

NBER TECHNICAL PAPER SERIES

CONSISTENT COVARIANCE MATRIX
ESTIMATION WITH CROSS-SECTIONAL
DEPENDENCE AND HETEROSKEDASTICITY
IN CROSS-SECTIONAL FINANCIAL DATA

Kenneth A. Froot

Technical Working Paper No. 62

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1987

The research reported here is part of the NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Consistent Covariance Matrix Estimation with Cross-Sectional Dependence
and Heteroskedasticity in Cross-Sectional Financial Data

ABSTRACT

This paper provides a simple method to account for heteroskedasticity and cross-sectional dependence in samples with large cross sections and relatively few time series observations. The estimators we derive are motivated by cross-sectional regression studies in finance and accounting. Simulation evidence suggests that these estimators are dependable in small samples and may be useful when generalized least squares is infeasible, unreliable, or computationally too burdensome. The approach allows a relatively small number of time series observations to yield a rich characterization of cross-sectional correlations. We also consider efficiency issues and show that in principle asymptotic efficiency can be improved using a technique due to Cragg (1983).

Kenneth A. Froot
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139

**Consistent Covariance Matrix Estimation with
Cross-Sectional Dependence and Heteroskedasticity
in Cross-Sectional Financial Data**

Kenneth A. Froot

Sloan School of Management,

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

Practically every empirical study in accounting and finance using cross-sectional data must in some way come to grips with the problem of contemporaneous correlation across firms (or portfolios). Such studies present special estimation problems because they typically employ panel data with a large number of firms, each with a less-than-desirable number of time series observations. With data sets of these dimensions, it is often impractical to implement standard techniques for the entire cross section and correct for contemporaneous correlation. Even when these techniques can be used, they often require homoskedasticity, a strong assumption in stock-based-return studies. Indeed, in spite of a rapidly growing literature documenting the severity of and proposing alternative solutions to these problems, there are no techniques designed for large cross-sections which address both contemporaneous correlation and heteroskedasticity simultaneously, and which effectively cope with the relative paucity of time series data.

This paper attempts to provide two relatively simple estimators to account simultaneously for contemporaneous correlation and heteroskedasticity in the cross section as well as heteroskedasticity over time. The first technique we develop, which combines elements of method-of-moments estimation pioneered by Hansen (1982) and the treatment of heteroskedasticity introduced by Eicker (1967) and White (1980a, b), is consistent and asymptotically efficient within the class of one-step estimators. The second technique uses a two-step instrumental variable estimator, similar to that in Cragg (1983). It can provide greater asymptotic efficiency than the more straightforward method-of-moments approach. Small-sample monte carlo simulations are performed for each of the estimators.

1. A Discussion of Alternative Estimation Techniques

It is worth stressing at the outset the very broad range of studies which must conduct inference in the presence of contemporaneous correlation and heteroskedasticity. Contemporaneous correlation appears most prominently in cross-sectional tests of the CAPM, going back to Black, Jensen and Scholes (1972) and continuing on through Gibbons (1982) and Brown and Weinstein (1983). These tests, which use generalized least-squares (GLS) or iterative GLS techniques, are extremely constrained in the time series dimension (the putative interval over which parameters of the market model are reasonably constant is five years).¹ Aggregation is also common in time-series, cross-section studies which use GLS, such as French, Ruback and Schwert (1983). Here the problem is not so much the lack of time series data as it is the computational difficulty of inverting very large contemporaneous correlation matrixes. Both contemporaneous correlation and heteroskedasticity are significant problems in the rapidly growing "event study" and "cross-sectional-return study" literatures. In these areas, the difficulty of accounting for both problems has resulted in some authors using techniques which ignore contemporaneous correlation (Brown and Warner (1980, 1985), Christie (1985), and Malatesta and Thompson (1985)) and others using techniques which ignore heteroskedasticity (Collins and Dent (1984), Shipper and Thompson (1982, 1983, 1985), and Malatesta (1986)).²

In the absence of heteroskedasticity, it is well known that GLS techniques (such as seemingly unrelated squares, SUR or multivariate regressions, MVRM) offer first-best solutions to problems of cross-sectional dependence. A researcher armed with a great deal of time series data may be fairly confident that the asymptotic distribution of the GLS estimator is a good approximation to the sample distribution. In practice, however, even when the homoskedasticity assumption is met, GLS is often infeasible, either because it requires estimation of too many parameters in the error covariance matrix, or because it involves the inversion of a very large matrix of contemporaneous correlations. When the cross section is small enough to make GLS feasible, inferences based on the asymptotic distribution of the regression coefficients are appropriate only if the error covariance

¹This five year limit usually necessitates the use of a small cross section (less than 30 firms or portfolios). The number of monthly time periods in five years is 60. Thus the error covariance matrix will be nonsingular only if the cross section (N) is composed of fewer than 60 portfolios. GLS, however, uses the inverted error covariance matrix. The expectation of this matrix exists only if $N \leq \frac{T-1}{2}$, i.e. if $N \leq 30$. See Press (1972).

²See Bernard (1986b), who draws the distinction between event and cross-sectional-return studies, for additional references.

matrix is a function of a relatively small number of parameters.³

When the cross section is, for whatever reason, "too large," results are usually obtained by aggregating individual firms into a small enough number of portfolios to permit estimation by GLS. Although this approach allows efficient estimation in the subsample, the associated loss in efficiency from aggregation removes any *a priori* argument for choosing GLS over less efficient estimators. Thus, while the estimators below are less efficient than GLS with an unlimited supply of time series data, they may nevertheless have more power because they may be applied to data sets for which GLS is infeasible, unreliable, or computationally too burdensome. As we shall see below, greater disaggregation is possible under the method-of-moments approach because: (i) the parameter estimates are linear functions of the residuals, so that one can have more confidence in the validity of the finite-sample inferences, (ii) the size of the cross section is not limited to the number of time series observations, and (iii) the estimators do not require formation or inversion of the error covariance matrix.

In many empirical studies in finance, the problem is not so much that the cross section is "too large," but that the time series is "too short:" no amount of aggregation will make GLS a reasonable procedure. This frequently occurs in what Bernard (1986b) terms "cross-sectional return" studies that use firm-specific information available on an annual or quarterly basis only.⁴ In these studies it is clearly most difficult to correct for cross-sectional correlation. Yet, at the same time, they seem to have a disproportionate need for such a correction. Bernard shows that the usual OLS standard errors of the coefficient estimates, which are biased in the presence of contemporaneous correlation, may be extremely misleading for these cases. In addition, Bernard demonstrates that this bias can only increase as the cross section is expanded. This finding dispels the notion that the bias will be attenuated by choosing a well-diversified portfolio. Unfortunately, no general method of accounting for cross-sectional dependence in these studies has previously emerged.

In all of these situations, a richer characterization of cross-sectional dependencies can be achieved only by imposing some restrictions on the data. We will make the assumption that the researcher can identify groups of firms within the cross section that are likely to exhibit little

³The finite-sample bias in the variance of the regression coefficients is of the order of the number of observations. Rothenberg (1984) notes that this bias may often be substantial; this will especially be the case when there are relatively many parameters in the error covariance matrix to be estimated.

⁴A partial listing of these studies would include Beaver, Eger, Ryan and Wolfson (1985), Bernard (1986a), Rayburn (1986) and Ricks (1986).

contemporaneous correlation in their residuals. That is, we assume that at any given time the cross-sectional correlation matrix of the residuals has a block diagonal structure. While this assumption is restrictive, it buys us considerable freedom in estimating intra-group (or intra-industry) correlations using method-of-moments techniques. For example, even with a single time series observation it will be possible to account for unrestricted intra-industry correlations in each industry, provided a reasonable number of industries are used. Indeed, for any given number of industries, there is no fixed upper limit on the number of firms within each industry.

The assumption that the error covariance matrix is block diagonal is common in a variety of contexts in finance. For example, several approximate arbitrage pricing models which allow the pricing error covariance matrix to include contemporaneous correlation across assets require that the cross-sectional idiosyncratic covariances be "nonpervasive." This slightly more general property often comes down to assuming block diagonality in practice.⁵ Also, empirical evidence on block diagonality is supportive: Bernard (1986b) finds that while intra-industry correlations of market model residuals are quite large in quarterly and annual data, the corresponding inter-industry correlations are near zero (the correlation coefficient is on average 0.06) and have a small effect on the standard errors in regression. Thus, our assumption appears natural for similar studies in finance and accounting.

While the issue of cross-sectional correlation is a difficult one, a priority is to develop estimators, such as those below, which allow simultaneously for cross-sectional correlation as well as conditional heteroskedasticity. In view of the large and predictable changes in variance frequently observed in stock returns, conditional heteroskedasticity is likely to be important. It is particularly relevant in event studies, in which the variance during event periods is likely to be systematically different than the variance in nonevent periods. The ability to permit both contemporaneous correlation and unrestricted heteroskedasticity is an important shortcoming of feasible GLS estimators.

