TECHNICAL WORKING PAPER SERIES

SMALL SAMPLE BIAS IN GMM ESTIMATION OF COVARIANCE STRUCTURES

Joseph G. Altonji
Lewis M. Segal

Technical Working Paper No. 156

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1994

Without implicating them we wish to thank Ian Domowitz, Art Goldberger, Bo Honoré, Costas Meghir, Bruce Meyer, Jörn-Steffen Pischke, Bent E. Sørensen, and seminar participants at Northwestern University (May 1991), a session of the Econometric Society Summer Meeting (June 1992) and the Federal Reserve Bank of Chicago (February 1994) for helpful comments on previous drafts of the paper. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #156
June 1994

SMALL SAMPLE BIAS IN GMM ESTIMATION OF COVARIANCE STRUCTURES

ABSTRACT

We examine the small sample properties of the GMM estimator for models of covariance structures, where the technique is often referred to as the optimal minimum distance (OMD) estimator. We present a variety of Monte Carlo experiments based on simulated data and on the data used by Abowd and Card (1987, 1990) in an examination of the covariance structure of hours and earnings changes. Our main finding is that OMD is seriously biased in small samples for many distributions and in relatively large samples for poorly behaved distributions. The bias is almost always downward in absolute value. It arises because sampling errors in the second moments are correlated with sampling errors in the weighting matrix used by OMD. Furthermore, OMD usually has a larger root mean square error and median absolute error than equally weighted minimum distance (EWMD).

We also propose and investigate an alternative estimator, which we call independently weighted optimal minimum distance (IWOMD). IWOMD is a split sample estimator using separate groups of observations to estimate the moments and the weights. IWOMD has identical large sample properties to the OMD estimator but is unbiased regardless of sample size. However, the Monte Carlo evidence indicates that IWOMD is usually dominated by EWMD.

Joseph G. Altonji
Department of Economics
Northwestern University
Evanston, IL 60208
and NBER

Lewis M. Segal
Economic Research, 11th Floor
Federal Reserve Bank of Chicago
230 S. La Salle Street
Chicago, IL 60604-1413

I. Introduction

Generalized method of moments estimators (GMM) have desirable asymptotic properties in many contexts but little is known about their small sample properties. In this paper we examine the small sample properties of GMM for models of covariance structures.[1] Sometimes models of covariance structures arise because a researcher is directly interested in the variances and covariances of an unobserved variable and has multiple measurements of the moments, possibly from different samples or from different years. Sometimes they arise because the researcher is estimating a model of the effects of a set of unobserved variables on the conditional mean of a set of observed variables. The model restricts the second moments of the data.[2] In these contexts, GMM minimizes the weighted distance between sample moments and the implied population moments, where the weighting matrix is the inverse of a consistent estimate of the covariance matrix of the sample moments. It is often referred to as the optimal minimum

---

[1] Malinvaud (1970), Chamberlain (1984) and especially Hansen (1982) are standard references on the large sample properties of GMM estimation in a variety of contexts but do not consider small sample properties. Tauchen (1986) examines the use of GMM to estimate parameters of nonlinear models of conditional means with endogenous variables in time series data. Horowitz and Neumann (1992) present a second order correction for bias in a GMM based test statistic of the proportional hazards assumption in duration analysis. They focus on problems related to Jensen's inequality that arise when one ignores estimation error in using weighted averages of nonlinear functions of estimated parameters (the model residuals in their case). Neither these studies nor a handful of other recent working papers referenced in Ogaki (forthcoming) raise the issue of bias from correlation between the moments under study and the weighting matrix, which is our focus. Very little research little theoretical research has been done on small sample approximations to GMM, particularly for models of second and higher order moments. Koenker, Machado, Skeels and Welsh (1994) provide a theoretical analysis of small sample properties of GMM in a somewhat different context than ours but focus on efficiency rather than bias. They suggest the use of robust estimation methods as an alternatives based on robust estimators rather than the least squares criteria that underlies conventional GMM. Arellano and Sargan (1992) discuss alternative small sample approximations to the distribution of estimators that may be written as functions of second moments, with references to earlier papers, but do not apply the approximations to the class of problems we consider.

[2] See for example, Abowd and Card's (1987) study of contract models of employment and earnings growth, Hall and Mishkin's (1982) and Altonji, Martins and Siow's studies of the permanent income hypothesis, Behrman, Rozenzweig and Taubman's (1992) study of the effects of individual endowments on earnings, own schooling, and spouse's schooling, or Griliches (1979) survey of sibling models.

distance estimator (OMD). Our study is motivated by the fact that several authors report difficulties in empirical applications based on OMD estimation of covariance models, including Abowd and Card (1989) and Altonji, Martins and Siow (1987).[3]

We present Monte Carlo evidence on the relative bias, variance, root mean squared error and median absolute error of the equally weighted minimum distance (EWMD) and OMD estimators for a set covariance models. To isolate the weighting procedure as the sole source of bias, we focus on linear models of covariance structures, although many of the issues carry over to nonlinear models of population moments. Our main finding is that OMD is seriously downward biased in absolute value in small samples for many distributions and in relatively large samples for poorly behaved distributions. The theoretical analysis of a specific model indicates that the bias is downward for most distributions in encountered in economics.[4] The bias arises because sampling errors in the second moments are correlated with sampling errors in the estimate of the covariance matrix of the sample moments. The latter is the weighting matrix for OMD. Furthermore, OMD usually has a larger root mean square error and median absolute error than EWMD, although the ranking depends upon the details of model and the characteristics of the data sample. Our general finding is that OMD outperforms EWMD in root mean squared error only in situations in which the root mean squared error of both estimators is small. This

[3] Finance is an area in which GMM has been widely adopted and researchers routinely work with models of higher moments. Lehmann (1990, p 84, 92) reports difficulties in using OMD to combine information from subsamples using an estimated weighting matrix. Shanken (1990) estimates models of the mean and variance of the return on portfolios by ordinary least squares. In his conclusion he notes that simultaneous estimation of the models for the mean return and variance offer advantages, but he reports in a footnote that "Attempts to incorporate the estimated residual variance relations in potentially more efficient WLS regression were unsuccessful and appear to induce spurious associations that I do not fully understand." Our analysis implies that serious biases might arise in applying feasible GLS to Shanken's model.

[4]We provide an example which is biased upwards in absolute value but it involves a symmetric distribution over finite support with most of the mass near the bounds of the support.

is true despite the fact that in a number of our experiments the theoretical advantage of OMD is quite large. By comparing feasible OMD to OMD based on the theoretically optimal weighting matrix, we show that estimating the weighting matrix typically involves a large increase in sampling variance and in root mean square error.

Our strongest evidence in favor of EWMD over OMD is based on the data used by Abowd and Card (1989, 1987). We begin with the "stationary model" of the growth in log earnings and log hours estimated by Abowd and Card and show that OMD leads to substantial underestimates (in absolute value) of population second moments. Since the change in hours and earnings probably are not covariance stationary, we also present a simulation in which we treat the Abowd and Card data as the population and use the moments of that population to define a model that is true by construction. We analyze the distribution of the various estimators by drawing samples (with replacement) from the Abowd and Card population and computing each estimator on the sample. The results strongly reinforce our main theme, which is that the OMD estimator of covariance structures suffers from serious downward bias in absolute value and is usually dominated by EWMD. The best choice for the Abowd and Card data is EWMD, which, incidentally, is what they chose.[5]

In addition to comparing EWMD and OMD we investigate an alternative estimator, which

---

[5] Another example suggesting that our concerns are empirically relevant is Schwert and Seguin (1990). They analyze the covariance in stock market returns in a micro econometric model relating the second moments of stock portfolio returns to a constant and a time varying measure of aggregate market volatility. They estimate the model individually for each second moment from five size ranked portfolios (5 variances and 10 covariances) by OLS. Then the OLS residuals are used in a Glejser style regression relating the square of the residuals to the regressors. Feasible weighted least squares is applied using the predicted values from the second stage regression. Ignoring the aggregate market volatility regressor, the Glesjer procedure is equivalent to estimating the variance of the sample second moments using the average squared deviation of the second moments from their average. Twenty nine of the thirty WLS coefficients are smaller in absolute value than the OLS coefficients even though both procedures yield consistent estimates. This is what our analysis predicts. However, the difference in parameter estimates is not large enough to effect the conclusions of the paper.

we call independently weighted optimal minimum distance (IWOMD). IWOMD is a split sample estimator that uses separate groups of observations to estimate the moments and the weights. The random partitioning of the data breaks the sampling covariance between the moments and the weights. Parameter estimates are computed separately for each partition, and then the estimates are averaged to form a final parameter estimate. IWOMD has identical large sample properties to the OMD estimator but is unbiased regardless of sample size. However, the Monte Carlo evidence indicates that IWOMD is usually dominated by EWMD.

The paper is organized as follows. In Section II we illustrate the class of models we are interested in with an example that underlies some of our simulations. We define the OMD and EWMD estimators for the problem, and present some initial Monte Carlo evidence indicating that OMD is downward biased. In section III we provide a theoretical discussion of the bias in OMD. The discussion points to an inverse relationship between the size of the bias and the precision in the second moments, which is strongly confirmed by the Monte Carlo evidence. In Section IV we present the IWOMD estimator. In section V we present more detailed Monte Carlo evidence on the performance of EWMD, OMD, and IWOMD. In section VI we present the empirical example based on the data used by Abowd and Card (1987, 1989). In section VII we consider statistical inference based on the usual asymptotic formula and on bootstrap methods. We close the paper with a brief summary and a research agenda.

II. Evidence Of Small Sample Bias in Optimal Minimum Distance

In the basic covariance model, multiple sample moments are combined into a single estimate of the population moments. Our examples focus on linear models relating sample variances (or covariances) to a single population parameter so that we isolate a particular source of small sample bias. One has data observations D_{pi}, where p=1,...,P indexes the variable used to compute sample moment p and i=1,...,N_p indexes the observations on that variable. For each variable p, the mean and variance are computed using the standard formulas

$$\bar{D}_p = \frac{1}{N_p} \sum_{i=1}^{N_p} D_{pi} \qquad (1)$$

$$m_p = \frac{1}{(N_p-1)} \sum_{i=1}^{N_p} (D_{pi}-\bar{D}_p)^2 \; ; \quad E(m_p) = \mu_p \qquad (2)$$

The second moment estimates are stacked into a (Px1) vector, *m*, and are related to a (Px1) vector of population moments, μ, through the model

$$m = \mu + \varepsilon = f(\theta) + \varepsilon . \qquad (3)$$

In (3), θ is the (Qx1) parameter vector one wishes to estimate and ε is a (Px1) vector of sampling errors. When f(θ) is linear in θ the model is

$$m = X\theta + \varepsilon . \qquad (4)$$

For example, suppose a researcher wishes to estimate a population variance from observations on a panel covering ten time periods. The vector *m* contains the ten estimates of the variance, one from each time period. The matrix X is a (10x1) vector of ones, and θ is the population variance (a scalar). Equally weighted minimum distance (EWMD) amounts to a least squares regression of *m* on X with the familiar solution

$$\theta_{EWMD} = \text{ArgMin}_\theta \, (m - f(\theta))'(m - f(\theta)) \qquad (5)$$

$$= (X'X)^{-1}(X'm) .$$

The EWMD estimator is not efficient if the elements of ε are heteroscedastic or correlated. Heteroscedasticity may arise as a result of unbalanced data (N_p may differ across p) or because the distributions of the D_{pi} are different.[6] The ε_p will be correlated if $Cov(D_{pi},D_{p'i})$ is not zero, which is likely in panel data applications. The variance of ε is not a scalar matrix in these cases. OMD takes this into account and is generalized least squares applied to (4). The covariance matrix of ε, Ω, is replaced with a consistent estimate obtained from the same data used to compute the sample second moments *m*. For example, a conventional estimator of the variance of m_p is

$$\hat{var}(m_p) = \frac{N_p^2}{(N_p-1)(N_p-2)^2}\left[\frac{1}{N_p}\sum_{i=1}^{N_p}(D_{pi} - \bar{D}_p)^4 - \left(\frac{1}{N_p}\sum_{i=1}^{N_p}(D_{pi} - \bar{D}_p)^2\right)^2\right] \qquad (6)$$

Comparable formulas exist for the covariance of second moment estimates. Let $\hat{\Omega}$ represent the estimated covariance matrix of *m*. The OMD estimator minimizes a quadratic form involving the sampling error in the moments and the estimated covariance matrix. In other words,

$$\theta_{OMD} = ArgMin_\theta \, (m - f(\theta))' \, \hat{\Omega}^{-1} \, (m - f(\theta)) \qquad (7)$$
$$= (X'\hat{\Omega}^{-1}X)^{-1} \, X'\hat{\Omega}^{-1}m$$

Malinvaud (1970), Chamberlain (1982,1984), and others show that under a relatively innocuous set of conditions $N^{1/2}(\theta_{OMD} - \theta) \rightarrow_D N(0, F' \Omega F)$ where F is $\partial f/\partial\theta$ evaluated at the true θ and that the OMD estimator is asymptotically efficient .

The replacement of Ω with a consistent estimate does not affect the asymptotic properties of θ_{OMD}, but the small sample distribution is another matter. Table 1 examines the EWMD and

[6] The higher moments of D_{pi} may differ across p even if the μ_p are the same across p. When the elements of $X'\theta$ differ it is unlikely that the higher moments will be the same.

OMD estimators of the ten variance model mentioned above based on several common distributions and several sample sizes. We consider the case in which the data D_{pi} are independent across i and p, which implies that the ten sample variances are independent. The distributions were chosen to include thick tailed symmetric distributions (student t(5) and student t(10)), long tailed asymmetric distributions (exponential and log normal), an asymmetric distribution (half normal), a bimodal distribution, and several "well behaved" distributions (normal and uniform).[7] All of the population distributions are scaled to have mean 0 and variance 1. For each distribution, a fixed number of observations are drawn at random and used to estimate the ten sample variances and their covariance. Both the diagonal and off diagonal elements of Ω are replaced with sample estimates, even though the elements of m are independent. This corresponds to the case in which the econometrician is unaware of the fact that the moments are independent. The EWMD and OMD estimates are computed and the entire process is replicated 1000 times.

