BOOTSTRAP METHODS FOR COVARIANCE STRUCTURES

by

Joel L. Horowitz
Department of Economics
University of Iowa
Iowa City, IA 52242

October 1996
Revised July 1997

ABSTRACT

The optimal minimum distance (OMD) estimator for models of covariance structures is asymptotically efficient but has much worse finite-sample properties than does the equally-weighted minimum distance (EWMD) estimator. This paper shows how the bootstrap can be used to improve the finite-sample performance of the OMD estimator. The theory underlying the bootstrap's ability to reduce the bias of estimators and errors in the coverage probabilities of confidence intervals is summarized. The results of numerical experiments and an empirical example show that the bootstrap often essentially eliminates the bias of the OMD estimator. The finite-sample estimation efficiency of the bias-corrected OMD estimator often exceeds that of the EWMD estimator. Moreover, the true coverage probabilities of confidence intervals based on the OMD estimator with bootstrap critical values are very close to the nominal coverage probabilities.

I thank Joseph Altonji for attracting my attention to the problem addressed in this paper and for making his data available. Richard Blundell and Gene Savin provided comments on earlier drafts. This research was supported in part by NSF grants SBR-9307677 and SBR-9617925.

BOOTSTRAP METHODS FOR COVARIANCE STRUCTURES

1. INTRODUCTION

Estimates of covariance structures are important in the analysis of a variety of economic processes. See, for example, Abowd and Card (1987, 1989), Behrman *et al*. (1994), Griliches (1979), and Hall and Mishkin (1982). Estimates of covariance structures can be obtained by minimizing the weighted distance between sample moments and the estimated population moments. Weighting all sample moments equally produces the equally-weighted minimum distance (EWMD) estimator, whereas choosing the weights to maximize asymptotic estimation efficiency produces the optimal minimum distance (OMD) estimator.

The OMD estimator dominates the EWMD estimator in terms of asymptotic efficiency, but it has been found to have poor finite-sample properties in applications (Abowd and Card 1989). Altonji and Segal (1994, 1996) carried out an extensive Monte Carlo investigation of the finite-sample performance of the OMD estimator. They found that the estimator is badly biased with samples of the sizes often found in applications and that its finite-sample root-mean-square estimation error (RMSE) often greatly exceeds the RMSE of the asymptotically inefficient EWMD estimator. Altonji and Segal also found that the true coverage probabilities of asymptotic confidence intervals based on the OMD estimator tend to be much lower than the nominal coverage probabilities. Thus, estimation and inference based on the OMD estimator can be highly misleading with finite samples.

This paper shows how the bootstrap can be used to overcome these problems. The bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling one's data. It amounts to treating the data as if they were the population for the purpose of evaluating the distribution of interest. If the data were, in fact, the population, the coverage probability of a confidence interval and the bias of an estimator could be computed with arbitrary accuracy by Monte Carlo simulation. Moreover, subtracting the bias from the estimator would yield an unbiased estimator with increased finite-sample efficiency. Since the data are not the population, the bootstrap provides only an approximation to the coverage probability of a confidence interval or the bias of an estimator. It turns out, however, that the bootstrap approximation is often more accurate than the approximation obtained from asymptotic distribution theory. As a result, the bootstrap often provides estimates of coverage probabilities of confidence intervals that are more accurate than estimates obtained from asymptotic distribution theory. Moreover, the use of the bootstrap to reduce the bias of an estimator often produces very large increases in finite-sample estimation efficiency.

Section 2 of this paper explains why the bootstrap often provides approximations that are more accurate than the approximations of asymptotic distribution theory. Section 3 explains how to use the bootstrap in OMD estimation of covariance structures. Section 4 presents the results of Monte Carlo experiments in which the bootstrap is applied to the estimation of covariance structures with simulated data. In experiments in which the kurtosis of the sampled population is low, the bootstrap essentially eliminates the finite-sample bias of the OMD estimator and the errors in the coverage probabilities of confidence intervals based on the OMD estimator. Moreover, with low-kurtosis populations, the finite-sample RMSE of the bias-corrected OMD estimator is less than that of the EWMD estimator. The bootstrap is less successful in removing bias and correcting coverage probabilities when the kurtosis of the sampled population is high. Section 5 shows that this problem can be overcome by modifying the OMD estimator in a way that improves the estimate of the OMD weight matrix. In the Monte Carlo experiments with high-kurtosis populations, application of the bootstrap to the modified OMD estimator produces estimates with little bias and RMSE's that are equal to or smaller than those of the EWMD estimates. The true coverage probabilities of confidence intervals obtained from the bootstrap and the modified OMD estimator are very close to the nominal probabilities. Section 6 presents an empirical example based on the data used by Altonji and Segal (1994, 1996). Section 7 presents concluding comments.

Sections 3c and 5b include step-by-step instructions for implementing the bootstrap procedures that are described. These instructions may be carried out without detailed study of the theoretical arguments that justify them.

2. THE THEORY OF THE BOOTSTRAP

This section explains how the bootstrap can be used to reduce the bias of an estimator and why the bootstrap often provides an approximation to the coverage probability of a confidence interval that is more accurate than the approximation of asymptotic distribution theory. The material in this section is not new, and the discussion is informal. Mathematically rigorous treatments are available in Beran (1988), Beran and Ducharme (1991), and Hall (1986, 1992).

To minimize the complexity of the discussion, it is assumed that the inferential problem is to obtain a point estimate and a confidence interval for a scalar parameter θ that can be expressed as a smooth function of a vector of population moments. It is also assumed that θ can be estimated consistently by substituting sample moments in place of population moments in the smooth function. Many important econometric estimators, including OMD estimators, are either functions of sample moments or can be approximated by functions of sample moments with an approximation error that approaches zero very rapidly as the sample size increases. Thus, the theory outlined in this section applies to a wide variety of estimators that are important in applications.

To be specific, let the vector $\mu \equiv E(X)$ be the population mean of a random vector X. Assume that the true value of θ is $\theta_0 = g(\mu)$, where g is a known, continuous function. Suppose that the data consist of a random sample $\{X_i: i = 1, ..., n\}$ of X. Define the vector $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Then θ is estimated consistently by

$$\theta_n = g(\bar{X}). \qquad (2.1)$$

If θ_n has a finite mean, then $E(\theta_n) = E[g(\bar{X})]$. However, $E[g(\bar{X})] \neq g(\mu)$ in general unless g is a linear function. Therefore, $E(\theta_n) \neq \theta_0$, and θ_n is a biased estimator of θ. In particular, $E(\theta_n) \neq \theta_0$ if θ_n is any of a variety of familiar maximum likelihood or generalized method of moments estimators, including the

OMD estimator. Section 2b explains how the bootstrap can be used to reduce the bias of θ_n.

Now consider the problem of obtaining a confidence interval for θ. Assume that g is continuously differentiable in a neighborhood of μ and that X has a finite, positive-definite covariance matrix V. Let G_1 denote the vector of first derivatives of g. Then as $n \to \infty$,

$$n^{1/2}(\theta_n - \theta_0) \to^d N(0,V_\theta), \tag{2.2}$$

where

$$V_\theta = G_1(\mu)'VG_1(\mu). \tag{2.3}$$

To obtain an asymptotic confidence interval for θ, let $V_{n\theta}$ be the consistent estimator of V_θ that is obtained by replacing μ with $\bar{X}$ and V with the sample covariance matrix of X in (2.3). Let T_n be the t statistic for testing the hypothesis H_0: $\theta = \theta_0$:

$$T_n = n^{1/2}(\theta_n - \theta_0)/V_{n\theta}^{1/2}.$$

It follows from (2.2) that $T_n \to^d N(0,1)$ as $n \to \infty$. Therefore,

$$\lim_{n \to \infty} P(|T_n| \le z_{\alpha/2}) = 1 - \alpha, \tag{2.4}$$

where $z_{\alpha/2}$ denotes the $1 - \alpha$ quantile of the standard normal distribution. Moreover, an asymptotic, $1 - \alpha$ confidence interval for θ is

$$\theta_n - z_{\alpha/2}(V_{n\theta}/n)^{1/2} \le \theta_0 \le \theta_n + z_{\alpha/2}(V_{n\theta}/n)^{1/2} \tag{2.5}$$