In sum, there are a number of advantages to the techniques presented below when applied to financial data. First, they are feasible when GLS and related techniques are not. Because they exploit the independence across industries and because they do not require the formation or inversion of the error covariance matrix, larger cross sections may be used. Second, by permitting

⁵See, for example, Ingersol (1984) and Chamberlain and Rothschild (1983).

more cross-sectional disaggregation than GLS, more independent information is incorporated in estimation. Consequently, these estimators need not be less efficient in practice than GLS. Third, the techniques are valid in the presence of stochastic regressors, and heteroskedastic and contemporaneously correlated residuals. (By comparison GLS is generally valid only in the presence of contemporaneous correlation.) Finally, the estimators and their standard errors are both easy to compute and dependable when the size of the cross section grows large relative to that of the time-series dimension.

The remainder of the paper proceeds as follows. In section 2 below, we lay out the method-of-moments covariance estimator and prove its asymptotic consistency. There we show that we can use the independence across industries to increase multiplicatively the number of independent observations, thereby improving the approximation to the asymptotic standard errors. In the third section, we consider an alternative two-step estimator which can increase the asymptotic efficiency of initial estimator. Section 4 then presents monte carlo simulations of both estimators for a variety of sample sizes, levels of contemporaneous correlation and conditional heteroskedasticity. Last, section 5 offers our conclusions.

2. The Heteroskedastic- and Contemporaneous-Correlation-Consistent

Covariance Estimator.

Consider the linear model

$$y_{nt} = x_{nt}\beta + \mu_{nt}, \tag{1}$$

where $n = 1, \dots, N$ and $t = 1, \dots, T$ index a particular industry and time period, respectively. Within each industry group are P firms, so that y_{nt} is a $P \times 1$ vector and x_{nt} is a $P \times K$ matrix:

$$y_{nt} = \begin{pmatrix} y_{1nt} \\ y_{2nt} \\ \vdots \\ y_{Pnt} \end{pmatrix} \quad x_{nt} = \begin{pmatrix} x_{1nt} \\ x_{2nt} \\ \vdots \\ x_{Pnt} \end{pmatrix}$$

We also make the following assumptions:

Assumption 1. (i) The pair (x_{pnt}, μ_{pnt}) is independently (but not necessarily identically) distributed over time, t , and industries, n . (ii) The regressors are taken to be uncorrelated with their associated residual, so $E(x_{pnt}^i \mu_{pnt}) = 0$, where $i = 1, \dots, K$ denotes the i th column of x_{pnt} . (iii) The $P \times 1$ residual vector, μ_{nt} , has a conditional expectation of zero, $E(\mu_{nt} | x_{nt}) = 0$, and $P \times P$ conditional covariance matrix given by $E(\mu_{nt} \mu_{nt}' | x_{nt}) = \Sigma_{nt}$, for all n and t .

Assumption 2. (i). The second moments of the residuals are bounded, in the sense that there exist positive constants, δ and Δ such that $E(\mu_{pnt}^2)^{1+\delta} < \Delta$. Also, each element of the $K \times K$ outer product of the regressors is bounded: $E(x_{pnt}^i x_{pnt}^j)^{1+\delta} < \Delta$, for all $i, j = 1, \dots, K$. (ii). The product of the second moments is assumed to be bounded, $E(\mu_{pnt}^2 x_{pnt}^i x_{pnt}^j)^{1+\delta} < \Delta$, as is the square of the sum of squared regressors, $E(x_a^2 x_b^2)^{2+2\delta} \leq \Delta$. (iii). $M_{NT} = (NT)^{-1} E(\mathbf{x}'_{NT} \mathbf{x}_{NT})$ is assumed to be nonsingular for sufficiently large NT , where \mathbf{x}_{NT} is a $PNT \times K$ vector of stacked regressors. (v). We define the average covariance matrix,

$$\Theta_{NT} = (NT)^{-1} \sum_{n=1}^N \sum_{t=1}^T x_{nt}' \mu_{nt} \mu_{nt}' x_{nt}. \quad (2)$$

Note that for the case in which the x 's are nonstochastic, Assumption 2(i) implies 2(ii).

We develop our estimation strategy allowing for stochastic regressors, which require a slightly more complicated notation than the fixed regressor case. The added difficulty is nevertheless worthwhile since it highlights the generality of the method; indeed, competing methods, such as GLS, generally yield incorrect inferences when the regressors are stochastic or when conditional heteroskedasticity is present.

Given our assumptions, it is straightforward to show that the OLS estimate of the unknown coefficient vector, $\hat{\beta}_{NT} = (\mathbf{x}'_{NT} \mathbf{x}_{NT})^{-1} \mathbf{x}'_{NT} \mathbf{y}_{NT}$, converges almost surely as $NT \rightarrow \infty$ to the true parameter vector in equation (1), β (for a proof under similar assumptions, see Lemma 2 in White, 1980a). The usual asymptotic OLS covariance matrix, $(PNT)^{-1} \sum_{p=1}^P \sum_{n=1}^N \sum_{t=1}^T E(\mu_{pnt}^2) M_{NT}^{-1}$, however, will be incorrect since we allow for contemporaneous correlation across firms in the same industry and for heteroskedasticity. In general, this type of correlation may be accounted for by

using a multivariate generalization of White's (1980a) estimator of the covariance matrix:

$$V(\hat{\beta}_{NT}) = (NT)^{-1} M_{NT}^{-1} \Theta_{NT} M_{NT}^{-1}. \quad (3)$$

Under the assumption that the inter-industry cross correlations are zero (Assumption 1(i)), the overall conditional error covariance matrix is $\Omega_{NT} = E(\mu_{NT} \mu'_{NT} | \mathbf{x}_{NT})$ (where μ_{NT} is an $PNT \times 1$ vector of residuals). This matrix has a block diagonal structure, with each block measuring the inter-firm contemporaneous correlations for a given industry at a given time:

$$\Omega_{NT} = \begin{pmatrix} \Sigma_{11} & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \Sigma_{1T} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & \Sigma_{N1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & \Sigma_{NT} \end{pmatrix} \quad (4)$$

where Σ_{nt} is the $P \times P$ conditional covariance matrix given in Assumption 1(iii), with (i, j) th element:

$$\sigma_{nt}^{ij} = E(\mu_{int} \mu_{jnt} | x_{nt}) \quad i, j = 1, \dots, P. \quad (5)$$

The difficulty clearly comes in estimating the error covariance matrix Ω_{NT} . With T large and under homoskedasticity, there is no problem; the industry covariance matrixes, Σ_{nt} , would then be independent of time and could be estimated consistently from the OLS residuals:

$$\hat{\sigma}_n^{ij} = \frac{1}{T} \sum_{t=1}^T \hat{\mu}_{int} \hat{\mu}_{jnt} : \quad i, j = 1, \dots, P. \quad (6)$$

where the $\hat{\mu}$'s are the OLS residuals from equation (1).

When T is not large in comparison with P , however, it would seem that there are far too many parameters in the error covariance matrix (P^2 elements for each Σ_{nt} , or $P^2 NT$ elements altogether) relative to the number of residuals (PNT) to allow the asymptotics to apply. Due to the independence of the regressors and disturbances across T and N , the number of parameters in the

error covariance matrix is irrelevant. We need be concerned only with the parameters in the average covariance matrix, Θ_{NT} , and not the estimation of each of the intra-industry contemporaneous correlation matrixes, Σ_{nt} . Thus, we must estimate only K^2 parameters.⁶ This intuition for handling contemporaneous correlation within a segment of the cross section is precisely analogous to the treatment of heteroskedasticity in White (1980a). To see this reasoning, note that the average covariance matrix can be written:

$$\Theta_{NT} = \frac{1}{N} \sum_{n=1}^N \left[\frac{1}{T} \sum_{t=1}^T \left[\sum_{i=1}^P \sum_{j=1}^P E(x_{int} x'_{jnt} \mu_{int} \mu_{jnt}) \right] \right]. \quad (7)$$

Equation (7) says that we average over the N industries and T time periods the outer products of the sums of intra-industry regressors weighted by the appropriate covariance. Thus, the average error covariance matrix depends only on the number of regressors K , instead of on the number of elements in Ω_{NT} or Σ_{nt} , or the number of time series observations. For the special case in which cross-sectional dependence does not occur, $T = P = 1$, equation (7) is the estimate of the average covariance matrix proposed by White (1980a) to handle conditional heteroskedasticity.

We can now state two propositions:

Lemma 1. Under assumptions 1 and 2, the standardized distribution of the OLS estimate, $\hat{\beta}_{NT}$, is:

$$\sqrt{NT} \Theta_{NT}^{-1/2} M_{NT} (\hat{\beta}_{NT} - \beta) \xrightarrow{d} N(0, \mathbf{I}_K).$$

Theorem 1. The OLS estimate of the coefficient covariance matrix converges almost surely to the coefficient covariance matrix in equation (3):

$$\left(\frac{\mathbf{x}'_{NT} \mathbf{x}_{NT}}{NT} \right)^{-1} \hat{\Theta}_{NT} \left(\frac{\mathbf{x}'_{NT} \mathbf{x}_{NT}}{NT} \right)^{-1} \xrightarrow{a.s.} M_{NT}^{-1} \Theta_{NT} M_{NT}^{-1}.$$

⁶The symmetry of Ω_{NT} and Σ_{nt} implies the usual reduction in the number of independent elements to $\frac{(m+1)m}{2}$, where m is the number of rows or columns.