The columns in the first panel of the table (rows 1 through 9) that are labelled "bias" report the average of $(\hat{\theta} - \theta)$ from 1000 replications of the experiment based on 50 observations per moment (N_p=50, p=1,...,10). The average of $(\hat{\theta} - \theta)$ for the EWMD estimates (column 1) is always near 0, demonstrating the lack of bias in EWMD. However, the bias in OMD (column 5) is always negative. When data are drawn from the standard normal distribution, the OMD estimates have a bias of -0.074, or 7.4 percent of the parameter value of 1. The mean of OMD when using data from the uniform distribution is still 0.013 below the true value.

The bias in OMD is larger for thick tailed distributions, which have higher fourth

---

[7] The bimodal distribution is generated by using a binomial to randomize between two normally distributed random variables. One has mean -2 and the other has mean 2; both have variance 1.

moments and therefore have more variable second moments. To see this, note that the bias declines with the number of degrees of freedom of the t distribution, and recall that the tails of the student t distribution become thinner as the degrees of freedom increase and the distribution approaches the normal. For example, when N is 50 the bias in the OMD estimator declines from -.199 to -.118 to -.100 as one moves from a t(5) distribution to a t(10) and then to a t(15) distribution. The log normal and exponential distributions produce the worst bias. The bias is -.616 and -.279 respectively when N is 50, which are very large relative to the true parameter value of 1.

The remaining panels of the table repeat the experiment increasing the amount of data used to estimate each moment. As N rises, EWMD remains unbiased and the bias in OMD declines. The additional data improve the accuracy of the variance estimates and the accuracy of the weights. This, in turn, improves the accuracy of the OMD estimator. For most distributions the bias is very small when 1,000 observations are available to estimate each sample moment. However, the bias is 23 percent in the log normal case even with 1,000 observations per moment.

Table 1 suggest three conclusions. First, OMD is consistently and seriously downward biased. Second, the bias dissipates with sample size. Third, the bias is worse in long tailed and thick tailed distributions.

Since there is no heteroscedasticity or serial correlation in the experiments of Table 1, the EWMD estimator is the optimally weighted estimator, and OMD and EWMD are asymptotically equivalent. It is not surprising that EWMD out performs OMD in these cases. However, we show in Section V that the three conclusions about bias hold when heteroscedasticity is introduced by varying the distribution of the data within an experiment (Table 2) and when OMD is used to estimate a parameter of the covariance matrix of a vector of correlated random variables (Table 3).

III. The Source of Bias in OMD

Feasible GLS estimation, including OMD, is justified by its asymptotic properties rather than small sample properties. In the case of the linear regression model, feasible GLS is unbiased in small samples if the estimator of the weighting matrix is an even function of the disturbances and the disturbances are from a symmetric distribution (provided that the mean of the estimator exists).[8] Unfortunately, bias will arise in most cases in which the estimate of the weighting matrix covaries with the regression model error.

This problem can be serious in the case of OMD estimation of covariance structure models. From (7) the bias for fixed N is

$$E(\theta_{OMD} - \theta) = E[\ (X'\hat{\Omega}^{-1}X)^{-1}\ X'\hat{\Omega}^{-1}\varepsilon\ ]\ . \qquad (8)$$

In the OMD procedure, the sample estimate, $\hat{\Omega}$, of the covariance matrix of the vector of sample moments, *m*, is the weighting matrix. The sample moments and their covariance are estimated from the same data and the resulting estimates are correlated. Individual observations that increase the sample estimate of a variance will also have an enlarging effect on the sample estimate of the variance of the variance, implying that the elements of $\hat{\Omega}$ and ε are correlated. This creates a correlation between the elements of $(X'\hat{\Omega}^{-1}X)^{-1}\ X'\hat{\Omega}^{-1}$ and the elements of ε. Thus, the bias is non-zero even though the expected error in the moments is zero.

---

[8] Kakawani (1967) discusses these conditions. When Ω does not have to be estimated the concerns raised in the present paper about bias are not relevant.

The consequences of correlation between the weights and the errors is easily illustrated in a model of independent but heteroscedastic sample moments. Let ω_p be the estimated variance of the p^{th} element of m, m_p. The bias of the OMD estimator may be written as

$$E(\theta_{OMD} - \theta) = E\left[(\sum_{p=1}^{P} \omega_p^{-1})^{-1} \sum_{p=1}^{P} \omega_p^{-1} \varepsilon_p\right] . \tag{9}$$

The sign of the bias in θ_{OMD} for a fixed sample size N requires an examination of the expectation in (9). To provide intuition, treat the first summation as a normalization and focus on the weighted sum of the ε_p in the second summation. The weight ω_p^{-1} (proportional to the inverse of the difference between the fourth moment and the square of the second sample moment) tends to be small when there are unusual observations in the sample used to compute the sample variance m_p, while m_p and its sampling error ε_p tend to be large. Consequently, the estimator gives less weight to positive sampling errors in ε_p than to negative sampling errors. As a result, the bias is negative. Evaluating the expectation on the right hand side of (9) is difficult because of the correlation between the terms of the first and second summations. However, the behavior of the probability limit of θ_{OMD} as the number P of independent sample variances of **fixed sample size** N goes to infinity may provide useful information about the bias of θ_{OMD} for a given sample size N, provided that P is large enough. Dividing each summation by P, the probability limit of θ_{OMD} as P goes to infinity is

$$\underset{P\to\infty}{\text{plim}} (\theta_{OMD} - \theta) = E(\omega_p^{-1})^{-1} E(\omega_p^{-1}\varepsilon_p) , \tag{10}$$

provided the expectations actually exist.

The random variable ω_p is greater than 0 with probability 1, and so the first expectation in (10) is positive. Thus the sign of the bias depends on the sign of the second expectation. Since ω_p is greater than 0 with probability 1, ω_p^{-1} is strictly decreasing in ω_p. This suggests that $E(\varepsilon_p/\omega_p)$ will be opposite in sign from $E(\varepsilon_p\omega_p)$, although the monotonicity of the transformation ω_p^{-1} is not sufficient to guarantee this. That is, if $Cov(\varepsilon_p,\omega_p)>0$ then $E(\varepsilon_p/\omega_p)$ is likely to be less than 0 and the bias in OMD is negative. In the appendix we show that the term $Cov(\varepsilon_p,\omega_p)$ is greater than 0 if the distribution of $(D_{ip}^2 - E(D_{ip}^2))$ has a positive skew. This condition holds for the standard distributions used in the economics literature, the distribution of the Abowd Card data, and for all of the distributions used in Table 1. Thus, the analysis of (10) supports the intuitive argument for a negative bias and the Monte Carlo evidence of Table 1.[9] We provide an example in the appendix in which $Cov(\varepsilon_p,\omega_p)<0$ and the bias in OMD is positive. The example involves a symmetric distribution with finite support and most of the mass near the minimum and maximum values.

What determines the size of the bias? The intuitive argument implies that it will depend on the degree to which ε_p covaries with ω_p, which suggests that the bias is lower for large N and higher for distributions with larger higher order moments, since these affect $Var(\varepsilon_p)$. To see the dependence on N, note that standard asymptotic analysis of OMD examines the behavior of the estimator for fixed P as N goes to infinity, and therefore implies that the bias in θ_{OMD} declines with the sample size per moment. The distributions of ω_p and m_p are implicitly indexed by N

[9] Our simulation results using the Abowd and Card data described below and other unreported monte carlo experiments indicate that OMD estimates of covariances, like variances, are often biased downwards in absolute value. To get some insight into this, we have formulated a comparable model with m_p redefined to be the covariance between two mean 0 random variables Z_{ip} and Y_{ip}. The sign of the bias continues to depend on $E(\varepsilon_p/\omega_p)$ where ε_p and ω_p are defined from the model of covariances. $E(\varepsilon_p/\omega_p)$ is likely to be of opposite sign to $Cov(\varepsilon_p,\omega_p)$. One may show that $Cov(\varepsilon_p, \omega_p)$ has the same sign as $E[(YZ - E(YZ))^3]$. Since the cross-products of positively (negatively) correlated variables are likely to have a positive (negative) skew, the simple theoretical analysis is consistent with the simulation results.

and the distribution of D_{pi}. As N goes to infinity, ε_p converges in probability to 0 and the covariance between ε_p and ω_p^{-1} dissipates. Therefore $E(\varepsilon_p/\omega_p)$ normalized by $E(1/\omega_p)^{-1}$ converges to 0. For a given N, $Var(\varepsilon_p)$ and $Cov(\varepsilon_p, \omega_p)$ are positive functions of the higher moments of the distribution, leading to larger bias for distributions with larger higher order moments. The bias may be small in situations in which the sampling errors ε_p are small, either because N is large or because the underlying distributions are well behaved. However, in these situations any improvements in precision from using OMD are also likely to be small in absolute magnitude. In Figure 1, we examine the implication that there is a relationship between $Var(\varepsilon_p)$ and the bias, $E(\theta_{OMD} - \theta)$, by plotting the theoretical standard deviation of the sample variance against the bias of θ_{OMD} for the Monte Carlo experiments of Table 1. Recall that in Table 1 the data underlying each of the P moments used to estimate θ are drawn from a single distribution.[10] The figure shows a strong negative relationship between the bias in θ_{OMD} and the standard deviation of the sample variances used by OMD to fit θ.[11] The solid regression line in the figure has a slope of -0.46 and an adjusted R-squared of 0.95. As the sample size declines and as the fourth moment of the underlying distribution increases holding the second moment fixed, the variance of the variance increases and the bias becomes worse. The tight fit in the figure suggests that the simple relationship between the bias and the theoretical standard

[10] The theoretical standard deviation is the square root of the variance of the variances and is computed from the sample size and the population moments of the particular distribution. The bimodal and half normal distributions are omitted from the table as we have not calculated the theoretical variance of the variance in these cases. Exclusion of the points corresponding to the log normal with 50 observations and the log normal with 100 observations, in which the theoretical standard deviation is much higher than the other cases (1.1 and 1.5 respectively), produces a regression equation of Bias = .03 - .56 (Standard Deviation) with an adjusted R-squared of .95.

[11] Note that in Table 1 the bias is larger in the exponential case than in the t(5) even though the two distributions have identical fourth moments. This provides limited evidence that the skewness of the distribution, as well as the variance of the second moments, affects bias.

deviation holds regardless of whether the variation in the theoretical standard deviation is due to variation in the distribution across experiments or due to variation in sample size across experiments.

IV. An Unbiased OMD Estimator

The bias in OMD arises because of a correlation between sample moments and the estimated weighting matrix $\hat{\Omega}$. Corrections based on bias approximations are cumbersome. However, (8) suggests a direct way to attack the source of the bias. Observe that estimates of the second moments based on part of the sample are statistically independent of estimates of the weighting matrix based on the other part.[12,13] Randomly partition the full data sample into G groups of equal size. Let the (Px1) vector m_g and the (PxP) matrix $\hat{\Omega}_g$ be the sample estimates of μ (recall that $\mu=f(\theta)$) and Ω using only the data in group g. Let $m_{(g)}$ and $\hat{\Omega}_{(g)}$ represent estimates based on the data excluding group g.[14] Define the independently weighted optimal

[12] In some applications the observations are dependent within clusters and are independent across clusters. For example, data may be available on several members of the same family or several students from the same school. In this case the sample separation should be based on clusters of observations.

[13] Splitting the sample to break statistical dependence is not a new idea in econometrics. Another recent example is Angrist and Krueger (1994) who use a split sample approach to address bias in instrumental variables estimation.

[14] Since the mean of the data typically is not known, the data are often centered prior to applying EWMD, OMD, or IWOMD either by simply taking means or by a multivariate regression using the full sample. The estimates m_g and $\hat{\Omega}_{(g)}$ will be dependent if a single sample mean (or regression function) is used to center the data. Consequently, IWOMD is not unbiased in small samples unless the data are centered separately for partition g and (g) with an appropriate degree of freedom adjustment in computing m_g and $\hat{\Omega}_{(g)}$. When we refer to IWOMD as unbiased, strictly speaking we are referring to the estimator where either the mean is known or the data are centered separately.

Estimation of the mean has no asymptotic effect because the mean converges more rapidly than the higher moments. Researchers often ignore the effects of estimating the means in actual applications. However, our simulations (not reported) indicate that the bias due to correlation through the mean may be substantial if the sample size is small and the distribution is heavily skewed, such as the log normal and the exponential distribution. Use of a single mean to center the data produces positive bias in the IWOMD estimator.

minimum distance estimator $\theta_{IWOMD(G)}$ as the split sample estimator that uses each m_g and $\hat{\Omega}_{(g)}$ pair to produce G separate parameter estimates, which are then averaged. The formal definition of the independently weighted estimator (IWOMD) is

$$\theta_{IWOMD(G)} = \frac{1}{G}\sum_{g=1}^{G} \text{ArgMin}_{\theta} \; (m_g - f(\theta))' \; \hat{\Omega}_{(g)}^{-1} \; (m_g - f(\theta)) \quad . \tag{11}$$

For $f(\theta)$ linear in θ, each term in the average is free of small sample bias because the sampling errors in the estimates of μ and Ω are independent. Hence the average is unbiased. In the linear case, θ_{OMD} and $\theta_{IWOMD(G)}$ are asymptotically equivalent estimators. This is easily demonstrated using (11). The basic argument is that the $\hat{\Omega}_{(g)}$ and $\hat{\Omega}$ used in OMD all converge to Ω. When $\hat{\Omega}_{(g)}$ is set to Ω for all g, θ_{IWOMD} is numerically identical to OMD based on the true Ω. Consequently, θ_{IWOMD} has the same limiting distribution as (feasible) OMD.

In Table 1 we report simulations of θ_{IWOMD} analogous to those of θ_{EWMD} and θ_{OMD} discussed above. We set G=2 throughout the paper.[15] The results in column 9 confirm that θ_{IWOMD} is unbiased.

---

[15] We investigated whether the number of partitions G matters much. Our initial intuition was that when $f(\theta)$ is linear it is best to set G relatively large, since this should provide the most precise estimate of Ω for each observation g. The number of groups affects the estimation by trading off the precision of the moment estimates and the weighting estimates. The precision loss in the moment estimates due to increased G is offset by the averaging of the parameter estimates as the last step in the procedure. The efficiency gain in estimating the weights tapers off quickly. An increase in G from 10 to N results in at most a 10 percent increase in the size of the samples use to compute $\hat{\Omega}_{(g)}$. When $f(\theta)$ is nonlinear in θ the average of the estimates of θ based on each of the subgroups of observations may be biased as a consequence of Jensen's inequality. One could use IWOMD to estimate the most restrictive linear model that nests the nonlinear model, and then to use EWMD to fit the nonlinear model to the parameter estimates of the linear model. In our Monte Carlo simulations (not reported) we compared the performance of the estimator with using 2, 5, 10, 25, and N groups when the data are not centered separately for each partition. The simulation results suggests that in most cases performance is not very sensitive to choice of G. However, we choose G=2 out of a concern that in actual applications centering the data separately for each partition may cause problems when N is small and G is as large as 10.

