September, 2005

Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches

Mitchell A. Petersen
Kellogg School of Management, Northwestern University
and NBER

Abstract

In both corporate finance and asset pricing empirical work, researchers are often confronted with panel data. In these data sets, the residuals may be correlated across firms and across time, and OLS standard errors can be biased. Historically, the two literatures have used different solutions to this problem. Corporate finance has relied on clustered standard errors, while asset pricing has used the Fama-MacBeth procedure to estimate standard errors. This paper examines the different methods used in the literature and explains when the different methods yield the same (and correct) standard errors and when they diverge. The intent is to provide intuition as to why the different approaches sometimes give different answers and give researchers guidance for their use.

I thank the Center for Financial Institutions and Markets at Northwestern University's Kellogg School for support. In writing this paper, I have benefitted greatly from discussions with Toby Daglish, Kent Daniel, Michael Faulkender, Wayne Ferson, Mariassunta Giannetti, John Graham, Chris Hansen, Wei Jiang, Toby Moskowitz, Joshua Rauh, Michael Roberts, Paola Sapienza, Georgios Skoulakis, Doug Staiger, and Annette Vissing-Jorgensen as well as the comments of seminar participants at the Federal Reserve Bank of Chicago, Northwestern University, Stanford University, and the Universities of California at Berkeley, Chicago, Columbia, and Iowa. The research assistance of Sungjoon Park, Nick Halpern, Casey Liang, and Amit Patel is greatly appreciated.

I) Introduction

It is well known that OLS standard errors are unbiased when the residuals are independent and identically distributed. When the residuals are correlated across observations, OLS standard errors can be biased and either over or underestimate the true variability of the coefficient estimates. Although the use of panel data sets (e.g. data sets that contain observations on multiple firms in multiple years) is common in finance, the ways that researchers have addressed possible biases in the standard errors varies widely and in many cases is incorrect. In recently published finance papers which include a regression on panel data, forty-two percent of the papers did not adjust the standard errors for possible dependence in the residuals.[1] Approaches for estimating the coefficients and standard errors in the presence of within cluster correlation varied among the remaining papers. Thirty-four percent of the papers estimated both the coefficients and the standard errors using the Fama-MacBeth procedure (Fama-MacBeth, 1973). Twenty-nine percent of the papers included dummy variables for each cluster (e.g. fixed effects). The next two most common methods used OLS (or an analogous method) to estimate the coefficients but reported standard errors adjusted for correlation within a cluster. Seven percent of the papers adjusted the standard errors using the

[1] I searched papers published in the *Journal of Finance*, the *Journal of Financial Economics*, and the *Review of Financial Studies* in the years 2001- 2004 for a description of how the coefficients and standard errors were estimated in a panel data set. I included both linear regressions as well as non-linear techniques such as logits and tobits in my survey. Panel data sets are data sets where observations can be grouped into clusters (e.g. multiple observations per firm, per industry, per year, or per country). I included only papers which report at least five observations in each dimension (e.g. firms and years). 207 papers met the selection criteria. Papers which did not report the method for estimating the standard errors, or reported correcting the standard errors only for heteroscedasticity (i.e. White standard errors which are not robust to within cluster dependence), were coded as not having corrected the standard errors for within cluster dependence. Where the paper's description was ambiguous, I contacted the authors.

Although White/OLS standard errors may sometimes be correct, many of the published papers report regressions where I would expect the residuals to be correlated across observations on the same firm in different years (e.g. bid-ask spread regressed on exchange dummies, stock price, volatility, and average daily volume or leverage regressed on the market to book ratio and firm size) or correlated across observations on different firms in the same year (e.g. equity returns regresses on earnings surprises). In these cases, the bias in the standard errors can be quite large. See Section VI for two illustrations.

Newey-West procedure (Newey and West, 1987) modified for use in a panel data set, while 23 percent of the papers reported clustered standard errors (Williams, 2000, Rogers, 1993, Andrews, 1991, Moulton, 1990, Arellano, 1987, Moulton, 1986) which are White standard errors adjusted to account for possible correlation within a cluster. These are also called Rogers standard errors.

Although the literature has used a diversity of methods to estimate standard errors in panel data sets, the chosen method is often incorrect and the literature provides little guidance to researchers as to which method should be used. Since the methods sometimes produce incorrect estimates, it is important to understand how the methods compare and how to select the correct one. That is this paper's objective.

There are two general forms of dependence which are most common in finance applications. They will serve as the basis for the analysis. The residuals of a given firm may be correlated across years (time series dependence) for a given firm. I will call this a firm effect. Alternatively, the residuals of a given year may be correlated across firms (cross-sectional dependence). I will call this a time effect. I will simulate panel data with both forms of dependence, first individually and then jointly. With the simulated data, I can estimate the coefficients and standard errors using each of the methods and compare their relative performance.

Section II contains the standard error estimates in the presence of a fixed firm effect. My results show that both OLS and the Fama-MacBeth standard errors are biased downward. The Newey-West standard errors, as modified for panel data, are also biased but the bias is small. Of the most common approaches used in the literature and examined in this paper, only clustered standard errors are unbiased as they account for the residual dependence created by the firm effect.

In Section III, the same analysis is conducted with a time effect instead of a firm effect. Since the Fama-MacBeth procedure is designed to address a time effect, not a firm effect, the Fama-MacBeth standard errors are unbiased. The intuition of these first two sections carries over to Section IV, were I simulate data with both a firm and a time effect.