The appendix contains the proofs of these propositions.⁷ Note that the asymptotics are conducted over NT . By allowing the number of industries and/or the number of time series observations to become large while holding fixed the number of firms within an industry, we get precise approximations to the asymptotic standard errors, even though the number of elements in the error covariance matrix increases with N and T . Note also that while P cannot grow with the sample size, there is no fixed upper limit to the number of firms per industry. As long as one can isolate a reasonable number of groups of firms for which the maintained hypothesis of no inter-group correlation is appropriate, consistent inferences may be drawn in the presence of contemporaneous correlation and heteroskedasticity, and large cross sections of data may be used. Indeed, when the size of the cross section is increased by raising the number of industry groups, the asymptotic approximation to the standard errors improves.

⁷After writing this paper, I became aware of important results in White (1984) which have received inadequate attention, and which are closely related to those here. See White (1984, Chapter 7).

3. More Efficient Estimation

A drawback of the estimator above is that it gives equal weight to each data point. Ideally we would like to form a more efficient estimator, one which weights data according to its precision. A more precise data point, i.e., an observation with a relatively small squared residual, deserves greater than equal weight. The problem, of course, is that we have only a single estimate of the true squared residual for each observation, $\hat{\mu}_{pnt}^2$, and this is a biased and inconsistent estimate of μ_{pnt}^2 . An instrumental variables solution to this problem in the presence of heteroskedasticity alone has been proposed by Cragg (1983). Here we show that Cragg's basic insight may be directly applied to improve asymptotic efficiency in the presence of contemporaneous correlation as well.

Consider a $PNT \times R$ matrix of instruments (where $K \leq R \leq NT$) for the regressors in equation (1), $\mathbf{q} = [\mathbf{x} \ \mathbf{p}]$, where \mathbf{q} includes the regressors, \mathbf{x} , in addition to other variables, \mathbf{p} . We make the following assumption concerning \mathbf{q} :

Assumption 3. Let Assumptions 1, 2(i) and 2(ii) hold when \mathbf{x} is replaced by \mathbf{q} , and define $R_{NT} = (NT)^{-1} E(\mathbf{q}'\mathbf{x})$, which is of rank K for sufficiently large NT . $P_{NT} = (NT)^{-1} E(\mathbf{p}'\mathbf{x})$ which is of rank $R - K$, and

$$\Theta_{NT}^{xp} = (NT)^{-1} \sum_{n=1}^N \sum_{t=1}^T \sum_{i=1}^P \sum_{j=1}^P E(x_{int} p'_{jnt} \mu_{int} \mu_{jnt}),$$

which is of rank K for sufficiently large NT .

Next consider the following transformation of equation (1):

$$\mathbf{y}^* = \mathbf{x}^* \boldsymbol{\beta} + \boldsymbol{\mu}^*. \quad (8)$$

where $\mathbf{y}^* = \mathbf{q}'\mathbf{y}$, $\mathbf{x}^* = \mathbf{q}'\mathbf{x}$, and $\boldsymbol{\mu}^* = \mathbf{q}'\boldsymbol{\mu}$. (Where it will not create confusion, we drop the NT subscript for the remainder of the text.) The error covariance matrix for equation (8) is now the $R \times R$ matrix $E(\boldsymbol{\mu}^* \boldsymbol{\mu}^{*\prime}) = E(\mathbf{q}' \boldsymbol{\mu} \boldsymbol{\mu}' \mathbf{q}) = \Theta^*$, which has uniformly bounded elements by Assumption

3 and has full rank. Given our boundedness and continuity assumptions, we can make the usual GLS transformation by premultiplying equation (8) by $(\Theta^*)^{-1/2}$, which then yields:

$$\mathbf{y}^{**} = \mathbf{x}^{**}\beta + \mu^{**}, \quad (9)$$

where $\mathbf{y}^{**} = (\Theta^*)^{-1/2}\mathbf{q}'\mathbf{y}$, $\mathbf{x}^{**} = (\Theta^*)^{-1/2}\mathbf{q}'\mathbf{x}$, $\mu^{**} = (\Theta^*)^{-1/2}\mathbf{q}'\mu$, and $E(\mu^{**}\mu^{**'}) = \mathbf{I}_R$. By running OLS on equation (9), we obtain the standard formula for the Two-Step Two-Stage Least Squares (2STLS) estimator due to Cumby, Huizinga and Obstfeld (1983):

$$\hat{\beta}^{**} = (\mathbf{x}^{**'}\mathbf{x}^{**})^{-1}\mathbf{x}^{**'}\mathbf{y}^{**} = (\mathbf{x}'\mathbf{q}(\Theta^*)^{-1}\mathbf{q}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{q}(\Theta^*)^{-1}\mathbf{q}'\mathbf{y}. \quad (10)$$

Given that \mathbf{x}^{**} and μ^{**} satisfy the boundedness conditions in Assumptions 2(i) and 2(ii), we can apply Lemma 1 to get the asymptotic distribution of $\hat{\beta}^{**}$:

$$\sqrt{NT} (\Theta^*)^{-1/2} P_{NT}(\hat{\beta}^{**} - \beta) \xrightarrow{d} N(0, \mathbf{I}_K). \quad (11)$$

The 2STLS estimator, $\hat{\beta}^{**}$, employs the average error covariance of the instruments, Θ^* , which can be estimated consistently from the data. To see this, note that we can use the estimated OLS residuals from equation (1) to form:

$$\hat{\Theta}_{NT}^* = (NT)^{-1} \sum_{n=1}^N \sum_{t=1}^T \sum_{i=1}^P \sum_{j=1}^P q_{int} q'_{jnt} \hat{\mu}_{int} \hat{\mu}_{jnt}. \quad (12)$$

The assumption of independence across N and T implies that, for NT sufficiently large, $\hat{\Theta}_{NT}^*$ is of full rank for $R \leq NT$.

This leads to a third proposition:

Theorem 2. The OLS estimate of the coefficient covariance matrix in equation (9), $\hat{V}(\hat{\beta}^{**})$ converges almost surely to the true 2STLS covariance matrix:

$$\left(\frac{\mathbf{x}'\mathbf{q}}{NT}\right)(\hat{\Theta}_{NT}^*)^{-1}\left(\frac{\mathbf{q}'\mathbf{x}}{NT}\right) \xrightarrow{a.s.} R'_{NT}(\Theta^*)^{-1}_{NT}R_{NT}$$

See the appendix for the proof, which follows directly from Assumption 3 and Theorem 1. The important intuition is that we can use the independence of the q_{pnt} 's and μ_{pnt} 's over T and N to estimate consistently the average error covariance matrix of the instrumental variables.

Now that we have established that the 2STOLS estimator, $\hat{\beta}^{**}$, will be feasible and provide consistent inferences, we consider the conditions for β^{**} to be asymptotically more efficient than the equally weighted OLS estimator from equation (1), $\hat{\beta}$. In other words, we ask when the difference between the asymptotic covariance matrixes,

$$M_{NT}^{-1}\Theta_{NT}M_{NT}^{-1} - (R'_{NT}(\Theta^*_{NT})^{-1}R_{NT})^{-1}, \quad (13)$$

will be positive semi-definite. Cragg (1983) and White (1984, proposition 4.49) show that, as long as \mathbf{q} contains \mathbf{x} and is of greater rank than \mathbf{x} , the asymptotic efficiency of $\hat{\beta}^{**}$ is at least as great as that of $\hat{\beta}$, and strictly greater if and only if:

$$\Theta_{NT}^{-1}\Theta_{NT}^{zp} \neq M_{NT}^{-1}P_{NT}. \quad (14)$$

The intuition behind this result can best be seen by rewriting equation (14) as:

$$\left(E(\mathbf{x}'\mu\mu'\mathbf{x})\right)^{-1}E(\mathbf{x}'\mu\mu'\mathbf{p}) \neq \left(E(\mathbf{x}'\mathbf{x})\right)^{-1}E(\mathbf{x}'\mathbf{p}). \quad (15)$$

The left-hand side of equation (15) is roughly the coefficient in a regression of $\mu'\mathbf{p}$ on $\mu'\mathbf{x}$, and the right-hand side is the coefficient in a regression of \mathbf{p} on \mathbf{x} . The interpretation is that we seek instruments \mathbf{p} which tell us something about the interaction of the regressors with the error terms. A "good" instrument is not judged by its high correlation with the regressors themselves, instead, it provides information about the true underlying residuals and their interaction with the regressors – information which cannot be obtained by looking at the correlations between the regressors and instruments alone.

Additional insight can be gained from considering more closely the case of fixed regressors, as in Cragg (1983). The assumption of nonstochastic regressors is sufficient to imply that GLS is asymptotically efficient. Cragg shows that the preferred instruments are those that explain the greatest amount of variation in the (inverted) GLS covariance matrix, $\mathbf{x}'\Omega\mathbf{x}$. Ideally, the instruments would be chosen to minimize the mean squared residual in the OLS regression:

$$\Omega^{-1/2}\mathbf{x} = \Omega^{1/2}\mathbf{q}\gamma + \epsilon \quad (16)$$

The estimated sum of square residuals from this regression, $\hat{\epsilon}'\hat{\epsilon}$, is a direct measure of the efficiency of the instrumental variable estimator since:

$$V(\hat{\beta}^{gls})^{-1} - V(\hat{\beta}^{**})^{-1} = \mathbf{x}'\boldsymbol{\Omega}^{-1}\mathbf{x} - \mathbf{x}'\mathbf{q}(\mathbf{q}'\boldsymbol{\Omega}\mathbf{q})^{-1}\mathbf{q}'\mathbf{x} = E(\hat{\epsilon}'\hat{\epsilon}). \quad (17)$$

Equation (16) may be rewritten,

$$\boldsymbol{\Omega}^{1/2}(\boldsymbol{\Omega}^{-1}\mathbf{x}) = \boldsymbol{\Omega}^{1/2}\mathbf{q}\gamma + \epsilon. \quad (18)$$

Equation (18) says that good instruments are those that will explain the variation in $\boldsymbol{\Omega}^{-1}\mathbf{x}$. To address the problem of heteroskedasticity in the μ 's, Cragg treats $\boldsymbol{\Omega}^{-1}\mathbf{x}$ as an arbitrary function of \mathbf{x} and suggests using a polynomial approximation. That is, the i th row of \mathbf{q} , or q_i , would contain the instruments $x_i, x_i x_i',$ etc.⁸ (see Cragg (1983) for more detail).

