of the OLS estimator is worse than that of the LAD estimator (463971952 and 12.789 respectively). Nevertheless, combining the LAD estimator with the OLS estimator makes an improvement over the LAD estimator. The improvements over the LAD estimator and the OLS estimator are about 0.6 - 2 % and 100 % respectively.

The LAD estimator is out-performed by the OLS estimator in terms of risk (398.972 and 121.770) when the regression error is Chi-square distribution with 12 degrees of freedom. However, the combination methods give positive weight to the LAD estimator. For the OWS estimator and the NRC estimator, the weight is very small (about 0.080 - 0.096). On the other hand, the JSC estimator gives a large positive weight to the LAD estimator (0.99), which clearly shows the inferiority of the JSC estimator when the regression error is not symmetric. The failure can be explained by the bias in the constant coefficient, which makes the distance between two estimators very large (384.778). This in turn makes the JS weight too large. All combination estimators are better than the LAD estimator, but worse than the OLS estimator. When the error

has the Rayleigh distribution with parameter 4, the result is basically the same as for the Chi-square distribution. However, the skewness is smaller than for the Chi-square distribution, and as a result, the bias in the constant term is much smaller. The OWS estimator and the NRC estimator now give a small negative weight to the LAD estimator.

The performance of the stable estimators are shown in the same table. The Ridge estimator gives smaller risk than the OLS estimator over all error distributions considered in the simulation including the normal distribution. This is a well-known standard result based on the trade-off between variance and bias. On the other hand, the other stable estimators (the Garrotte estimator and the Non-Negative Garrotte estimator) are not better than the OLS estimator, which might at first seem surprising. However, Breiman (1995, 1996) showed that the Garrotte estimator and the Non-Negative Garrotte estimator give smaller prediction mean squared error than the OLS estimator when irrelevant variables appear in the model. Here none of our variables are irrelevant. Despite this, the additional risk associated with the Garrotte and Non-Negative Garrotte is small.

6. EMPIRICAL STUDY: OUT-OF-SAMPLE PREDICTION

In this section we investigate the out-of-sample predictive ability of the combination estimators using actual data. Let y be a $T \times 1$ vector of out-of-sample actual values and let e be a $T \times 1$ vector of the prediction errors where T is the number of out-of-sample observations. In order to evaluate forecasting performance, we use the following forecasting error measurements: prediction mean squared error $PMSE(e) \equiv e'e/T$ and prediction mean absolute error $PMAE(e) \equiv T^{-1}\sum_{t=1}^{T}|e_t|$. We also use R^2 type prediction measures: $R^2 \equiv 1 - PMSE(e)/S^2(y)$

and $R_A^2 \equiv 1 - PMAE(e)/MAE(y)$ where $S^2(y)$ is the sample variance of y and $MAE(y)$ is the mean absolute error of y. The data set contains daily stock market returns for ADC TeleCom Co. and HomeStake Co., stocks that have been randomly chosen from the DATASTREAM database. The sample period covers January 1, 1990 through March 31, 1996 which gives us 1630 observations. We model daily excess returns, computed by subtracting the 3-month US T-bill rate from daily returns. Table 3 provides summary statistics.

Table 3. Summary Statistics for Daily Excess Stock Returns (in percent)

	Mean	Median	Max	Min	Std. Dev.	Skew.	Excess Kurtosis
ADC TelCom	0.15	-0.01	11.92	-22.13	2.94	-0.17	3.93
HomeStake	0.01	-0.02	11.25	-12.45	2.52	0.08	1.96

Our forecasting model for excess returns is

$$r_t = a + \sum_{i=1}^{k_1} b_i r_{t-i} + \sum_{i=1}^{k_2} g_i r_{mt-i} + e_t \qquad (12)$$

where r_{mt} is the daily excess returns on the S&P500 index and $k_1 = k_2 = 1$. The simple efficient market hypothesis requires that $a = b = g = 0$, so that the best predictor is zero. We call this the Random Walk predictor and we include this in our comparison study. We use a fixed rolling window method to estimate the coefficients, and set the size of estimation window to be 520, which is about a two year sample period. We repeat the entire exercise identically for each of the 8 estimators and for each of the target variables.