V. Evaluating the Performance of EWMD, OMD and IWOMD

In this section we provide a broader assessment of the performance of EWMD and OMD. We consider the bias, root mean square error, median absolute error and standard deviation of the estimators.[16] We begin with the Monte Carlo experiments in which a variance parameter is estimated from sample moments that are drawn from identical distributions. Second we consider cases in which a variance parameter is estimated from moments drawn from different distributions. Third we consider a case involving correlated moments. We defer discussion of IWOMD until the end of the section.

V.1 Models with identically distributed moments

In Table 1 we report the standard deviation (Std), root mean squared error (RMSE) and median absolute error (MAE) of EWMD and OMD estimators of the variance parameter θ as well as the bias. In these experiments EWMD is OMD based on the true weights. The table shows that OMD is not only biased, but always has a RMSE and MAE that are as large or larger than EWMD . For example, in the case of the t(5) with 50 observations the bias in OMD is -0.199. It has an RMSE of 0.223, about double the RMSE of EWMD.

Given that EWMD is "true" OMD in these experiments, it is initially surprising that for the t(5) and the log normal cases, OMD has a smaller standard deviation than EWMD. The source of bias in OMD—the correlation between the weight and the second moment—is expected to reduce the sampling variance of OMD when the kurtosis of the underlying data is large (Koenker et al, 1994). These are typically the cases where bias is large.

---

[16] Although not reported in tables 1, 2 and 3, we have also computed the 5^{th} and 95^{th} percentile values of the empirical distribution of each estimator. The values confirm that the distribution of the OMD estimator is shifted to the left relative to the location of θ, and that the OMD estimator often has a tighter distribution than EWMD or IWOMD. See table 4 for limited evidence.

V.2 Fitting Variance Parameters to Second Moments Based on Different Distributions

In Table 2 we report Monte Carlo experiments when 5 moments come from one distribution and 5 from another. We report the bias, standard deviation, RMSE, and MAE of θ_{OMD}. Since θ_{EWMD} and θ_{IWOMD} are unbiased estimators, there is no difference between the standard deviation and RMSE. Therefore we only report the RMSE and MAE for these estimators. In column 11 we present the theoretical asymptotic variance of EWMD relative to OMD for each experiment. The values range from 1 when the exponential and t(5) are paired to 35.7 when the log normal and the uniform are paired. There are a number of general conclusions from the table. First, the bias in OMD is serious when the sample size is relatively small and at least one of the distributions is badly behaved. For example, in row 6 we set N to 50 and drew five moments from a normal distribution and five from a t(5) distribution. While the theoretic asymptotic efficiency of EWMD relative to OMD is 1.56, OMD estimation imparts a bias of -.132. The ratio of the observed variance of EWMD to the observed variance of OMD is only 1.12, so the theoretical efficiency gain is not realized either. The log normal-uniform case is another example. In this case the bias in θ_{OMD} is -.246 when N is 50 and -.026 when N is 300. When both distributions are badly behaved, the bias is severe for small sample sizes, as evidenced by the log normal-exponential cases and the t(5)-exponential cases. This result is not surprising given the results of Table 1.

Second, EWMD frequently outperforms OMD in RMSE and MAE when both distributions are badly behaved.[17] EWMD does significantly worse than OMD in RMSE only when a very well behaved distribution (the uniform) is paired with badly behaved distributions such as the log

---

[17] For example, consider the exponential-t(5) for all sample sizes and the log normal-t(5) and exponential-log normal when N is 500 or less.

normal, the t(5), or the exponential. In most of the remaining cases EWMD dominates OMD in RMSE and MAE. We also find that the relative performance of OMD improves substantially with N for pairs of distributions that imply a large superiority for OMD in asymptotic efficiency.[18] This makes sense, because the bias falls and the accuracy of the weights improves with N. Of course, the improvement with N is not a strong point in favor of OMD, because when N is large the choice of estimator makes less difference in absolute terms.

Figure 2a summarizes the evidence on this point. For the simulation experiments of Table 2 with similar theoretical efficiency gains (between 1.125 and 1.5) we plot the bias squared (solid line) and difference between the variance of EWMD and OMD (dashed line) against the average variance of the sample variance. The MSE of the estimators are the same where the two lines intersect. For a fixed efficiency gain, the OMD bias increases much faster than the difference in variance as the precision of the sample moments increases. The figure suggests that the realized efficiency gain of OMD dominates the bias only for very well behaved distributions with small fourth moments. In Figure 2b we exclude experiments with high values for the average variance of the sample variance to be sure that these are not dominating the figure. Figures 2c and 2d are analogous to 2a and 2b (respectively) but consider cases in which the theoretical efficiency gain is between 3.03 and 4.04.[19] The story is the same in all four figures.

The figures as well as the detailed results in the tables suggest that the advantage of OMD in RMSE over EWMD (when there is an advantage) is usually only a small fraction of the

---

[18] For example, compare the RMSE and the standard deviations of EWMD and OMD as N increases from 50 to 1000 in the log normal- normal case in rows 31-35 and in the exponential-uniform case in rows 96-100.

[19] The pairs of distributions are the uniform-t(5), the uniform-exponential, the t(5)-lognormal, and the exponential-log normal.

theoretical advantage. To get further insight we compare the performance of feasible OMD to the performance of "true" OMD based on a known, optimal weighting matrix (columns 9 and 10 of Table 2). When the sample size is 100 or more, the efficiency of EWMD relative to true OMD is close to the value implied by the asymptotic theory.[20] However, when the sample size is 300 or less, the RMSE of OMD is usually substantially larger than the RMSE of true OMD. The gap in RMSE between OMD and true OMD arises both because of bias in OMD and because in most cases OMD has a higher sampling variance. Both bias and sampling variance are more important when the theoretical efficiency gain from OMD is high (e.g., the log normal-normal and log normal-uniform cases.) Bias is more important when the theoretical efficiency gain is low but the distributions are badly behaved (e.g, the exponential-t(5).)

The high cost of estimating the weighting matrix suggests that there are large gains from a priori information about the relative precision of different sample moments. A dramatic way to make this point is to note that in some situations, researchers would be better off "throwing away" some moments rather than including them with estimated weights. For example, the OMD estimator in the log normal-normal experiment with 100 observations has a bias of 0.188 and a RMSE of 0.227. EWMD also has a large RMSE in this case. If the 5 moments based on the log normal are ignored and only the 5 moments based on the normal random variables are used in OMD estimation, the bias and RMSE fall to .03 and .005 respectively. The optimal weights are close to zero for the log normal data. Prior information that one distribution is "bad" and the other is "good" could lead to a substantial improvement in OMD relative to EWMD. However,

---

[20] When N is 100, and the normal-t(5) pair of distributions are used, $Var(\theta_{EWMD})/Var(\theta_{OMD})$ is 1.57 while the asymptotic ratio is 1.56. For the uniform-t(5) the corresponding ratios are 3.432 and 3.025. In the log normal-normal case the ratios are 16.25 and 14.59. We obtain the variances from columns 1 and 4 of Table 2.

estimating the weights leads to bias and in most cases a higher RMSE in OMD than EWMD.

V.3 Experiments involving Correlated Moments.

We next consider the behavior of the estimators when the sample moments are correlated. There are two conclusions from this discussion. The efficiency gains become substantial only when the correlations (or covariances) contain information on the parameters and the correlations are included in the model. Second, EWMD continues to dominate OMD. Third, increasing the size of the model while holding constant the number of moments that contain information on a given sample moment leads to additional small sample bias.

Since the efficiency gains of OMD are surprisingly small when the moments are homoscedastic (but correlated) unless there are restrictions across the parameters of the model, the interesting model for OMD involves correlated moments where the structure of the correlation is known and aides in the identification of the parameter.[21] Generate $D_p = (Z_p + \rho Z_{p+1}) /(1 + \rho)^{0.5}$ for p = 1 to 10 from a set of mean 0, variance θ, i.i.d. random variables $Z_1,\ldots,Z_{11}$. The D variables are mean 0 with variance θ, but D_p and D_{p+1} have a covariance of $\theta\rho/(1+\rho^2)$. We assume ρ is known, so the 10 sample variances and 9 first order autocovariances are functions

[21] Suppose one is estimating the variance of the annual change in earnings from a 10 years of panel data under the assumption that the variance in the change in earnings is the same in the 10 years. This model for the earnings variance is a subset of the stationary model we investigate using the Abowd and Card data below. Serial correlation in the data will produce a correlation in the variance estimates from adjacent years. However, a correlation of 0.5 in the sample variances from adjacent years produces an efficiency gain of only 4 percent when the underlying data are normally distributed. (The efficiency gain depends on the distribution and is lower for badly behaved distributions.) Simultaneous estimation of a parameter for the earnings variances and a parameter for the first order autocovariances without restrictions across parameters does not alter the theoretical efficiency gains. For example, in the context of Abowd and Card's "stationary model" discussed below, it appears that there is no gain in asymptotic efficiency from simultaneously estimating the parameters of the model, which is standard practice. We should note that we derived the theoretical efficiency gains for a case in which the sample variances are homoscedastic and the covariances are homoscedastic. This does not rule out the possibility that when there is heteroscedasticy as well as serial correlation, then simultaneous estimation of the parameters of the stationary model leads to efficiency gains over separate estimation of the individual parameters from the moments directly relevant for the individual parameters.

of the single unknown population variance θ, which is set to 1 in our experiments. The mix of variances and covariances in the vector of sample moments, m, implies heteroscedasticity in addition to serial correlation.

Table 3 contains the simulation results from this model applied to a variety of distributions, sample sizes, and implied values for $Cov(D_p, D_{p+1})$. The last column of the table reports the theoretical efficiency of EWMD relative to OMD and suggests that the relative efficiency gain from OMD is larger when the Z_p are from a well behaved distribution. For example, in the log normal case the relative efficiency is 1.08 when $Cov(D_p,D_{p+1})$=.5, while in the normal and uniform cases relative efficiency is 1.42 and 1.78. We were surprised by the extent to which the relative efficiency depends on the distribution of the Z_p given that the distribution is the same for all ten random variables.

The results confirm our earlier finding that the bias in OMD is larger when the sample size is small or the distribution is poorly behaved. For example, in the log normal case when the sample size is 100 and ρ is .5, the bias is -.551, while the bias is -.074 in the corresponding case for the normal. In the log normal case with N=300 and ρ=.5, the bias falls to -.338 (row 69). The table also shows that OMD is not only biased, but has a higher RMSE and MAE than EWMD in every case except those involving the uniform with N greater than 100. In summary, the experiments with correlated data reinforce our other experiments.

A comparison of the experiments in Table 1 with the experiments in Table 3 when ρ=0 provide some information about whether the size of the model, holding constant the amount of information on a particular population moment, affects the bias in OMD that arises from estimation of Ω. When ρ=0 the experiments in Table 3 differ from those in Table 1 only in the fact that the dimension of m and X is 19x1 in Table 3 and 10x1 in Table 1. In Table 3 the 9 elements of X corresponding to the $Cov(D_p, D_{p+1})$ are 0. Since Ω is diagonal when $\rho=0$, the OMD estimators of Var(D) with and without the 9 covariances are **numerically identical** if the true Ω is used.[22] In practice, the bias in θ_{OMD} is worse for every distribution and sample size that we checked, including all of those that appear in both of the tables. For example, the normal case when N is 50, the bias is -.074 in Table 1 (see row 4, column 5) and -.140 in Table 3 (row 31, column 3). In the exponential case when N is 50 the bias is -.279 in Table 1 and -.387 in Table 3. Evidently, the additional elements of $\hat{\Omega}$ and ε add to the problem of bias arising from covariance between $\hat{\Omega}$ and ε. These results are obviously limited to the experiments we performed, but provide some suggestion that researchers should avoid unrestricted estimation of the weighting matrix Ω in large models.

[22] Let ε_v denote the sampling error vector of the 10 variances and ε_c denote the sampling error vector for the 9 covariances. Let Ω_v denote the 10x10 upper left submatrix of Ω corresponding to $var(m_v)$ and Ω_c denote the 9x9 lower right submatrix of Ω corresponding to $var(m_c)$. Let $\hat{\theta}_{19}$ be the estimator based on m_v and m_c, and let $\hat{\theta}_{10}$ be the estimator based on m_v only. The sampling error in $\hat{\theta}_{19}$ is

$$\hat{\theta}_{19} - \theta = ([X_v' \quad 0]\, \hat{\Omega}^{-1} [X_v' \quad 0]')^{-1} [X_v' \quad 0] \hat{\Omega}^{-1} [\varepsilon_v' \quad \varepsilon_c']'$$

The sampling error when only m_v is used (Table 1) is

$$\hat{\theta}_{10} - \theta = [X_v' \hat{\Omega}_v^{-1} X_v]^{-1} X_v' \hat{\Omega}_v^{-1} \varepsilon_v$$

Since the variances and covariances are uncorrelated when $\rho=0$, Ω is a block diagonal matrix. Consequently, when the true Ω is used

$$\begin{aligned}\hat{\theta}_{19} - \theta &= \left(\begin{bmatrix} X_v' & 0 \end{bmatrix} \begin{bmatrix} \Omega_v^{-1} & 0 \\ 0 & \Omega_c^{-1} \end{bmatrix} \begin{bmatrix} X_v \\ 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} X_v' & 0 \end{bmatrix} \begin{bmatrix} \Omega_v^{-1} & 0 \\ 0 & \Omega_c^{-1} \end{bmatrix} \begin{bmatrix} \varepsilon_v \\ \varepsilon_c \end{bmatrix} \\ &= (X_v' \, \Omega_v^{-1} X_v)^{-1} X_v' \, \Omega^{-1} \varepsilon_v \\ &= (\hat{\theta}_{10} - \theta) \ .\end{aligned}$$

Thus the "true OMD" versions of the two estimators are numerically identical.