Equation (2.4) insures that the nominal coverage probability of the confidence interval (2.5), $1 - \alpha$, is arbitrarily close to the true coverage probability if n is sufficiently large. In samples of the sizes found in applications, however, the distribution of T_n may not be close to N(0,1), and $P(|T_n| \le z_{\alpha/2})$ may not be close to $1 - \alpha$. Therefore, the nominal and true coverage probabilities of (2.5) may be greatly different. Section 2c shows how the bootstrap can be used to reduce the difference between the nominal and true coverage probabilities of a confidence interval.

a. The Bootstrap Sampling Procedure

The bootstrap treats the estimation data as if they were the population. Bootstrap "data" are generated by sampling the estimation data randomly with replacement. In other words, the bootstrap samples the empirical distribution of the estimation data. Let $\{X_i^*: \ i = 1, ..., n\}$ be a bootstrap sample that is obtained this way. Define $\overline{X^*} = n^{-1}\sum_{i=1}^{n} X_i^*$ to be the vector of bootstrap sample means. Then the bootstrap estimator of θ is

$$\theta_n^* = g(\overline{X^*}). \tag{2.6}$$

Conditional on the original estimation data $\{X_i\}$, the probability distribution of θ_n^* can be determined with arbitrary accuracy through a Monte Carlo simulation that consists of repeatedly drawing bootstrap samples $\{X_i^*\}$ and computing the value of θ_n^*. Therefore, the distribution of θ_n^* conditional on the data can be treated as known.

The distribution of θ_n^* is induced by sampling the empirical distribution of X, whereas the distribution of θ_n is induced by sampling the population distribution of X. The empirical and population distributions of X are not the same, so the distributions of θ_n^* and θ_n are not the same. However, the empirical distribution of X converges to the population distribution as $n \rightarrow \infty$. Therefore, the distribution of θ_n^* mimics the distribution of θ_n, at least if n is large. In fact, for reasons that will now be explained, this mimicry can provide better approximations to the bias of θ_n and to the coverage probabilities of confidence intervals than does asymptotic distribution theory.

b. Using the Bootstrap to Reduce Bias

This section explains how the bootstrap can be used to reduce the bias of θ_n. The theoretical results are illustrated with a simple numerical example.

Suppose that g is three times continuously differentiable in a neighborhood of μ. Let $\|\bullet\|$ denote the Euclidean norm. Define G_1 as in (2.3), and let G_2 denote the matrix of second partial derivatives of g. Then a Taylor series expansion of the right-hand side of (2.1) about $\overline{X} = \mu$ gives

$$\theta_n - \theta_0 = G_1(\mu)'(\bar{X} - \mu) + (1/2)(\bar{X} - \mu)'G_2(\mu)(\bar{X} - \mu)$$
$$+ O(\|\bar{X} - \mu\|^3). \qquad (2.7)$$

If the components of X all have finite third absolute moments, then taking expectations on both sides of (2.7) gives

$$E(\theta_n - \theta_0) = (1/2)E[(\bar{X} - \mu)'G_2(\mu)(\bar{X} - \mu)] + O(n^{-3/2}). \qquad (2.8)$$

The first term on the right-hand side of (2.8) has size $O(n^{-1})$. Therefore, through $O(n^{-1})$ the bias of θ_n is

$$B_n = (1/2)E[(\bar{X} - \mu)'G_2(\mu)(\bar{X} - \mu)]. \qquad (2.9)$$

Now consider the bootstrap. The bootstrap samples a distribution whose true mean is $\bar{X}$. Therefore, $\bar{X}$ is the bootstrap analog of μ, and $\theta_n = g(\bar{X})$ is the bootstrap analog of θ_0. The bootstrap analog of (2.7) is

$$\theta_n{}^* - \theta_n = G_1(\bar{X})'(\bar{X}^* - \bar{X}) + (1/2)(\bar{X}^* - \bar{X})'G_2(\bar{X})(\bar{X}^* - \bar{X})$$
$$+ O(\|\bar{X}^* - \bar{X}\|^3). \qquad (2.10)$$

Let E* denote the expectation under bootstrap sampling, that is the expectation relative to the empirical distribution of the estimation data. Let $B_n{}^* \equiv E^*(\theta_n{}^* - \theta_n)$ denote the bias of $\theta_n{}^*$ as an estimator of θ_n. Taking E* expectations on both sides of (2.10) shows that

$$B_n{}^* = E^*[(\bar{X}^* - \bar{X})'G_2(\bar{X})(\bar{X}^* - \bar{X})] + O(n^{-3/2}) \qquad (2.11)$$

almost surely. Because the distribution that the bootstrap samples is known, $B_n{}^*$ can be computed with arbitrary accuracy by Monte Carlo simulation. Thus, $B_n{}^*$ is a feasible estimator of the bias of θ_n. The details of the simulation procedure are described below.

By comparing (2.9) and (2.11), it can be seen that the only differences between B_n and the leading term of $B_n{}^*$ are that $\bar{X}$ replaces μ in $B_n{}^*$ and the empirical expectation, E*, replaces the population expectation, E. This observation together with arguments based on the strong law of large numbers leads to the result that $B_n{}^* = B_n + O(n^{-3/2})$ almost surely. Therefore, through

$O(n^{-1})$ use of the bootstrap bias estimate B_n^* almost surely provides the same bias reduction that would be obtained if the infeasible population value B_n could be used. This is the source of the bootstrap's ability to reduce the bias of θ_n. The resulting bias-corrected estimator of θ is $\theta_n - B_n^*$.

The Monte Carlo procedure for computing B_n^* is as follows:

B1. Use the estimation data to compute θ_n.

B2. Generate a bootstrap sample of size n by sampling the estimation data randomly with replacement. Compute θ_n^* by using (2.6).

B3. Compute $E^*\theta_n^*$ by averaging the results of many repetitions of step B2. Set $B_n^* = E^*\theta_n^* - \theta_n$.

To implement this procedure it is necessary to choose the number of repetitions, m, of step B2. It suffices to choose m sufficiently large that the estimate of $E^*\theta_n^*$ does not change significantly if m is increased further.

The following simple numerical example illustrates the bootstrap's ability to reduce bias. Examples that are more realistic but also more complicated are presented in Sections 4-6. Let $X \sim N(0,6)$ and $n = 10$. Let $g(\mu) = \exp(\mu)$. Then $\theta_0 = 1$, and $\theta_n = \exp(\bar{X})$. B_n and the bias of $\theta_n - B_n^*$ can be found through the following Monte Carlo procedure:

MC1. Generate an estimation data set of size n by sampling from the N(0,6) distribution. Use this data set to compute θ_n.

MC2. Compute B_n^* by carrying out steps B1-B3. Form $\theta_n - B_n^*$.

MC3. Estimate $E(\theta_n - \theta_0)$ and $E(\theta_n - B_n^* - \theta_0)$ by averaging the results of many repetitions of steps MC1-MC2. Estimate the mean square errors of θ_n and $\theta_n - B_n^*$ by averaging the realizations of $(\theta_n - \theta_0)^2$ and $(\theta_n - B_n^* - \theta_0)^2$.

The following are the results obtained with 1000 Monte Carlo replications and 100 repetitions of step B2 at each Monte Carlo replication:

	Bias	Mean-Square Error
θ_n	0.356	1.994
$\theta_n - B_n^*$	-0.063	1.246

In this example, the bootstrap reduces the magnitude of the bias of the estimator of θ by nearly a factor of 6. The mean-square estimation error is reduced by 38 percent.

c. Using the Bootstrap to Obtain a Confidence Interval

This section explains how the bootstrap can be used to reduce the difference between the true and nominal coverage probabilities of a confidence interval for θ. A simple numerical example is presented to illustrate the theory.