The outcomes are summarized Figures 1 and 2. We can represent an estimator as a point in PMAE-PMSE space.

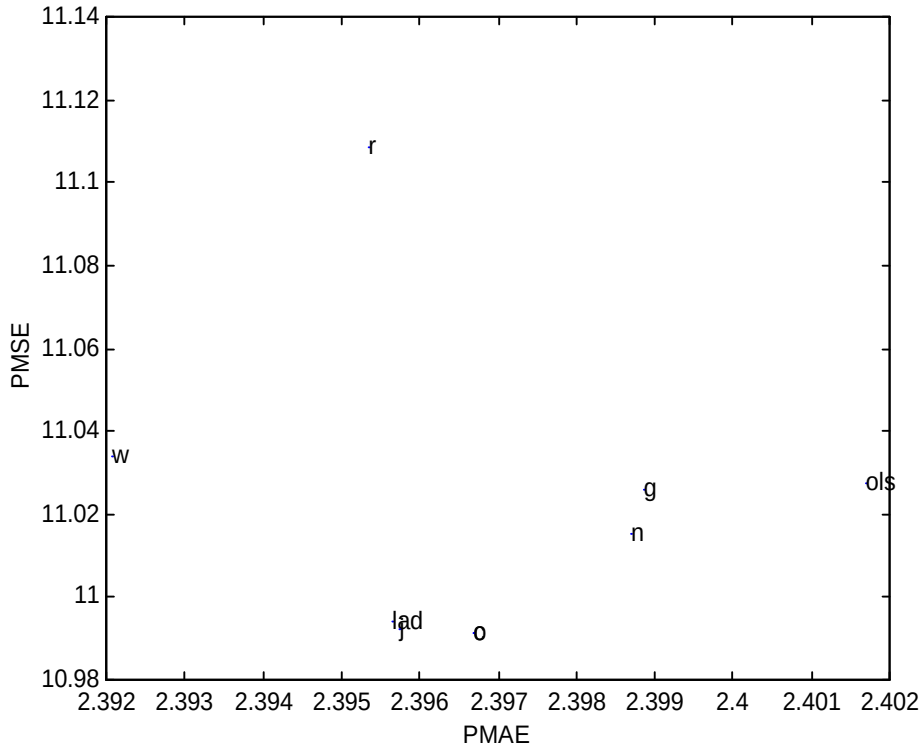

Figure 1. Out-of-Sample Prediction Performance for ADC TeleCom.

In this diagram, we prefer estimators located closer to the origin because the PMAE and the PMSE can be treated as "bad" commodities. We represent combination and stable estimators by their first initial in PMAE-PMSE space except the NRC estimator denoted by "c" and the Random Walk predictor denoted by "w". For example, "r" stands for the Ridge estimator, "g" for the Garrotte estimator, and so on. In the case of ADC TeleCom (Figure 1), all combination estimators outperform both the LAD and the OLS estimators in terms of PMSE, but the improvement over the LAD estimator is very small. The performance of the NRC and OWS are almost identical. They also achieve better performance than the stable estimators. Interestingly

all our estimators beat the Random Walk predictor in terms of PMSE, but the Random Walk predictor beats all estimators in terms of PMAE.