Thus far, I have specified the firm effect as a constant (e.g. it does not decay over time). In practice, the firm effect may decay and so the correlation between residuals declines as the time between them grows. In Section V, I simulate data with a more general correlation structure. This allows me to compare OLS, clustered, and Fama-MacBeth standard errors in a more general setting. I show that OLS and Fama-MacBeth standard errors are biased and clustered standard errors are unbiased. Simulating the temporary firm effect also allows me to examine the relative accuracy of two additional methods for adjusting standard errors: fixed effects (firm dummies) and an adjusted Fama-MacBeth standard error whose use is becoming more popular. I show that including fixed effects eliminates the bias in OLS standard errors only when the firm effect is fixed. I also show that even after adjusting Fama-MacBeth standard errors, as suggested by some authors, they are still biased (Cochrane, 2001).

Most papers do not report standard errors estimated by multiple methods. Thus in Section VI, I apply the various estimation techniques to two real data sets and compare their relative performance. This serves two purposes. First, it demonstrates that the methods used in some published papers produce biases in the standard errors and t-statistics which are significant. This is why using the correct method to estimate standard errors is important. Examining actual data also allows me to show how differences in standard error estimates (e.g. White versus clustered standard error) can provide information about the deficiency in a model and directions for improving them.

II) Estimating Standard Errors in the Presence of a Fixed Firm Effect.

A) Clustered Standard Error Estimates.

To provide intuition on why the standard errors produced by OLS are incorrect and how clustered standard errors correct this problem, it is helpful to very briefly review the expression for the variance of the estimated coefficients. The standard regression for a panel data set is:

$$Y_{it} = X_{it}\,\beta + \varepsilon_{it} \qquad \textbf{(1)}$$

where we have observations on firms (i) across years (t). X and ε are assumed to be independent of each other and to have a zero mean. The zero mean is without loss of generality and allows us to calculate variances as sums of the squares of the variable. The estimated coefficient is:

$$\begin{aligned}\hat{\beta}_{OLS} &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\,Y_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\,(X_{it}\,\beta + \varepsilon_{it})}{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2} \\ &= \beta + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\,\varepsilon_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2}\end{aligned} \qquad \textbf{(2)}$$

and, taking the regressors as fixed, the variance of the coefficient is:

$$\begin{aligned}Var[\,\hat{\beta}_{OLS} - \beta\,] &= E\left[\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\,\varepsilon_{it}\right)^2 \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2\right)^{-2}\right] \\ &= E\left[\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2\,\varepsilon_{it}^2\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2\right)^{-2}\right] \\ &= NT\,\sigma_X^2\,\sigma_\varepsilon^2\,(\,NT\,\sigma_X^2\,)^{-2} \\ &= \frac{\sigma_\varepsilon^2}{NT\,\sigma_X^2}\end{aligned} \qquad \textbf{(3)}$$

This is the standard OLS formula and is based on the assumption that the errors are independent and identically distributed (Green, 2000). The independence assumption is used to move from the first to the second line in equation (3) (i.e., the covariance between residuals is zero). The assumption of an identical distribution (e.g., homoscedastic errors) is used to move from the second to the third line.[2] The independence assumption is often violated in panel data and this is the focus of the paper.

In relaxing the assumption of independent errors, I initially assume the data has a fixed firm effect. Thus the residuals consist of a firm specific component (γ_i) as well as a component which is unique to each observation (η_{it}). The residuals can be specified as:

$$\varepsilon_{it} = \gamma_i + \eta_{it} \tag{4}$$

Assume that the independent variable X also has a firm specific component.

$$X_{it} = \mu_i + \nu_{it} \tag{5}$$

Each of the components of X (μ and ν) and ε (γ and η) are independent of each other. This is necessary for the coefficient estimates to be consistent.[3] Both the independent variable and the residual are correlated across observations of the same firm, but are independent across firms.

[2] Clustered standard errors are robust to heteroscedasticity. Since this is not my focus, I assume the errors are homoscedastic. I use White standard errors as my baseline estimates when analyzing actual data in Section VI, since the residuals are not homoscedastic in those data sets (White, 1984).

[3] I am assuming that the model is correctly specified. I do this to focus on estimating the standard errors. In actual data sets, this assumption does not necessarily hold and would need to be tested.

$$
\begin{aligned}
\operatorname{corr}(X_{it}, X_{js}) &= 1 && \text{for } i=j \text{ and } t = s \\
&= \rho_X = \sigma_\mu^2 / \sigma_X^2 && \text{for } i=j \text{ and all } t \neq s \\
&= 0 && \text{for all } i \neq j \\
\operatorname{corr}(\varepsilon_{it}, \varepsilon_{js}) &= 1 && \text{for } i=j \text{ and } t = s \\
&= \rho_\varepsilon = \sigma_\gamma^2 / \sigma_\varepsilon^2 && \text{for } i=j \text{ and all } t \neq s \\
&= 0 && \text{for all } i \neq j
\end{aligned}
\tag{6}
$$

Given this data structure [equations (1), (4), and (5)], I can calculate the true standard error of the OLS coefficient. Since the residuals are no longer independent within cluster, the square of the summed residuals is not equal to the sum of the squared residuals. The same statement can be made about the independent variable. The co-variances must be included as well. The variance of the OLS coefficient estimate is:

$$
\begin{aligned}
\operatorname{Var}[\hat{\beta}_{OLS} - \beta] &= E\left[\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\,\varepsilon_{it}\right)^2 \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2\right)^{-2}\right] \\
&= E\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T} X_{it}\,\varepsilon_{it}\right)^2 \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2\right)^{-2}\right] \\
&= E\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T} X_{it}^2\,\varepsilon_{it}^2 + 2\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} X_{it}X_{is}\,\varepsilon_{it}\,\varepsilon_{is}\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2\right)^{-2}\right] \\
&= \left(N\,T\,\sigma_X^2\,\sigma_e^2 + N\,T\,(T-1)\,\rho_x\,\sigma_X^2\,\rho_e\,\sigma_e^2\right)\left(NT\,\sigma_X^2\right)^{-2} \\
&= \frac{\sigma_\varepsilon^2}{NT\,\sigma_X^2}\left(1 + (T-1)\,\rho_X\,\rho_\varepsilon\right)
\end{aligned}
\tag{7}
$$