V.4 The Performance of IWOMD

We have already noted that the experiments with independent, homoscedastic moments in Table 1 confirm that IWOMD is unbiased. We now provide a broader assessment of the performance of this estimator. Columns 9-12 of Table 1 report the bias, standard deviation, MAD, and RMSE of IWOMD for the homoscedastic model. The MAE and RMSE of IWOMD are less than or equal to that of OMD in 33 of 36 cases and 30 of 36 of cases respectively. For the t(5) distribution with 50 observations the MAE values are 0.201 for OMD and 0.125 for IWOMD. The RMSE values are .223 and .202 respectively. On the other hand, the RMSE of the IWOMD estimator is always greater than EWMD.

Comparison of the RMSE and MAE between EWMD and IWOMD provides evidence of the cost of estimating the weighting matrix that is uncontaminated by the correlation between the weights and the second moments. For poorly behaved distributions and small sample sizes, the costs can be quite large. For example, for the t(5) distribution with N=50 the median absolute error for EWMD is .071 and .125 for IWOMD. The difference corresponds to an increase of more than 75 percent and the RMSE rises from .125 to .202. The difference dissipates as sample size increases but there is still a 50 percent difference in RMSE for the same experiment using 500 observations to estimate each moment.

In Table 2 we report Monte Carlo experiments for IWOMD based on ten moments from two different distributions. The detailed results show that IWOMD never outperforms EWMD in RMSE or MAE when N is 50 or 100. As the sample size increases the two estimators perform similarly, with IWOMD having an advantage in larger sample sizes for pairs of distributions that imply a substantial asymptotic efficiency gain for IWOMD and OMD. The sampling variance of IWOMD is larger than OMD even though the asymptotic variances of IWOMD and OMD are identical. As in the single distribution case, the cost of estimating the weighting matrix is high. The experiments based on serially correlated data in Table 3 show that IWOMD often dominates OMD in MAD and RMSE but usually does not perform as well as EWMD unless the efficiency gain is above 1.4 and the sample size is above 300. In the experiments based on the Abowd and Card data IWOMD performs much better than OMD but is dominated by EWMD.

V.5 Summary

The experiments based on independent and homoscedastic second moments in Table 1, moments from 2 different distributions in Table 2, and correlated moments in Table 3 all indicate first that θ_{OMD} suffers from serious small sample bias in many cases, second that it is almost always dominated by θ_{EWMD} in RMSE and MAE, and third that there is a large cost to having to estimate the weighting matrix. θ_{EWMD} also typically dominates θ_{IWOMD}, which is unbiased but has a larger sampling variance.

VI. An Empirical Example

It is difficult to know what to assume about the data and models used in practice. Therefore we supplement our Monte Carlo analysis by applying the various estimators to Abowd and Card's (1987, 1989) analysis of the covariance of changes in log earnings and changes in log hours from the Panel Study of Income Dynamics. The data are based on an eleven year (1969-1979) sample of male heads of household. The sample consists of 1536 individuals with complete data for the period, annual hours continuously above 0 but less than or equal to 4680, and average hourly wages continually less than $100..[23] All dollar values have been adjusted to 1967 levels using the Consumer Price Index. Log hours and log earnings are adjusted for labor market experience and year effects prior to computing variances.[24]

The assumption of stationarity in first differences implies that the autocovariances and cross covariances do not depend on time. The raw data for 1969-1979 provide 10 observations per person on the changes in log earnings and in log hours. There are 210 unrestricted moment estimates for various years and lags (10 hours variances, 45 hours autocovariances, 10 earning variances, 45 hours autocovariances, and 100 hours/earnings covariances). The stationary model has 39 unique parameters (1 hours variance, 9 hours autocovariances, 1 earnings variance, 9 earnings autocovariances and 19 hours/earnings covariances). Table 4 presents the EWMD and OMD estimates of the covariance at time lags of 0, 1 and 2. The results of Abowd and Card are presented as well.

The difference between EWMD and OMD is striking. There is a systematic tendency for the absolute value of the EWMD estimates to exceed that of the OMD estimates. For example, the OMD estimate of the variance of the change in log earnings is .086, which is less than half of the EWMD estimate. The OMD estimate of the covariance of the change in hours and the first lag of the change in log earnings is .026 which is less than a third of the EWMD estimate of .080. Altonji, Martins and Siow (1987) report the same phenomena for similar data on

[23] In constructing our sample we followed the data appendix in Abowd and Card (1987), who worked with a sample size of 1448. With their assistance we attempted to track down the differences between their sample and ours. We gave up after a modest effort because the EWMD and OMD estimates are very similar for the two samples. We thank John Abowd and David Card for their assistance.

[24] The adjustment consists of separate regressions of the change on log earnings and the change in log hours on the 1967 potential experience level (age - education attainment - 5) and time dummies. The remainder of the procedure is performed using the residuals from the experience regression, which now have mean zero.

earnings and hours as well as for data on changes in family income, wage rates, consumption, and the hours of unemployment. The fact that the OMD estimates are consistently below the EWMD estimates in absolute value is disturbing.

Note however that the IWOMD estimates are somewhat smaller in absolute value than EWMD despite the fact that both estimators are unbiased. This may indicate that the covariances are in fact nonstationary, since in the nonstationary case the true weights are likely to be positively related to the absolute values of the variances and covariances. As a result, OMD and IWOMD produce weighted averages of the true variances and covariances that will be less than the simple average produced by EWMD.

We would like to compare the three estimators using a model of the Abowd-Card data that we know is true by construction. To this end, we perform an experiment in which we treat the Abowd-Card sample as a population and re-sample from it with replacement. We specify the following "model" of the sample moments. Consider first the variance in hours. Let $m_{t,t-\tau}$ be the estimate of $Cov(\Delta Hours_t, \Delta Hours_{t-\tau})$ for a sample drawn from the population with replacement. Then

$$m_{t,t-\tau} = X_{t,t-\tau}\,\theta_\tau + \varepsilon_{t,t-\tau} \qquad t=1970...1979,\ \tau=0,...,9 \qquad (12)$$

where $X_{t,t-\tau}$ is $Cov(\Delta Hours_t, \Delta Hours_{t-\tau})$ for the "population". The true parameter value for θ_τ is 1 and the sampling error $\varepsilon_{t,t-\tau}$ has mean 0. We construct a corresponding model for the variance of earnings and the covariance of hours and earnings (leads and lags). The observations are stacked into the model

$$m = X\,\theta + \varepsilon \qquad (13)$$

where m is (210x1), X is a (210x39) matrix of constructed "explanatory variables", θ is a (39x1) vector of parameters, and ε is a (210x1) vector of sampling errors. All of the elements of the true θ equal 1.[25] The dimensions of the constructed model match the Abowd and Card stationary model.

We draw samples of 500 observations (with replacement) from the population of 1536 individuals, and estimate the 39 elements of θ by EWMD and OMD. In Table 5 we report the bias, standard deviation, RMSE, and MAE for a subset of the moments. We focus our attention on the results for autocovariances and cross covariances at lags 0, 1, and 2. The results are quite striking. OMD is badly biased and has a much larger RMSE and MAE than EWMD. The bias is typically between -.7 and -.85, which is very large relative to the true parameter value of 1. These results indicate that main reason that the OMD estimates of the "stationary model" of hours and earnings in Table 4 are smaller than the EWMD and IWOMD estimates is bias, not mis-specification.

The RMSE of the OMD estimator typically exceeds the RMSE for EWMD by a factor between 2 and 7. The MAE is also much larger for OMD. OMD does usually have a smaller sampling variance, but its advantage in this dimension is not enough to make up for the large bias. Using the diagonal of $\hat{\Omega}$, instead of the complete $\hat{\Omega}$, in the stationary model in Table 4 and the constructed model of Table 5 leads to a substantial reduction in the bias. (Not reported.) We speculate that OMD performs particularly poorly when Ω is a large matrix and the random variables are correlated, which is consistent with the findings in Section V.3.

In summary, the results suggest that EWMD is the best estimator for fitting linear models

[25] For example, if we order the observations so that the first 10 correspond to the observations on Var($\Delta hours_t$,$\Delta hours_{t-\tau}$), then the first 10 elements of *m* in are the "sample" estimates of the "population" parameters Var($\Delta hours_{1970}$,$\Delta hours_{1970}$) ... Var($\Delta hours_{1979}$,$\Delta hours_{1979}$). The first 10 rows of the first column of X are Var($\Delta hours_{1970}$,$\Delta hours_{1970}$) ... Var($\Delta hours_{1979}$,$\Delta hours_{1979}$) for the "population".

of the covariance structure of the PSID data on earnings and hours. It provides at least some evidence that the small sample bias in OMD is a serious problem in applications. If anything, the problem seems far worse in real world data with a large model than in our simulations.

VII. Statistical Inference

We conclude that EWMD is almost always preferred to OMD when Ω must be estimated. However, the fact that sampling error in $\hat{\Omega}$ leads to bias in OMD raises concern about the use of the conventional asymptotic t-statistics, standard errors, and confidence intervals as the basis of statistical inference in the EWMD case, because these rely heavily on $\hat{\Omega}$. Also, to the extent that departures from normality in the second moments and the estimators are substantial, confidence intervals and test statistics based on the asymptotic standard errors must be treated cautiously.[26] In this section we begin by examining the distribution of the estimators and then consider bootstrap methods and conventional asymptotic t-statistics as a basis for statistical inference.

In Table 6 we examine inference about the parameter θ for a subset of the "two distribution" experiments of Table 2. The distribution of the estimators is determined for each

[26] Graphs of the kernel smoothed density of the EWMD and OMD estimators, and of the second moment estimates suggest substantial departures from normality for the exponential, t(5), and especially the log normal when N=100. Aside from the question of statistical inference, this raises the issues of whether in some situations alternatives to the least squares criteria used by EWMD, IWOMD, and OMD deserve consideration given that least squares is inefficient when there are large departures from normality. There would seem to be two directions in which we might proceed. The first is to use more robust methods to combine the second moments. The lower sampling variance of OMD in some cases noted above may reflect the fact that it implicitly provides less weight than the least squares criteria dictates to sample realizations for elements of m that are in the right tail. We have experimented with some modifications to IWOMD that involve symmetric trimming of the smallest and largest estimate of θ attained for the various groups prior to averaging them. Trimming reduces the dispersion of IWOMD but simultaneously introduces bias due to the asymmetric distribution of IWOMD. The second direction, suggested by Koenker et al (1994), is to fit the models to statistics that are more robust than the second sample moments. We leave this to future research.

design from 5,000 replications of the experiment. The first column corresponds to the experiment involving 5 moments from a t(5) distribution and 5 moments from a normal distribution with a sample size of 100. The symmetric 90 % sampling interval is 0.90 to 1.11. The corresponding confidence interval for IWOMD is .87 to 1.15. Both intervals are roughly centered around the true value of 1, although the IWOMD interval is slightly skewed. The OMD interval is from 0.83 to 1.027. Note that it is smaller than either of the other intervals but it is shifted to the left. The 90 percent interval barely contains the true value of 1. In the log normal-exponential case with 100 observations per moment, the OMD interval runs from .45 to .90. These values confirm the negative bias in OMD.

For each replication of the experiment in Table 6, we compute an asymptotic t-test of the null hypothesis that the OMD or EWMD parameter equals 1. The rows labeled "Reject t test" report the fraction of time the test rejects when a 90% confidence level is used. The rejection rate is close to but slightly above 10% in the case of EWMD for both sample sizes of 100 and sample sizes of 500, except when the log normal is one of the two distributions used in the experiment. In these cases the rejection rates are between 20 and 25 percent when N is 100 and between 16 percent and 19 percent when N is 500. The asymptotic t-test for the OMD estimator rejects 93 percent of the time in the log normal-exponential case with N=100. It seriously over-rejects in most of the other cases, even when N is 500.

The table also presents evidence for EWMD on the performance of a test based on a 90% symmetric bootstrap confidence interval, which is one alternative to inference based on the asymptotic t-statistics.[27] When N is 100 the EWMD bootstrap confidence interval excludes the true value of 1 more than ten percent of the time in all cases and in a few cases performs substantially worse than the asymptotic t-test. For example, in the t(5)-normal case with N=100, the bootstrap based test rejects 16.2 percent of the time and the test based on the asymptotic t-statistics rejects 11.8 percent of the time. When N=500 the two tests perform about equally well.

There are two key lessons from the table. First, inference based on asymptotic normality and the use of $\hat{\Omega}$ is satisfactory in most cases and is almost always better than inference based on the bootstrap. Difficulties arise in the cases involving the log normal. Second, the bias of the OMD estimator leads to substantial over rejection.[28]

We have also investigated inference based on the asymptotic formulae and on the bootstrap for the experiments based on the Abowd-Card data. In the case of the stationary model in Table 4, the estimated standard errors based on the two approaches are very close. (Not reported). For the covariance model of hours and earnings in Table 5, we report the rejection rates for tests at the .10 significance level of the null hypothesis that the parameter is 1 for the asymptotic t-statistic (column 3). In the case of Var(ΔEarnings$_t$), the actual size of the conventional asymptotic t test is .142. The size is typically about .125 for other moment parameters. Column 4 of Table 5 reports on the performance of a test based on a 90%

---

[27] We constructed the bootstrap confidence interval for EWMD as follows. For each of 1,000 Monte Carlo samples of size N, we draw with replacement 500 different bootstrap samples of size N. We then computed EWMD estimator using each of the 500 bootstrap samples. The 5th and 95th values of the EWMD estimator form the 90 percent confidence interval estimate for the particular Monte Carlo sample. The entry " boot test" for EWMD in Table 5 is the fraction of Monte Carlo samples for which the 90 percent confidence interval estimate does not contain the true value of 1. We use an analogous bootstrap procedure to produce the "boot test" result for IWOMD in Table 6. Hall and Horowitz (1994) provides a discussion of bootstrapping methods in the GMM setting.

[28] We also include IWOMD in our examination of statistical inference in Table 6. We find that inference from IWOMD with a bootstrapped confidence interval performs substantially better than OMD but not as well as EWMD with either a bootstrap confidence interval or an asymptotic t-test.

symmetric bootstrap confidence interval. In the case of Var(ΔEarnings$_t$), the actual size of the test is .112, and the typical result is that bootstrap based tests also reject a bit too frequently. The results based on the Abowd-Card data are similar to the Monte Carlo experiments in Table 6 in that both the bootstrap and asymptotic confidence intervals over reject. Whether or not the departure of the actual size from the true size is serious would seem to depend on the application.[29] We view these discrepancies as small enough to permit one to use either method as the basis for inference in most situations.

