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

Kenneth A. Froot*



## 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 method is 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. We also consider efficiency issues and show that, in principle, asymptotic efficiency can be improved using a technique due to Cragg (1983).

## I. Introduction

Practically every empirical study in accounting and finance using cross-sectional data must come to grips with the problem of contemporaneous correlation across firms (or portfolios). Such studies present special estimation problems because they employ panel data with a large number of firms but relatively few 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 that address both contemporaneous correlation and heteroskedasticity, and that effectively cope with the relative paucity of time-series data.

This paper outlines two relatively simple estimators to account for contemporaneous correlation and heteroskedasticity in the cross section as well as heteroskedasticity over time. The first technique, which combines elements of method-of-moments estimation pioneered by Hansen (1982) and the treatment of

* Graduate School of Business Administration, Harvard University, and Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139. The author is grateful to Vic Bernard, Ian Domowitz, Dick Meese, Danny Quah, and Jeff Wooldridge for helpful discussions, to the Alfred Sloan and John Olin Foundations for generous research support, and to an anonymous *JFQA* referee for useful comments.

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.

## II. A Discussion of Alternative Estimation Techniques

It is worth stressing the broad range of studies that 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.[1] Often cross-sectional data must be aggregated, either because of the lack of time-series data or because of the computational difficulty in diverting very large contemporaneous covariance matrices. 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 that ignore contemporaneous correlation (Brown and Warner (1980), (1985), Christie (1985), and Malatesta and Thompson (1985)) and others using techniques that ignore heteroskedasticity (Collins and Dent (1984), Shipper and Thompson (1982), (1983), (1985), and Malatesta (1986)).

In the absence of heteroskedasticity, it is well known that GLS techniques (such as seemingly unrelated squares or multivariate regressions) in principle offer first-best solutions to problems of cross-sectional dependence. In practice, however, 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 matrix is a function of a relatively small number of parameters.[2]

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.

[1] The putative interval over which parameters of the market model are reasonably constant is five years. This 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 $N \leq (T-1)/2$, i.e., if $N \leq 30$. See Press (1972).

[2] The finite-sample bias in the variance of the regression coefficients is of the order of the number of observations. See Rothenberg (1984).

Thus, while the estimators below are less efficient than GLS with an unlimited supply of time-series data, they nevertheless may have more power because they may be applied to data sets for which GLS is infeasible, unreliable, or computationally too burdensome.

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. 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, he demonstrates that this bias can only increase as the cross section is expanded. 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 exhibit little contemporaneous correlation in their residuals. That is, we assume the cross-sectional correlation matrix of the residuals is block diagonal. While this assumption is restrictive, it buys 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.[3] 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, the block diagonality assumption may be useful for studies in finance and accounting.

In addition to cross-sectional correlation, many financial data sets exhibit 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

---

[3] See, for example, Ingersol (1984) and Chamberlain and Rothschild (1983).

variance during event periods is likely to be systematically different from the variance in nonevent periods. The ability to permit both contemporaneous correlation and unrestricted heteroskedasticity is an important advantage of GMM over feasible GLS estimators.

The remainder of the paper proceeds as follows. In Section III 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 Section IV, we consider an alternative two-step estimator that can increase the asymptotic efficiency of the initial estimator. Section V then presents Monte Carlo simulations of both estimators for a variety of sample sizes, levels of contemporaneous correlation, and degrees of conditional heteroskedasticity. Last, Section VI offers our conclusions.

## III. The Heteroskedasticity and Contemporaneous Correlation-Consistent Covariance Estimator

Consider the linear model

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

where $n = 1,\ldots, N$ and $t = 1,\ldots, T$ 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,[4]

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

We also make the following assumptions.

*Assumption 1*. (i) The pair $(x_{pnt}, \mu_{pnt})$ 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 $E(x^i_{pnt}\mu_{pnt}) = 0$, where $i = 1,\ldots K$ denotes the $i$th column of $x_{pnt}$. (iii) The $P \times 1$ residual vector, $\mu_{nt}$, has a conditional expectation of zero, $E(\mu_{nt} \mid x_{nt}) = 0$, and the $P \times P$ conditional covariance matrix given by $E(\mu_{nt}\mu'_{nt} \mid x_{nt}) = \Sigma_{nt}$.

*Assumption 2*. (i) The second moments of the residuals are bounded, in the sense that there exist positive constants, $\delta$ and $\Delta$, such that $E(\mu^2_{pnt})^{1+\delta} < \Delta$. Also, each element of the $K \times K$ outer product of the regressors is bounded: $E(x^i_{pnt}x^j_{pnt})^{1+\delta} < \Delta$, for all $i, j = 1,\ldots K$. (ii) The product of the second moments is assumed to be bounded, $E(\mu^2_{pnt}x^i_{pnt}x^j_{pnt})^{1+\delta} < \Delta$, as is the square of the product of squared

[4] For notational convenience, we assume that the number of firms per industry is the same across all industries. It is straightforward to relax this assumption in deriving the results below.

regressors, $E((x^{i}_{pnt})^2(x^{j}_{pnt})^2)^{2+2\delta} \leq \Delta$. (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. (iv) The average covariance matrix is

$$\Theta_{NT} = (NT)^{-1}E\left(\sum_{n=1}^{N}\sum_{t=1}^{T} x'_{nt}\mu_{nt}\mu'_{nt}x_{nt}\right). \tag{2}$$