Let $t_{n\alpha}$ be the exact solution to

$$P(|T_n| \le t_{n\alpha}) = 1 - \alpha.$$

Observe that $t_{n\alpha}$ is the exact finite-sample critical value of a symmetrical t test of the hypothesis H_0: $\theta - \theta_0$. If $t_{n\alpha}$ were known, an exact $1 - \alpha$ confidence interval for θ could be obtained by replacing $z_{\alpha/2}$ with $t_{n\alpha}$ in (2.5). However, $t_{n\alpha}$ is unknown because the exact finite-sample distribution of T_n is unknown. One way to overcome this problem is to approximate the finite-sample distribution of T_n by its asymptotic distribution, thereby leading to the asymptotic confidence interval (2.5). The bootstrap can provide a better approximation to the finite-sample distribution of T_n, a better approximation to $t_{n\alpha}$ and, therefore, a confidence interval whose true coverage probability is closer to $1 - \alpha$.

To see how the bootstrap does this, let the exact finite-sample cumulative distribution function (CDF) of T_n be $H_n(z,F) \equiv P(T_n \le z)$, where F is the CDF of X. Let F_n be the empirical distribution function (EDF) of the estimation data $\{X_i\}$. Then $t_{n\alpha}$ satisfies $H_n(t_{n\alpha},F) - H_n(-t_{n\alpha},F) = 1 - \alpha$. Under regularity conditions (see, e.g., Hall 1992), H_n has an Edgeworth expansion of the form

$$H_n(z,F) = \Phi(z) + n^{-1/2}p_1(z,F) + n^{-1}p_2(z,F) + o(n^{-1}) \qquad (2.12)$$

uniformly over z, where Φ is the standard normal CDF, p_1 and p_2 are functionals of (z,F), $p_1(z,F)$ is an even function of z for each F, $p_2(z,F)$ is an odd function of z, and $p_2(z,F_n) \to p_2(z,F)$ almost surely as $n \to \infty$ uniformly over z. It follows from (2.12) and the symmetry of Φ, p_1, and p_2 that for any $z > 0$

$$P(|T_n| > z) = 2[1 - \Phi(z) - n^{-1}p_2(z,F)] + o(n^{-1}). \qquad (2.13)$$

Now consider the bootstrap. The bootstrap samples a population whose CDF is F_n. Let $T_n{*}$ be the bootstrap analog of T_n: $T_n{*} = n^{1/2}(\theta_n{*} - \theta_n)/(V_{n\theta}{*})^{1/2}$, where $V_{n\theta}{*}$ is bootstrap analog of V_θ. $V_{n\theta}{*}$ is obtained from V_θ by replacing μ with $\overline{X{*}}$ and V with the bootstrap sample covariance matrix in (2.3). Denote the CDF of $T_n{*}$ conditional on the estimation data by $H_n{*}(z,F_n)$. Under regularity conditions, $H_n{*}(z,F_n)$ has the following asymptotic expansion:

$$H_n{*}(z,F_n) = \Phi(z) + n^{-1/2}p_1(z,F_n) + n^{-1}p_2(z,F_n) + o(n^{-1}) \qquad (2.14)$$

uniformly over z almost surely. The leading terms of (2.12) and (2.14) are identical. Therefore, because p_1 is even and $p_2(z,F_n) \to p_2(z,F)$ almost surely, the distributions of $|T_n|$ and $|T_n{*}|$ are almost surely identical through $O(n^{-1})$. This is the source of the bootstrap's ability to provide an improved approximation to $t_{n\alpha}$ and improved coverage probabilities for confidence intervals.

To obtain the improved approximation to $t_{n\alpha}$, let $t_{n\alpha}{*}$ be the solution to

$$P{*}(|T_n{*}| \le t_{n\alpha}{*}) = 1 - \alpha.$$

The quantity $t_{n\alpha}{*}$ is the critical value of the bootstrap symmetrical t test of the hypothesis $H_0{*}$: $\theta = \theta_n$. The value of $t_{n\alpha}{*}$ can be computed with arbitrary accuracy by the following Monte Carlo procedure:

C1. Use the estimation data to compute θ_n.

C2. Generate a bootstrap sample of size n by sampling the estimation data randomly with replacement. Compute $T_n{*}$ from this sample.

C3. Use the results of many repetitions of C2 to compute the EDF of $|T_n{*}|$. Set $t_{n\alpha}{*}$ equal to the $1 - \alpha$ quantile of this EDF.

The number of replications of step C2 can be chosen by finding a value sufficiently large that further increases have no effect on $t_{n\alpha}{*}$.

It follows from (2.14) and $p_2(t_{n\alpha}{*},F_n) - p_2(t_{n\alpha}{*},F) \to 0$ almost surely that

$$2[1 - \Phi(t_{n\alpha}*) - n^{-1}p_2(t_{n\alpha}*,F)] = \alpha + o(n^{-1}) \tag{2.15}$$

almost surely. Combining (2.13) and (2.15), yields the result that

$$t_{n\alpha}* = t_{n\alpha} + o(n^{-1})$$

almost surely. Thus, $t_{n\alpha}*$ and $t_{n\alpha}$ differ by $o(n^{-1})$ almost surely. In contrast, $t_{n\alpha} - z_{\alpha/2} = O(n^{-1})$. Therefore, the bootstrap approximation to $t_{n\alpha}$ is more accurate than the approximation provided by asymptotic distribution theory.

It can also be shown that

$$P[\theta_n - t_{n\alpha}*(V_{n\theta}/n)^{1/2} \leq \theta_0 \leq \theta_n + t_{n\alpha}*(V_{n\theta}/n)^{1/2}] = 1 - \alpha + O(n^{-2}).$$

Thus the difference between the true and nominal coverage probabilities of a symmetrical confidence interval for θ based on the bootstrap critical value $t_{n\alpha}*$ is $O(n^{-2})$ as $n \to \infty$. In contrast, the difference is $O(n^{-1})$ when the asymptotic critical value $z_{\alpha/2}$ is used. Therefore, the bootstrap reduces the error in the coverage probability of a symmetrical confidence interval. This is because first order (or "ordinary") asymptotic approximations drop the terms involving p_1 and p_2 in (2.12) and (2.14).

Similar arguments can be used to show that the bootstrap reduces the error in the coverage probability of a one-sided confidence interval. The error is usually $O(n^{-1})$ when the bootstrap is used, whereas it is $O(n^{-1/2})$ when the confidence interval is obtained from first-order asymptotic approximations.

The foregoing results depend critically on the fact that T_n is an asymptotically pivotal statistic. That is, its asymptotic distribution does not depend on unknown population parameters. The bootstrap reduces the error in the coverage probability of a confidence interval only if the interval is based on an asymptotically pivotal statistic. If a statistic that is not asymptotically pivotal is used ($\theta_n - \theta_0$, for example), then the error in the coverage probability of a bootstrap confidence interval has the same size as the error in the coverage probability of an asymptotic confidence interval.

It may seem that one can reduce the error in the coverage probability of a confidence interval by replacing θ_n in (2.5) with the bias-corrected estimator $\theta_n - B_n^*$. However, this procedure does not account for differences between the finite-sample and asymptotic values of the variance, skewness, and kurtosis of T_n. Therefore, the error in coverage probability remains $O(n^{-1})$ when θ_n is replaced by $\theta_n - B_n^*$ in (2.5). This is the size of the error made with the unmodified (2.5). Of course numerical values of the errors made with the two versions of (2.5) may differ.

Another possibility is to replace θ_n with $\theta_n - B_n^*$ in the bootstrap procedure for obtaining a confidence interval. This does not improve the accuracy of the coverage probability beyond that available with θ_n. The reason is that $t_{n\alpha}^*$ automatically accounts for bias through $O(n^{-1})$, which is the size of the correction provided by B_n^*. Moreover, use of $\theta_n - B_n^*$ does not reduce errors associated with higher-order cumulants of T_n. However, $t_{n\alpha}^*$ is not an $O(n^{-1})$ correct critical value for a confidence interval centered at $\theta_n - B_n^*$. Obtaining such a critical value requires bootstrapping based on $\theta_n - B_n^*$.

The bootstrap's ability to reduce the error in the coverage probability of a confidence interval can be illustrated with an extension of the numerical example of Section 2b. The extension consists of carrying out a Monte Carlo experiment to evaluate the coverage probability of a nominal 95% confidence interval for θ. The Monte Carlo procedure is:

MC4. Generate an estimation data set of size n by sampling from the N(0,6) distribution. Use this data set to compute θ_n.