Our foremost concern in this paper is with the additional problems posed by the presence of contemporaneous correlation in financial data. Fortunately, additional information on cross-sectional dependence is relatively easy to find. In many finance and accounting studies, the dependent variable, \mathbf{y} , is an asset return or a suitably defined excess return. For this variable, there is usually a large amount of data available prior to the period of estimation. So we might include in \mathbf{p} the variables $(\boldsymbol{\Omega}^H)^{-1}\mathbf{x}$, where $\boldsymbol{\Omega}^H$ is the (historical) correlation matrix of the dependent variable calculated during an earlier period. Indeed, one frequently hears informal recommendations in support of using $\boldsymbol{\Omega}^H$ to form the OLS covariance matrix in equation (3). Of course, there is little basis in theory for such a substitution. There is no reason why the covariance matrix of the dependent variable should be similar to the covariance matrix of the true residuals, even with *contemporaneous* data. Nevertheless, these informal recommendations reflect an implicit frustration about the inability to use such an obvious source of information. Thus, the 2STOLS technique has the important advantage of incorporating the information contained in historical correlations in a systematic way, so that the asymptotic efficiency of the standard errors may be improved.

Bernard (1986b), for example, presents evidence on inter- and intra-industry contemporaneous correlations of market model residuals from stock market data. While he finds that inter-industry correlations are on average small, ranging from 0.01 (daily data) to 0.06 (annual data), he reports intra-industry correlations which average from 0.04 to 0.30. The markedly higher correlations

⁸If a constant is included in the columns of \mathbf{x} , we assume that it is excluded in the higher powers, so that $\mathbf{q}'\mathbf{q}$ is not singular.

within industries corresponds closely with what we might expect to find in the error covariances themselves. Market model residuals may thus contain potent information about the true error covariance structure, which could be exploited using our 2STLS technique.

4. Simulation Experiments

Now that the theoretical properties of our estimators are clear, we wish to obtain some evidence about their performance in finite samples under a variety of circumstances. The simulations below are designed to address three issues regarding the estimators we have proposed. First, suppose we were to circumvent the problem of cross-sectional correlation by aggregating the cross section over intra-industry observations (so that the cross section would consist of only N aggregated industries). We would like to know how much can be gained by using the method-of-moments estimator on the entire cross section in comparison with a naïve OLS estimator on the aggregated sample. Second, how efficient are the estimators presented in sections 2 and 3 in comparison with GLS, and how sensitive is the efficiency differential to different levels of N , T and P ? Third, can we hope to improve efficiency in small samples by using the 2STLS estimator in section 3 instead of the simpler method-of-moments estimator in section 2? Finally, since all of our results are based on asymptotic distribution theory, we also wish to determine whether there are systematic finite-sample biases in either of these estimators.

A linear model with a single exogenous regressor was chosen for the simulations:

$$y_{pnt} = \beta_0 + \beta x_{pnt} + \mu_{pnt}. \quad (19)$$

The x_{pnt} were drawn randomly from a log-normal distribution, and are independent across p , n and t . In the simulation and the notation that follows, we treat the regressors as fixed. A set of independently distributed primitive disturbance terms, η_{pnt} , were drawn from a standardized normal distribution, then modified to generate contemporaneous correlation and heteroskedasticity in the residuals, μ_{pnt} . Contemporaneous intra-industry correlation was assumed to be identical across industries, and was introduced by transforming the $PNT \times 1$ column vector, η :

$$\mu = (\mathbf{I}_{NT} \otimes \Delta)\eta, \quad (20)$$

where \mathbf{I}_k is a $k \times k$ identity matrix, and Δ is a nonstochastic $P \times P$ matrix:

$$\Delta = \begin{pmatrix} 1 & \lambda & \dots & \lambda \\ \lambda & 1 & \dots & \lambda \\ \vdots & \vdots & \ddots & \vdots \\ \lambda & \lambda & \dots & 1 \end{pmatrix},$$

and λ is a function of the correlation coefficient between firms in a given industry, ρ , and the number of firms per industry, P :

$$\lambda = \frac{1 - ((1 - \rho)(1 + \rho(P - 1)))^{1/2}}{(P - 1)(\rho - 1) + 1}. \quad (21)$$

Note that this specification of the residuals implies homoskedasticity both over time and across the panel. Since the error in estimating the parameters is not a function of the parameters themselves, β_0 and β can be chosen arbitrarily without affecting the results. We selected $\beta_0 = 0$ and $\beta = 1$. For each experiment, 1000 draws of η were used to compute 1000 sets of dependent variables and, using a single set of regressors, 1000 sets parameter estimates were obtained. Only the estimates for the slope parameter are reported in order to conserve space.

Table 1 reports the results for the method-of-moments estimator in section 2 for a variety of sample sizes when the correlation coefficient in equation (21), ρ , is set to one-half. For purposes of comparison, the variance estimates are reported as proportions of the OLS asymptotic variance for the data set aggregated up to the industry level:

$$V(\hat{\beta}_{OLS}^A) = \mathbf{i}'_{PNT}(\mathbf{I}_{NT} \otimes \Delta)\mathbf{i}_{PNT}(\mathbf{i}'_P(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})\mathbf{i}_P)^{-1}, \quad (22)$$

where $\tilde{\mathbf{x}}$ is a $NT \times P$ matrix containing the elements of \mathbf{x} , and \mathbf{i}_k is a $k \times 1$ vector of ones.⁹ The column labeled "asymptotic" is the appropriate diagonal element of the asymptotic method-of-moments variance divided by the corresponding element of the aggregated OLS asymptotic variance.¹⁰

$$\frac{(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'(\mathbf{I}_{NT} \otimes \Delta)\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}}{\mathbf{i}'_{PNT}(\mathbf{I}_{NT} \otimes \Delta)\mathbf{i}_{PNT}(\mathbf{i}'_P(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})\mathbf{i}_P)^{-1}}. \quad (23)$$

⁹ Under our extreme assumption that the regressors are independent across p , the usual OLS variance of the parameters using the disaggregated data, $\sum_{i=1}^{PNT} \mu_i(\mathbf{x}'\mathbf{x})^{-1}$, will be asymptotically consistent. We use the OLS variance on the intraindustry aggregation of the data for comparison, however, since in practice this assumption is not likely to hold.

¹⁰ To simplify the notation, we assume that the mean of both x and y are zero, so that the variance matrix of the parameters can be treated as a scalar.

Clearly, when the number of firms per industry, P , is equal to one, there is no asymptotic gain from the method-of-moments approach, since the two estimators are identical. Here the method-of-moments estimator is precisely that of White (1980a). For completeness, the last column in Table 1 reports the asymptotic GLS variance as a proportion of the aggregated OLS variance:

$$\frac{(\mathbf{x}'(\mathbf{I}_{NT} \otimes \Delta^{-1})\mathbf{x})^{-1}}{\mathbf{i}'_{PNT}(\mathbf{I}_{NT} \otimes \Delta)\mathbf{i}_{PNT}(\mathbf{i}'_P(\bar{\mathbf{x}}'\bar{\mathbf{x}})\mathbf{i}_P)^{-1}} \quad (24)$$

With $P = 1$ there is also no gain from using GLS since there is no cross-sectional correlation or heteroskedasticity to exploit.

When P is increased to 5, however, substantial improvements in efficiency from the method-of-moments estimator appear. For $N = 12$ and $T = 1$, the relative asymptotic efficiency of the method-of-moments estimator falls to 0.037 and that of the GLS estimator falls to 0.025. This gain in efficiency over OLS comes from two separate effects. The first effect is a result of aggregating the data before running OLS. Since the cross section for OLS contains only 1/5th of the total number of observations, we would expect the disaggregated data to improve the relative asymptotic efficiency by a factor of about 5.¹¹ The second effect, which explains the increase in efficiency from .20 to .037, is that the method-of-moments variance exploits the cross-sectional dependence within industries. This second source of efficiency gain is positively related to the degree of cross-sectional correlation, with no gains occurring in the complete absence of cross-sectional correlation. It also becomes increasingly important relative to the aggregation effect as P becomes larger. The degree of reduction in the relative variance of the method-of-moments estimator attributable solely to contemporaneous correlation increases from a factor of approximately 5 (.20/.0364) when $P = 5$, to a factor of approximately 25 (.05/.00193) when $P = 20$. Finally, while GLS is most efficient in all cases, its additional efficiency is not much greater when the industry groups are larger. Unfortunately, the gains in efficiency from GLS are not generally feasible with so few time series observations.