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 (1), $\beta$.[5] The usual asymptotic OLS covariance matrix, $(PNT)^{-1}\Sigma_{p=1}^{P}\Sigma_{n=1}^{N}\Sigma_{t=1}^{T}E(\mu^2_{pnt})M_{NT}^{-1}$, however, will be incorrect since we allow for contemporaneous correlation across firms in the same industry and for heteroskedasticity. The correct standard errors come out of a multivariate generalization of White's (1980a) estimator of the covariance matrix

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

Under the assumption that the inter-industry cross-correlations are zero (Assumption 1(i)), the overall conditional error covariance matrix is $\mathbf{\Omega}_{NT} = E(\mu_{NT}\mu'_{NT} \mid \mathbf{x}_{NT})$ (where $\mu_{NT}$ is a $PNT \times 1$ 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

$$\mathbf{\Omega}_{NT} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \mathbf{\Sigma}_{1T} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & \mathbf{\Sigma}_{N1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & \mathbf{\Sigma}_{NT} \end{pmatrix}, \tag{4}$$

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

$$\sigma^{ij}_{nt} = E\left(\mu_{int}\mu_{jnt} \mid x_{nt}\right) \quad i,j = 1,\ldots,P. \tag{5}$$

The difficulty clearly comes in estimating the error covariance matrix $\mathbf{\Omega}_{NT}$. With $T$ large and under homoskedasticity, there is no problem; the industry covariance matrices, $\mathbf{\Sigma}_{nt}$, would then be independent of time and could be estimated consistently from the OLS residuals,

$$\hat{\sigma}^{ij}_{n} = \frac{1}{T}\sum_{t=1}^{T}\hat{\mu}_{int}\hat{\mu}_{jnt}: \quad i,j = 1,\ldots,P, \tag{6}$$

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

---

[5] For a proof under similar assumptions, see Lemma 2 in White (1980a).

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 $((P+1)P)/2$ elements for each $\Sigma_{nt}$, or $((P+1)PNT)/2$ 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 $((K+1)K)/2$ parameters in the average covariance matrix, $\Theta_{NT}$. 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

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

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 approproriate covariance. Thus, the average error covariance matrix depends only on the number of regressors $K$. For the special case in which $P = 1$, (7) is the estimate of the average covariance matrix proposed by White (1980a).

We can now state two propositions:

*Proposition 1*. Under Assumptions 1 and 2, the standard distribution of the OLS estimate, $\hat{\beta}_{NT}$, is

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

*Proposition 2*. The OLS estimate of the coefficient covariance matrix converges almost surely to the coefficient covariance matrix in (3),

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

The Appendix contains the proofs of these propositions, which are similar to those in White (1984), Chapter 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.

## IV. 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 that weights data according to their 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, $\hat{\mu}^2_{pnt}$, and this is a biased and inconsistent estimate of $E(\mu^2_{pnt})$. 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 $NT > R > K$) for the regressors in (1), $\mathbf{q} = [\mathbf{x}\ \mathbf{p}]$, which includes the regressors, $\mathbf{x}$, as well as 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

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

which is of rank $R-K$ for sufficiently large $NT$.

Next, consider the following transformation of (1),

$$\mathbf{y}^* = \mathbf{x}^*\beta + \mu^*, \tag{8}$$

where $\mathbf{y}^* = \mathbf{q}'\mathbf{y}$, $\mathbf{x}^* = \mathbf{q}'\mathbf{x}$, and $\mu^* = \mathbf{q}'\mu$.[6] The error covariance matrix for (8) is now the $R \times R$ matrix $E(\mu^*\mu^{*\prime}) = E(\mathbf{q}'\mu\mu'\mathbf{q}) = \Theta^*$, 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 (8) by $(\Theta^*)^{-1/2}$, which then yields

$$\mathbf{y}^{**} = \mathbf{x}^{**}\beta + \mu^{**}, \tag{9}$$

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 $E(\mu^{**}\mu^{**\prime}) = \mathbf{I}_R$. By running OLS on (9), we obtain the standard formula for the two-step, two-stage least squares (2STSLS) estimator due to Cumby, Huizinga, and Obstfeld (1983),

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

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

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

[6] Where it will not create confusion, we drop the $NT$ subscript for the remainder of the text.

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

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

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.

*Proposition 3.* The OLS estimate of the coefficient covariance matrix in (9), $\hat{V}(\hat{\beta}^{**})$ converges almost surely to the true 2STSLS covariance matrix,

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

See the Appendix for the proof, which follows directly from Assumption 3 and Proposition 2. The important intuition is that we can use the independence of the $q_{pnt}$'s and $\mu_{pnt}$'s over $T$ and $N$ to estimate consistently the average error covariance matrix of the instrumental variables.