I use the assumption that residuals are independent across firms in deriving the second line. Given the assumed data structure, the within cluster correlations of both X and ε are positive and are equal to the fraction of the variance which is attributable to the fixed firm effect. When the data have a fixed firm effect, the OLS standard errors will always understate the true standard error if and only if both ρ_X or ρ_ε are non-zero.[4] The magnitude of the error is also increasing in the number of years in the data (see Bertrand, Duflo, and Mullainathan, 2004). To understand this intuition, consider the extreme case where the independent variables and residuals are perfectly correlated across time (i.e. $\rho_X = 1$ and $\rho_\varepsilon = 1$). In this case, each additional year provides no additional information and will have no effect on the true standard error. However, the OLS standard errors will assume each additional year provides N additional observations and the estimated standard error will shrink accordingly and incorrectly.

The correlation of the residuals within cluster is the problem the clustered standard errors (White standard errors adjusted for clustering) are designed to correct.[5] By squaring the sum of $X_{it}\varepsilon_{it}$ within each cluster, the covariance between residuals within cluster is estimated (see Figure 1). This correlation can be of any form; no parametric structure is assumed. However, the squared sum of $X_{it}\varepsilon_{it}$ is assumed to have the same distribution across the clusters. Thus these standard errors are consistent as the number of clusters grows (Donald and Lang, 2001; and Wooldridge, 2002). I return

[4] If the firm effect is not fixed, the variance of the coefficient estimate is a weighted sum of the correlations between ε_t and ε_{t-k} times the correlation between X_t and X_{t-k}, for all k<T and is equal to:

$$\mathrm{Var}[\hat{\beta}_{OLS} - \beta] = \frac{\sigma_\varepsilon^2}{NT\,\sigma_X^2}\left(1 + \frac{2}{T}\sum_{k=1}^{T}(T-k)\,\rho_{x,k}\,\rho_{\varepsilon,k}\right)$$

The auto-correlations can be positive or negative. It is thus possible for the OLS standard error to under or over-estimate the true standard error. I will address auto-correlations which decline as the lag length (k) increases in Section V. If the panel is unbalanced (different T for each i), the true standard error and the bias in the OLS standard errors is even larger than equation (7) (see Moulton, 1986).

[5] The exact formula for the clustered standard error is:

$$S^2(\beta) = \frac{N\,(NT-1)\sum_{i=1}^{N}\left(\sum_{t=1}^{T} X_{it}\,\varepsilon_{it}\right)^2}{(NT-k)\,(N-1)\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2\right)^2}$$

to this issue in Section III.

B) Testing the Standard Error Estimates by Simulation.

I simulated a panel data set and then estimated the slope coefficient and its standard error. By doing this multiple times we can observe the true standard error as well as the average estimated standard errors.[6] In the first version of the simulation, I included a fixed firm effect but no time effect in the independent variable and the residual. Thus the data are simulated as described in equations (4) and (5). Across simulations I assumed that the standard deviation of the independent variable and the residual are both constant at one and two respectively. This will produce an R^2 of 20 percent. Across different simulations, I altered the fraction of the variance in the independent variable which is due to the firm effect. This fraction ranges from zero to seventy-five percent in twenty-five percent increments (see Table 1). I did the same for the residual. This allows me to demonstrate how the magnitude of the bias in the OLS standard errors varies with the strength of the firm effect in both the independent variable and the residual.

The results of the simulations are reported in Table 1. The first two entries in each cell are the average value of the slope coefficient and the standard deviation of the coefficient estimate. The standard deviation is the true standard error of the coefficient and ideally the estimated standard error will be close to this number. The average standard error estimated by OLS is the third entry in each cell and is the same as the true standard error in the first row of the table. When there is no firm effect in the residual (i.e. the residuals are independent across observations), the standard error

[6] Each simulated data set contains 5,000 observations (500 firms and 10 years per firm). The components of the independent variable (μ ν) and the residual (γ η) are independent of each other and normally distributed with zero means. For each data set, I estimated the coefficients and standard errors using each method described below. The reported means and standard deviations reported in the tables are based on 5,000 simulations. The basic program which I used to simulate the data and estimate the coefficients and standard errors is posted on my web site. I have also posted the code for estimating the different standard errors which are discussed in this paper.

estimated by OLS is correct (see Table 1, row 1). When there is no firm effect in the independent variable (i.e. the independent variable is independent across observations), the standard errors estimated by OLS are also unbiased, even if the residuals are highly correlated (see Table 1, column 1). This follows from the intuition in equation (7). The bias in the OLS standard errors is a product of the dependence in the independent variable (ρ_X) and the residual (ρ_ε). When either correlation is zero, OLS standard errors are unbiased.

When there is a firm effect in both the independent variable and the residual, then the OLS standard errors underestimate the true standard errors, and the magnitude of the underestimation can be large. For example, when fifty percent of the variability in both the residual and the independent variable is due to the fixed firm effect ($\rho_X = \rho_\varepsilon = 0.50$), the OLS estimated standard error is one half of the true standard error (0.557 = 0.0283/0.0508).[7] The standard errors estimated by OLS do not rise as the firm effect increases across either the columns (i.e. in the independent variable) or across the rows (i.e. in the residual). The true standard error does rise.

When I estimate the standard error of the coefficient using clustered standard errors, the estimates are very close to the true standard error. These estimates rise along with the true standard error as the fraction of variability arising from the firm effect increases. The clustered standard errors correctly account for the dependence in the data common in a panel data set (Rogers, 1993, Williams, 2000) and produce unbiased estimates.