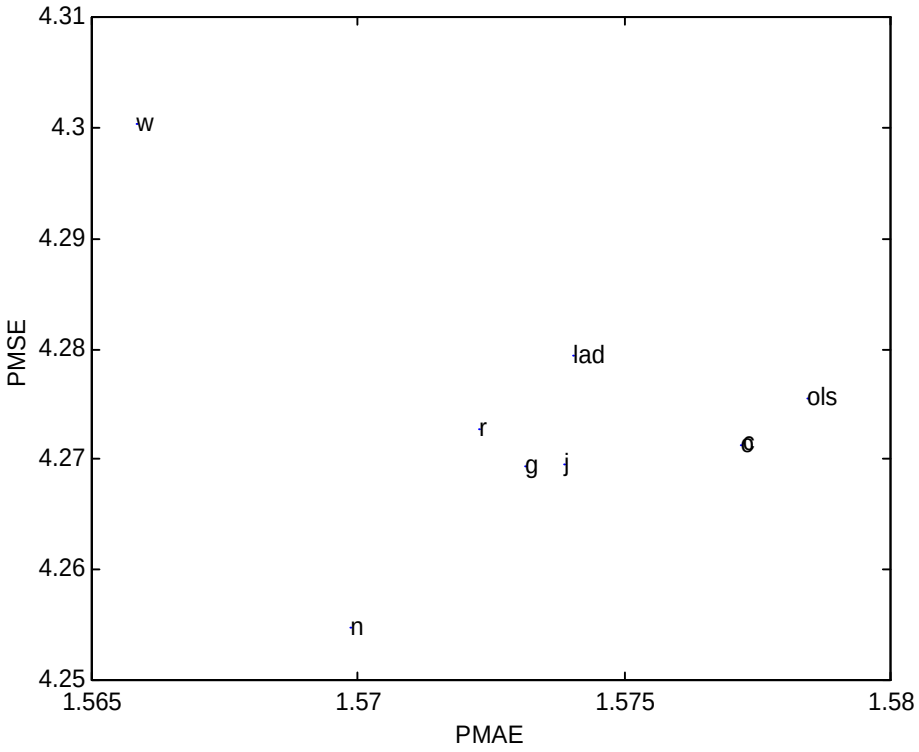

Figure 2. Out-of-Sample Prediction Performance for HomeStake.

The out-of-sample prediction result for HomeStake stock is given in Figure 2. The Non-negative Garrotte, Garrotte, Ridge and JSC estimators outperform all other estimators in terms of both PMSE and PMAE. The behavior of Random Walk predictor is similar to the results seen for ADC TeleCom stock. The combinations estimators generally achieve better performance than the LAD and OLS estimators but the magnitude of improvement is very small.

Prediction performance measured by prediction R^2 is summarized in Table 4. The prediction R^2 is not necessarily positive because out-of-sample predictions are not guaranteed to be orthogonal to out-of-sample residuals. The prediction R^2 compares the performance of a

predictor to the imaginary situation where we know in advance the sample mean of the target variable over the entire out-of-sample period and use it as our predictor. Therefore, a positive prediction R^2 indicates that the predictor is better in terms of PMSE than the sample mean assumed known in advance. According to the summary statistics in Table 4, the return on ADC TeleCom Co. is more difficult to predict than that for HomeStake Co.. Nevertheless, all combination estimators and the LAD estimator give positive prediction R^2's.

Table 4. Out-of-Sample Prediction Performance

	ADC TeleCom Co.		HomeStake Co.	
	R^2	R_A^2	R^2	R_A^2
OLS	-0.001410	0.000879	0.005622	-0.00955
LAD	0.001613	0.003402	0.004698	-0.00674
NRC	0.001871	0.002967	0.006569	-0.00878
JSC	0.001782	0.003366	0.006987	-0.00663
OWS	0.001858	0.002964	0.006606	-0.00875
RIDGE	-0.000540	0.003894	0.006282	-0.00561
GAR	-0.001280	0.002069	0.007043	-0.00616
NNGAR	-0.000320	0.002136	0.010452	-0.00405
Random Walk	-0.001999	0.004883	-0.000147	-0.00149

7. CONCLUSION

We have proposed an extension of James-Stein type estimators in a direction that preserves their risk improvement when the sample size goes to infinity. This extension supports use of James-Stein type estimators when one has a moderate or large number of observations. This is important because large data sets are becoming more and more easily available. We permit the data-dependent point towards which we shrink our base estimator to be asymptotically biased or asymptotically correlated with the base estimator, in contrast to previous work. Our results thus suggest that many other interesting estimators are potential candidates for use as a data-dependent point.