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

$$M^{-1}_{NT}\Theta_{NT}M^{-1}_{NT} - \left(R'_{NT}\left(\Theta^*_{NT}\right)^{-1} R_{NT}\right)^{-1} , \tag{13}$$

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

$$\Theta^{-1}_{NT}\Theta^{xp}_{NT} \neq M^{-1}_{NT}P_{NT} . \tag{14}$$

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

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

The left-hand side of (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 $\mathbf{p}$ that 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 that 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, $\mathbf{x}'\mathbf{\Omega x}$. Ideally, the instruments would be chosen to minimize the mean squared residual in the OLS regression

$$\mathbf{\Omega}^{-1/2}\mathbf{x} = \mathbf{\Omega}^{1/2}\mathbf{q}\gamma + \epsilon\ . \tag{16}$$

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

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

Equation (16) may be rewritten,

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

Equation (18) says that good instruments are those that will explain the variation in $\mathbf{\Omega}^{-1}\mathbf{x}$. To address the problem of heteroskedasticity in the $\mu$'s, Cragg treats $\mathbf{\Omega}^{-1}\mathbf{x}$ 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.[7] (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 $(\mathbf{\Omega}^H)^{-1}\mathbf{x}$, where $\mathbf{\Omega}^H$ is the historical correlation matrix of the dependent variable calculated during an earlier period.

## V. Simulation Experiments

In this section, we perform simulations of the estimators above. We try to address three issues in these simulations. 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 III and IV 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 2STSLS estimator in Section IV instead of the simpler method-of-moments estimator in Section

[7] If 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.

III? 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

$$y_{pnt} = \beta_0 + \beta x_{pnt} + \mu_{pnt} \; . \tag{19}$$

The $x_{pnt}$ were drawn randomly from a log-normal distribution, and are independent across $p$, $n$, and $t$.[8] In the simulation and the notation that follow, we treat the regressors as fixed. A set of independently distributed primitive disturbance terms, $\eta_{pnt}$, was drawn from a standardized normal distribution, then modified to generate contemporaneous correlation and heteroskedasticity in the residuals, $\mu_{pnt}$.[9] Contemporaneous intra-industry correlation was assumed to be identical across industries, and was introduced by transforming the $PNT \times 1$ column vector, $\eta$,

$$\mu = \left(\mathbf{I}_{NT} \otimes \boldsymbol{\Delta}\right)\eta \; , \tag{20}$$

where $\boldsymbol{\Delta}$ is a nonstochastic $P \times P$ matrix,

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

and $\lambda$ is a function of the correlation coefficient between firms in a given industry, $\rho$, and the number of firms per industry, $P$,

$$\lambda = \frac{1 - \left((1-\rho)(1+\rho(P-1))\right)^{1/2}}{(P-1)(\rho-1) + 1} \; . \tag{21}$$

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, $\beta_0$ and $\beta$ can be chosen arbitrarily without affecting the results. We selected $\beta_0 = 0$ and $\beta = 1$. For each experiment, 1000 draws of $\eta$ were used to compute 1000 sets of dependent variables and, using a single set of regressors, 1000 sets of 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 III for a variety of sample sizes when the correlation coefficient in (21), $\rho$, 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,

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

[8] The same vector, $x$, was used for all of the simulations that follow.

[9] We use the fast acceptance-rejection algorithm proposed by Kinderman and Ramage (1976).

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.[10] 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,[11]

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

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,

$$\frac{\left(\mathbf{x}'\left(\mathbf{I}_{NT}\otimes\mathbf{\Delta}^{-1}\right)\mathbf{x}\right)^{-1}}{\mathbf{i}'_{PNT}\left(\mathbf{I}_{NT}\otimes\mathbf{\Delta}\right)\mathbf{i}_{PNT}\left(\mathbf{i}'_P(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})\mathbf{i}_P\right)^{-1}}. \tag{24}$$

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 0.027. 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 ⅕th of the total number of observations, we would expect the disaggregated data to improve the relative asymptotic efficiency by a factor of about 5.[12] The second effect, which explains the increase in efficiency from 0.20 to 0.037, 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 (0.20/0.0364), when $P = 5$, to a factor of approximately 25 (0.05/0.00193), 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.

---

[10] Under our extreme assumption that the regressors are independent across $p$, the usual OLS variance of the parameters using the disaggregated data, $\Sigma_{i=1}^{PNT}\mu_i(\mathbf{x}'\mathbf{x})^{-1}$, 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.

[11] To simplify the notation, we assume that the means of both $x$ and $y$ are zero, so that the variance matrix of the parameters can be treated as a scaler.

[12] This improvement in efficiency is a function of the cross-sectional correlation of the regresors. 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.

TABLE 1

**Simulation Results: Variance of Method-of-Moments Estimator as a Fraction of Asymptotic OLS Variance under Cross-Sectional Homoskedasticity**

| | | Time Series Observations: $T = 1$ | | | | Time Series Observations: $T = 10$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Variance Measure | | | | | | | |
| Number of Obser- vations per Group | Number of Groups | Asymp- totic | Simple | Esti- mated | GLS | Asymp- totic | Simple | Esti- mated | GLS |
| $P = 1$ | $N = 12$ | 1.0000 | 0.97359 | 0.79635 | 1.0000 | 1.0000 | 1.00438 | 0.87800 | 1.0000 |
| | | | (0.04328) | (0.02029) | | | (0.04886) | (0.01190) | |
| | $N = 25$ | 1.0000 | 0.99973 | 0.92386 | 1.0000 | 1.0000 | 1.00926 | 0.96258 | 1.0000 |
| | | | (0.04424) | (0.01457) | | | (0.04325) | (0.00968) | |
| $P = 5$ | $N = 12$ | 0.03654 | 0.03689 | 0.03604 | 0.02743 | 0.05313 | 0.05391 | 0.05213 | 0.0333 |
| | | | (0.00169) | (0.00060) | | | (0.00228) | (0.00051) | |
| | $N = 25$ | 0.05912 | 0.05786 | 0.05504 | 0.03634 | 0.06573 | 0.06265 | 0.06284 | 0.0395 |
| | | | (0.00243) | (0.00087) | | | (0.00269) | (0.00060) | |
| $P = 20$ | $N = 12$ | 0.00193 | 0.00193 | 0.00196 | 0.00129 | 0.00431 | 0.00421 | 0.00420 | 0.0023 |
| | | | (0.00009) | (0.00003) | | | (0.00019) | (0.00003) | |
| | $N = 25$ | 0.00242 | 0.00243 | 0.00244 | 0.00153 | 0.00377 | 0.00366 | 0.00374 | 0.0021 |
| | | | (0.00012) | (0.00003) | | | (0.00015) | (0.00002) | |