VIII Conclusion

In this paper we provide a theoretical argument and Monte Carlo evidence showing OMD is biased in small samples. For a given sample size the bias depends on the distribution of the underlying data. The bias is worse when the data are drawn from distributions with heavy tails. The problem goes away as the sample size gets large but does not go away (and may get worse) as the number of moments available to fit a model increases or as the size of the model increases holding constant the number of moments that are informative about a given parameter. Our findings using the Abowd-Card data are particularly striking. We also present an estimator called IWOMD, which is an unbiased split sample alternative to conventional OMD and is asymptotically equivalent to it. However, in most cases we consider, the asymptotic efficiency gain of IWOMD relative to EWMD is overwhelmed by the extra noise introduced by estimated weights.

[29] We did compute the actual 5th and 95 percentile values of the EWMD estimator and compare them to the estimators based on the asymptotic standard errors and the bootstrap. While it is easy to look for bias in these estimators, it is not clear how to summarize the accuracy of interval estimators.

upon EWMD and OMD by using a priori information about which sets of moments are likely to be highly correlated or particularly noisy to reduce the dimensionality of $\hat{\Omega}$. Third, both IWOMD and EWMD have instrumental variables interpretations that suggest it may be beneficial to look for a way to combine the two estimators into an IV estimator that converges to EWMD as the number of moment conditions p becomes large with N fixed and to IWOMD and OMD as N becomes large with p fixed.

Fourth, our Monte Carlo evidence is focussed on cases in which the model of the second moments is linear in parameters. Many applications of OMD involve models in which the parameters of interest are nonlinear functions of the second moments. (For example, Abowd and Card estimate factor models that are nonlinear in the second moments.) In these cases, bias is likely, but the sign and severity is not likely to be easy to predict. For example, if a factor loading parameter is estimated as the covariance divided by a variance, then downward bias (in absolute value) in the two moments may partially cancel out, and the direction and size of the parameter bias will depend on the relative bias in the two moments. A large scale Monte Carlo study using some of the nonlinear models that are popular in the literature is warranted.

Fifth, Monte Carlo evidence is needed on the validity of model mis-specification tests based on OMD and EWMD, an issue that we have not explored. Six, we suspect similar biases in maximum likelihood and quasi-maximum likelihood estimation of covariance structures. This should be investigated.

We wish to emphasize that our basic concern—feasible GMM estimators are biased in small samples because of correlation between the moments used to fit the model and the weight matrix—applies in situations involving moments of any order. This is not a new point, since it

Our conclusion is that EWMD is almost always preferable to using OMD when the optimal weighting matrix is unknown, especially when bias is an important concern. This is true even in situations in which OMD is far superior in asymptotic efficiency. At a minimum, researchers should estimate models by both OMD and EWMD or both OMD and IWOMD and worry about bias in OMD if the parameter estimates differ substantially. Discrepancies between OMD and EWMD may be a risky basis for Hausman type model mis-specification tests for models of second moments.

If one is going to use EWMD because of small sample considerations, the issue of how to do statistical inference arises. The limited Monte Carlo evidence in Table 5 and Table 6 suggests, perhaps surprisingly given our evidence on bias in OMD, that in most cases the standard asymptotic formula provides a satisfactory basis for inference when using EWMD. The use of the asymptotic formula seems to be superior to using bootstrap methods, although the bootstrap performs slightly better in the examples based on the Abowd and Card data. Hypothesis testing in the case of OMD is a disaster in small samples with poorly behaved distributions because the sampling distribution of OMD is shifted far away from the true parameter value.

Throughout the text we identify a number of extensions of our Monte Carlo studies. Six broader issues deserve mention. First, perhaps the use of robust estimation methods to estimate the weighting matrix or to estimate the moments being modelled (see Koenker et al, 1994) may lead to an OMD estimator that is superior to conventional OMD and IWOMD. Second, it is clear from the performance of true OMD that a priori information about the appropriate weighting matrix is valuable. There may be many situations in which researchers may improve is well known, for example, that feasible generalized least squares is biased in small samples in many situations. However, it has not gotten the attention it deserves in the burgeoning literature that involves GMM.

References

Abowd, John M. and David Card, "Intertemporal Labor Supply and Long-Term Employment Contracts," *American Economic Review*, March 1987, 77(1), 50-68.

_____, "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, March 1989, 57(2), 411-445.

Altonji, Joseph G., Ana P. Martins, and Aloysius Siow, "Dynamic Factor Models of Consumption, Hours and Income," 1987, NBER Working Paper 2155.

Angrist, Joshua D. and Alan B. Krueger, "Split Sample Instrumental Variables," January 1994, NBER Technical Working Paper No. 150.

Arrellano M. & J. D. Sargan, "Imhof Approximations to Econometric Estimators," *The Review of Economic Studies*, October 1990, 57(4), 627-646.

Behrman, Jere, Mark Rosenzweig and Paul Taubman, "The Allocation of Schooling in the Family and in the Marriage Markets," unpublished, 1993.

Chamberlain, Gary, "Multivariate Regression Models For Panel Data," *Journal of Econometrics*, 1982, 18, 5-46.

_____, "Panel Data" in *Z. Griliches and M.D. Intriligator (eds.), Handbook of Econometrics, Volume II*, Elsevier Science Publishers, 1984.

Hall, Robert E. and Frederick Mishkin, "The Sensitivity of Consumption to Transitive Income: Estimates from Panel Data on Households," *Econometrica* 50, 461-480.

Hall, Peter and Joel L. Horowitz, "Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators," February 1994, University of Iowa Working Paper Series #94-07.

Hansen, Lars P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1988, 50(4), 1029-1054.

Horowitz, Joel L. and George R. Neumann, "A Generalized Moments Specification Test of the Proportional Hazards Model," *Journal of the American Statistical Association*, 1992, 87, 234-240.

Kakawani, N.C. "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equation Estimators," *Journal of the American Statistical Association* 82 (1967): 141-142.

Koenker, Roger, Jose A.F. Machado, Christopher L. Skeels and Alan H. Welsh, "Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimation," *Econometric Theory*, January 1994, 10(1).

Lehmann, Bruce N., "Residual Risk Revisited", *Journal of Econometrics*, July/August 1990, 45(1/2), 71-98.

Malinvaud, E., *Statistical methods of econometrics*, Amersterdam: North-Holland, 1970.

Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes, *Introduction to the theory of statistics*, New York: McGraw-Hill, 1973.

Ogaki, M., "Generalized Method of Moments: Econometric Applications". *Handbook of Statistics*, Vol. 11: Econometrics, edited by G.S. Maddala, C.R. Rao, and H.D. Vinod. Amsterdam: North-Holland (forthcoming)

Schwert, G. William and Paul J. Seguin, "Heteroskedasticity in Stock Returns," *Journal of Finance*, September 1990, 45(4), 1129-1154.

Shanken, Jay, "Intertemporal Asset Pricing: An Empirical Investigation," *Journal of Econometrics*, July/August 1990, 45(1/2), 99-120.

Tauchen, George, "Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics* 4 (1986): 397-425.

Appendix

This appendix establishes that $Cov(\varepsilon_p,\omega_p)$ is negative provided that the distribution of D_{ip}^2 is skewed to the right, with $E[(D_{ip}^2 - E(D_{ip}^2))^3] > 0$. Assume that the underlying data, D_{ip}, are known to have mean zero. Let μ_{jp} represent the j^{th} population moment of D_{ip} and let m_{jp} be the j^{th} sample moment of D_{ip}, computed as

$$m_{jp} = \frac{1}{N}\sum_{i=1}^{N} D_{ip}^{j} \quad . \tag{1}$$

To simplify the notation we suppress the p subscripts in what follows, so that m_1 is a sample mean and m_2 is a sample variance. The notation m_j should not be confused with m_p used in the text to represent the p^{th} element of the vector of second moments. Note that

$$E(m_j) = \mu_j \tag{2}$$

and

$$m_2 = \mu_2 + \varepsilon \quad , \tag{3}$$

where ε is the sampling error. The variance of m_j is

$$Var(m_j) = \frac{1}{N}(\mu_{2j} - \mu_j^2) \quad . \tag{4}$$

The variance of m_2 is

$$Var(m_2) = \frac{1}{N}(\mu_4 - \mu_2^2) \quad , \tag{5}$$

which is estimated as

$$\omega = \frac{1}{N}(m_4 - m_2^2) \quad . \tag{6}$$

The covariance between ε and ω depends on the skewness of D^2 as shown by

$$\begin{aligned} Cov(\varepsilon,\omega) &= E(\varepsilon\omega)-E(\varepsilon)E(\omega) \\ &= E(\varepsilon\omega) \\ &= E(m_2\omega) - E(\mu_2\omega) \\ &= \frac{1}{N}E(m_2 m_4 - m_2^3) - \frac{(N-1)}{N^2}(\mu_2\mu_4 - \mu_2^3) \\ &= \frac{(N-1)}{N^3}(\mu_6 - 3\mu_2\mu_4 + 2\mu_2^3) \\ &= \frac{(N-1)}{N^3} E\left[(D^2-\mu_2)^3\right]. \end{aligned} \tag{9}$$

where the last equality follows from the fact that

$$\begin{aligned} &\mu_6 - 3\mu_2\mu_4 + 2\mu_2^3 \\ &\quad = E[\ (D^4-2D^2\mu_2)\ (D^2-\mu_2)\ ] \\ &\quad = E[\ (D^4-2D^2\mu_2 + \mu_2^2)\ (D^2-\mu_2)\ ] - E[\mu_2^2(D^2-\mu_2)] \\ &\quad = E\left[(D^2-\mu_2)^3\right] \end{aligned} \tag{10}$$

Thus, the expectation of Cov(ε,ω) is positive if the distribution of D^2 is skewed to the right. This restriction is satisfied for all of the distributions we consider. It is not satisfied if the density of D is symmetric and the density is concentrated near the minimum and maximum of D. For example, consider the density $dF(D) = c|D|^{\alpha}$ when $-D_b \leq 0 \leq D_b$ and 0 otherwise, where c normalizes the CDF to 1 and $\alpha>0$. When $\alpha>0$ the density has a minimum at 0 and maximums at $-D_b$ and D_b. We established numerically that the sign of $E[(D^2 - \mu_2))^3]$ is less than zero when $\alpha>1$ and that there is a small positive bias in θ_{OMD} when this distribution is used in an experiment analogous to those in Table 1. The intuition is that when α is large the "unusual" observations on D are those near the mean of 0. These correspond to low values of D^2 but substantial values of $(D^2-\mu_2)^2$, which is the contribution to the weighting matrix. Consequently, second moments with negative sampling errors receive too little weight, and θ_{OMD} is biased upward.

Table 1 Performance of Estimators of a Variance Parameter Using 10 Sample Variances From 1 Distribution

Row	Distribution	Obs.	Equally Weighted (EWMD) Bias (1)	Std. (2)	RMSE (3)	MAE (4)	Optimally Weighted (OMD) Bias (5)	Std. (6)	RMSE (7)	MAE (8)	Independently Weighted (IWOMD) Bias (9)	Std. (10)	RMSE (11)	MAE (12)
1	t(5)	50	-.000	0.126	0.125	0.071	-.199	0.101	0.223	0.201	-.011	0.202	0.202	0.125
2	t(10)	50	-.001	0.079	0.079	0.052	-.118	0.089	0.148	0.121	0.000	0.136	0.136	0.080
3	t(15)	50	-.002	0.073	0.073	0.050	-.100	0.083	0.130	0.101	-.001	0.116	0.116	0.077
4	Normal	50	-.002	0.063	0.063	0.041	-.074	0.075	0.105	0.079	-.003	0.099	0.099	0.065
5	Uniform	50	-.001	0.042	0.042	0.029	-.013	0.048	0.049	0.032	-.002	0.055	0.055	0.036
6	LogNormal	50	-.001	0.395	0.394	0.206	-.616	0.128	0.629	0.625	-.024	0.899	0.899	0.377
7	Exp	50	-.001	0.125	0.125	0.085	-.279	0.131	0.309	0.281	-.012	0.255	0.256	0.165
8	HalfNormal	50	-.001	0.074	0.074	0.050	-.118	0.088	0.147	0.120	-.001	0.131	0.131	0.084
9	Bimodal	50	0.000	0.039	0.039	0.027	-.021	0.044	0.049	0.034	-.001	0.053	0.053	0.035
10	t(5)	100	0.003	0.083	0.083	0.057	-.124	0.070	0.143	0.126	0.003	0.126	0.126	0.077
11	t(10)	100	-.002	0.054	0.054	0.038	-.065	0.059	0.087	0.067	0.000	0.073	0.073	0.050
12	t(15)	100	-.002	0.051	0.051	0.035	-.052	0.055	0.076	0.056	-.001	0.068	0.068	0.046
13	Normal	100	0.001	0.043	0.043	0.029	-.034	0.047	0.058	0.039	0.003	0.055	0.055	0.037
14	Uniform	100	-.000	0.029	0.028	0.020	-.006	0.031	0.031	0.021	-.001	0.033	0.033	0.021
15	LogNormal	100	-.004	0.344	0.344	0.156	-.474	0.115	0.488	0.480	0.016	0.752	0.752	0.254
16	Exp	100	-.004	0.085	0.085	0.055	-.165	0.093	0.189	0.169	-.003	0.147	0.147	0.095
17	HalfNormal	100	0.000	0.052	0.052	0.034	-.062	0.059	0.085	0.066	-.001	0.074	0.074	0.051
18	Bimodal	100	-.001	0.027	0.027	0.018	-.011	0.029	0.031	0.020	-.001	0.031	0.031	0.021
19	t(5)	500	0.002	0.042	0.042	0.025	-.041	0.034	0.054	0.043	0.002	0.067	0.067	0.031
20	t(10)	500	0.001	0.025	0.025	0.017	-.014	0.025	0.029	0.021	0.001	0.027	0.027	0.018
21	t(15)	500	-.000	0.023	0.023	0.016	-.011	0.024	0.026	0.018	0.000	0.025	0.025	0.018
22	Normal	500	0.001	0.019	0.019	0.012	-.007	0.020	0.021	0.013	0.000	0.020	0.020	0.013
23	Uniform	500	0.000	0.012	0.012	0.008	-.001	0.012	0.012	0.008	0.000	0.012	0.012	0.008
24	LogNormal	500	0.005	0.147	0.147	0.080	-.233	0.084	0.247	0.236	0.008	0.238	0.238	0.118
25	Exp	500	-.000	0.041	0.041	0.025	-.043	0.043	0.061	0.047	-.001	0.051	0.051	0.033
26	HalfNormal	500	0.001	0.024	0.024	0.017	-.012	0.024	0.027	0.020	0.002	0.027	0.027	0.019
27	Bimodal	500	-.000	0.012	0.012	0.008	-.002	0.012	0.012	0.008	-.000	0.012	0.012	0.008
28	t(5)	1000	0.000	0.029	0.029	0.018	-.026	0.025	0.036	0.027	-.001	0.033	0.033	0.021
29	t(10)	1000	-.000	0.018	0.018	0.012	-.008	0.018	0.020	0.013	-.000	0.019	0.019	0.013
30	t(15)	1000	-.001	0.016	0.016	0.011	-.006	0.016	0.017	0.012	-.001	0.017	0.017	0.011
31	Normal	1000	-.000	0.014	0.014	0.009	-.004	0.014	0.015	0.010	-.000	0.014	0.014	0.010
32	Uniform	1000	-.000	0.009	0.009	0.006	-.001	0.009	0.009	0.006	-.000	0.009	0.009	0.006
33	LogNormal	1000	0.004	0.100	0.100	0.060	-.161	0.067	0.175	0.165	0.002	0.151	0.150	0.085
34	Exp	1000	0.002	0.030	0.030	0.020	-.020	0.031	0.037	0.026	0.003	0.035	0.035	0.023
35	HalfNormal	1000	0.000	0.017	0.017	0.011	-.006	0.017	0.018	0.013	0.001	0.018	0.018	0.012
36	Bimodal	1000	0.001	0.008	0.008	0.006	-.000	0.008	0.008	0.006	0.001	0.008	0.008	0.006

Notes
Estimates based on 1,000 replications.