[7] All of the regressions contained a constant whose true value is zero. The above intuition carries over to the intercept estimation. The estimated slope coefficient averages -0.0003 with a standard deviation of 0.0669, when $\rho_X = \rho_\varepsilon = 0.50$. The OLS standard errors are biased (0.0283) and the clustered standard errors are unbiased (0.0663).

The simulated residuals are homoscedastic, so calculating standard errors which are robust to heteroscedasticity is not necessary. When I estimated White standard errors in the simulation they have the same bias as the OLS standard errors. For example, the average White standard error was 0.0283 compared to the OLS estimate of 0.0283 and a true standard error of 0.0508 when $\rho_X = \rho_\varepsilon = 0.50$.

An alternative way to examine the magnitude of the bias is to examine the empirical distribution of the simulated t-statistics. The t-statistics based on the OLS standard errors are too large in absolute value (see Figure 2-A). 15.3 percent of the OLS t-statistics are statistically significant at the 1 percent level (i.e. greater than 2.58). This is the intuition we saw in Tables 1. The clustered standard errors are unbiased (see Table 1), and the empirical distribution of the t-statistics is also correct (see Figure 2-B). 0.9 percent of the clustered t-statistics are significant at the one percent level. The reason the t-statistics give us the same intuition as the standard errors is because the standard errors are estimated very precisely. For example, the mean OLS standard error is 0.0282 with a standard deviation of 0.0007 and the mean clustered standard error is 0.0508 with a standard deviation of 0.0027 (when $\rho_X = \rho_\varepsilon = 0.50$).[8]

The bias in OLS standard errors is highly sensitive to the number of time periods (years) used in the estimation as well. As the number of years doubles, OLS assumes a doubling of the information. However if the independent variable and the residual are correlated within the cluster, the amount of information (independent variation) increases by less than a factor of two. The bias rises from about 30 percent when there are five years of data per firm to 73 percent when there are 50 years (when $\rho_X=\rho_\varepsilon=0.50$, see Figure 3). The robust standard errors are consistently close to the true standard errors independent of the number of time periods (see Figure 3).

Most of the simulations in the paper are based on linear regressions. To evaluate the performance of the standard error estimates in a non-linear setting, I simulated data according to equations 4 and 5. I took y as the latent variable and either censored the bottom 25% of the data (y<-

[8] I do not report the MSE of the standard error estimates, since they add no additional information beyond what is reported in Tables 1 and 2. Since the variances of the standard error estimates is extremely small, the MSE are essentially equal to the bias squared. Tables of MSE are available from the author.

1.5) or created a dummy variable (equal to one if y is positive, and zero otherwise). With this data I estimated a tobit and a probit model. The usual ("OLS") standard errors are too small and the clustered standard errors are unbiased. The magnitude of the underestimate is the same as reported in Table 1 for the tobit model and slightly smaller in the probit model. The results are available from the author.

C) Fama-MacBeth Standard Errors: The Equations

An alternative way to estimate the regression coefficients and standard errors when the residuals are not independent is the Fama-MacBeth approach (Fama and MacBeth, 1973). In this approach, the researcher runs T cross sectional regressions. The average of the T estimates is the coefficient estimate.

$$\hat{\beta}_{FM} = \sum_{t=1}^{T} \frac{\hat{\beta}_t}{T} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\sum_{i=t}^{N} X_{it} Y_{it}}{\sum_{i=t}^{N} X_{it}^2} \right) = \beta + \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\sum_{i=t}^{N} X_{it} \varepsilon_{it}}{\sum_{i=t}^{N} X_{it}^2} \right) \qquad \textbf{(8)}$$

and the estimated variance of the Fama-MacBeth estimate is calculated as:

$$S^2(\hat{\beta}_{FM}) = \frac{1}{T} \sum_{t=1}^{T} \frac{(\hat{\beta}_t - \hat{\beta}_{FM})^2}{T - 1} \qquad \textbf{(9)}$$

The variance formula, however, assumes that the yearly estimates of the coefficient (β_t) are independent of each other. This is only correct if $X_{it}\ \varepsilon_{it}$ is uncorrelated with $X_{is}\ \varepsilon_{is}$ for $t \neq s$. As discussed above, this is not true when there is a firm effect in the data (i.e. $\rho_X\ \rho_\varepsilon \neq 0$). Thus, Fama-MacBeth variance estimate is too small in the presence of a firm effect. In the presence of a firm effect, the true variance of the Fama-MacBeth estimate is:

$$\begin{aligned} \text{Var}(\hat{\beta}_{FM}) &= \frac{1}{T^2}\,\text{Var}\Big(\sum_{t=1}^{T}\hat{\beta}_t\Big) \\ &= \frac{\text{Var}(\hat{\beta}_t)}{T} + \frac{2\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\text{Cov}(\hat{\beta}_t,\hat{\beta}_s)}{T^2} \\ &= \frac{\text{Var}(\hat{\beta}_t)}{T} + \frac{T\,(T-1)}{T^2}\,\text{Cov}(\hat{\beta}_t,\hat{\beta}_s) \end{aligned} \tag{10}$$