*Notes:* The column labeled "Asymptotic" reports the asymptotic method-of-moments variance divided by the OLS asymptotic variance (see Equation (23)). The asymptotic OLS variance is constructed by aggregating the data over the cross section into $N$ groups. The column labeled "Simple'" gives the mean-squared deviation of the estimated method-of-moments estimate scaled by the asymptotic OLS variance (see Equation (25)). The column labeled "Estimated" reports the average estimated method-of-moments variance divided by the asymptotic OLS variance (see Equation (26)). Cross-sectional correlation is homogeneous across groups, and cross-sectional and time-series homoskedasticity is assumed. Standard errors are in parenthesis. All data are constructed using 1000 replications for given regressors. The intra-group correlation coefficient = 0.5.

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

$$\frac{\left(\hat{\beta}-1\right)^{2}}{\mathbf{i}'_{PNT}\left(\mathbf{I}_{NT}\otimes\mathbf{\Delta}\right)\mathbf{i}_{PNT}\left(\mathbf{i}'_{P}(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})\mathbf{i}_{P}\right)^{-1}}\,. \tag{25}$$

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,

$$\frac{(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\left(\mathbf{I}_{T}\otimes\left(\mathbf{I}_{N}\otimes\hat{\mathbf{\Delta}}\right)\right)\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}}{\mathbf{i}'_{NT}\left(\mathbf{I}_{NT}\otimes\mathbf{\Delta}\right)\mathbf{i}_{NT}\left(\mathbf{i}'_{P}(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})\mathbf{i}_{P}\right)^{-1}}\,, \tag{26}$$

where $\hat{\boldsymbol{\Delta}}$ is the estimate of the $P \times P$ contemporaneous correlation matrix $\boldsymbol{\Delta}$.[13]

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, $\Sigma_{i=1}^{NT} x_i x_i' \hat{\mu}_i^2$ is a biased estimator of $\Sigma_{i=1}^{NT} E(\mu_i^2) \Sigma_{i=1}^{NT} x_i x_i'$, even when the residuals are homoskedastic. It may be corrected using Rao's MINQUE procedure. This bias in estimated method-of-moments covariance matrices when the residuals are heteroskedastic is well known. 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 in which 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 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 preselection. 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,

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

[13] In computing (26), we do not impose the restriction that the $\hat{\boldsymbol{\Delta}}$ matrix is equal across industries.

In the simulations of Table 2, $\delta_0$ is set equal to 1, and $\delta_1 = 0.2$. Thus the asymptotic covariance matrix of the residuals becomes

$$\mathbf{I}_T \otimes \tilde{\mathbf{\Delta}} = \mathbf{I}_T \otimes \begin{pmatrix} \mathbf{\Delta}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{\Delta}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{\Delta}_N \end{pmatrix}, \tag{27'}$$

where $\tilde{\mathbf{\Delta}}$ is $NP \times NP$, and $\mathbf{\Delta}_n$ 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.

TABLE 2

Simulation Results: Variance of Method-of-Moments Estimator as a Fraction of Asymptotic OLS Variance under Cross-Sectional Heteroskedasticity

| | | Number of Time-Series Observations: $T = 1$ | | | | Number of Time-Series Observations: $T = 10$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Variance Measure | | | | | | | |
| Number of Observations per Group | Number of Groups | Asymptotic | Simple | Estimated | GLS | Asymptotic | Simple | Estimated | GLS |
| $P = 1$ | $N = 12$ | 1.0000 | 1.01094 (0.04737) | 0.67817 (0.01949) | 0.90301 | 1.0000 | 1.06510 (0.04748) | 0.63808 (0.01927) | 0.0934 |
| | $N = 25$ | 1.0000 | 1.03679 (0.04838) | 0.86988 (0.01867) | 0.69589 | 1.0000 | 1.01401 (0.04699) | 0.86700 (0.01702) | 0.1555 |
| $P = 5$ | $N = 12$ | 0.03928 | 0.04295 (0.00189) | 0.03662 (0.00086) | 0.01672 | 0.05615 | 0.05731 (0.00266) | 0.04950 (0.00103) | 0.0095 |
| | $N = 25$ | 0.05714 | 0.06077 (0.00273) | 0.03565 (0.00143) | 0.00276 | 0.06745 | 0.06650 (0.00292) | 0.05432 (0.00118) | 0.0051 |
| $P = 20$ | $N = 12$ | 0.00181 | 0.00176 (0.00007) | 0.00159 (0.00004) | 0.00042 | 0.00438 | 0.00459 (0.00020) | 0.00421 (0.00005) | 0.0016 |
| | $N = 25$ | 0.00314 | 0.00333 (0.00015) | 0.00301 (0.00005) | 0.00073 | 0.00388 | 0.00369 (0.00017) | 0.00380 (0.00004) | 0.0009 |