Table 2 Performance of Estimators of a Variance Parameter Using 10 Sample Variances From 2 Distributions

Row	Distribution	Obs.	EWMD Std, RMSE (1)	EWMD MAE (2)	OMD Bias (3)	OMD Std. (4)	OMD RMSE (5)	OMD MAE (6)	IWOMD Std, RMSE (7)	IWOMD MAE (8)	'True' OMD Std, RMSE (9)	'True' OMD MAE (10)	Rel. Eff. (11)
1	t(10),t(5)	50	0.098	0.064	-.157	0.096	0.184	0.159	0.158	0.094	0.094	0.063	1.26
2	t(10),t(5)	100	0.068	0.043	-.091	0.064	0.112	0.097	0.099	0.064	0.064	0.041	1.26
3	t(10),t(5)	300	0.044	0.028	-.037	0.037	0.053	0.040	0.049	0.030	0.039	0.026	1.26
4	t(10),t(5)	500	0.034	0.022	-.024	0.030	0.039	0.027	0.037	0.023	0.030	0.019	1.26
5	t(10),t(5)	1000	0.021	0.014	-.015	0.019	0.025	0.017	0.022	0.014	0.019	0.013	1.26
6	Normal,t(5)	50	0.094	0.056	-.132	0.089	0.159	0.137	0.161	0.086	0.079	0.053	1.56
7	Normal,t(5)	100	0.065	0.041	-.071	0.060	0.093	0.072	0.095	0.053	0.055	0.039	1.56
8	Normal,t(5)	300	0.038	0.024	-.030	0.033	0.044	0.032	0.041	0.026	0.032	0.022	1.56
9	Normal,t(5)	500	0.031	0.020	-.018	0.025	0.031	0.022	0.030	0.019	0.025	0.017	1.56
10	Normal,t(5)	1000	0.023	0.014	-.010	0.017	0.020	0.014	0.021	0.013	0.018	0.012	1.56
16	Uniform,t(5)	50	0.094	0.056	-.069	0.073	0.100	0.067	0.120	0.071	0.054	0.036	3.03
17	Uniform,t(5)	100	0.067	0.038	-.033	0.044	0.055	0.036	0.064	0.037	0.039	0.026	3.03
18	Uniform,t(5)	300	0.038	0.024	-.012	0.023	0.026	0.018	0.028	0.018	0.022	0.015	3.03
19	Uniform,t(5)	500	0.032	0.018	-.007	0.018	0.019	0.013	0.020	0.013	0.018	0.011	3.03
20	Uniform,t(5)	1000	0.021	0.013	-.005	0.012	0.013	0.009	0.013	0.009	0.012	0.008	3.03
21	Uniform,Normal	50	0.054	0.036	-.031	0.060	0.067	0.045	0.073	0.046	0.049	0.032	1.23
22	Uniform,Normal	100	0.037	0.025	-.016	0.037	0.040	0.027	0.042	0.028	0.034	0.023	1.23
23	Uniform,Normal	300	0.021	0.014	-.006	0.020	0.021	0.014	0.020	0.014	0.019	0.013	1.23
24	Uniform,Normal	500	0.017	0.011	-.004	0.015	0.016	0.010	0.016	0.011	0.015	0.010	1.23
25	Uniform,Normal	1000	0.012	0.008	-.002	0.011	0.011	0.008	0.011	0.008	0.011	0.007	1.23
26	LogNormal,t(5)	50	0.357	0.148	-.443	0.169	0.474	0.436	0.753	0.301	0.169	0.094	4.04
27	LogNormal,t(5)	100	0.226	0.115	-.276	0.127	0.303	0.267	0.354	0.171	0.118	0.071	4.04
28	LogNormal,t(5)	300	0.126	0.071	-.127	0.073	0.146	0.123	0.185	0.079	0.066	0.045	4.04
29	LogNormal,t(5)	500	0.101	0.056	-.085	0.051	0.100	0.083	0.139	0.057	0.051	0.033	4.04
30	LogNormal,t(5)	1000	0.077	0.041	-.051	0.036	0.062	0.049	0.061	0.035	0.039	0.024	4.04
31	LogNormal,Normal	50	0.326	0.134	-.344	0.185	0.390	0.324	0.869	0.290	0.088	0.059	14.59
32	LogNormal,Normal	100	0.242	0.110	-.188	0.127	0.227	0.176	0.423	0.150	0.064	0.043	14.59
33	LogNormal,Normal	300	0.138	0.074	-.064	0.052	0.083	0.061	0.121	0.051	0.036	0.024	14.59
34	LogNormal,Normal	500	0.099	0.052	-.037	0.035	0.051	0.037	0.076	0.032	0.028	0.019	14.59
35	LogNormal,Normal	1000	0.080	0.042	-.018	0.022	0.029	0.020	0.039	0.019	0.020	0.013	14.59
36	LogNormal,Uniform	50	0.351	0.132	-.246	0.174	0.301	0.211	1.208	0.260	0.057	0.038	35.72
37	LogNormal,Uniform	100	0.391	0.106	-.107	0.098	0.145	0.087	0.296	0.110	0.039	0.028	35.72
38	LogNormal,Uniform	300	0.115	0.069	-.026	0.031	0.041	0.027	0.055	0.027	0.023	0.015	35.72
39	LogNormal,Uniform	500	0.092	0.059	-.015	0.022	0.027	0.018	0.031	0.018	0.018	0.013	35.72
40	LogNormal,Uniform	1000	0.096	0.045	-.007	0.014	0.016	0.010	0.019	0.011	0.013	0.009	35.72
41	Exp,t(5)	50	0.121	0.081	-.235	0.121	0.264	0.234	0.249	0.152	0.121	0.081	1.00
42	Exp,t(5)	100	0.086	0.059	-.140	0.082	0.162	0.140	0.134	0.082	0.086	0.059	1.00
43	Exp,t(5)	300	0.053	0.033	-.062	0.049	0.079	0.065	0.072	0.043	0.053	0.033	1.00
44	Exp,t(5)	500	0.053	0.026	-.042	0.038	0.057	0.045	0.060	0.034	0.053	0.026	1.00
45	Exp,t(5)	1000	0.027	0.018	-.024	0.026	0.035	0.026	0.031	0.020	0.027	0.018	1.00
46	Exp,Normal	50	0.101	0.066	-.162	0.118	0.200	0.158	0.193	0.120	0.081	0.053	1.56
47	Exp,Normal	100	0.071	0.050	-.081	0.071	0.107	0.081	0.101	0.063	0.057	0.041	1.56
48	Exp,Normal	300	0.041	0.029	-.028	0.036	0.046	0.033	0.041	0.027	0.034	0.023	1.56
49	Exp,Normal	500	0.032	0.021	-.017	0.027	0.031	0.022	0.029	0.020	0.026	0.018	1.56
50	Exp,Normal	1000	0.022	0.014	-.008	0.018	0.019	0.013	0.018	0.012	0.017	0.011	1.56
51	Exp,Uniform	50	0.095	0.065	-.086	0.092	0.126	0.078	0.163	0.091	0.054	0.036	3.03
52	Exp,Uniform	100	0.066	0.044	-.036	0.048	0.060	0.041	0.069	0.041	0.038	0.026	3.03
53	Exp,Uniform	300	0.039	0.025	-.010	0.024	0.026	0.018	0.026	0.018	0.023	-0.015	3.03
54	Exp,Uniform	500	0.030	0.021	-.007	0.018	0.019	0.012	0.019	0.012	0.017	0.011	3.03
55	Exp,Uniform	1000	0.021	0.014	-.002	0.012	0.012	0.008	0.012	0.008	0.012	0.008	3.03
56	Exp,LogNormal	50	0.316	0.154	-.471	0.163	0.498	0.472	0.698	0.320	0.170	0.122	4.04
57	Exp,LogNormal	100	0.241	0.105	-.323	0.137	0.351	0.321	0.503	0.198	0.119	0.075	4.04
58	Exp,LogNormal	300	0.139	0.073	-.154	0.085	0.175	0.154	0.237	0.100	0.070	0.047	4.04
59	Exp,LogNormal	500	0.108	0.060	-.100	0.063	0.118	0.100	0.129	0.066	0.055	0.036	4.04
60	Exp,LogNormal	1000	0.067	0.041	-.057	0.043	0.071	0.056	0.067	0.041	0.037	0.024	4.04

Notes
1. Estimates based on 1,000 replications.
2. The relative efficiency of EWMD (column 11) is the ratio of the asymptotic variances of EWMD and OMD.

Table 3 Performance of Estimators of a Variance Parameter Using Correlated Variances and Covariances

				EWMD		OMD				IWOMD		Rel. Eff.
Row	Dist.	Obs.	Cov	Std., RMSE (1)	MAE (2)	Bias (3)	Std. (4)	RMSE (5)	MAE (6)	Std., RMSE (7)	MAE (8)	(9)
1	t(5)	50	0.00	0.117	0.075	-.280	0.099	0.297	0.283	0.332	0.160	1.00
2	t(5)	50	0.25	0.118	0.074	-.284	0.097	0.300	0.287	0.258	0.160	1.06
3	t(5)	50	0.50	0.140	0.078	-.282	0.099	0.299	0.286	0.286	0.159	1.17
4	t(5)	100	0.00	0.098	0.055	-.172	0.071	0.186	0.177	0.133	0.081	1.00
5	t(5)	100	0.25	0.088	0.058	-.174	0.070	0.188	0.174	0.133	0.087	1.06
6	t(5)	100	0.50	0.091	0.059	-.173	0.073	0.188	0.174	0.124	0.079	1.17
7	t(5)	300	0.00	0.055	0.031	-.081	0.042	0.091	0.081	0.079	0.042	1.00
8	t(5)	300	0.25	0.052	0.032	-.081	0.041	0.091	0.082	0.063	0.040	1.06
9	t(5)	300	0.50	0.054	0.034	-.078	0.044	0.090	0.081	0.062	0.039	1.17
10	t(5)	500	0.00	0.038	0.024	-.056	0.033	0.065	0.057	0.047	0.032	1.00
11	t(5)	500	0.25	0.041	0.025	-.056	0.033	0.065	0.057	0.051	0.031	1.06
12	t(5)	500	0.50	0.045	0.027	-.053	0.034	0.063	0.054	0.050	0.031	1.17
13	t(5)	1000	0.00	0.029	0.018	-.033	0.024	0.041	0.034	0.034	0.021	1.00
14	t(5)	1000	0.25	0.030	0.018	-.033	0.024	0.041	0.035	0.034	0.021	1.06
15	t(5)	1000	0.50	0.031	0.019	-.031	0.025	0.040	0.033	0.034	0.020	1.17
16	t(10)	50	0.00	0.076	0.050	-.191	0.091	0.212	0.189	0.194	0.121	1.00
17	t(10)	50	0.25	0.082	0.056	-.191	0.092	0.212	0.189	0.192	0.127	1.15
18	t(10)	50	0.50	0.093	0.061	-.197	0.095	0.219	0.196	0.204	0.126	1.32
19	t(10)	100	0.00	0.055	0.036	-.106	0.062	0.123	0.104	0.084	0.054	1.00
20	t(10)	100	0.25	0.059	0.039	-.108	0.060	0.123	0.108	0.084	0.054	1.15
21	t(10)	100	0.50	0.067	0.045	-.111	0.065	0.128	0.115	0.088	0.061	1.32
22	t(10)	300	0.00	0.030	0.021	-.040	0.032	0.051	0.041	0.037	0.025	1.00
23	t(10)	300	0.25	0.035	0.024	-.040	0.033	0.052	0.042	0.039	0.026	1.15
24	t(10)	300	0.50	0.038	0.025	-.041	0.035	0.054	0.042	0.041	0.028	1.32
25	t(10)	500	0.00	0.023	0.016	-.025	0.024	0.035	0.026	0.028	0.019	1.00
26	t(10)	500	0.25	0.026	0.018	-.026	0.025	0.036	0.027	0.029	0.020	1.15
27	t(10)	500	0.50	0.029	0.020	-.025	0.027	0.037	0.028	0.029	0.020	1.32
28	t(10)	1000	0.00	0.017	0.012	-.011	0.017	0.021	0.015	0.019	0.013	1.00
29	t(10)	1000	0.25	0.018	0.012	-.012	0.018	0.021	0.015	0.019	0.013	1.15
30	t(10)	1000	0.50	0.021	0.015	-.013	0.019	0.023	0.016	0.020	0.014	1.32
31	Normal	50	0.00	0.061	0.041	-.140	0.081	0.162	0.139	0.159	0.099	1.00
32	Normal	50	0.25	0.073	0.050	-.143	0.085	0.166	0.140	0.162	0.102	1.22
33	Normal	50	0.50	0.081	0.054	-.157	0.090	0.181	0.157	0.170	0.111	1.42
34	Normal	100	0.00	0.046	0.031	-.074	0.053	0.091	0.072	0.069	0.047	1.00
35	Normal	100	0.25	0.049	0.032	-.070	0.053	0.088	0.071	0.068	0.047	1.22
36	Normal	100	0.50	0.060	0.041	-.081	0.057	0.099	0.084	0.074	0.047	1.42
37	Normal	300	0.00	0.026	0.018	-.024	0.028	0.037	0.028	0.030	0.021	1.00
38	Normal	300	0.25	0.028	0.020	-.024	0.027	0.036	0.026	0.030	0.021	1.22
39	Normal	300	0.50	0.034	0.023	-.027	0.030	0.040	0.031	0.033	0.021	1.42
40	Normal	500	0.00	0.020	0.013	-.015	0.021	0.026	0.018	0.022	0.015	1.00
41	Normal	500	0.25	0.022	0.015	-.014	0.021	0.025	0.018	0.022	0.015	1.22
42	Normal	500	0.50	0.026	0.017	-.016	0.023	0.028	0.020	0.025	0.017	1.42
43	Normal	1000	0.00	0.014	0.010	-.007	0.014	0.016	0.011	0.015	0.011	1.00
44	Normal	1000	0.25	0.015	0.010	-.007	0.014	0.016	0.011	0.015	0.010	1.22
45	Normal	1000	0.50	0.018	0.011	-.009	0.016	0.018	0.012	0.016	0.011	1.42