The second column for each set of simulation results in Table 1, marked "simple", gives the mean squared deviation of the estimated method-of-moments parameter (obtained by running OLS

¹¹ This improvement in efficiency is a function of the cross-sectional correlation of the regressors. If there is positive correlation across the regressors, the improvement factor will be less than 5, and if there is negative correlation, the factor will be greater than 5.

on the disaggregated cross section) from the true value, scaled by the asymptotic OLS variance:

$$\frac{(\hat{\beta} - 1)^2}{\mathbf{i}'_{PNT}(\mathbf{I}_{NT} \otimes \Delta)\mathbf{i}_{PNT}(\mathbf{i}'_P(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})\mathbf{i}_P)^{-1}} \quad (25)$$

On average the method-of-moments estimator appears to achieve asymptotic efficiency in our small samples, regardless of the relative sizes of P , N and T . All of the estimates of the simple variance measure are easily within two standard deviations of the asymptotic variances.

Finally, the third column in Table 1 reports "estimated" variance measures, which are calculated as the mean estimated method-of-moments variance relative to the asymptotic OLS variance:

$$\frac{(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'(\mathbf{I}_T \otimes \hat{\Delta})\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}}{\mathbf{i}'_{NT}(\mathbf{I}_{NT} \otimes \Delta)\mathbf{i}_{NT}(\mathbf{i}'_P(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})\mathbf{i}_P)^{-1}} \quad (26)$$

where $\hat{\Delta}$ is the estimate of the $NP \times NP$ contemporaneous correlation matrix, $(\mathbf{I}_N \widehat{\otimes} \Delta)$.¹²

In the simulations with the smallest samples, the estimated variance measure is significantly downward biased. Such finite sample bias occurs because the average covariance matrix, $\sum_{i=1}^{NT} x_i x_i' \hat{\mu}_i^2$ is a biased estimator of $\sum_{i=1}^{NT} E(\mu_i^2) \sum_{i=1}^{NT} x_i x_i'$, even when the residuals are homoskedastic. It may be corrected using Rao's MINQUE procedure. Several authors have noted the presence of bias in estimated method-of-moments covariance matrices when the residuals are heteroskedastic. The fact that the bias persists even under homoskedasticity reveals a hidden danger in the practice of reporting heteroskedasticity-consistent covariance estimators without prior evidence that conditional heteroskedasticity is present in the data. Thus, the White (1980a) correction for heteroskedasticity should be applied *only* in those cases where the null hypothesis of no conditional heteroskedasticity is rejected. When the null hypothesis cannot be rejected, only the usual OLS standard errors are both asymptotically valid and free of finite-sample bias.

The results in Table 1 suggest, however, that for any given number of independent observations, the bias of the method-of-moments estimated variance is actually smaller when P is larger. Indeed, the bias is more severe for the case in which $T = 10$, $N = 25$ and $P = 1$ (with a total of 250 independent observations), than in the case in which $T = 1$, $N = 12$ and $P = 5$ (with a total of only 12 independent observations)! The bias contracts when P is larger because of the averaging over P of each of the NT independent observations used to construct the estimated average covariance

¹² In computing equation (26), we do not impose the restriction that the Δ matrix is equal across industries.

matrix. When P is raised to 20, the small-sample bias can no longer be detected in the simulations, even with a very low number of independent observations ($T = 1, N = 12$). These results suggest that there is no reason to limit the number of firms per industry through aggregation or pre-selection. The estimated method-of-moments variance seems to become more reliable when it is most needed: in data sets with very large cross-sections, but not many industry groups or time series observations.

Of course, the results reported in Table 1 are for the ideal case in which the data contain no conditional heteroskedasticity across industry groups or over time. While these assumptions may be a reasonable first approximation, in many data sets they are likely to be rejected. We therefore relax these assumptions in Table 2, where the simulated data contain cross-sectional heteroskedasticity. Only this type of heteroskedasticity is analyzed here because its presence is exactly analogous to heteroskedasticity in the time domain, under our assumptions of independence across industries and over time. We add heteroskedasticity by assuming that the residual variances are linearly related to the sum of squared regressors in each industry group:

$$\sigma_n^2 = E(\mu_{jnt}^2) = \delta_0 + \delta_1 \sum_{i=1}^P x_{int}^2, \quad \forall j = 1, \dots, P, \quad \forall t = 1, \dots, T. \quad (27)$$

In the simulations of Table 2, δ_0 is set equal to 1, and $\delta_1 = 0.2$. Thus the asymptotic covariance matrix of the residuals becomes:

$$\mathbf{I}_T \otimes \tilde{\Delta} = \mathbf{I}_T \otimes \begin{pmatrix} \Delta_1 & 0 & \dots & 0 \\ 0 & \Delta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Delta_N \end{pmatrix}. \quad (27')$$

where $\tilde{\Delta}$ is $NP \times NP$, and Δ_n is the $P \times P$ contemporaneous correlation matrix for the n th group. Since the asymptotic OLS variance employed in Table 1 is not appropriate under these conditions, we use instead the White (1980a) asymptotic correction for heteroskedasticity on the aggregated data as a scaling measure for the method-of-moments variances.

Turning to the results in Table 2, note first that the asymptotic variances yield similar improvements in efficiency over the White technique when they are compared with the relative asymptotic

variances reported in Table 1. Second, in contrast with the homoskedastic case, GLS provides larger improvements in efficiency in Table 2. Third, the simple variance measure in the second column is consistently within two standard deviations of the asymptotic variances in all of the simulations. Even in small samples, the method-of-moments estimator exploits much of the information about the residuals. Fourth, the actual estimated variance still contains a distinct downward bias. The heteroskedasticity makes this bias somewhat more severe than that reported in Table 1. In the simulations with larger data sets, a statistically significant bias of between 3 and 10 percent remains. The bias is inversely related to the size of P . Once again, the MINQUE technique of Rao could be applied to eliminate the bias in the estimated variance measure.

Table 3 is intended to help gauge the sensitivity of the results given in Tables 1 and 2 to alternative assumptions regarding both the severity of the contemporaneous correlation within industries and the degree of heteroskedasticity. Three different levels of contemporaneous correlation are considered, with correlation coefficient, ρ , set to 0.1, 0.5 and 0.9. The degree of conditional heteroskedasticity is adjusted by allowing δ_1 in equation (27) to take on the values 0.0, 0.2, 1.0 and 4.0 indicated by heteroskedasticity levels 1 through 4, respectively, in Table 3. For purposes of comparison, a single sample size ($T = 1$, $N = 25$ and $P = 5$) is chosen.

Table 3 shows that the asymptotic gains in efficiency from the method-of-moments approach do not increase substantially as the heteroskedasticity becomes more extreme, but that they do increase with the level of contemporaneous correlation. In contrast, the relative GLS asymptotic variance noticeably improves when either contemporaneous correlation or the level conditional heteroskedasticity increases. The amount of bias in the estimated method-of-moment variance also displays a clear positive response to higher levels of heteroskedasticity and cross-sectional correlation. When the level of heteroskedasticity is 4 and $\rho = 0.9$, the average estimated variance is only slightly more than 50 percent of its asymptotic value. Once again, this bias is less pronounced when either the number of industries or the number of firms per industry is increased.

We turn next to simulations of the 2STSLS estimator. Here there is a preliminary step in which instruments for $\Omega^{-1}\mathbf{x}$ are constructed. Since in many financial studies, additional observations of the dependent variable are usually available, for each simulation we construct a fake set of "historical" observations on y_{pnt} according to equation (19). As in the earlier simulations, the

contemporaneous correlation in the historical disturbances is induced by using equation (20), and cross-sectional heteroskedasticity is induced by using equation (27). We then form the historical covariance matrix of the dependent variable for each of the industry groups:

$$\hat{\Sigma}_n^H = \frac{1}{T^H} \sum_{t=1}^{T^H} \mathbf{y}_n \mathbf{y}_n' \quad \forall n = 1, \dots, N. \quad (28)$$

where T^H is the number of previous time series observations on the dependent variable, and \mathbf{y}_n is a $P \times 1$ vector. To be conservative, we assumed relatively few past observations are available, setting $T^H = 2p + 1$. The entire historical covariance matrix is then:

$$\hat{\Omega}^H = \mathbf{I}_T \otimes \begin{pmatrix} \hat{\Sigma}_1^H & 0 & \dots & 0 \\ 0 & \hat{\Sigma}_2^H & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\Sigma}_N^H \end{pmatrix}. \quad (29)$$

The $PNT \times 2$ estimated matrix of instruments becomes:

$$\hat{\mathbf{q}} = [\mathbf{x} \quad (\hat{\Omega}^H)^{-1} \mathbf{x}]. \quad (30)$$

Table 4 reports measures of the relative performance of the 2STOLS estimator. We use the same levels of cross-sectional correlation ($\rho = 0.5$) and conditional heteroskedasticity ($\delta_1 = 0.2$) as in Table 2. All variance measures are reported relative to the asymptotic variance of the method-of-moments estimator. Thus the column marked asymptotic gives the ratio:

$$\frac{\mathbf{x}' \hat{\mathbf{q}} (\hat{\mathbf{q}}' (\mathbf{I}_T \otimes \tilde{\Delta}) \hat{\mathbf{q}})^{-1} \hat{\mathbf{q}}' \mathbf{x}}{(\mathbf{x}' \mathbf{x})^{-1} \mathbf{x}' (\mathbf{I}_T \otimes \tilde{\Delta}) \mathbf{x} (\mathbf{x}' \mathbf{x})^{-1}}. \quad (31)$$

Note that the asymptotic variance uses the *estimated* instruments (i.e. the estimated historical covariance matrix of the dependent variables), denoted by $\hat{\mathbf{q}}$. The results in Table 4 suggest that gains in efficiency may be obtained by using 2STOLS even with relatively few industry groups and time series observations. For example, With $T = 1$, $N = 12$ and $P = 10$, the asymptotic 2STOLS estimator is 11 (1/.1109) times more efficient than the method-of-moments estimator. Indeed, these

gains seem to be a substantial portion of the asymptotic efficiency gains from using GLS, which are reported in the fourth column.