*Notes:* The column labeled "Asymptotic" reports the asymptotic method-of-moments variance divided by the OLS asymptotic variance. The asymptotic OLS variance is constructed by aggregating the data over the cross section into $N$ groups. The column labeled "Simple" gives the mean-squared deviation of the estimated method-of-moments estimate scaled by the asymptotic OLS variance The column labeled "Estimated" reports the average estimated method-of-moments variance divided by the asymptotic OLS variance (see Equation (26)). Cross-sectional correlation is homogeneous across groups, and time-series homoskedasticity is assumed. Cross-sectional heteroskedasticity is given by Equation (27). Standard errors are in parentheses. All data are constructed using 1000 replications for given regressors. The intra-group correlation coefficient = 0.5.

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.[14] 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, $\rho$, set to 0.1, 0.5, and 0.9. The degree of conditional heteroskedasticity is adjusted by allowing $\delta_1$ in (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 of 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 $\mathbf{\Omega}^{-1}\mathbf{x}$ 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 (19). As in the earlier simulations, the contemporaneous correlation in the historical disturbances is induced by using (20), and cross-sectional heteroskedasticity is induced by using (27). We then form the historical covariance matrix of the dependent variable for each of the industry groups,

$$\hat{\Sigma}_n^H = \frac{1}{T^H}\sum_{t=1}^{T^H} y_n y_n' \quad \forall n = 1,\ldots,N, \tag{28}$$

where $T^H$ is the number of previous time-series observations on the dependent variable, and $y_n$ is a $P \times 1$ vector. To be conservative, we assumed relatively few

[14] This bias can be seen by comparing the numbers in the column marked "simple" with those in the "estimated" column.

TABLE 3

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

| | Intra-Group Correlation = 0.1 | | | | Intra-Group Correlation = 0.5 | | | | Intra-Group Correlation = 0.9 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Level of Cross-Sectional Hetero-skedasticity | Asymp-totic | Simple | Esti-mated | GLS | Asymp-totic | Simple | Esti-mated | GLS | Asymp-totic | Simple | Esti-mated | GLS |
| 1 | 0.1308 | 0.1231 | 0.1224 | 0.1270 | 0.0591 | 0.0546 | 0.0546 | 0.0363 | 0.0373 | 0.0343 | 0.0338 | 0.0048 |
| | | (0.0059) | (0.0018) | | | (0.0025) | (0.0008) | | | (0.0016) | (0.0055) | |
| 2 | 0.1231 | 0.1197 | 0.0749 | 0.0088 | 0.0571 | 0.0557 | 0.0339 | 0.0028 | 0.0371 | 0.0363 | 0.0211 | 0.0004 |
| | | (0.0051) | (0.0028) | | | (0.0024) | (0.0013) | | | (0.0016) | (0.0008) | |
| 3 | 0.1223 | 0.1222 | 0.0774 | 0.0034 | 0.0569 | 0.0548 | 0.0342 | 0.0011 | 0.0370 | 0.0349 | 0.0208 | 0.0002 |
| | | (0.0053) | (0.0029) | | | (0.0024) | (0.0013) | | | (0.0016) | (0.0008) | |
| 4 | 0.1221 | 0.1248 | 0.0756 | 0.0024 | 0.0569 | 0.0566 | 0.0326 | 0.0008 | 0.0370 | 0.0358 | 0.0197 | 0.0001 |
| | | (0.0052) | (0.0028) | | | (0.0024) | (0.0013) | | | (0.0015) | (0.0008) | |

*Notes:* The levels of heteroskedasticity 1–4, are given by setting $\delta_1$ in Equation (27) to 0.0, 0.2, 1.0, and 4.0, respectively. The column labeled "Asymptotic" reports the asymptotic method-of-moments variance divided by the OLS asymptotic variance. The asymptotic OLS variance is constructed by aggregating the data over the cross section into $N$ groups. The column labeled "Simple" gives the mean-squared deviation of the estimated method-of-moments estimate scaled by the asymptotic OLS variance. The column labeled "Estimated" reports the average estimated method-of-moments variance divided by the asymptotic OLS variance. All assume $N = 25$, $P = 5$, and $T = 1$. All data use 1000 replications for given regressors. Standard errors are in parentheses. All data are constructed using 1000 replications for given regressors.

past observations are available, setting $T^H = 2p+1$. The entire historical covariance matrix is then

$$\widehat{\boldsymbol{\Omega}}^H = \mathbf{I}_T \otimes \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_1^H & 0 & \dots & 0 \\ 0 & \widehat{\boldsymbol{\Sigma}}_2^H & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \widehat{\boldsymbol{\Sigma}}_N^H \end{pmatrix}. \tag{29}$$

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

$$\hat{\mathbf{q}} = \left[\mathbf{x}\left(\widehat{\boldsymbol{\Omega}}^H\right)^{-1}\mathbf{x}\right]. \tag{30}$$

Table 4 reports measures of the relative performance of the 2STSLS 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

$$\frac{\mathbf{x}'\hat{\mathbf{q}}\left(\hat{\mathbf{q}}'\left(\mathbf{I}_T \otimes \widetilde{\boldsymbol{\Delta}}\right)\hat{\mathbf{q}}\right)^{-1}\hat{\mathbf{q}}'\mathbf{x}}{(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\left(\mathbf{I}_T \otimes \widetilde{\boldsymbol{\Delta}}\right)\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}}. \tag{31}$$

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

2STSLS even with relatively few industry groups and time-series observations. For example, with $T = 1, N = 12$, and $P = 5$, the asymptotic 2STSLS estimator is about 10 ($(0.1022)^{-1}$) 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.