APPENDIX

Proof of Lemma 1. First note that $n^{1/2}(b_n - g_n) = n^{1/2}(b_n - b^0) - n^{1/2}(g_n - b^0) \xrightarrow{d} U_1 - U_2$ by Assumption 2. One can also show that $\| b_n - g_n \|_n^2 \xrightarrow{d} \| U_1 - U_2 \|^2$ by Assumption 2 where $\| U_1 - U_2 \|^2 \equiv (U_1 - U_2)Q(U_1 - U_2)$. Combining these observations, we have $n^{1/2}(d_{c_1}^{JS}(b_n, g_n) - b^0) = \{1 - c_1 / \| b_n - g_n \|_n^2\} n^{1/2}(b_n - g_n) + n^{1/2}(g_n - b^0) \xrightarrow{d} \{1 - c_1 / \| U_1 - U_2 \|^2\}(U_1 - U_2) + U_2$ which is $d_{c_1}^{JS}(U_1, U_2)$. The loss function for $d_{c_1}^{JS}(b_n, g_n)$ is given by $L(d_{c_1}^{JS}(b_n, g_n), b^0) = \| d_{c_1}^{JS}(b_n, g_n) - b^0 \|_n^2$ which converges to $\| d_{c_1}^{JS}(U_1, U_2) - 0 \|^2$. Therefore, $AR(\{ d_{c_1}^{JS}(b_n, g_n)\}, b^0) = E\{\| d_{c_1}^{JS}(U_1, U_2) - 0 \|^2\} = R(d_{c_1}^{JS}(U_1, U_2), 0)$. □

Proof of Theorem 1. By Lemma 1 $AR(\{ d_{c_1}^{JS}(b_n, g_n)\}, b^0) = w\, c_1^2 - 2n\, c_1 + k$. The first and second derivatives of the asymptotic risk with respect to c_1 are $2(w\, c_1 - n)$ and $2w$. Since $w > 0$ by Assumption 3, $AR(\{ d_{c_1}^{JS}(b_n, g_n)\}, b^0)$ is strictly convex in c_1. By setting the first derivative to zero and solving for c_1, we have $c_1^* = n / w$. Plugging c_1^* into the asymptotic risk, we have the minimum asymptotic risk $AR(\{ d_{c_1^*}^{JS}(b_n, g_n)\}, b^0) = -n^2 / w + k$. Since $k \equiv AR(\{b_n\}, b^0)$, $AR(\{ d_{c_1^*}^{JS}(b_n, g_n)\}, b^0) \leq AR(\{b_n\}, b^0)$, where the equality holds only when $n = 0$. The last result follows from the strict convexity of the asymptotic risk. □

Remark A.1. Proofs of Lemma 2 and Lemma 3 follow the same arguments in the proof of Lemma 1. Also, Theorem 2 can be proved in an analogous way as in Theorem 1. We therefore omit these proofs.

Proof of Theorem 3. It follows from Lemma 3 that $AR(\{d_l^{OW}(b_n, g_n)\}, b^0) = a\, l_1^2 - 2r\, l_1 + 2\, l_1\, l_2 + w\, l_2^2 - 2n\, l_2 + k$. The Hessian is given by $\begin{bmatrix} 2a & 0 \\ 0 & 2w \end{bmatrix}$, which is positive definite. Hence, $AR(\{d_l^{OW}(b_n, g_n)\}, b^0)$ is strictly convex in l. By setting the first derivative to zero and solving for l, we have $l_1^* = (aw-1)^{-1}(rw-n)$ and $l_2^* = (aw-1)^{-1}(an-r)$. Plugging l^* into the asymptotic risk, we have the minimum asymptotic risk $AR(\{d_{l^*}^{OW}(b_n, g_n)\}, b^0) = (aw-1)^{-2}\{-ar^2w^2 - (2arn - a^2n^2 + r^2)w + (an^2 - 2rn)\} + k$. For the last result, we define $h(w) \equiv -\{-ar^2w^2 - (2arn - a^2n^2 + r^2)w + (an^2 - 2rn)\}$. We want to show $h(w) \geq 0$ for all w, which delivers the desired result. (Case 1) $r = 0$ and $n = 0$. Then $h(w) = 0$. (Case 2) $r = 0$ and $n \neq 0$. Then $h(w) = an^2(aw-1) > 0$ because $a > 0$, $n \neq 0$ and $aw > 1$. (Case 3) $r \neq 0$ and $n = 0$. Then $h(w) = r^2 w(aw-1) > 0$ because $w > 0$, $r \neq 0$ and $aw > 1$. (Case 4) $r \neq 0$ and $n \neq 0$. Define $w^* \in \{w \mid h(w) = 0\}$. Suppose that $an - r = 0$. Then $w^* = -(a^2n^2 - r^2 - 2arn)/2ar^2$. It can be shown that $w > w^*$, because $aw > 1$ and $r \neq 0$. This implies that $h(w) \geq 0$ for all w. Now consider the case that $an - r \neq 0$. Define $w_-^* \equiv n(2r-a)/r^2$ and $w_+^* \equiv 1/a$. It follows that $w_-^* < w_+^*$ and $w_+^* < w$ because $aw > 1$. This implies that $h(w) \geq 0$ for all w. □