MC5. Compute $t_{n,0.05}^*$ by carrying out steps C1-C3. Determine whether θ_0 is contained in the asymptotic confidence interval (2.5) and in the bootstrap interval that is formed by replacing $z_{0.025}$ with $t_{n,0.05}^*$ in (2.5).

MC6. Determine the empirical coverage probabilities of the asymptotic and bootstrap confidence intervals from the results of 1000 repetitions of steps MC4 and MC5.

The empirical coverage probability of the asymptotic confidence interval was 0.886 in this experiment, whereas the empirical coverage probability of the bootstrap interval was 0.943. The asymptotic coverage probability is statistically significantly different from the nominal probability of 0.95 ($p < 0.01$), whereas the bootstrap coverage probability is not ($p > 0.10$).

3. APPLICATION TO THE OMD ESTIMATOR

This section specializes the results of Section 2 to OMD estimation of covariance structures. Section 3a provides a brief description of the EWMD and OMD estimators. Section 3b explains how to form a bootstrap analog of the OMD estimator. Section 3c gives step-by-step procedures for using the bootstrap to reduce the bias of the OMD estimator and to obtain improved approximations to the coverage probabilities of confidence intervals based on OMD estimates.

a. The EWMD and OMD estimators

Let X be a $\ell \times 1$ random variable with mean μ and covariance matrix V. In general, V has $\ell(\ell + 1)/2$ unique elements. However, some of these elements may be assumed to equal zero *a priori*. Let $q \le \ell(\ell + 1)/2$ be the number of unique elements of V that are not assumed to equal zero. In the applications of interest in this paper, further *a priori* restrictions on the covariance structure of X reduce the number of unique, non-zero elements of V to $r < q$. These restrictions set some elements of V equal to others but do not specify their numerical values. For example, the ℓ components of X may be assumed to have equal but unspecified variances. Let θ denote the $r \times 1$ vector of unique elements of V under the *a priori* restrictions that are imposed, excluding any elements of V that are assumed to equal zero. The statistical problem of interest in this section is to obtain point and interval estimates of θ.

To obtain the EWMD and OMD estimators, let $\{X_i: i = 1,\ldots,n\}$ be a random sample of X. Let V_n denote the sample covariance matrix:

$$V_n = (n - 1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})', \tag{3.1}$$

where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Let S be a $q \times 1$ vector containing the unique elements of V that are not *a priori* zero. Let e be the $q \times r$ matrix of zeros and ones that maps elements of θ into elements of S. Then $e\theta = S$. For example, suppose the data consist of observations of a scalar random variable X in each two time periods. Then the unique elements of V are the variances of X in each of the periods, V_{11} and V_{22}, and the covariance of X across time periods, V_{12}. Let $S = (V_{11}, V_{22}, V_{12})'$. If the data generation process is stationary, then $V_{11} = V_{22}$ and θ is a 2×1 matrix. Let $\theta_1 = V_{11} = V_{22}$ and $\theta_2 = V_{12}$. Then

$$e = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Now let S_n be the vector that is obtained from S by replacing the elements of V with the corresponding elements of V_n. The EWMD and OMD estimators of θ have the form

$$\theta_n = (e'\Omega_n e)^{-1} e'\Omega_n S_n, \tag{3.2}$$

where Ω_n is a q×q, positive-definite matrix. The EWMD estimator, $\theta_{n,EWMD}$, is obtained by setting $\Omega_n = I_{q\times q}$, the q×q identity matrix. The OMD estimator, $\theta_{n,OMD}$, is obtained by setting Ω_n equal to the inverse of an estimator of the population covariance matrix of S_n. One such estimator is $\Omega_n = \Sigma_n^{-1}$, where Σ_n is a matrix with elements of the form

$$n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^{(r)}(X_i - \overline{X})^{(s)}(X_i - \overline{X})^{(t)}(X_i - \overline{X})^{(u)} \tag{3.3}$$

$$- \left[n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^{(r)}(X_i - \overline{X})^{(s)}\right]\left[n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^{(t)}(X_i - \overline{X})^{(u)}\right],$$

$(X_i - \overline{X})^{(r)}$ denotes the r'th element of $X_i - \overline{X}$, and (r,s) and (t,u) are the indices of non-zero elements of V.

The EWMD estimator is unbiased, though inefficient, because S_n is an unbiased estimator of the unique, nonzero elements of V. The OMD estimator is biased in finite samples, though it is asymptotically unbiased and efficient, because the OMD version of Ω_n is a random variable that is not independent of S_n. The bias of the OMD estimator is a source of finite-sample (but not asymptotic) inefficiency. The finite-sample efficiency of the OMD estimator could be increased if its bias could be reduced without greatly increasing its variance. Section 3c gives the bootstrap procedure for accomplishing this.

The EWMD and OMD estimators are both asymptotically normally distributed with means of zero and covariance matrices that can be estimated consistently. Accordingly, each can be used to construct asymptotic confidence intervals for components of θ. However, Monte Carlo evidence presented by Altonji and Segal (1994, 1996) shows that the true coverage probabilities of asymptotic confidence intervals based on the OMD estimator often are far below the nominal coverage probabilities. Section 3c gives the bootstrap procedure for alleviating this problem.

Before presenting the procedures for reducing bias and errors in coverage probabilities, it is necessary to explain how to construct a suitable bootstrap analog of $\theta_{n,OMD}$.

b. The Bootstrap Sampling and Estimation Procedures

Section 2 outlined the theory underlying the bootstrap's ability to reduce the bias of an estimator and provide an improved approximation to the coverage probability of a confidence interval. This section explains how to form a bootstrap analog of $\theta_{n,OMD}$ to which the theory applies. As was discussed in Section 2, a bootstrap sample $\{X_i^*: i = 1,..., n\}$ is generated by sampling the estimation data randomly with replacement. The most obvious way to obtain a bootstrap version of $\theta_{n,OMD}$ is to replace $\{X_i\}$ with $\{X_i^*\}$ in (3.1)-(3.3). This method is unsatisfactory, however.

To see why, observe that the population value of θ, θ_0, is identified by the moment condition $E(S_n - e\theta) = 0$, which holds when $\theta = \theta_0$ but not otherwise. In bootstrap sampling, the population expectation E is replaced by E*, the expectation with respect to the empirical distribution of the data. The population parameter value θ_0 is replaced by $\theta_{n,OMD}$. S_n is replaced by S_n^*, which is obtained from S_n by replacing X_i and $\bar{X}$ with X_i^* and $\bar{X}^* = n^{-1}\sum_{i=1}^{n} X_i^*$ in V_n. It is not difficult to show that $E^*S_n^* = [(n - 1)/n]S_n$. Therefore, the bootstrap analog of the population moment condition that identifies θ is $E^*(S_n^* - e\theta) = [(n - 1)/n]S_n - e\theta = 0$. It is clear from (3.2), however, that $[(n - 1)/n]S_n -$

$e\theta_{n,OMD} \neq 0$ except, possibly, in special cases. Therefore, replacing $\{X_i\}$ with $\{X_i^*\}$ in (3.1)-(3.3) yields a bootstrap estimator of θ that is based on a moment condition that does not hold in the population that the bootstrap samples. As a result, this version of the bootstrap does not reduce bias or differences between the true and nominal coverage probabilities of confidence intervals.

This problem can be solved by recentering the bootstrap moment condition so that it holds at $\theta = \theta_{n,OMD}$. See Brown and Newey (1995) for an alternative approach based on empirical likelihood. The recentered moment condition is

$$E^*(S_n^* - e\theta - R_n) = 0, \tag{3.4}$$

where $R_n = [(n - 1)/n]S_n - e\theta_{n,OMD}$. OMD estimation of θ based on (3.4) yields the following bootstrap OMD estimator:

$$\theta_{n,OMD}^* = (e'\Omega_n^*e)^{-1}e'\Omega_n^*(S_n^* - R_n), \tag{3.5}$$

where Ω_n^* is obtained from (3.3) by replacing X_i with X_i^* and $\bar{X}$ with $\bar{X}^*$. The theory of the bootstrap and the procedures for implementing it that were described in Section 2 apply to the OMD estimator when $\theta_{n,OMD}^*$ is used as the bootstrap analog of $\theta_{n,OMD}$.

c. Implementation of the Bootstrap in OMD Estimation

This section presents step-by-step procedures for implementing the bootstrap with the OMD estimator.