TABLE 4

Simulation Results: Variance of 2STSLS Estimator as a Fraction of Asymptotic Method-of-Moments Variance

| | | Time Series Observations: $T = 1$ | | | | Time Series Observations: $T = 10$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Variance Measure | | | | | | | |
| Number of Observations per Group | Number of Groups | Asymptotic | Simple | Estimated | GLS | Asymptotic | Simple | Estimated | GLS |
| $P = 1$ | $N = 12$ | 0.9602 | 1.0768 (0.0478) | 0.5500 (0.0153) | 0.9030 | 0.1173 | 0.2005 (0.0091) | 0.1361 (0.0018) | 0.0935 |
| | $N = 25$ | 0.9682 | 1.0753 (0.0482) | 0.7525 (0.0164) | 0.6959 | 0.3357 | 0.4538 (0.0198) | 0.3152 (0.0031) | 0.1555 |
| $P = 5$ | $N = 12$ | 0.8166 | 0.8746 (0.0404) | 0.6242 (0.0145) | 0.4255 | 0.3081 | 0.3326 (0.0148) | 0.2911 (0.0033) | 0.1695 |
| | $N = 25$ | 0.1022 | 0.1805 (0.0080) | 0.1360 (0.0032) | 0.0484 | 0.1666 | 0.1561 (0.0074) | 0.1673 (0.0014) | 0.0769 |

*Notes:* The column labeled "Asymptotic" reports the asymptotic 2STSLS variance divided by the OLS asymptotic variance. The asymptotic OLS variance is constructed by aggregating the data over the cross section into $N$ groups. The column labeled "Simple" gives the mean-squared deviation of the estimated 2STSLS estimate scaled by the asymptotic OLS variance. The column labeled "Estimated" reports the average estimated 2STSLS variance divided by the asymptotic OLS variance. The 2STSLS estimator is contructed using the historical covariance matrix of the dependent variable, as in Equation (29). Cross-sectional correlation is homogeneous across groups, 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 parentheses. All data are constructed using 1000 replications for given regressors. The intra-group correlation coefficient = 0.5.

The "simple" variance measure is the mean squared deviation of the estimated 2STSLS 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 (19) are obtained by running OLS. A consistent estimate of the average covariance matrix, $\widetilde{\mathbf{\Delta}}$, is formed and then used 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

$$\frac{\mathbf{x}'\hat{\mathbf{q}}\left(\hat{\mathbf{q}}'\left(\mathbf{I}_T \otimes \hat{\widetilde{\mathbf{\Delta}}}\right)\hat{\mathbf{q}}\right)^{-1}\hat{\mathbf{q}}'\mathbf{x}}{(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\left(\mathbf{I}_T \otimes \widetilde{\mathbf{\Delta}}\right)\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}}. \tag{32}$$

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 2STSLS estimator investigated by Cragg (1983). Cragg also reports that the estimated variances of the 2STSLS 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 2STSLS 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 2STSLS 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.

TABLE 5

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

| | Intra-Group Correlation = 0.1 | | | | Intra-Group Correlation = 0.5 | | | | Intra-Group Correlation = 0.9 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Level of Cross-Sectional Hetero-skedasticity | Asymp-totic | Simple | Esti-mated | GLS | Asymp-totic | Simple | Esti-mated | GLS | Asymp-totic | Simple | Esti-mated | GLS |
| 1 | 1.0000 | 1.1000 | 0.7908 | 0.9710 | 0.9532 | 0.9546 | 0.7795 | 0.6146 | 0.4226 | 0.5812 | 0.4321 | 0.1312 |
| | | (0.0507) | (0.0127) | | | (0.0412) | (0.0124) | | | (0.0278) | (0.0064) | |
| 2 | 0.3734 | 0.3646 | 0.2762 | 0.0718 | 0.1299 | 0.2240 | 0.1506 | 0.0484 | 0.0292 | 0.0764 | 0.0984 | 0.0104 |
| | | (0.0162) | (0.0078) | | | (0.0098) | (0.0036) | | | (0.0048) | (0.0034) | |
| 3 | 0.1198 | 0.1743 | 0.1529 | 0.0279 | 0.0618 | 0.1116 | 0.1194 | 0.0194 | 0.0096 | 0.0600 | 0.0798 | 0.0042 |
| | | (0.0082) | (0.0049) | | | (0.0055) | (0.0037) | | | (0.0036) | (0.0034) | |
| 4 | 0.0488 | 0.1242 | 0.1082 | 0.0199 | 0.0242 | 0.0820 | 0.0873 | 0.0140 | 0.0076 | 0.0405 | 0.0718 | 0.0030 |
| | | (0.0057) | (0.0032) | | | (0.0045) | (0.0030) | | | (0.0024) | (0.0029) | |