Table 3 Performance of Estimators of a Variance Parameter Using Correlated Variances and Covariances

				EWMD		OMD				IWOMD		Rel. Eff.
Row	Dist.	Obs.	Cov	Std., RMSE (1)	MAE (2)	Bias (3)	Std. (4)	RMSE (5)	MAE (6)	Std., RMSE (7)	MAE (8)	(9)
46	Uniform	50	0.00	0.041	0.027	-.052	0.056	0.076	0.057	0.101	0.062	1.00
47	Uniform	50	0.25	0.050	0.033	-.052	0.056	0.076	0.051	0.102	0.065	1.53
48	Uniform	50	0.50	0.068	0.046	-.082	0.068	0.106	0.084	0.118	0.076	1.78
49	Uniform	100	0.00	0.028	0.019	-.022	0.033	0.039	0.028	0.038	0.026	1.00
50	Uniform	100	0.25	0.038	0.025	-.023	0.034	0.041	0.029	0.038	0.026	1.53
51	Uniform	100	0.50	0.049	0.032	-.037	0.044	0.058	0.042	0.051	0.034	1.78
52	Uniform	300	0.00	0.017	0.011	-.007	0.018	0.019	0.014	0.019	0.012	1.00
53	Uniform	300	0.25	0.021	0.014	-.007	0.017	0.018	0.012	0.018	0.012	1.53
54	Uniform	300	0.50	0.027	0.019	-.012	0.022	0.025	0.017	0.023	0.016	1.78
55	Uniform	500	0.00	0.013	0.009	-.005	0.013	0.014	0.009	0.013	0.009	1.00
56	Uniform	500	0.25	0.016	0.010	-.004	0.013	0.014	0.009	0.014	0.009	1.53
57	Uniform	500	0.50	0.021	0.015	-.007	0.017	0.018	0.012	0.017	0.011	1.78
58	Uniform	1000	0.00	0.009	0.006	-.002	0.009	0.009	0.006	0.009	0.006	1.00
59	Uniform	1000	0.25	0.011	0.007	-.002	0.009	0.010	0.007	0.010	0.006	1.53
60	Uniform	1000	0.50	0.015	0.009	-.003	0.011	0.012	0.007	0.011	0.007	1.78
61	LogNormal	50	0.00	0.415	0.217	-.688	0.096	0.695	0.694	1.146	0.390	1.00
62	LogNormal	50	0.25	0.375	0.208	-.683	0.097	0.690	0.691	0.900	0.388	1.01
63	LogNormal	50	0.50	0.401	0.214	-.660	0.093	0.667	0.671	0.833	0.365	1.08
64	LogNormal	100	0.00	0.277	0.169	-.551	0.098	0.560	0.552	0.559	0.269	1.00
65	LogNormal	100	0.25	0.382	0.163	-.550	0.100	0.559	0.558	0.562	0.267	1.01
66	LogNormal	100	0.50	0.320	0.164	-.535	0.094	0.543	0.546	0.470	0.254	1.08
67	LogNormal	300	0.00	0.192	0.103	-.360	0.086	0.371	0.368	0.330	0.153	1.00
68	LogNormal	300	0.25	0.200	0.105	-.352	0.086	0.362	0.357	0.325	0.149	1.01
69	LogNormal	300	0.50	0.175	0.102	-.338	0.082	0.348	0.343	0.253	0.140	1.08
70	LogNormal	500	0.00	0.136	0.081	-.283	0.074	0.293	0.285	0.197	0.113	1.00
71	LogNormal	500	0.25	0.138	0.087	-.280	0.075	0.290	0.286	0.206	0.118	1.01
72	LogNormal	500	0.50	0.165	0.078	-.269	0.073	0.278	0.274	0.244	0.107	1.08
73	LogNormal	1000	0.00	0.094	0.055	-.197	0.061	0.207	0.199	0.138	0.084	1.00
74	LogNormal	1000	0.25	0.110	0.060	-.196	0.062	0.205	0.198	0.138	0.081	1.01
75	LogNormal	1000	0.50	0.190	0.058	-.188	0.059	0.197	0.190	0.256	0.081	1.08
76	Exp	50	0.00	0.128	0.082	-.387	0.124	0.406	0.393	0.309	0.198	1.00
77	Exp	50	0.25	0.129	0.087	-.377	0.121	0.396	0.378	0.320	0.202	1.06
78	Exp	50	0.50	0.132	0.090	-.353	0.113	0.370	0.357	0.296	0.187	1.17
79	Exp	100	0.00	0.090	0.065	-.236	0.089	0.253	0.237	0.155	0.097	1.00
80	Exp	100	0.25	0.092	0.063	-.232	0.090	0.249	0.233	0.170	0.113	1.06
81	Exp	100	0.50	0.100	0.066	-.225	0.093	0.243	0.230	0.152	0.093	1.17
82	Exp	300	0.00	0.052	0.034	-.100	0.054	0.114	0.104	0.072	0.049	1.00
83	Exp	300	0.25	0.051	0.032	-.098	0.052	0.111	0.096	0.072	0.048	1.06
84	Exp	300	0.50	0.056	0.039	-.094	0.053	0.108	0.095	0.074	0.048	1.17
85	Exp	500	0.00	0.040	0.025	-.065	0.042	0.077	0.065	0.052	0.035	1.00
86	Exp	500	0.25	0.040	0.026	-.064	0.043	0.077	0.066	0.055	0.039	1.06
87	Exp	500	0.50	0.045	0.029	-.061	0.042	0.074	0.062	0.052	0.034	1.17
88	Exp	1000	0.00	0.028	0.019	-.035	0.029	0.045	0.035	0.035	0.022	1.00
89	Exp	1000	0.25	0.030	0.021	-.035	0.030	0.046	0.035	0.034	0.024	1.06
90	Exp	1000	0.50	0.031	0.021	-.033	0.029	0.044	0.033	0.035	0.024	1.17

Notes to Table 3

1. Estimates based on 1,000 replications.
2. The relative efficiency of EWMD (column 9) is the ratio of the asymptotic variance of EWMD to the asymptotic variance of OMD.
3. Data are generated as $D_p = (Z_p + \rho Z_{p+1}) * (1 + \rho)^{-0.5}$ for p = 1 to 10 from a set of mean 0, variance 1, i.i.d. random variables Z from the specified distribution. The column labelled 'Cov' reports the $Cov(D_p, D_{p+1}) = \theta\rho/(1+\rho^2)$.
4. We assume ρ is known, so the 10 sample variances and 9 first order covariances are functions of the single unknown population variance θ.

Table 4 Estimates of the Abowd-Card Stationary Covariance Structure of the Changes in Log Hours and Log Earnings for a PSID Eleven Year Sample

	LAG: $\tau=0$	LAG: $\tau=1$	LAG: $\tau=2$
$Cov(\Delta Earnings_t, \Delta Earnings_{t-\tau})$			
EWMD	0.175	-0.060	-0.008
OMD	0.086	-0.026	-0.008
IWOMD	0.149	-0.060	-0.007
EWMD (Abowd and Card)	0.172	-0.060	-0.007
$Cov(\Delta Hours_t, \Delta Hours_{t-\tau})$			
EWMD	0.131	-0.047	-0.006
OMD	0.060	-0.022	-0.005
IWOMD	0.111	-0.043	-0.005
EWMD (Abowd and Card)	0.117	-0.035	-0.011
$Cov(\Delta Hours_t, \Delta Earnings_{t-\tau}))$			
EWMD	0.080	-0.026	-0.008
OMD	0.026	-0.002	-0.005
IWOMD	0.060	-0.024	-0.006
EWMD (Abowd and Card)	0.073	-0.023	-0.006
$Cov(\Delta Earnings_t, \Delta hours_{t-\tau})$			
EWMD	0.080	-0.024	-0.002
OMD	0.026	-0.005	-0.003
IWOMD	0.060	-0.016	-0.001
EWMD (Abowd and Card)	0.073	-0.020	-0.002

Table 5 Simulation of a True Covariance Model of the Change in Log Hours and Earnings Based on Eleven Year PSID Sample

	EWMD				OMD				IWOMD	
	Std., RMSE (1)	MAE (2)	t-Test (3)	BootTest (4)	Bias (5)	Std. (6)	RMSE (7)	MAE (8)	Std., RMSE (9)	MAE (10)
Var(ΔEarnings$_t$)	0.100	0.066	0.142	0.112	-.722	0.042	0.723	0.726	0.249	0.161
Cov(ΔEarnings$_t$,ΔEarnings$_{t-1}$)	0.150	0.099	0.134	0.120	-.737	0.048	0.738	0.742	0.289	0.193
Cov(ΔEarnings$_t$,ΔEarnings$_{t-2}$)	0.501	0.348	0.098	0.106	-.811	0.147	0.824	0.820	1.057	0.638
Var(ΔHours$_t$)	0.113	0.076	0.122	0.112	-.727	0.043	0.729	0.730	0.246	0.168
Cov(ΔHours$_t$,ΔHours$_{t-1}$)	0.156	0.107	0.126	0.130	-.726	0.048	0.727	0.728	0.279	0.195
Cov(ΔHours$_t$,ΔHours$_{t-2}$)	0.410	0.288	0.098	0.096	-.770	0.113	0.778	0.775	0.806	0.520
Var(ΔHours$_t$)	0.163	0.113	0.134	0.128	-.802	0.047	0.803	0.806	0.292	0.190
Cov(ΔHours$_t$,ΔEarnings$_{t-1}$)	0.254	0.180	0.120	0.116	-.794	0.058	0.797	0.799	0.376	0.262
Cov(ΔHours$_t$,ΔEarnings$_{t-2}$)	0.434	0.302	0.124	0.124	-.857	0.098	0.863	0.864	0.685	0.412
Var(ΔEarnings$_t$,ΔHours$_{t-1}$)	0.283	0.197	0.126	0.132	-.814	0.058	0.816	0.818	0.354	0.245
Cov(ΔEarnings$_t$,ΔHours$_{t-2}$)	0.597	0.419	0.138	0.142	-.846	0.106	0.852	0.851	0.757	0.503

Notes

1. Estimates based on 500 replications.
2. The column labeled "t-Test" reports the rejection rate for an asymptotic t-test of the hypothesis that the estimate equals the true value at the 90 % confidence level.
3. The column labeled "Boot Test" reports the rejection rate for a test of whether a 90% symmetric bootstrap confidence interval contains the true parameter.

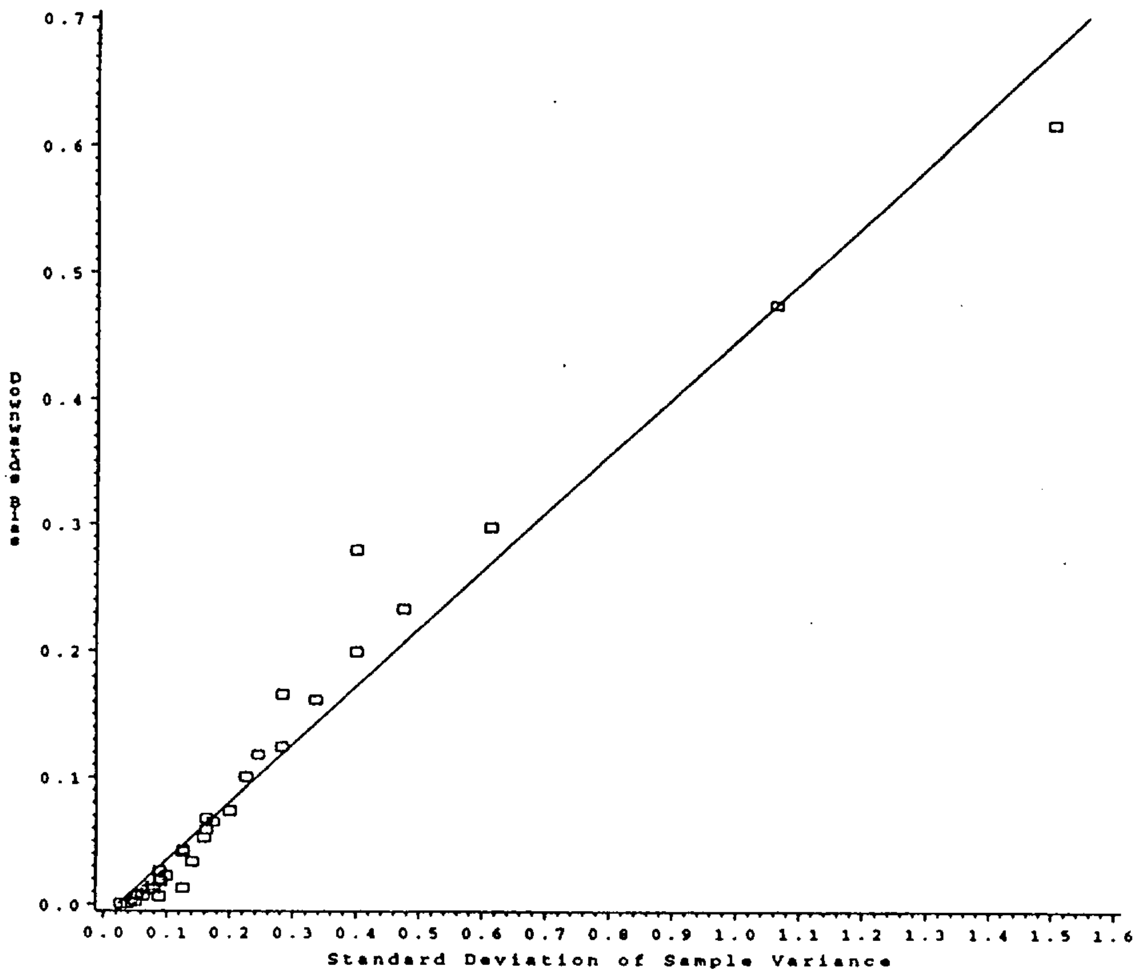
Table 6 Comparison of Inference in OMD, EWMD and IWOMD Estimation
Simulations using 10 Moments from Two Distributions