First consider bias reduction. Let θ_j be the j'th component of θ. Let $\theta_{nj,OMD}$ and $\theta_{nj,OMD}^*$, respectively, be the j'th components of $\theta_{n,OMD}$ and $\theta_{n,OMD}^*$. Define $B_j^* = E^*(\theta_{nj,OMD}^* - \theta_{nj,OMD})$, where E^* denotes the mean under bootstrap sampling conditional on the estimation data. The bootstrap reduced-bias OMD estimator of θ_j is $\theta_{nj,OMD} - B_j^*$. B_j^* can be computed with arbitrary accuracy by using the Monte Carlo procedure described in Section 2b. For the specific case of OMD estimation, this procedure is:

OMD1. Use the estimation data to compute $\theta_{nj,OMD}$.

OMD2. Generate a bootstrap sample of size n by sampling the estimation data randomly with replacement. Compute $\theta_{nj,OMD}*$ by applying (3.5) to the bootstrap sample.

OMD3. Compute $E*\theta_{nj,OMD}*$ by averaging the results of many repetitions of step OMD2. Set $B_j* = E*\theta_{nj,OMD}* - \theta_{nj,OMD}$.

Now consider a confidence interval for θ_j. A bootstrap symmetrical confidence interval is obtained by replacing $z_{\alpha/2}$ with $t_{n\alpha}*$ in (2.5). The Monte Carlo procedure for estimating $t_{n\alpha}*$ is:

OMD4. Use the estimation data to compute $\theta_{nj,OMD}$.

OMD5. Generate a bootstrap sample of size n by sampling the estimation data randomly with replacement. Compute $T_{nj}*$ from this sample. $T_{nj}*$ is computed the same way as T_n in Section 2c but with $\theta_{nj,OMD}$, $\theta_{nj,OMD}*$, and the OMD variance estimator in place of θ_n, θ_n*, and $V_{n\theta}$.

OMD6. Use the results of many repetitions of OMD5 to compute the EDF of $|T_{nj}*|$. Set $t_{n\alpha}*$ equal to the $1 - \alpha$ quantile of this EDF.

4. MONTE CARLO EXPERIMENTS

This section reports the results of a Monte Carlo investigation of the numerical accuracy of the bootstrap as a means of reducing the bias of the OMD estimator and the differences between the true and nominal coverage probabilities of confidence intervals based on this estimator.

The theory outlined in Section 2 shows that the bootstrap can be used to improve the finite-sample performance of the OMD estimator. However, the bootstrap provides only approximations to the finite-sample bias and distribution of an estimator. Although these approximations are more accurate than those of asymptotic theory (in the sense that the errors of the bootstrap approximations converge to zero more rapidly as $n \to \infty$), nothing in the theory guarantees that the bootstrap will provide high numerical accuracy with a sample of fixed size. The Monte Carlo results provide information about numerical accuracy.

The data-generation processes used in the experiments are taken from the correlated-moments processes of Altonji and Segal (1994). These are used instead of the independent-moments processes of Altonji and Segal (1994, 1996) because $\theta_{n,OMD}$ is asymptotically more efficient than $\theta_{n,EWMD}$ with the correlated-moments processes but not the independent-moments ones. In each experiment, X has 10 components, and n = 500. The j'th component of X, $X^{(j)}$ (j = 1,...,10), is generated by $X^{(j)} = (Z^{(j)} + \rho Z^{(j+1)})/(1 + \rho^2)^{1/2}$, where $Z^{(1)},\ldots, Z^{(11)}$ are iid random variables with means of zero and variances of 1, and $\rho = 0.5$. The Z's are sampled from five different distributions, depending on the experiment. These are the uniform, normal, Student t with 10 degrees of freedom, exponential, and lognormal, all standardized to have means of zero and variances of one. As in Altonji and Segal (1994), it is assumed that ρ is known and that the components of X are known to be identically distributed and to follow MA(1) processes. The estimation problem is to infer the scalar parameter θ that is identified by the moment conditions $\mathrm{Var}(X^{(j)}) = \theta$ (j = 1,..., 10) and $\mathrm{Cov}(X^{(j)}, X^{(j-1)}) = \rho\theta/(1 + \rho^2)$ (j = 2,..., 10). Accordingly, S_n is a vector containing the 10 sample variances and 9 sample covariances of $X^{(1)},\ldots, X^{(10)}$.

Experiments were carried out using both the EWMD and OMD estimators. Since the EWMD estimator is unbiased, the experiments with it consisted of computing its empirical RMSE and the empirical coverage probability of a nominal 95% confidence interval for θ based on the asymptotic critical value of the t statistic. In the experiments with the OMD estimator, the empirical bias, RMSE, and coverage probability of a nominal 95% confidence interval based on the asymptotic critical value were computed. In addition, the empirical bias and RMSE of the OMD estimator following bootstrap bias reduction were computed as was the empirical coverage probability of a nominal 95% confidence interval based on the bootstrap critical value. The experiments with bias reduction followed a modified version of steps MC1-MC3 in Section 2, and the experiments with confidence intervals followed a modified version of steps MC4-MC6. There were three modifications in each case. First, the data were generated by sampling the distributions described in this section rather than the N(0,6) distribution. Second, there was no bootstrapping in the experiments with the EWMD estimator. Third, bootstrapping was carried out by applying steps OMD1-

OMD3 and OMD4–OMD6 in Section 3c rather than steps B1–B3 and B4–B6. There were 1000 Monte Carlo replications per experiment. Bootstrap bias corrections and critical values were computed using 500 bootstrap resamples at each replication.

The results of the experiments are shown in Table 1. $\theta_{n,OMD}$ is biased and its RMSE exceeds that of $\theta_{n,EWMD}$ for all distributions of Z except the uniform. Moreover, the coverage probabilities of confidence intervals based on $\theta_{n,OMD}$ with asymptotic critical values are far below the nominal value of 0.95 except in the experiment with uniform Z's. In contrast, the coverage probabilities of confidence intervals based on $\theta_{n,EWMD}$ with asymptotic critical values are close to nominal in all of the experiments except the one with lognormal Z's. In the lognormal experiment, the error in coverage probability is much smaller with $\theta_{n,EWMD}$ than with $\theta_{n,OMD}$. These results are consistent with those of Altonji and Segal (1994, 1996).

Table 1 also shows that bootstrap bias reduction greatly reduces both the bias and RMSE of $\theta_{n,OMD}$. In addition, the use of bootstrap critical values greatly reduces the errors in the coverage probabilities of confidence intervals based on $\theta_{n,OMD}$. In the experiments with normal, Student t, or uniform Z's, the bootstrap essentially eliminates the bias of $\theta_{n,OMD}$ and the errors in the coverage probabilities of the confidence intervals. Moreover, the RMSE of the bias-corrected $\theta_{n,OMD}$ in these experiments is 12–50% less than that of $\theta_{n,EWMD}$. In the experiments with exponential or lognormal Z's, the bootstrap reduces but does not eliminate the bias of $\theta_{n,OMD}$ and the errors in the coverage probabilities of the confidence intervals. The bootstrap also reduces the RMSE of $\theta_{n,OMD}$. However, with exponential or lognormal Z's, the RMSE and error in coverage probability are larger for $\theta_{n,OMD}$ with bootstrap bias reduction and bootstrap critical values than for $\theta_{n,EWMD}$.

The relatively poor performance of the bootstrap in the experiments with exponential or lognormal Z's is due to the poor precision of Σ_n in (3.3) as an estimator of the covariance matrix of S_n and, therefore, of Ω_n as an estimator of the inverse covariance matrix, when Z is exponentially or lognormally distributed. This problem is illustrated by the coefficient of variation of

$e'\Omega_n e$ in the experiments. It is 0.066, 0.083, 0.10, 0.16, and 0.35 in the experiments with uniform, normal, Student t, exponential, and lognormal Z's, respectively. Section 5 describes an estimator of the covariance matrix of S that is more precise in finite samples. The numerical performance of the bootstrap in the experiments with exponential and lognormal Z's is greatly improved when this estimator is used in place of Σ_n.