*Notes:* The levels of heteroskedasticity 1–4 are given by setting $\delta_1$ in Equation (27) to 0.0, 0.2, 1.0, and 4.0, respectively. The column labeled "Asymptotic" reports the asymptotic 2STSLS variance divided by the OLS asymptotic variance. The asymptotic OLS variance is constructed by aggregating the data over the cross section into $N$ groups. The column labeled "Simple" gives the mean-squared deviation of the estimated 2STSLS estimate scaled by the asymptotic OLS variance. The column labeled "Estimated" reports the average estimated 2STSLS variance divided by the asymptotic OLS variance. All assume $N = 25$, $P = 5$, and $T = 1$. All data use 1000 replications for given regressors. Standard errors are in parentheses. All data are constructed using 1000 replications for given regressors

## VI. 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 that 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 the above approach. Clearly, the need to find uncorrelated industry groups is reduced in cases in which sufficient time-series data are available.

Several conclusions emerge from the simulations presented in Section V. 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 cost 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. Finally, there is a caveat. When correlation patterns across the $N$ groups are highly dissimilar, we would expect 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.

## Appendix

*Proposition 1.* Under Assumptions 1 and 2

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

*Proof.* By the Hölder inequality, Assumption 2(ii) implies that for every $i, j = 1, \ldots, K$, $E(|x_a^i x_b^j \mu_a \mu_b|)^{1+\delta} < \boldsymbol{\Delta}$ (where $a = (t-1)PN + (n-1)P + i$ and $b = (t-1)PN + (n-1)P + j$), and thus, for sufficiently large $NT$, $\boldsymbol{\Theta}_{NT}$ 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.

*Proposition 2*. Assumptions 1 and 2 also imply

$$\left(\frac{\mathbf{X}'_{NT}\mathbf{X}_{NT}}{NT}\right)^{-1}\widehat{\Theta}_{NT}\left(\frac{\mathbf{X}'_{NT}\mathbf{X}_{NT}}{NT}\right)^{-1}\overset{a.s.}{\rightarrow} M_{NT}^{-1}\Theta_{NT}M_{NT}^{-1}\ .$$

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

For each element $\beta^i$ of the vector $\beta$, there exists $\overline{\beta}^i$ within a finite compact neighborhood of $\beta^i$ given by $\Psi$, and there exists a finite element $\widetilde{\beta}^i$, such that $(\overline{\beta}^i-\beta^i)^2 \leqslant (\widetilde{\beta}^i-\beta^i)^2$ for all $\overline{\beta}^i \in \Psi$. Note that for all $j, k = 1,\ldots,K$, and for all $a, b = 1,\ldots,PNT$,

$$\left(y_a - x'_a\beta\right)\left(y_b - x'_b\beta\right)x_a^jx_b^k = \left(\mu_a - x'_a(\overline{\beta}-\beta)\right)\left(\mu_b - x'_b(\overline{\beta}-\beta)\right)x_a^jx_b^k$$

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

$$\left|\left(\mu_a - x'_a(\overline{\beta}-\beta)\right)\left(\mu_b - x'_b(\overline{\beta}-\beta)\right)x_a^jx_b^k\right|$$

$$\leqslant \sum_{i=1}^{K}\left|\left(\mu_a^2+\mu_b^2+\left(\left(x_a^i\right)^2+\left(x_b^i\right)^2\right)\left(\overline{\beta}^i-\beta^i\right)^2\right)x_a^jx_b^k\right|$$

$$\leqslant \sum_{i=1}^{K}\left|\left(\mu_a^2+\mu_b^2+\left(\left(x_a^i\right)^2+\left(x_b^i\right)^2\right)\left(\widetilde{\beta}^i-\beta^i\right)^2\right)x_a^jx_b^k\right| = m_{ab}\ ,$$

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

$$\begin{aligned} E\left(m_{ab}\right)^{1+\delta} &\leqslant k_0E\left(\left(\left|\mu_a^2x_a^ix_b^j\right|\right)^{1+\delta}\right) + k_1E\left(\left(\left|\mu_b^2x_a^jx_b^k\right|\right)^{1+\delta}\right) \\ &+ \sum_{i=2}^{K+1}k_iE\left(\left|\left(x_a^i\right)^2+\left(x_b^i\right)^2x_a^jx_b^k\right|\right)^{1+\delta}\left(\widetilde{\beta}^i-\beta^i\right)^{2\delta+2} \\ &\leqslant k_0E\left(\left(\left|\mu_a^2x_a^jx_b^k\right|\right)^{1+\delta}\right) + k_1E\left(\left(\left|\mu_b^2x_a^jx_b^k\right|\right)^{1+\delta}\right) \\ &+ \sum_{i=2}^{K+1}k_iE\left(\left(x_a^i\right)^2+\left(x_b^i\right)^2\right)^{2+2\delta}\left(\widetilde{\beta}^i-\beta^i\right)^{2\delta+2}, \end{aligned}$$