The “simple” variance measure is the mean squared deviation of the estimated 2STLS coefficient from its true value, $(\hat{\beta}^{**} - 1)^2$ divided by the asymptotic method-of-moments variance. A two-step procedure is required to calculate $\hat{\beta}^{**}$. Consistent estimates of the residuals in equation (19) are obtained by running OLS. A consistent estimate of the average covariance matrix, $\hat{\Delta}$, is formed, and then used to as a weighting matrix to compute $\hat{\beta}^{**}$. In about half of the simulations in Table 4, simple variances are significantly greater than the asymptotic variances.

The “estimated” variances in the third columns are given by:

$$\frac{\mathbf{x}'\hat{\mathbf{q}}(\hat{\mathbf{q}}'(\mathbf{I}_T \otimes \hat{\Delta})\hat{\mathbf{q}})^{-1}\hat{\mathbf{q}}'\mathbf{x}}{(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'(\mathbf{I}_T \otimes \hat{\Delta})\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}} \quad (32)$$

In the smaller data sets, the reported values are significantly less than both the simple and the asymptotic variance measures. It is worth pointing out that when there is no contemporaneous correlation in the cross section (i.e., when $P = 1$), we have the univariate case of the 2STLS estimator investigated by Cragg (1983). Cragg also reports that the estimated variances of the 2STLS estimator were significantly less than their asymptotic values. However, Table 4 shows that when P is greater than 1, this downward bias is again reduced. The results suggest that systematic downward bias is not a serious problem when the cross section is expanded to allow for 5 or 10 firms per industry,

To see whether any of the conclusions one might draw from Table 4 are sensitive to alternative parameter values, we report simulations for a single data set in Table 5 (with $T = 1$, $N = 25$ and $P = 5$), using the same levels of contemporaneous correlation and conditional heteroskedasticity as in Table 3. The gains in efficiency provided by 2STLS grow with both the degree of cross-sectional correlation and heteroskedasticity. The estimated variances are often significantly less than the simple variance measures for lower cross-sectional correlation, but *greater* than the simple variance measures when $\rho = 0.9$. Also, when more heteroskedasticity is present in the data, the estimated variances increase relative to the asymptotic and simple variances. Table 5 therefore suggests not only that 2STLS can provide efficiency gains over method-of-moments in small samples where both heteroskedasticity and cross sectional dependence are present, but also that the estimates of

the parameter covariance matrix will on average be reliable estimates of the asymptotic covariance matrix.

5. Conclusions

We have presented two method-of-moments estimators useful for conducting inference in finance and accounting studies with relatively large cross sections and few time-series observations. The important insight is that we can exploit cross-sectional dependence in the regression residuals without depending on a great deal of time-series data relative to the size of the cross section. Instead we derive asymptotic results by letting the number of industries times the number of time-series observations, NT , become large while holding fixed the number of firms within an industry, P . Under the assumption that the correlations across industry groups are zero, completely unrestricted cross-sectional dependence and conditional heteroskedasticity is permitted within each industry.

One obvious difficulty is that, in practice, our results may rely on the existence of a large enough number of industries or groups which do not exhibit contemporaneous correlation. There is some empirical evidence in stock market data that market model residuals across industry groups are in fact uncorrelated. Nevertheless, analysis on a case by case basis is needed, because more pervasive contemporaneous correlation in the residuals can vitiate the validity of approach above. Clearly, the need to find uncorrelated industry groups is reduced in cases where sufficient time-series data are available.

Several conclusions emerge from the simulations presented in section 4. First, aggregating the data in an attempt to avoid the issue of cross-sectional correlation can be costly in terms of a loss in asymptotic efficiency. By contrast, the costs of employing one of the above techniques is low, both in terms of additional computational requirements and in terms of the reliability of the approximation to the asymptotic standard errors. Second, additional efficiency gains can be achieved in small samples using the 2STSLS estimator. Third, in the smaller samples ($N = 12$ or 25 , $P = 1$ and $T = 1$) the univariate estimated variances of White (1980a) and Cragg (1983) are consistently biased downward. The multivariate extensions considered above actually mitigate the bias by taking averages of disaggregated cross sections. Thus our approach kills three birds with one stone. By using some of the data which under other estimation strategies would be discarded or aggregated together we: i) allow for contemporaneous correlation; ii) improve the asymptotic approximation to the standard errors; and iii) obtain nonnegligible gains in asymptotic efficiency. Finally, there is a caveat. When correlation patterns across the N groups are highly dissimilar, we

would expect the all of these gains to materialize, but more slowly, as P is increased.

A topic left for future study is the determination of the relative performance of the foregoing estimators versus GLS when the data are aggregated just enough to make GLS feasible. The potential for greater disaggregation of the data and the attraction of a linear estimator in small samples, may offset the greater asymptotic efficiency of GLS over a comparable cross section, making the method-of-moments technique above more desirable. Also, larger cross sections may be used because the estimators above do not require inversion of the error covariance matrix. All the offsetting advantages and disadvantages must be addressed on a case-by-case basis in simulation and bootstrap studies. From such work, we might hope that rules of thumb will emerge to guide the choice of alternative estimation strategies.

6. Appendix

Lemma 1. Under Assumptions 1 and 2:

$$\sqrt{NT} \Theta_{NT}^{-1/2} M_{NT}(\hat{\beta}_{NT} - \beta) \xrightarrow{d} N(0, \mathbf{I}_K).$$

Proof. By the Hölder inequality, Assumption 2(ii) implies that for every $i, j = 1, \dots, K$, $E(|x_a^i x_b^j \mu_a \mu_b|)^{1+\delta} < \Delta$ (where a and b are as given in the text) and thus, for sufficiently large NT , Θ_{NT} will be positive definite. Given the boundedness conditions in Assumptions 1 and 2, White's (1980a) Lemma 2 may be applied directly, yielding our result.

Theorem 1. Assumptions 1 and 2 also imply:

$$\left(\frac{\mathbf{x}'_{NT} \mathbf{x}_{NT}}{NT}\right)^{-1} \hat{\Theta}_{NT} \left(\frac{\mathbf{x}'_{NT} \mathbf{x}_{NT}}{NT}\right)^{-1} \xrightarrow{a.s.} M_{NT}^{-1} \Theta_{NT} M_{NT}^{-1}.$$

Proof. This proof is similar to that of White (1980b, Theorem 1.). We first show that $\hat{\Theta}_{NT} \xrightarrow{a.s.} \Theta_{NT}$.

For each element β^i of the vector β , there exists $\bar{\beta}^i$ within a finite compact neighborhood of β^i given by Ψ , and there exists a finite element $\tilde{\beta}^i$, such that $(\bar{\beta}^i - \beta^i)^2 \leq (\tilde{\beta}^i - \beta^i)^2$ for all $\bar{\beta}^i \in \Psi$. Note that for all $j, k = 1, \dots, K$, and for all $a, b = 1, \dots, PNT$:

$$(y_a - x'_a \beta)(y_b - x'_b \beta) x_a^j x_b^k = (\mu_a - x'_a(\bar{\beta} - \beta))(\mu_b - x'_b(\bar{\beta} - \beta)) x_a^j x_b^k$$

is finite for all $\bar{\beta} \in \Psi$. By the Hölder inequality,

$$\begin{aligned} |(\mu_a - x'_a(\bar{\beta} - \beta))(\mu_b - x'_b(\bar{\beta} - \beta)) x_a^j x_b^k| &\leq \sum_{i=1}^K |((\mu_a^2 + \mu_b^2 + ((x_a^i)^2 + (x_b^i)^2)(\bar{\beta}^i - \beta^i)^2) x_a^j x_b^k| \\ &\leq \sum_{i=1}^K |((\mu_a^2 + \mu_b^2 + ((x_a^i)^2 + (x_b^i)^2)(\tilde{\beta}^i - \beta^i)^2) x_a^j x_b^k| = m_{ab} \end{aligned}$$

where the second inequality is given by the above assumption about $\hat{\beta}^i$. From the fundamental inequality $|c + d|^r \leq 2^{r-1}|c|^r + 2^{r-1}|d|^r$, we have that:

$$\begin{aligned} E(m_{ab})^{1+\delta} &\leq k_0 E((|\mu_a^2 x_a^i x_b^j|)^{1+\delta}) + k_1 E((|\mu_b^2 x_a^j x_b^k|)^{1+\delta}) \\ &\quad + \sum_{i=2}^{K+1} k_i E((|(x_a^i)^2 + (x_b^i)^2) x_a^j x_b^k|)^{1+\delta} (\hat{\beta}^i - \beta^i)^{2\delta+2} \\ &\leq k_0 E((|\mu_a^2 x_a^j x_b^k|)^{1+\delta}) + k_1 E((|\mu_b^2 x_a^j x_b^k|)^{1+\delta}) + \sum_{i=2}^{K+1} k_i E((x_a^i)^2 + (x_b^i)^2)^{2+2\delta} (\hat{\beta}^i - \beta^i)^{2\delta+2} \end{aligned}$$

where the second inequality is given by the Hölder inequality. By the fundamental inequality above, k_0 , k_1 and k_i are constants, and $(\hat{\beta}^i - \beta^i)^2$ is finite by assumption. In addition, Assumptions 1 and 2 in the text guarantee that $E((|\mu_a^2 x_a^j x_b^k|)^{1+\delta})$, $E((|\mu_b^2 x_a^j x_b^k|)^{1+\delta})$, and $E((x_a^i)^2 + (x_b^i)^2)^{2+2\delta}$ are bounded. Therefore $E(m_{ab})^{1+\delta}$ must be bounded. Since m_{ab} weakly dominates $|((y_a - x_a' \bar{\beta})(y_b - x_b' \bar{\beta})) x_a^j x_b^k|$, and is independent over t and n , we can apply the strong law of large numbers, so that:

$$\begin{aligned} &N^{-1} \sum_{n=1}^N T^{-1} \sum_{t=1}^T \left(\sum_{i=1}^P \sum_{j=1}^P (y_a - x_a' \beta)(y_b - x_b' \beta) x_a^i x_b^j \right) \\ &- N^{-1} \sum_{n=1}^N T^{-1} \sum_{t=1}^T \left(\sum_{i=1}^P \sum_{j=1}^P E((y_a - x_a' \beta)(y_b - x_b' \beta) x_a^i x_b^j) \right) \xrightarrow{a.s.} 0. \end{aligned}$$

Given that $\hat{\beta}_N \xrightarrow{a.s.} \beta$, the above equation and Lemma 2.6 of White (1980b) imply:

$$\begin{aligned} &N^{-1} \sum_{n=1}^N T^{-1} \sum_{t=1}^T \left(\sum_{i=1}^P \sum_{j=1}^P (y_a - x_a' \hat{\beta}_N)(y_b - x_b' \hat{\beta}_N) x_a^i x_b^j \right) \\ &- N^{-1} \sum_{n=1}^N T^{-1} \sum_{t=1}^T \left(\sum_{i=1}^P \sum_{j=1}^P E(\mu_a \mu_b x_a^i x_b^j) \right) \xrightarrow{a.s.} 0. \end{aligned}$$

or $\hat{\Theta}_{NT} - \Theta_{NT} \xrightarrow{a.s.} 0$.

By the boundedness of the regressors given in assumption 1, and their independence over NT , we then have that:

$$\left(\frac{\mathbf{x}'_{NT} \mathbf{x}_{NT}}{NT} \right) - M_{NT} \xrightarrow{a.s.} 0.$$

The continuity of the matrix inverse therefore implies:

$$\left(\frac{\mathbf{x}'_{NT} \mathbf{x}_{NT}}{NT} \right)^{-1} - M_{NT}^{-1} \xrightarrow{a.s.} 0.$$

By continuity and the boundedness of each term, we have that:

$$\left(\frac{\mathbf{x}'_{NT} \mathbf{x}_{NT}}{NT} \right)^{-1} \hat{\Theta}_{NT} \left(\frac{\mathbf{x}'_{NT} \mathbf{x}_{NT}}{NT} \right)^{-1} \xrightarrow{a.s.} M_{NT}^{-1} \Theta M_{NT}^{-1}.$$

This proves Theorem 1.

Theorem 2. The OLS estimate of the coefficient covariance matrix in equation (9), $\hat{V}(\hat{\beta}^{**})$, converges almost surely to the true coefficient covariance matrix:

$$\left(\frac{\mathbf{x}'_{NT} \mathbf{q}_{NT}}{NT} \right) \left(\hat{\Theta}_{NT}^* \right)^{-1} \left(\frac{\mathbf{q}'_{NT} \mathbf{x}_{NT}}{NT} \right) \xrightarrow{a.s.} R'_{NT} \Theta_{NT}^{*-1} R_{NT}$$

Part 1 of Theorem 1 implies that $\hat{\Theta}_{NT} \xrightarrow{a.s.} \Theta_{NT}$. By Assumption 3, the second moments of \mathbf{q} and its cross products with the μ 's are bounded in the sense of Assumption 2. Part 1 of Theorem 1 therefore also implies that:

$$\hat{\Theta}_{NT}^* \xrightarrow{a.s.} \Theta_{NT}^*$$

Since Θ_{NT}^* is of full rank, the matrix $(\Theta_{NT}^*)^{-1/2}$ is defined such that $((\Theta_{NT}^*)^{-1/2})^2 = (\Theta_{NT}^*)^{-1}$. Thus,

$$(\hat{\Theta}_{NT}^*)^{-1/2} \xrightarrow{a.s.} (\Theta_{NT}^*)^{-1/2}.$$

The boundedness of the instruments and their independence over N and T imply that:

$$\left(\frac{\mathbf{q}'_{NT} \mathbf{x}_{NT}}{NT} \right) - R_{NT} \xrightarrow{a.s.} 0.$$

Thus, by continuity and the uniform boundedness of each term we have that:

$$\left(\frac{\mathbf{x}'_{NT}\mathbf{q}_{NT}}{NT}\right)\left(\hat{\Theta}_{NT}^*\right)^{-1}\left(\frac{\mathbf{q}'_{NT}\mathbf{x}_{NT}}{NT}\right)\xrightarrow{a.s.} R_{NT}^{-1}\Theta_{NT}^{*-1}R_{NT}^{-1},$$

which proves Theorem 2.

7. References

- Beaver, W.H., C. Eger, S. Ryan, and M. Wolfson, "Financial Reporting and the Structure of Bank Share Prices," Stanford University, July 1985.
- Bernard, Victor L., "Unanticipated Inflation and the Value of the Firm," *Journal of Financial Economics* Vol. 15, No. 3, 1986.
- , "Cross-sectional Dependence and Problems in Inference in Market Based Accounting Research," University of Michigan, October 1986b.
- Black, F., M. C. Jensen, and M. Scholes, "The Capital Asset Pricing Model: Some Empirical Tests," in *Studies in the Theory of Capital Markets*, ed. M. C. Jensen (Praeger, New York), pp. 79-121, 1972.
- Brown, S., and J. Warner, "Measuring Security Price Performance," *Journal of Financial Economics*, pp. 205-258, September 1980.
- , "Using Daily Stock Returns: The Case of Event Studies," *Journal of Financial Economics*, pp. 3-32, March 1985.
- Brown, S. and M. Weinstein, "A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm," *Journal of Finance*, Vol. 38, pp. 711-743, 1983.
- Chamberlain, G. and M. Rothschild, "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, Vol. 51, No. 5, pp. 1281-1304, September 1983.
- Christie, A. "On Cross-Sectional Analysis in Accounting Research," University of Southern California, 1985.
- Collins, D. and W. Dent, "A Comparison of Alternative Testing Methodologies Used in Capital Markets Research," *Journal of Accounting Research*, pp. 48-84, Spring 1984.
- Cragg, J. G., "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica* Vol. 51, No. 3, pp. 751-763, May 1983.
- Cumby, R., J. Huizinga and M. Obstfeld, "Two-Step Two-Stage Least Squares Estimation in Models with Rational Expectations," *Journal of Econometrics* Vol. 21, pp. 333-355, 1983.
- Eicker, F., "Limit Theorems for Regressions with Unequal and Dependent Errors," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. by L. LeCam and J. Neyman, Berkeley, University of California Press, pp. 59-82, 1967.
- French, K. R., R. S. Ruback and G. W. Schwert, "Effects of Nominal Contracting on Stock

- Returns," *Journal of Political Economy*, Vol 91, No. 1, pp. 70-96, 1983.
- Gibbons, M., "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics* pp. 3-27, March 1982.
- Hansen, Lars P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* Vol. 50, pp. 1029-1054, 1982.
- Ingersol, J., "Some Results in the Theory of Arbitrage Pricing," *Journal of Finance*, Vol. 39, pp. 1021-1039, September 1984.
- Jaffe, J., "Special Information and Insider Trading," *Journal of Business*, pp. 410-428, July 1974.
- Malatesta, P., "Measuring Abnormal Performance: The Event Parameter Approach Using Joint Generalized Least Squares," *Journal of Financial and Quantitative Analysis*, pp. 27-38, March 1986.
- Press, S., *Applied Multivariate Analysis*, New York. Holt, Rinehart and Winston, 1972.
- Rao, C., "Estimation of Heteroskedastic Variances in Linear Models," *Journal of the American Statistical Association* Vol. 65, pp. 161-172, 1970.
- Sefcik, S. and R. Thompson, "An Approach to Statistical Inference in Cross-Sectional Regression with Security Abnormal Returns as Dependent Variable," Working Paper No. 1017, University of British Columbia, January 1985, and forthcoming, *Journal of Accounting Research*.
- Shipper, K., and R. Thompson, "The Impact of Merger-Related Regulations on the Shareholders of Acquiring Firms," *Journal of Accounting Research*, pp. 184-221, Spring 1983.
- _____, "The Impact of Merger-Related Regulations Using Exact Distributions of Test Statistics," *Journal of Accounting Research*, pp. 408-415, Spring 1985.
- Thompson, Rex., "Conditioning the Return-Generating Process on Firm-Specific Events: A Discussion of Event Study Methodologies," *Journal of Financial and Quantitative Analysis*, pp 151-172, June 1985.
- White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, vol. 48 no 4., pp. 817-838, 1980a.
- _____, "Nonlinear Regression on Cross-Section Data," *Econometrica*, vol. 48 no 3., pp. 721-746, 1980b.
- _____, *Asymptotic Theory for Econometricians*, Academic Press, (New York), 1984.