5. AN IMPROVED COVARIANCE-MATRIX ESTIMATOR

a. Description of the Estimator

The exponential and lognormal distributions both have high kurtosis (6 and 111, respectively, compared to −1.2, 0 and 1 for the uniform, normal, and Student t distributions). In the experiments based on exponential and lognormal distributions, Σ_n is an imprecise estimator of the covariance matrix of S_n because it is strongly influenced by outlier values of X. Koenker, *et al.* (1994) provide a theoretical analysis of this problem. The problem can be alleviated by trimming observations of X to reduce the influence of outliers.

To describe the resulting covariance-matrix estimator, let $\{a_n\}$ be a sequence of positive real numbers that increases to ∞ at a suitable rate as $n \to \infty$. The required rate of increase is discussed in Section 5b. Let $X_i^{(j)}$ denote the j'th component of the i'th observation of X, and let $\overline{X}^{(j)}$ be the sample average of the j'th component. That is, $\overline{X}^{(j)} = n^{-1}\sum_{i=1}^{n} X_i^{(j)}$. Define

$$\delta_i(a_n) = \begin{cases} 1 \text{ if } \sup_j |X_i^{(j)} - \overline{X}^{(j)}| \le a_n \\ 0 \text{ otherwise} \end{cases}$$

and

$$\tilde{n} = \sum_{i=1}^{n} \delta_i(a_n).$$

The trimmed covariance estimator is the version of Σ_n that is obtained by applying (3.3) to observations X_i for which $\delta_i(a_n) = 1$. Denote this estimator by Σ_n. Σ_n consists of terms of the form

$$n^{-1}\sum_{i=1}^{n}\delta_i(a_n)(X_i - \bar{X})^{(r)}(X_i - \bar{X})^{(s)}(X_i - \bar{X})^{(t)}(X_i - \bar{X})^{(u)}$$

$$- \left[n^{-1}\sum_{i=1}^{n}\delta_i(a_n)(X_i - \bar{X})^{(r)}(X_i - \bar{X})^{(s)}\right]\left[n^{-1}\sum_{i=1}^{n}\delta_i(a_n)(X_i - \bar{X})^{(t)}(X_i - \bar{X})^{(u)}\right].$$

Let $\theta_{n,OMD}$ denote the OMD estimator that is obtained by using Σ_n instead of Σ_n to estimate the covariance matrix of S_n. $\theta_{n,OMD}$ is asymptotically equivalent to $\theta_{n,OMD}$ but has better finite-sample properties. The arguments leading to this conclusion are outlined in the Appendix.

The usefulness of the trimmed covariance-matrix estimator is illustrated in Table 2, which shows the results of the experiments with exponential and lognormal Z's when a_n is selected by Monte Carlo to minimize the RMSE of the estimator. There were no trimming experiments with uniform, normal, or Student-t Z's because Table 1 shows that good results can be obtained with these distributions by making a_n sufficiently large. With exponential Z's, the estimator obtained using $\theta_{n,OMD}$, bootstrap bias reduction, and bootstrap critical values, is essentially unbiased and its RMSE is the same as that of $\theta_{n,EWMD}$. The error in the coverage probability of a nominal 95% confidence interval is not statistically significantly different from zero. With lognormal Z's, the RMSE of the estimator based on $\theta_{n,OMD}$ is 9% below that of $\theta_{n,EWMD}$. The coverage probability of the OMD-based confidence interval is close to the nominal value and closer to nominal than the coverage probability of the EWMD-based interval. Thus, the bootstrap combined with the improved covariance-matrix estimator largely eliminates the poor finite-sample performance of the OMD estimator.

The values of a_n in Table 2 are in the range 2-2.5. Since X has a standard deviation of 1, only observations that are 2-2.5 standard deviations from the mean are trimmed in the computation of Σ_n. Nonetheless, trimming eliminates 21% of the observations from the computation of Σ_n when Z is exponentially distributed and 27% when Z is lognormally distributed. This reflects the relatively high frequency of outliers with these high-kurtosis distributions.

Of course, the Monte Carlo method used to choose a_n in the experiments is not available in applications. However, it is possible to construct a Monte

Carlo procedure that mimics the one used in the experiments and is feasible in applications. This procedure is described in Section 5b.

b. Choosing a_n in Applications

This section describes an empirical method for selecting the trimming parameter a_n in Σ_n. The discussion is informal and the arguments heuristic.

The procedure consists of using Monte Carlo simulation to compute a_m for some $m < n$ and then multiplying a_m by a scale factor to obtain a_n. Suppose for the moment that θ is scalar. Then a_m is obtained by solving the problem

$$\underset{a}{\text{minimize}} \ |B_m(a)|, \tag{5.1}$$

where $B_m(a)$ is obtained as follows:

A1. Use the full estimation data set to compute $\theta_{n,EWMD}$.

A2. Generate m observations of X by sampling the estimation data randomly <u>without replacement</u>. Compute $\theta_{m,OMD}$ using the m observations.

A3. Average the values of $\theta_{m,OMD}$ that are obtained by repeating step A2 many times. Denote the resulting average by $E_m(\theta_{m,OMD})$

A4. Set $B_m(a) = E_m(\theta_{m,OMD}) - \theta_{n,OMD}$.

To understand the rationale for this procedure, note that each sample obtained in step A2 is a random sample of size m from the true but unknown process that generated the estimation data. Therefore, for the determination of a_m, the only differences between steps A1–A4 and the infeasible Monte Carlo procedure used in Section 5a are that (1) A1–A4 chooses a_m to minimize the bias of $\theta_{m,OMD}$ relative to $\theta_{n,EWMD}$ rather than relative to the true but unknown population value of θ, and (b) E_m is an average over finitely many, non-independent samples of size m. Accordingly, a_m reflects random sampling error in $\theta_{n,EWMD}$ and in E_m. However, the sampling error in $\theta_{n,EWMD}$ becomes negligible as $n \to \infty$, and sampling error in E_m becomes negligible as $m \to \infty$ in such a way that $(m/n) \to 0$. Thus, the procedure for selecting a_m is justified asymptotically. A similar argument would hold if $\theta_{n,OMD}$ were used instead of $\theta_{n,EWMD}$. The random sampling error of the computed a_m would be larger, however, because $\theta_{n,OMD}$ has a larger finite-sample RMSE than does $\theta_{n,EWMD}$.

If θ is a vector, so is $B_m(a)$, and (5.1) can be replaced by minimization of the distance of $B_m(a)$ from zero in the desired metric.

It remains to determine how a_m should be scaled to yield an estimate of a_n. The Appendix presents an argument showing that the rate of increase of a_n as $n \to \infty$ is less than $n^{1/4}$. Therefore,

$$a_m \leq a_n < (n/m)^{1/4} a_m \tag{5.2}$$

Choosing m requires balancing two considerations. If m is too large, the random sampling error in a_m is large, whereas if m is too small, the asymptotic approximations that justify (5.2) are inaccurate. I have carried out Monte Carlo experiments based on the exponential and lognormal designs of Section 4 with n = 1500 (approximately the sample size used in the empirical example in Section 6) and m = 500. The resulting average values of a_m were in the range 2.5–3 with standard deviations of 0.5–0.6. Thus, the values of a_m obtained from steps A1–A4 are close to those obtained with the infeasible Monte Carlo procedure of Section 5a. With n/m = 3, (5.2) yields $a_m \leq a_n < 1.32a_m$. This range of uncertainty in a_n is acceptable in the empirical example of Section 6. There is little variation in the results as a_n varies over this range.

6. AN EMPIRICAL EXAMPLE

This section illustrates the bootstrap methods discussed in Sections 2–5 by using them to estimate the covariance structure of year-to-year changes in the logarithms of annual earnings and working hours of male heads of households in the Panel Study of Income Dynamics (PSID). The data are those used by Altonji and Segal (1994, 1996) and are similar to those of Abowd and Card (1987, 1989), who also used data from the PSID. The data are based on observations of annual earnings and hours worked by 1536 individuals over an 11-year period (1969–1979). The logarithms of earnings and hours have been adjusted for labor-market experience and year effects, and dollar values have been adjusted to 1967 levels using the Consumer Price Index. See Altonji and Segal (1994, 1996) for further details on the preparation of the data.