Given our specification of the data structure (equations 4 and 5), the covariance between the coefficient estimates of different years is independent of t-s (which justifies the simplification in the last line of equation 10) and can be calculated as follows for $t \neq s$:

$$\begin{aligned} \text{Cov}(\hat{\beta}_t,\hat{\beta}_s) &= E\left[\left(\sum_{i=1}^{N} X_{it}^2\right)^{-1}\left(\sum_{i=1}^{N} X_{it}\,\varepsilon_{it}\right)\left(\sum_{i=1}^{N} X_{is}\,\varepsilon_{is}\right)\left(\sum_{i=1}^{N} X_{is}^2\right)^{-1}\right] \\ &= (N\,\sigma_X^2)^{-2}\,E\left[\left(\sum_{i=1}^{N} X_{it}\,\varepsilon_{it}\right)\left(\sum_{i=1}^{N} X_{is}\,\varepsilon_{is}\right)\right] \\ &= (N\,\sigma_X^2)^{-2}\,E\left[\sum_{i=1}^{N} X_{it}\,X_{is}\,\varepsilon_{it}\,\varepsilon_{is}\right] \\ &= (N\,\sigma_X^2)^{-2}\,N\,\rho_X\,\sigma_X^2\,\rho_\varepsilon\,\sigma_\varepsilon^2 \\ &= \frac{\rho_X\,\rho_\varepsilon\,\sigma_\varepsilon}{N\,\sigma_X^2} \end{aligned} \tag{11}$$

Combining equations (10) and (11) gives us an expression for the true variance of the Fama-MacBeth coefficient estimates.

$$
\begin{aligned}
\operatorname{Var}(\hat{\beta}_{FM}) &= \frac{\operatorname{Var}(\hat{\beta}_t)}{T} + \frac{T\,(T-1)}{T^2}\operatorname{Cov}(\hat{\beta}_t\,,\,\hat{\beta}_s) \\
&= \frac{1}{T}\left(\frac{\sigma_\varepsilon^2}{N\,\sigma_X^2}\right) + \frac{T\,(T-1)}{T^2}\left(\frac{\rho_X\,\rho_\varepsilon\,\sigma_\varepsilon^2}{N\,\sigma_X^2}\right) \\
&= \frac{\sigma_\varepsilon^2}{NT\,\sigma_X^2}\left(1 + (T-1)\,\rho_X\,\rho_\varepsilon\right)
\end{aligned}
\qquad \textbf{(12)}
$$

This is same as our expression for the variance of the OLS coefficient (see equation 7). The Fama-MacBeth standard error are biased in exactly the same way as the OLS estimates. In both cases, the magnitude of the bias is a function of the serial correlation of both the independent variable and the residual within a cluster and the number of time periods per firm.

D) Simulating Fama-MacBeth Standard Errors.

To document the bias of the Fama-MacBeth standard error estimates, I calculate the Fama-MacBeth estimate of the slope coefficient and the standard error in each of the 5,000 simulated data sets which were used in Table 1. The results are reported in Table 2. The Fama-MacBeth estimates are consistent and as efficient as OLS (the correlation between the two is consistently above 0.99). The standard deviation of the two coefficient estimates is also the same (compare the second entry in each cell of Table 1 and 2). These results demonstrate that both OLS and Fama-MacBeth standard errors are biased downward (see Table 2). However the Fama-MacBeth standard errors have a larger bias than the OLS standard errors. For example, when both ρ_X and ρ_ε are equal to 75 percent, the OLS standard error has a bias of 60% (0.595 = 1 - 0.0283/0.0698, see Table I) and the Fama-MacBeth standard error has a bias of 74 percent (0.738 = 1 - 0.0699/0.0183, see Table II). Moving down the diagonal of Table 2 from upper left to bottom right, the true standard error increases but the standard error estimated by Fama-MacBeth shrinks. As the firm effect becomes larger (ρ_X ρ_ε

increases), the OLS bias grows, and the Fama-MacBeth bias grows even faster.[9] The incremental bias of the Fama-MacBeth standard errors is due to the way in which the estimated variance is calculated. To see this we need to expand the expression of the estimated variance (equation 9).

$$\begin{aligned} Var[\beta_{FM}] &= \frac{1}{T(T-1)} \sum_{t=1}^{T} \left[\frac{\sum_{i=1}^{N} X_{it}\ \varepsilon_{it}}{\sum_{i=1}^{N} X_{it}^2} - \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\sum_{i=1}^{N} X_{it}\ \varepsilon_{it}}{\sum_{i=1}^{N} X_{it}^2} \right) \right]^2 \\ &= \frac{1}{T(T-1)} \sum_{t=1}^{T} \left[\frac{\sum_{i=1}^{N} (\mu_i + \nu_{it})(\gamma_i + \eta_{it})}{\sum_{i=t}^{N} (\mu_i + \nu_{it})^2} - \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\sum_{i=1}^{N} (\mu_i + \nu_{it})(\gamma_i + \eta_{it})}{\sum_{i=1}^{N} (\mu_i + \nu_{it})^2} \right) \right]^2 \end{aligned} \quad \textbf{(13)}$$

The true variance of the Fama-MacBeth coefficients is a measure of how far each yearly coefficient estimate deviates from the true coefficient (one in our simulations). The estimated variance, however, measures how far each yearly estimate deviates from the sample average. Since the firm effect influences both the yearly coefficient estimate and the sample average of the yearly coefficient estimates, it does not appear in the estimated variance. Thus increases in the firm effect (increases in $\rho_X\rho_\varepsilon$) actually reduce the estimated Fama-MacBeth standard error at the same time it increases the true standard error of the estimated coefficients. To make this concrete, take the extreme example where $\rho_X\rho_\varepsilon$ is equal to one; the true standard error is $(\sigma_\varepsilon/N\sigma_X)^{½}$ while the estimated Fama-MacBeth standard error is zero. This additional source of bias shrinks as the number of years increases since the estimated slope coefficient will converge to the true coefficient (see Figure 3).

Although I have just demonstrated that the Fama-MacBeth standard errors are biased in the

[9] The distribution of empirical t-statistics is even wider for the Fama-MacBeth than for OLS (see Figures 2-A and 2-C). 25 percent of the Fama-MacBeth t-statistics are statistically significant at the 1 percent level compared to 15 percent of the OLS t-statistics (when $\rho_X = \rho_\varepsilon = 0.50$).

presence of a firm effect, they are often used to measure statistical significance in published papers when the underlying regression contains a firm effect. As part of my literature survey, I looked for papers which regressed a persistent firm characteristics on other persistent firm characteristics. This is the type of data that will likely contain a firm effect. Since I am not able to replicate each of the studies, I will discuss several examples of where Fama-MacBeth standard errors have been used with such data.