Distribution 1	t(5)	Normal	Normal	Uniform	LogNorm.	t(5)	Normal	Normal	Uniform	LogNorm.
Distribution 2	Normal	Uniform	Exp.	Exp.	Exp.	Normal	Uniform	Exp.	Exp.	Exp.
Observations	100	100	100	100	100	500	500	500	500	500
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
EQUALLY WEIGHTED										
True SE	0.0662	0.0376	0.0699	0.0650	0.2420	0.0303	0.0168	0.0321	0.0297	0.1002
True 5 %tile	0.9023	0.9387	0.8923	0.8992	0.7596	0.9554	0.9722	0.9496	0.9529	0.8731
True 95 %tile	1.1117	1.0611	1.1227	1.1111	1.3468	1.0522	1.0283	1.0537	1.0497	1.1702
Mean Boot SE	0.0594	0.0370	0.0679	0.0621	0.1720	0.0287	0.0167	0.0313	0.0290	0.0867
Mean Boot 5 %tile	0.8935	0.9273	0.8808	0.8898	0.7622	0.9512	0.9689	0.9478	0.9498	0.8745
Mean Boot 95 %tile	1.0884	1.0489	1.1038	1.0938	1.3167	1.0454	1.0238	1.0506	1.0452	1.1559
Reject t test	0.1184	0.0960	0.1110	0.1116	0.2014	0.0942	0.1050	0.1062	0.1064	0.1634
Reject Boot test	0.1620	0.1340	0.1200	0.1500	0.2440	0.0920	0.1180	0.1060	0.0980	0.1600
INDEPENDENTLY WEIGHTED										
True SE	0.0934	0.0404	0.0999	0.0676	0.4529	0.0297	0.0157	0.0289	0.0187	0.1240
True 5 %tile	0.8719	0.9318	0.8485	0.8895	0.5783	0.9545	0.9746	0.9529	0.9689	0.8478
True 95 %tile	1.1497	1.0656	1.1759	1.1101	1.6520	1.0503	1.0260	1.0474	1.0304	1.2068
Mean Boot SE	0.0791	0.0422	0.1002	0.0752	0.3479	0.0283	0.0157	0.0290	0.0189	0.1048
Mean Boot 5 %tile	0.8394	0.9204	0.8287	0.8932	0.5670	0.9366	0.9679	0.9361	0.9607	0.8171
Mean Boot 95 %tile	1.0156	1.0473	1.0149	1.0314	0.7974	1.0269	1.0267	1.0311	1.0265	0.9807
Reject Boot test	0.2520	0.1460	0.2120	0.1840	0.3120	0.1420	0.1240	0.1360	0.1460	0.2840
OPTIMALLY WEIGHTED										
True 5 %tile	0.8299	0.9233	0.8047	0.8758	0.4534	0.9412	0.9721	0.9402	0.9640	0.7930
True 95 %tile	1.0257	1.0424	1.0322	1.0373	0.9010	1.0221	1.0226	1.0272	1.0226	1.0019
Reject t test	0.5030	0.1874	0.4948	0.3350	0.9254	0.2280	0.1180	0.2086	0.1484	0.6868

Notes
1. True values correspond to the average across 5,000 replications. All other estimates are based on 1,000 replications.
2. The row labeled 't-Test' reports the rejection rate for an asymptotic t-test of the hypothesis that the estimate equals the true value at the 90% confidence level.
3. The row labeled 'Boot Test' reports the rejection rate for a test of whether a 90% symmetric bootstrap confidence interval contains the true parameter.

Figure 1 Relationship Between Sampling Variance and OMD Bias

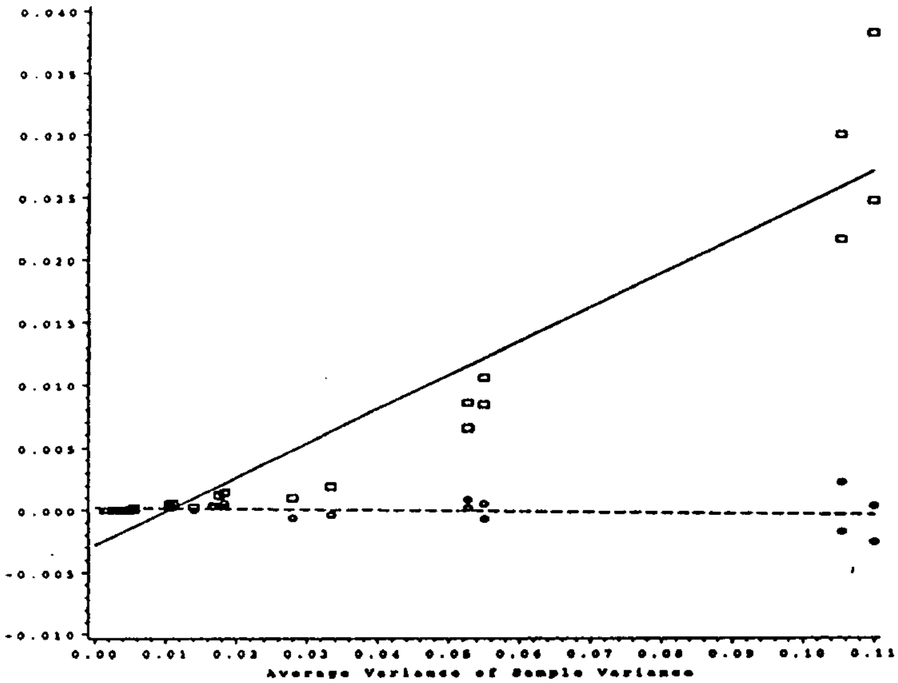
Based on 10 Moments from 7 Distributions and 5 Sample Sizes



Predicted Bias = -0.01 + 0.46 * Standard Deviation of Sample Variance
Adjusted R Squared 0.95

Figure 2a Decomposition of the Mean Squared Error of EWMD and OMD

Experiments with efficiency gains between 1.125 and 1.5



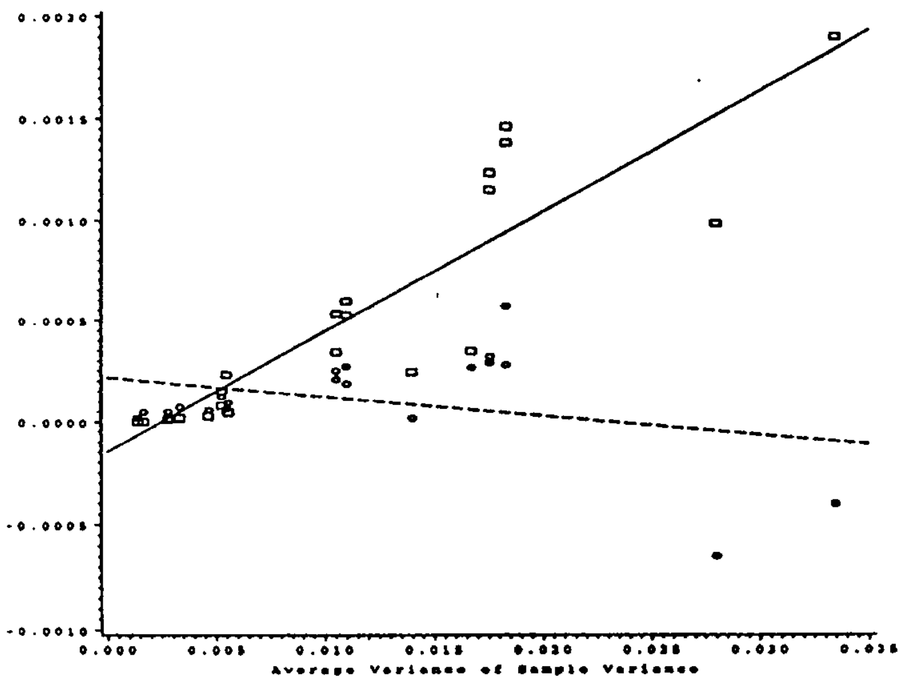
Square: Solid Line = Bias Squared

Circle: Dashed Line = Difference in Variance = V(EWMD)-V(OMD)

Figure 2b Decomposition of the Mean Squared Error of EWMD and OMD

Experiments with efficiency gains between 1.125 and 1.5

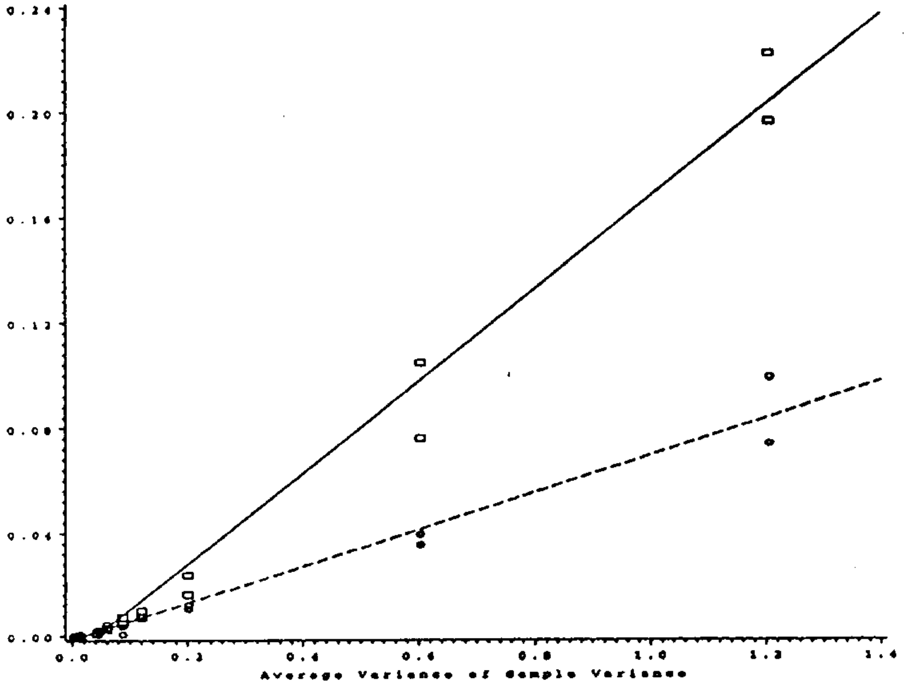
Average variance of variance between 0.0 and 0.04



Square: Solid Line = Bias Squared

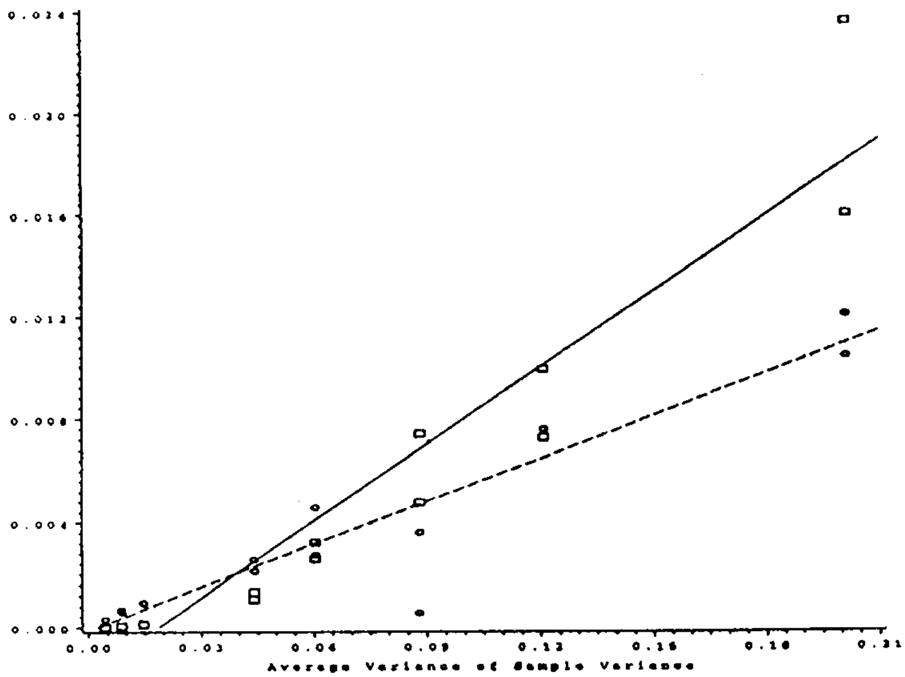
Circle: Dashed Line = Difference in Variance = V(EWMD)-V(OMD)

Figure 2c Decomposition of the Mean Squared Error of EWMD and OMD
Experiments with efficiency gains between 3.03 and 4.04



Square: Solid Line = Bias Squared
Circle: Dashed Line = Difference in Variance = V(EWMD)-V(OMD)

Figure 2d Decomposition of the Mean Squared Error of EWMD and OMD
Experiments with efficiency gains between 3.03 and 4.04
Average variance of variance between 0.0 and 0.2[illegible]



Square: Solid Line = Bias Squared
Circle: Dashed Line = Difference in Variance = V(EWMD)-V(OMD)