Table 1

Simulation Results: Variance of Method-of-Moments Estimator as a Percentage of Asymptotic OLS Variance with Cross-Sectional Homoskedasticity

Number of Time Series Observations:		T = 1				T = 10			
Variance Measure:		<u>Asymptotic</u>	<u>Simple</u>	<u>Estimated</u>	<u>GLS</u>	<u>Asymptotic</u>	<u>Simple</u>	<u>Estimated</u>	<u>GLS</u>
P = 1	# of obsv./ group								
	# of groups								
P = 1	N = 12	1.0000	.97359 (.04328)	.79635 (.02029)	1.0000	1.0000	1.00438 (.04886)	.87800 (.01190)	1.0000
	N = 25	1.0000	.99973 (.04424)	.92386 (.01457)	1.0000	1.0000	1.00926 (.04325)	.96258 (.00968)	1.0000
P = 5	N = 12	.03654	.03689 (.00169)	.03604 (.00060)	.02473	.05313	.05391 (.00228)	.05213 (.00051)	.03337
	N = 25	.05912	.05786 (.00243)	.05504 (.00087)	.03634	.06573	.06265 (.00269)	.06284 (.00060)	.03955
P = 20	N = 12	.00193	.00193 (.00009)	.00196 (.00003)	.00129	.00431	.00421 (.00019)	.00420 (.00003)	.00233
	N = 25	.00242	.00243 (.00012)	.00244 (.00003)	.00153	.00377	.00366 (.00015)	.00374 (.00002)	.00212

NOTES: Asymptotic OLS variance is constructed using data aggregated over the cross-section into N groups. Cross-sectional correlation is homogeneous across groups, and cross-sectional and time-series homoskedasticity is assumed. Standard errors are in parenthesis. All data reported are constructed using 1000 replications for given regressors. Intra-group correlation coefficient = 0.5.

Table 2

Simulation Results: Variance of Method-of-Moments Estimator as a Percentage of Asymptotic OLS Variance with Cross-Sectional Homoskedasticity

Number of Time Series Observations:		T = 1				T = 10			
		<u>Asymptotic</u>	<u>Simple</u>	<u>Estimated</u>	<u>GLS</u>	<u>Asymptotic</u>	<u>Simple</u>	<u>Estimated</u>	<u>GLS</u>
Variance Measure:	# of obsv./ group								
	# of groups								
P = 1	N = 12	1.0000	1.01094 (.04737)	.67817 (.01949)	.90301	1.0000	1.06510 (.04748)	.63808 (.01927)	.09349
	N = 25	1.0000	1.03679 (.04838)	.86988 (.01867)	.69589	1.0000	1.01401 (.04699)	.86700 (.01702)	.15553
P = 5	N = 12	.03928	.04295 (.00189)	.03662 (.00086)	.01672	.05615	.05731 (.00266)	.04950 (.00103)	.00952
	N = 25	.05714	.06077 (.00273)	.03565 (.00143)	.00276	.06745	.06650 (.00292)	.05432 (.00118)	.00518
P = 20	N = 12	.00181	.00176 (.00007)	.00159 (.00004)	.00042	.00438	.00459 (.00020)	.00421 (.00005)	.00162
	N = 25	.00314	.00333 (.00015)	.00301 (.00005)	.00073	.00388	.00369 (.00017)	.00380 (.00004)	.00099

NOTES: Asymptotic OLS variance is constructed using data aggregated over the cross-section into N groups. Cross-sectional correlation is homogeneous, and time-series homoskedasticity is assumed. Cross-sectional heteroskedasticity is given by equation (27). Standard errors are in parenthesis. All data reported are constructed using 1000 replications for given regressors. Intra-group correlation coefficient = 0.5.

Table 3

**Simulation Results: Sensitivity of Method-
of-Moments Estimator to Alternative Levels of
Intra-Group Correlation and Cross-Sectional Heteroskedasticity**

Intra-Group
Correlation:

Level of Cross-
Sectional Hetero-
skedasticity

	P = 0.1				P = 0.5				P = 0.9			
	Asymp.	Simple	Estim.	GLS	Asymp.	Simple	Estim.	GLS	Asymp.	Simple	Estim.	GLS
1	.1308	.1231 (.0059)	.1224 (.0018)	.1270	.0591	.0546 (.0025)	.0546 (.0008)	.0363	.0373	.0343 (.0016)	.0338 (.0055)	.0049
2	.1231	.1197 (.0051)	.0749 (.0028)	.0088	.0571	.0557 (.0024)	.0339 (.0013)	.0028	.0371	.0363 (.0016)	.0211 (.0008)	.0004
3	.1223	.1222 (.0053)	.0774 (.0029)	.0034	.0569	.0548 (.0024)	.0342 (.0013)	.0011	.0370	.0349 (.0016)	.0208 (.0008)	.0002
4	.1221	.1248 (.0052)	.0756 (.0028)	.0024	.0569	.0566 (.0024)	.0326 (.0013)	.0008	.0370	.0358 (.0015)	.0197 (.0008)	.0001

NOTES: Variance measures are expressed as a percentage of OLS asymptotic variance. All assume N = 25, P = 5, and T = 1. All data reported are constructed using 1000 replications for a single set of randomly chosen regressors.

Table 4

**Simulation Results: Variance of 2STSLs Estimator
as a Percentage of Asymptotic Method-of-Moments Variance**

Number of Time Series Observations:		T = 1				T = 10			
		<u>Asymptotic</u>	<u>Simple</u>	<u>Estimated</u>	<u>GLS</u>	<u>Asymptotic</u>	<u>Simple</u>	<u>Estimated</u>	<u>GLS</u>
# of obsv./ group	# of groups								
P = 1	N = 12	.9602	1.0768 (.0478)	.5500 (.0153)	.9030	.1173	.2005 (.0091)	.1361 (.0018)	.0935
	N = 25	.9682	1.0753 (.0482)	.7525 (.0164)	.6959	.3357	.4538 (.0198)	.3162 (.0031)	.1555
P = 5	N = 12	.8166	.8746 (.0404)	.6242 (.0145)	.4255	.3081	.3326 (.0148)	.2911 (.0033)	.1695
	N = 25	.1022	.1805 (.0080)	.1360 (.0032)	.0484	.1666	.1561 (.0074)	.1673 (.0014)	.0769
P = 10	N = 12	.1109	.2499 (.0111)	.1757 (.0067)	.0415	.3360	.3599 (.0154)	.3283 (.0028)	.1830
	N = 25	.1962	.2319 (.0099)	.2090 (.0033)	.0882	.4775	.4954 (.0230)	.4432 (.0038)	.1844

NOTES: 2STSLs estimator is constructed using the historical covariance matrix of the dependent variable, as in equation (29). Cross-sectional correlation is homogeneous, and time-series homoskedasticity is assumed. Cross-sectional heteroskedasticity is given by equation (27), with $\delta_0 = 1$ and $\delta_1 = 0.2$. Standard errors are in parenthesis. All data reported are constructed using 1000 replications for given regressors. Intra-group correlation coefficient is set to 0.5.

Table 5

Simulation Results: Sensitivity of 2SLS Estimator to Alternative Levels of Intra-Group Correlation and Cross-Sectional Heteroskedasticity

Intra-Group Correlation:

Level of Cross-Sectional Heteroskedasticity	P = 0.1				P = 0.5				P = 0.9			
	Asymp.	Simple	Estim.	GLS	Asymp.	Simple	Estim.	GLS	Asymp.	Simple	Estim.	GLS
1	1.0000	1.1000 (.0507)	.7908 (.0127)	.9710	.9532	.9546 (.0412)	.7795 (.0124)	.6146	.4226	.5812 (.0278)	.4321 (.0064)	.1312
2	.3734	.3646 (.0162)	.2762 (.0078)	.0718	.1299	.2240 (.0098)	.1506 (.0037)	.0484	.0292	.0764 (.0048)	.0984 (.0036)	.0104
3	.1198	.1743 (.0082)	.1529 (.0049)	.0279	.0618	.1116 (.0055)	.1194 (.0037)	.0194	.0096	.0600 (.0036)	.0798 (.0034)	.0042
4	.0488	.1242 (.0057)	.1082 (.0032)	.0199	.0242	.0820 (.0045)	.0873 (.0030)	.0140	.0076	.0405 (.0024)	.0718 (.0029)	.0030

NOTES: Variance measures are expressed as a percentage of Asymptotic Method-of-Moments variance. All assume N = 25, P = 5, and T = 1. All data reported are constructed using 1000 replications for given regressors.