The first example is a logit estimate of whether a firm pays a dividend (a highly persistent variable) on firm characteristics such as the firm's market to book ratio, the earnings to assets ratio, and relative firm size (Fama and French, 2001). Another example are the papers which examine how the market values firms by regressing a firm's market to book ratio on firm characteristics such as the firm's age, a dummy for whether it pays a dividend, leverage, and firm size (Pastor and Veronesi, 2003, and Kemsley and Nissim, 2002).[10] A third example are the papers which run capital structure regressions. In these papers, the authors try to explain a firm's use of leverage by regressing the firm's debt to asset ratio on firm characteristics such as the firm's market to book ratio, the ratio of property, plant, and equipment to total assets, the earnings to assets, depreciation to asset ratio, R&D to assets ratio, and firm size (see for example Baker and Wurgler, 2002, Fama and French, 2002, and Johnson, 2003).[11] Since both the left and right hand side variables in these

[10] Both papers correct the Fama-MacBeth standard errors for the first order auto-correlation of the estimated slopes. Pastor and Veronesi (2003) report that this does not change their answer. I will show in Section V-C that this correction still produces biased standard errors and this may explain Pastor and Veronesi's finding that the adjustment has little effect on their estimated standard errors.

[11] Baker and Wurgler (2002) estimate both White and Fama-MacBeth standard errors but do not report the Fama-MacBeth standard errors since they are the same as the White standard errors. Fama and French (2002) acknowledge that Fama-MacBeth standard errors may understate the true standard errors and so report adjusted Fama-MacBeth standard errors which I will discuss in Section V-C ("We use a less formal approach. We assume the standard errors of the average slopes... should be inflated by a factor of 2.5").

three regressions are highly persistent, this is the kind of data which likely contains a significant firm effect. In Section VI-B, I will estimate a capital structure regression and show that the magnitude of the bias is indeed large. Despite the presence of a firm effect and the resulting bias of Fama-MacBeth standard errors documented above, they are still used in the literature.

The literature is a teaching tool. Authors read published papers to learn which econometric methods are appropriate in which situations. Thus when readers see published papers using the Fama-MacBeth standard errors in the kinds of regressions I have listed, they believe (incorrectly) that this approach is correct. The problem is actually worse. The finance literature also contains (incorrect) advice that the Fama-MacBeth approach corrects the standard errors for the residual correlation in the presence of a firm effect (e.g. $\rho_X \neq 0$ and $\rho_\varepsilon \neq 0$). For example, Wu (2004) uses "...the Fama and MacBeth (1973) method to account for the lack of independence because of multiple yearly observations per company." Denis, Denis, and Yost (2002) argue that the "...pooling of cross-sectional and time-series data in our tests creates a lack of independence in the regression models. This results in the deflated standard errors and, therefore, inflated t-statistics. To address the importance of this bias we estimate the regression model separately for each of the 14 calendar years in our sample... The coefficients and statistical significance of the other control variables are similar to those in the pooled cross-sectional time series data."

In the presence of a firm effect, Fama-MacBeth and OLS standard errors are both biased, and as discussed above the estimates can be quite close to each other even when the bias is large (compare equations 7 and 12). The problem isn't with the Fama-MacBeth method, per se, only with its use. It was developed to account for correlation between observations on different firms in the same year, not to account for correlation between observations on the same firm in different years.

It is now being used and recommended in cases where it produces biased estimates and overstated significance levels. Given the Fama-MacBeth approach was designed to deal with time effects in a panel data set, not firm effects, I turn to this data structure in the next section.

E) Newey-West Standard Errors.

An alternative approach for addressing the correlation of errors across observation is the Newey-West procedure (Newey and West, 1987). This procedure was initially designed to account for serial correlation of unknown form in the residuals of a single time series. It has been modified for use in a panel data set by estimating only correlations between lagged residuals in the same cluster (see Bertrand, Duflo, and Mullainathan, 2004, Doidge, 2004, MacKay, 2003, Brockman and Chung, 2001). The problem of choosing a lag length is simplified in a panel data set, since the maximum lag length is one less than the maximum number of years per firm.[12] To examine the relative performance of the Newey-West, I simulated 5,000 data sets where the fixed firm effect is assumed to account for twenty-five percent of the variability of both the independent variable and the residual.

The standard error estimated by the Newey-West is an increasing function of the lag length in the simulation. When the lag length is set to zero, the estimated standard error is numerically identical to the White standard error, which is only robust to heteroscedasticity (White, 1984). This is the same as the OLS standard error in my simulation. Not surprisingly, this estimate significantly underestimates the true standard error (see Figure 4). As the lag length is increased from 0 to 9, the

[12] In the standard application of Newey-West, a lag length of M implies that the correlation between ε_t and ε_{t-k} are included for k running from -M to M. When Newey-West has been applied to panel data sets, correlations between lagged and leaded values are only included when they are drawn from the same cluster. Thus a cluster which contains T years of data per firm uses a maximum lag length of T-1 and would include t-1 lags and T-t leads for the t^{th} observation where t runs from 1 to T.

standard error estimated by the Newey-West rises from the OLS/White estimate of 0.0283 to 0.0328 when the lag length is 9 (see Figure 4). In the presence of a fixed firm effect, an observation of a given firm is correlated with all observations for the same firm no matter how far apart in time the observations are spaced. Thus having a lag length of less than the maximum (T-1), will cause the Newey-West standard errors to underestimate the true standard error when the firm effect is fixed. However, even with the maximum lag length of 9, the Newey-West estimates have a small bias – underestimating the true standard error by 8% [0.084 = 1-0.0328/0.0358].