Proof of Corollary 1. Because of strict convexity, $AR(\{d_{l^*}^{OW}(b_n, g_n)\}, b^0) \leq AR(\{d_l^{OW}(b_n, g_n)\}, b^0)$ for any l. Choose $l = (0, l_2^*)'$. It can be shown with some algebra that $AR(\{d_l^{OW}(b_n, g_n)\}, b^0) = -n^2/w + k$, which is $AR(\{d_{c_1^*}^{JS}(b_n, g_n)\}, b^0)$. Since l^* is the unique and global solution, the strict inequality holds if l_1^* is not equal to zero. For the second claim, choose $l = (l_1^*, 0)'$. Then, $AR(\{d_l^{OW}(b_n, g_n)\}, b^0) = -r^2/a + k$, which is equal to $AR(\{d_{c_2^*}^{NR}(b_n, g_n)\}, b^0)$. The same argument applies to the strict inequality. □

Proof of Lemma 4. It is trivial to show that $|a| < \infty$ and $|r| < \infty$ because these are obtained by adding variances and covariances of normal random variables. In order to show that $|w| < \infty$, we define $V \equiv U_1 - U_2$, $R \equiv P'QP$ where P is a square root matrix of Σ_{22}. Note that $\frac{1}{V'QV} = \frac{1}{Y'RY} = \frac{1}{Y'Y}\frac{Y'Y}{Y'RY}$ where $Y \equiv P^{-1}V$. It can be shown that $\frac{1}{l_M} \leq \frac{Y'Y}{Y'RY} \leq \frac{1}{l_m}$ where l_M and l_m are the largest and smallest eigenvalues of R (hence, of Q) respectively. This implies that $\frac{1}{l_M}E\left[\frac{1}{Y'Y}\right] \leq w \leq \frac{1}{l_m}E\left[\frac{1}{Y'Y}\right]$. If $k > 2$, then $\frac{1}{l_M(k-2)} \leq w \leq \frac{1}{l_m(k-2)}$. For the last claim, note that $\nu^2 \leq E\left[(U_1'QV)^2\right]E\left[\frac{1}{(V'QV)^2}\right]$ by the Cauchy-Schwarz inequality. Since U_1 and V are normal random variables, $E\left[(U_1'QV)^2\right] < \infty$. Hence, we have $\frac{1}{l_M^2}E\left[\frac{1}{(Y'Y)^2}\right] \leq E\left[\frac{1}{(V'QV)^2}\right] \leq \frac{1}{l_m^2}E\left[\frac{1}{(Y'Y)^2}\right]$. If $k \neq 2$ and $k \neq 4$, then $\frac{1}{l_M^2(k-2)(k-4)} \leq E\left[\frac{1}{(V'QV)^2}\right] \leq \frac{1}{l_m^2(k-2)(k-4)}$. Therefore, $|\nu| < \infty$. □