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 $(\widetilde{\beta}^i-\beta^i)^2$ is finite by assumption. In addition, Assumptions 1 and 2 in the text guarantee that $E((|\mu_a^2x_a^jx_b^k|)^{1+\delta})$, $E((|\mu_b^2x_a^jx_b^k|)^{1+\delta})$, 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

$$N^{-1}\sum_{n=1}^{N} T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{P}\sum_{j=1}^{P}\left(y_a - x_a'\beta\right)\left(y_b - x_b'\beta\right)x_a^l x_b^k\right)$$
$$-N^{-1}\sum_{n=1}^{N} T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{P}\sum_{j=1}^{P}E\left(\left(y_a - x_a'\beta\right)\left(y_b - x_b'\beta\right)x_a^l x_b^k\right)\right)\overset{a.s.}{\to} 0 .$$

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

$$N^{-1}\sum_{n=1}^{N} T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{P}\sum_{j=1}^{P}\left(y_a - x_a'\hat{\beta}_N\right)\left(y_b - x_b'\hat{\beta}_N\right)x_a^l x_b^k\right)$$
$$-N^{-1}\sum_{n=1}^{N} T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{P}\sum_{j=1}^{P}E\left(\mu_a \mu_b x_a^l x_b^k\right)\right)\overset{a.s.}{\to} 0 ,$$

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

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

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

The continuity of the matrix inverse therefore implies

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

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

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

This proves Proposition 2.

*Proposition 3*. The OLS estimate of the coefficient covariance matrix in (9), $\hat{V}(\hat{\beta}^{**})$, converges almost surely to the true coefficient covariance matrix

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

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

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

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

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

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

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

which proves Proposition 3.

## References

Beaver, W. H.; C. Eger; S. Ryan; and M. Wolfson. "Financial Reporting and the Structure of Bank Share Prices." Working Paper, Stanford Univ. (July 1985).

Bernard, V. L. "Unanticipated Inflation and the Value of the Firm." *Journal of Financial Economics,* 15 (March 1986a), 285–321.

——————. "Cross-Sectional Dependence and Problems in Inference in Market-Based Accounting Research." Working Paper, Univ. of Michigan (Oct. 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,* M. C. Jensen, ed. New York: Praeger (1972), 79–121.

Brown, S., and J. Warner. "Measuring Security Price Performance." *Journal of Financial Economics,* 8 (Sept. 1980), 205–258.

——————. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics,* 14 (March 1985), 3–32.

Brown, S., and M. Weinstein. "A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm." *Journal of Finance,* 38 (June 1983), 711–743.

Chamberlain, G., and M. Rothschild. "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets." *Econometrica,* 51 (Sept. 1983), 1281–1304.

Christie, A. "On Cross-Sectional Analysis in Accounting Research." Working Paper, Univ. of Southern California (1985).

Collins, D, and W. Dent. "A Comparison of Alternative Testing Methodologies Used in Capital Markets Research." *Journal of Accounting Research,* 22 (Spring 1984), 48–84.

Cragg, J. G. "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form." *Econometrica,* 51 (May 1983), 751–763.

Cumby, R.; J. Huizinga; and M. Obstfeld. "Two-Step Two-Stage Least Squares Estimation in Models with Rational Expectations." *Journal of Econometrics,* 21 (April 1983), 333–355.

Eicker, F. "Limit Theorems for Regressions with Unequal and Dependent Errors." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* L. LeCam and J. Neyman, eds. Berkeley, CA: Univ. of California Press (1967), 59–82.

Gibbons, M. "Multivariate Tests of Financial Models: A New Approach." *Journal of Financial Economics,* 10 (March 1982), 3–27.

Hansen, L. P. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica,* 50 (July 1982), 1029–1054.

Ingersol, J. "Some Results in the Theory of Arbitrage Pricing." *Journal of Finance,* 39 (Sept. 1984), 1021–1039.

Jaffe, J. "Special Information and Insider Trading." *Journal of Business,* 47 (July 1974), 410–428.

Kinderman, A. J., and J. G. Ramage. "Computer Generation of Normal Random Numbers." *Journal of the American Statistical Association,* 71 (Dec. 1976), 893–896.

Malatesta, P. "Measuring Abnormal Performance: The Event Parameter Approach Using Joint Generalized Least Squares." *Journal of Financial and Quantitative Analysis,* 21 (March 1986), 27–38.

Malatesta, P., and R. Thompson. "Partially Anticipated Events: A Model of Stock Price Reaction with an Application to Corporate Acquisitions." *Journal of Financial Economics,* 14 (June 1985), 237–250.

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,* 65 (Jan. 1970), 161–172.

Rothenberg, T. "Approximate Normality of Generalized Least Squares Estimates." *Econometrica,* 52 (July 1984), 811–825.

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, Univ. of British Columbia (Jan. 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,* 21 (Spring 1983), 184–221.

________________. "The Impact of Merger-Related Regulations Using Exact Distributions of Test Statistics." *Journal of Accounting Research,* 23 (Spring 1985), 408–415.

Thompson, R. "Conditioning the Return-Generating Process on Firm-Specific Events: A Discussion of Event Study Methodologies." *Journal of Financial and Quantitative Analysis,* 20 (June 1985), 151–172.

White, H. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica,* 48 (May 1980a), 817–838.

________________. "Nonlinear Regression on Cross-Section Data." *Econometrica,* 48 (April 1980b), 721–746.

________________. *Asymptotic Theory for Econometricians.* New York: Academic Press (1984).