As the simulation demonstrates, the Newey-West approach to estimating standard errors, as applied to panel data, does not yield the unbiased estimates produced by the clustered standard errors. The difference between the two estimates is due to the weighting function used by Newey West. When estimating the standard errors, Newey-West multiplies the covariance of lag j (e.g. $\varepsilon_t \varepsilon_{t-j}$) by the weight [1-j/(M+1)], where M is the specified maximum lag. If I set the maximum lag equal to T-1, then the central matrix in the variance equation of the Newey-West standard error is:

$$
\begin{aligned}
\sum_{i=1}^{N}\left(\sum_{t=1}^{T} X_{it}\ \varepsilon_{it}\right)^2 &= \sum_{i=1}^{N}\left(\sum_{t=1}^{T} X_{it}^2\,\varepsilon_{it}^2 + 2\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} w(t-s)\,X_{it}\,X_{is}\,\varepsilon_{it}\,\varepsilon_{is}\right) \\
&= \sum_{i=1}^{N}\left(\sum_{t=1}^{T} X_{it}^2\,\varepsilon_{it}^2 + 2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t} w(j)\,X_{it}\,X_{it-j}\,\varepsilon_{it}\,\varepsilon_{it-j}\right) \\
&= \sum_{i=1}^{N}\left(\sum_{t=1}^{T} X_{it}^2\,\varepsilon_{it}^2 + 2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}\left(1-\frac{j}{T}\right) X_{it}\,X_{it-j}\,\varepsilon_{it}\,\varepsilon_{it-j}\right)
\end{aligned}
\qquad \textbf{(14)}
$$

This is identical to the clustered standard error formula (see footnote 5) except for the weighting function [w(j)]. The clustered standard errors use a weighting function of one for all co-variances. The Newey-West procedure was originally designed for a single time-series and the weighting function was necessary to make the estimate of this matrix positive semi-definite. For fixed j the weight w(j) approaches 1 as the maximum lag length (M) grows. Newey and West show that if M is allowed to grow at the correct rate with the sample size (T), then their estimate is consistent. However, in the panel data setting, the number of time periods is often small. The consistency of the clustered standard error is based on the number of clusters (N) being large, opposed to the number of time periods (T). Thus the Newey-West weighting function is unnecessary and leads to standard error estimates which have a small bias in a panel data setting.[13]

III) Estimating Standard Errors in the Presence of a Time Effect.

To demonstrate how the techniques work in the presence of a time effect I generated data sets which contain only a time effect (observations on different firms within the same year are correlated). This is the data structure for which the Fama-MacBeth approach was designed (see Fama-MacBeth, 1973). If I assume that the panel data structure contains only a time effect, the equations I derived above are essentially unchanged. The expressions for the standard errors in the presence of only a time effect are correct once I exchange N and T.

A) Clustered Standard Error Estimates.

Simulating the data with only a time effect means the dependent variable will still be specified by equation (1), but now the error term and independent variable are specified as:

$$\begin{aligned} \varepsilon_{it} &= \delta_t + \eta_{it} \\ X_{it} &= \zeta_t + \nu_{it} \end{aligned} \qquad \textbf{(15)}$$

[13] Although the bootstrap method of estimating standard errors was rarely used in the articles which I surveyed, it is an alternative way to estimate the standard errors in a panel data set (see for example Kayhan and Titman, 2004 and Efron and Tibshirani, 1986). To test its relative performance, I drew 100 samples with replacement and re-estimated the regression for each simulated data set. When I drew observations independently (e.g. I drew 5,000 firm-year), the estimated standard errors are the same as the OLS standard errors reported in Table I (e.g. 0.0282 for the bootstrap versus 0.0283 for OLS when $\rho_X = \rho_\varepsilon = 0.50$). When I drew observations as a cluster (e.g. I drew 500 firms with replacement and took all 10 years for any firm which was drawn), the estimated standard errors are the same as the clustered standard errors (e.g. 0.0505 for bootstrap versus 0.0508 for clustered). The results were essentially the same when I drew 1,000 samples for each simulated data set.

As before, I simulated 5,000 data sets of 5,000 observations each. I allowed the fraction of variability in both the residual and the independent variable which is due to the time effect to range from zero to seventy-five percent in twenty-five percent increments. The OLS coefficient, the true standard error, as well as the OLS and clustered standard errors are reported in Table 3. There are several interesting findings to note. First, as with the firm effect results, the OLS standard errors are correct when there is no time effect in either the independent variable ($\sigma(\zeta)=0$) or the residual ($\sigma(\delta)=0$). As the time effect in the independent variable and the residual rise, so does the magnitude by which the OLS standard errors underestimate the true standard errors. When half of the variability in both comes from the time effect, the true standard error is eleven times the OLS estimate [11.0 = 0.3105/0.0282, see Table 3].

The clustered standard errors are much more accurate, but unlike our results with the firm effect, they underestimate the true standard error. The magnitude of the underestimate is small, ranging from 13 percent [1-0.1297/0.1490] when the time effect accounts for 25 percent of the variability to 19 percent [1-0.3986/0.4927] when the time effect accounts for 75 percent of the variability. The problem arises due to the limited number of clusters (e.g. years). When I estimated the standard errors in the presence of the firm effects, I had 500 firms (clusters). When I estimated the standard errors in the presence of a time effect, I have only 10 years (clusters). Since the clustered standard error places no restriction on the correlation structure of the residuals within a cluster, its consistency depends upon having a sufficient number of clusters. Based on these results, 10 clusters is too small and 500 is sufficient (see Kezdi, 2004, and Bertrand, Duflo, and Mullainathan, 2004, Hansen, 2005).