The data provide 10 observations per individual on year-to-year changes in the logarithms of earnings and hours. Accordingly, the matrix V in (3.1) contains 210 separate moments for the various years and lags. However, Abowd and Card (1987, 1989) found that the covariances of observations separated by more than two years are negligible. Accordingly, the estimates reported here are based on a vector S consisting of q = 98 variances and covariances with time lags of up to 2 years. The estimated model is stationary, so θ consists of r = 11 parameters. These are the hours variance, the earnings variance, 2 hours autocovariances, 2 earnings autocovariances, and 5 hours/earnings covariances.

Point estimates and confidence intervals for θ were obtained using EWMD and OMD with asymptotic critical values, OMD with bootstrap bias reduction and bootstrap critical values, and the modified OMD estimator of Section 5 with bootstrap bias reduction and bootstrap critical values. Five hundred bootstrap replications were used to compute bootstrap estimates of bias and critical values. The trimming parameter a_n required for the modified OMD estimator was computed using the procedure described in Section 5b with m = 500. This yielded $a_m = 0.5$, so $0.5 \leq a_n < 0.66$ by (5.2). The sample standard deviations of the changes in the logarithms of earnings and hours are in the range 0.28–0.47. Thus, in the computation of Σ_n, an observation is trimmed away if it is 1.06–2.36 standard deviations from the mean. The exact trimming threshold in standard deviation units depends on the year, whether the observation is of earnings or hours, and the value of a_n within the range 0.5–0.66.

The estimation results are shown in Table 3. The EWMD and unbootstrapped OMD estimates (columns 1 and 2) are similar to those of Altonji and Segal (1994, 1996) but not identical because Altonji and Segal used all available lags in estimation. As in Altonji and Segal (1994, 1996), the EWMD and OMD point estimates are very different. Moreover, the half-widths of the OMD confidence intervals are only about half those of the EWMD intervals, which suggests that the OMD estimates are much more precise than the EWMD estimates. In fact, the OMD parameter estimates and the half-widths of the OMD confidence intervals are misleading. In a Monte Carlo experiment in which the estimation data were sampled with replacement, Altonji and Segal found that the OMD estimator is badly biased and that the true coverage probabilities of confidence intervals based on asymptotic critical values are far below the nominal probabilities.

The results of OMD estimation with bootstrap bias reduction and critical values but not trimming are shown in column 3 of Table 3. The OMD parameter estimates with bias reduction are closer to the EWMD estimates than are the OMD estimates without bias reduction. However, the OMD estimates with bias reduction are closer to those without bias reduction than to the EWMD estimates. This suggests that the bootstrap has reduced but not eliminated the bias of the OMD estimates. This finding is not surprising. The sample kurtoses of the changes in the logarithms of earnings and hours are 17-47. The Monte Carlo results discussed in Section 4 indicate that the bootstrap, by itself, cannot eliminate bias in the presence of such high kurtosis. The half widths of the OMD-based confidence intervals are much larger with bootstrap critical values than with asymptotic ones. However, the high kurtosis of the estimation data and the bootstrap's inability to remove the bias of the estimates suggest that the coverage probabilities of these intervals may be below the nominal 0.95.

Columns 4 and 5 of Table 3 show the results of OMD estimation with the modified weight matrix and bootstrap bias reduction and critical values. Estimates are shown for values of a_n at the lower and upper end of the range given by (5.2). There is little difference between the estimates obtained with the two values of a_n. Moreover, both sets of values are close to the EWMD estimates. Most of the differences between the EWMD and modified OMD estimates are less than the half-widths of the confidence intervals of either. Moreover, most of the modified OMD estimates are much closer to the EWMD estimates than to the unmodified OMD estimates with or without bootstrap bias reduction. The half-widths of the confidence intervals based on the modified OMD estimator with bootstrap critical values are slightly larger than the half-widths based on the EWMD estimator. The estimation results do not reveal the reason for this. One possibility is that the finite-sample efficiency of OMD is less than that of EWMD with these data. Another possibility is that the EWMD confidence intervals with asymptotic critical values are too narrow. This interpretation is suggested by the Monte Carlo results of Section 4, in which the true coverage probability of the EWMD confidence interval is less than the nominal probability when the kurtosis of the population distribution is high. Altonji and Segal (1994, 1996) obtained a similar results in the experiments in which they resampled the estimation data randomly with replacement.

In summary, the estimation results suggest that with the PSID data, there is no great advantage to using OMD instead EWMD. The results do suggest, however, that using the modified OMD weight matrix together with bootstrap bias reduction and critical values overcomes the poor finite-sample performance of the standard OMD estimator. This finding is consistent with the Monte Carlo results reported in Sections 4 and 5.

6. CONCLUSIONS

The OMD estimator of covariance structures is asymptotically efficient, but its finite-sample performance is poor. With samples of the sizes encountered in applications, it is often badly biased, its RMSE can be much larger than that of the asymptotically inefficient EWMD estimator, and there can be large differences between the true and nominal coverage probabilities of confidence intervals.

This paper has demonstrated the ability of the bootstrap to reduce the finite-sample bias and RMSE of the OMD estimator and to reduce the differences between the true and nominal coverage probabilities of confidence intervals based on the OMD estimator. In numerical experiments in which the sampled population had low kurtosis, the bootstrap essentially eliminated the finite-sample bias of the OMD estimator and the errors in the coverage probabilities of confidence intervals. Moreover, the RMSE of the OMD estimator after bootstrap bias reduction was below that of the EWMD estimator. Similar results were obtained in experiments with high-kurtosis populations when the estimator of the OMD weight matrix was modified to reduce the influence of outlier observations. Thus, in the data-generation processes used in the experiments, the bootstrap combined with the modified weight-matrix estimator overcomes the poor finite-sample performance of the OMD estimator.

These results are subject to the usual qualifications about attempting to generalize from numerical evidence. Certainly, there is no guarantee that the OMD estimator with bootstrapping and the modified weight-matrix estimator will always perform as well as they did in the experiments described here. There is also no guarantee that bootstrapping and weight-matrix modification will always produce an OMD estimator that is more efficient in finite samples than the EWMD estimator. The results presented here do show, however, that there are situations in which OMD estimation with bootstrapping is considerably more efficient in finite samples than EWMD estimation.

Altonji and Segal (1994, 1996) concluded that EWMD estimation is almost always preferable to standard OMD estimation with samples of the sizes found in applications. Nothing in this paper contradicts that conclusion. This paper has shown, however, that there are situations in which OMD estimation with bootstrapping may be preferable to EWMD.

APPENDIX

This appendix outlines the arguments leading to (5.2). To minimize the complexity of the discussion, it is assumed that the components of X are known by the analyst to be independently distributed with means of zero and common variance θ. Only the scalar parameter θ has to be estimated, and the OMD weight matrix is diagonal. The arguments for the general case are similar, but the notation and algebra are much more complex.