Proof of Theorem 4. Some elementary linear algebra gives that $a = tr[(A - \Delta - \Delta' + B)Q]$. Since trace is a continuous function, $\hat{a}_n \equiv tr[(\hat{A}_n - \hat{\Delta}_n - \hat{\Delta}_n' + \hat{B}_n)\hat{Q}_n]$ is consistent. By the same reasoning, $\hat{r}_n \equiv tr[(\hat{A}_n - \hat{\Delta}_n')\hat{Q}_n]$ is consistent. The argument for $\hat{w}_n$ and $\hat{n}_n$ is more involved. We define $g(\Sigma_{22}Q,t) \equiv \det(I + 2t\Sigma_{22}Q)^{-1/2}$. We want to show that there exists a dominating function $d(t)$ such that (i) $|g(\Sigma_{22}Q,t)| \le d(t)$ for all Σ_{22} and Q in a compact parameter space and (ii) $\int_0^\infty d(t)dt < \infty$. Using the relationship between determinant and eigenvalues of a matrix, we can express $g(\cdot)$ in terms of eigenvalues; $g(\Sigma_{22}Q,t) = \left\{\prod_{i=1}^{k} l_i\right\}^{1/2}$ where l_i is an eigenvalue of the inverse matrix of $I + 2t\Sigma_{22}Q$. Using some linear algebra, we can obtain an upper bound given by $g(\Sigma_{22}Q,t)^2 \le \left\{\frac{1}{2|\bar{k}|t+1}\right\}^k$ where $\bar{k}$ is the minimum (in absolute value) eigenvalue of $\Sigma_{22}Q$. Hence the natural candidate for the dominating function is $d(t) = \left\{\frac{1}{2|\bar{k}|t+1}\right\}^{k/2}$. As long as $k > 2$ which is assumed in the corollary, the dominating function $d(t)$ will satisfy the second condition, $\int_0^\infty d(t)dt < \infty$. Hence we have the desired result. The proof for the consistency of $\hat{n}_n$ is more complicated, but the key step is again to find a dominating function. Since the argument is similar, we just give the dominating function: $D(t) = 2k|\bar{x}|\left\{\frac{1}{2|\bar{k}|t+1}\right\}^{k+1}$ where $\bar{k}$

is the minimum eigenvalue (in absolute value) of M and $\bar{x}$ is the maximum (in absolute value) eigenvalue of M_1^2. □

Proof of Corollary 2. Both $\hat{l}_{1n}$ and $\hat{l}_{2n}$ are continuous function of the consistent estimators. Since the limit of continuous function of consistent estimators is the value of function evaluated at the limit of the consistent estimators, we have the desired results: $\hat{l}_{1n} \xrightarrow{p} l_1^*$ and $\hat{l}_{2n} \xrightarrow{p} l_2^*$. The consistency of the estimated weights together with the Slutsky Theorem delivers $n^{1/2}(d_{\hat{l}_n}^{OW}(b_n, g_n) - b^0) \xrightarrow{d} d_{l^*}^{OW}(U_1, U_2)$ which in turn implies that $AR(\{d_{\hat{l}_n}^{OW}(b_n, g_n)\}, b^0) = AR(\{d_{l^*}^{OW}(b_n, g_n)\}, b^0)$. □

REFERENCES

Arthanari, T. S. and Dodge, Y. (1982), "*Mathematical Programming in Statistics*," New York: John Wiley.

Barrodale, I. and Roberts, F.D.K. (1974), "Algorithm 478: Solution of an Over-Determined System of Equations in the L_1 Norm," *Communications of the Association for Computing Machinery*, 17, 319-320.

Bates, C. E. and White, H. (1993), "Determination of Estimators with Minimum Asymptotic Covariance Matrices," *Econometric Theory*, 9, 633-648.

Breiman, L. (1995), "Better Subset Regression Using the Nonnegative Garrote," *Technometrics*, 37, 373-384.

Breiman, L. (1996), "Heuristics of Instability and Stabilization in Model Selection," *The Annals of Statistics*, 24, 2350-2383.

Cohen, A. (1976), "Combining Estimates of Location," *Journal of the American Statistical Association*, 71, 172-175.

Green, E. J. and Strawderman W. E. (1991), "James-Stein Type Estimator for Combining Unbiased and Possibly Biased Estimators," *Journal of the American Statistical Association*, 86, 1001-1006.

James, W. and Stein, C. (1960), "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.

Laplace (1818), Deuxieme Supplement a la Theorie Analytique des Probabilites.

Saleh, A. K. M. E. and Sen, P. K. (1985), "On Shinkage M-estimators of Location Parameters," *Communications in Statistics: Theory and Methods*, 14, 2313-2329.

Schmoyer, R. and Arnold, S. (1989), "Shrinking Techniques for Robust Regression," in *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin*, ed. L. J. Gleser, New York: Springer-Verlag, pp. 368-384.

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

Stein, C. (1955), "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.

Taylor, L. D. (1974), "Estimation by Minimizing the Sum of Absolute Errors," in *Frontiers in Econometrics*, ed. Zarembka, P., New York: Academic Press.

Ullah, A. (1990), "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, A. Ullah. New York: Springer-Verlag, pp. 242-292.