Under the foregoing assumptions, the j'th component of the q-vector S is

$$S^{(j)} = n^{-1}\sum_{i=1}^{n}(X_i^{(j)})^2 .$$

Ω_n is the q×q diagonal matrix whose (j,j) component is Σ_{nj}^{-1}, where $\Sigma_{nj} = \beta_{nj} - \gamma_{nj}^2$,

$$\beta_{nj} = n^{-1}\sum_{i=1}^{n}\delta_i(a_n)(X_i^{(j)})^4 ,$$

$$\gamma_{nj} = n^{-1}\sum_{i=1}^{n}\delta_i(a_n)(X_i^{(j)})^2,$$

$\delta_i(a_n) = I(\sup_j |X_i^{(j)}| \leq a_n)$, and $I(\bullet)$ is the indicator function. The modified OMD estimator of θ satisfies

$$\theta_{n,OMD} - \theta_0 = \left[\sum_{j=1}^{q}(\beta_{nj} - \gamma_{nj}^2)^{-1}\right]^{-1}\sum_{j=1}^{q}(\beta_{nj} - \gamma_{nj}^2)^{-1}(S^{(j)} - \theta_0). \quad \text{(A1)}$$

Assume that the components of X have moments through order 8. Suppose for the moment that $a_n = \infty$ for all n. Then, it follows from the theory of Edgeworth expansions (see, e.g., Hall, 1992) that the bias and variance of $n^{1/2}\theta_{n,OMD}$ have asymptotic expansions of the forms $b/n^{1/2} + O(n^{-3/2})$ and $v_0 + v_1/n + O(n^{-2})$, respectively, where b, v_0 and v_1 are constants. Therefore, the mean-square error of $n^{1/2}\theta_{n,OMD}$ has the expansion

$$MSE = v_0 + (v_1 + b^2)/n + O(n^{-2}). \quad \text{(A2)}$$

Making a_n a finite, increasing function of n adds an extra term to (A2). A Taylor series approximation to (A1) shows that the extra term has the form

$$\Delta(a_n) = \sum_{j=1}^{q} c_j\beta_{nj}^*,$$

where c_j (j = 1,...,q) is a constant and

$$\beta_{nj}^* = E\{(X_1^{(j)})^4[1 - \delta_1(a_n)]\}.$$

Let $f_j(\bullet)$ denote the probability density function of $X^{(j)}$, and suppose that $f_j(x) \sim |x|^{-d}$ as $|x| \to \infty$ for some $d > 5$. Then $\beta_{nj}^* = O(a_n^{5-d})$ as $n \to \infty$.

The procedure of Section 5 aims at selecting a_n to minimize the $O(n^{-1})$ term in the asymptotic expansion (A2). This requires $\Delta(a_n) = O(n^{-1})$ and, therefore, $\beta_{nj}^* = O(n^{-1})$, which happens if $a_n \sim n^{1/(d-5)}$ as $n \to \infty$. Existence of the expansion (A2) requires the components of X to have moments through order 8, so $d > 9$. Accordingly, the rate of increase of a_n as $n \to \infty$ is slower than $n^{1/4}$. If $f_j(x)$ decreases at an exponential or faster rate as $|x| \to \infty$, then a similar argument shows that a_n has a logarithmic rate of increase as $n \to \infty$. The rate of increase of a_n is still slower than $n^{1/4}$.

TABLE 1: RESULTS OF MONTE CARLO EXPERIMENTS[1]

	EWMD Estimator		OMD Estimator without Bootstrap			OMD Estimator with Bootstrap		
Distribution	RMSE	Coverage Prob. of Nominal 95% Conf. Interval with Asymptotic Critical Value	Bias	RMSE	Coverage Prob. of Nominal 95% Conf. Interval with Asymptotic Critical Value	Bias	RMSE	Coverage Prob. of Nominal 95% Conf. Interval with Bootstrap Critical Value
Uniform	0.019	0.96	0.005	0.015	0.93	0.002	0.014	0.96
Normal	0.024	0.96	0.016	0.025	0.85	0.0	0.021	0.95
Student t	0.029	0.94	0.024	0.034	0.79	0.002	0.026	0.95
Exponential	0.042	0.95	0.061	0.073	0.54	0.014	0.048	0.91
Lognormal	0.138	0.86	0.274	0.285	0.03	0.136	0.173	0.76

[1] Based on 1000 replications. Empirical coverage probabilities between 0.94 and 0.96 are not statistically significantly different from the nominal probability at the 0.05 level.

TABLE 2: RESULTS OF MONTE CARLO EXPERIMENTS WITH TRIMMED COVARIANCE ESTIMATOR[1]

			OMD Estimator with Bootstrap		
Distribution	a_n	$\tilde{n}/n$	Bias	RMSE	Coverage Prob. of Nominal 95% Conf. Interval with Bootstrap Critical Value
Exponential	2.5	0.78	0.004	0.042	0.96
Lognormal	2.0	0.73	0.046	0.126	0.91

[1] Based on 1000 replications. Empirical coverage probabilities between 0.94 and 0.96 are not statistically significantly different from the nominal probability at the 0.05 level.

TABLE 3: ESTIMATES OF COVARIANCE STRUCTURE OF CHANGES IN LOGARITHMS OF EARNINGS AND HOURS

Quantities in Parentheses are Half-Widths of Nominal 95% Symmetrical Confidence Intervals

Covariance Parameter	EWMD with Asymp. Crit. Vals.	OMD with Asymp. Crit. Vals.	OMD with Bootstrap Bias Reduction and Crit. Vals.	Modified OMD with Bootstrap Bias Reduction and Crit. Vals. $a_n = 0.5$	$a_n = 0.66$
	(1)	(2)	(3)	(4)	(5)
E(t), E(t)	0.175 (0.020)	0.107 (0.011)	0.129 (0.050)	0.173 (0.025)	0.167 (0.024)
E(t), E(t−1)	−0.060 (0.011)	−0.035 (0.005)	−0.043 (0.021)	−0.056 (0.012)	−0.057 (0.011)
E(t), E(t−2)	−0.008 (0.005)	−0.009 (0.003)	−0.009 (0.006)	−0.010 (0.008)	−0.013 (0.008)
H(t), H(t)	0.131 (0.016)	0.072 (0.008)	0.088 (0.035)	0.118 (0.018)	0.114 (0.019)
H(t), H(t−1)	−0.047 (0.008)	−0.027 (0.004)	−0.032 (0.014)	−0.039 (0.009)	−0.040 (0.009)
H(t), H(t−2)	−0.006 (0.004)	−0.005 (0.002)	−0.006 (0.005)	−0.012 (0.007)	−0.010 (0.008)
E(t), H(t)	0.081 (0.014)	0.036 (0.007)	0.046 (0.029)	0.077 (0.016)	0.074 (0.015)
E(t), H(t−1)	−0.024 (0.008)	−0.008 (0.003)	−0.011 (0.011)	−0.021 (0.005)	−0.021 (0.008)
E(t), H(t−2)	−0.002 (0.002)	−0.004 (0.003)	−0.005 (0.005)	−0.001 (0.008)	−0.008 (0.008)
H(t), E(t−1)	−0.026 (0.008)	−0.012 (0.004)	−0.015 (0.029)	−0.023 (0.010)	−0.020 (0.009)
H(t), E(t−2)	−0.008 (0.004)	−0.006 (0.003)	−0.007 (0.006)	−0.015 (0.007)	−0.015 (0.008)

REFERENCES

Abowd, J.M. and D. Card (1987). Intertemporal Labor Supply and Long-Term Employment Contracts, *American Economic Review*, 77, 50-68.

Abowd, J.M. and D. Card (1989). On the Covariance Structure of Earnings and Hours Changes, *Econometrica*, 57, 411-445.

Altonji, J.G. and L.M. Segal (1994). Small Sample Bias in GMM Estimation of Covariance Structures, NBER Technical Working Paper no. 156, National Bureau of Economic Research, Cambridge, MA.

Altonji, J.G. and L.M. Segal (1996). Small Sample Bias in GMM Estimation of Covariance Structures, *Journal of Business and Economic Statistics*, 14, 353-366.

Behrman, J., M. Rosenzweig, and P. Taubman (1994). Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment, *Journal of Political Economy*, 102, 1131-1174.

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

Beran, R. and G.R. Ducharme (1991). *Asymptotic Theory for Bootstrap Methods in Statistics*, Les Publications CRM, Centre de recherches mathematiques, Universite de Montreal, Montreal, Canada.

Brown, B.W. and W.K. Newey (1995). Bootstrapping for GMM, unpublished manuscript, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA.

Griliches, Z. (1979). Sibling Models and Data in Economics: Beginnings of a Survey, *Journal of Political Economy*, 87, S37-S64.

Hall, P. (1986). On the Bootstrap and Confidence Intervals, *Annals of Statistics*, 14, 1431-1452.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, R.E. and F. Mishkin (1982). The Sensitivity of Consumption to Transitive Income: Estimates from Panel Data on Households, *Econometrica*, 50, 461-481.

Koenker, R., J.A.F. Machado, C.L. Skeels, and A.H. Welsh (1994). Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimation, *Econometric Theory*, 10, 172-197.