To explore this issue, I simulated data sets of 5,000 observations with the number of years (or clusters) ranging from 10 to 100. In all of the simulations, 25 percent of the variability in both the independent variable and the residual is due to the time effect [i.e. $\rho_X = \rho_\varepsilon = 0.25$]. The bias in the clustered standard error estimates declines with the number of clusters, dropping from 13 percent when there are 10 years (or clusters) to 4 percent when there are 40 years to under 1 percent when there are 100 years (see Figure 5). The bias in the clustered standard error estimates is a product of the small number of clusters. Since panel data sets of 10 or 20 years are not uncommon in finance, however, this may be a concern in practice.

B) Fama-MacBeth Standard Errors.

When there is only a time effect, the correlation of the estimated slope coefficients across years is zero and the standard errors estimated by Fama-MacBeth are unbiased (see equation 9 and 12). This is what I find in the simulation (see Table 4). The estimated standard errors are extremely close to the true standard errors and the confidence intervals are the correct size. In addition to producing unbiased standard error estimates, Fama-MacBeth also produces more efficient estimates than OLS. For example, when 25 percent of the variability of both the independent variable and the residual is due to the time effect, the standard deviation of the Fama-MacBeth coefficient estimate is 81 percent [1-0.0284/0.1490] smaller than the standard deviation of the OLS estimate (compare Table 3 and 4). The improvement in efficiency arises from the way in which Fama-MacBeth accounts for the time effect. By running cross sectional regressions for each year the intercept absorbs the time effect. Since the variability due to the time effect is no longer in the residual, the residual variability is significantly smaller. The lower residual variance leads to less variable coefficient estimates and greater efficiency. I will revisit this issue in the next section when I consider the presence of both a firm and a time effect.

IV) Estimating Standard Errors in the Presence of a Fixed Firm Effect and Time Effect.

According to the simulation results thus far, the best method for estimating the coefficient and standard errors in a panel data set depends upon the source of the dependence in the data. If the panel data only contains a firm effect, the standard errors clustered by firm are superior as they produce unbiased standard errors. If the data has only a time effect, the Fama-MacBeth estimates perform better than standard errors clustered by time when there are few years (clusters) and equally well when the number of years (clusters) is sufficiently large.

Although the above results are instructive, they are unlikely to be completely descriptive of actual data confronted by empirical financial researchers. Panel data sets may include both a firm effect and a time effect. Thus to provide guidance on which method to use I need to assess their relative performance when both effects are present. In this section, I will simulate a data set where both the independent variable and the residual have both a firm and a time effect.

A conceptual problem with using these techniques (Clustered or Fama-MacBeth standard errors) is neither is designed to deal with correlation in two dimensions (e.g. across firms and across time).[14] The clustered standard error approach allows us to be agnostic about the form of the correlation within a cluster. However, the cost of this is the residuals must be uncorrelated across clusters. Thus if we cluster by firm, we must assume there is no correlation between residuals of different firms in the same year. In practice, empirical researchers account for one dimension of the cross observation correlation by including dummy variables and account for the other dimension by

[14] It is possible to estimate standard errors accounting for clustering in multiple dimensions, but only if there are a sufficient number of observations within each cluster. For example, if a researcher has observations on firms in industries across multiple years, she could cluster by industry and year (i.e. each cluster would be a specific year and industry, see Lipson and Mortal, 2004). In this case, since there are multiple firms in a given industry in each year, clustering would be possible. If clustering was done by firm and year, since there is only one observation within each cluster, this is numerically identical to White standard errors.

clustering on that dimension. Since most panel data sets have more firms than years, the most common approach is to include dummy variables for each year (to absorb the time effect) and then cluster by firm (Anderson and Reeb, 2004, Gross and Souleles, 2004, Petersen and Faulkender, 2005, Sapienza, 2004, and Lamont and Polk, 2001). I use this approach in my simulations.

A) Clustered Standard Error Estimates.

To test the relative performance of the two methods, 5,000 data sets were simulated with both a firm and a time effect. Across the simulations, the fixed firm effect accounts for either 25 or 50 percent of the variability. The fraction of the variability due to the time effect is also assumed to be 25 or 50 percent of the total variability. This gives us three possible scenarios for the independent variable [(25,25),(25,50), and (50,25)]. The same three scenarios were used for simulating the residual, for a total of nine simulations (see Table 5).

The results in the presence of both a firm and time effect (Table 5) are qualitatively similar to what we found in the presence of only a fixed firm effect (Table 1). The OLS standard errors underestimate the true standard errors whereas the standard errors clustered by firm are unbiased independent of how I specify the firm and time effects. As we saw above, the bias in the OLS standard errors increases as the firm effect becomes larger.

B) Fama-MacBeth Estimates

The statistical properties of the OLS and Fama-MacBeth coefficient estimates are quite similar. The means and the standard deviations of the estimates are almost identical (see Table 5 and Table 6), and the correlation between the two estimates is never less than 0.999 in any of the simulations. Once I include a set of time dummies in the OLS regression, the difference in efficiency I found in Tables 3 and 4 disappears. The OLS estimates are now as efficient as the Fama-MacBeth, even in the presence of a time effect. The standard errors estimated by Fama-MacBeth, however, are once again too small, just as I found in the absence of a firm effect (Table 2). When 50 percent of the variability is from the firm effect and 25 percent comes from a time effect, the true standard error is three times the Fama-MacBeth estimate [3.1 = 0.0632/0.0206].

Most of the intuition from the earlier tables carries over. In the presence of a fixed firm effect both OLS and Fama-MacBeth standard error estimates are biased down significantly. Clustered standard errors which account for clustering by firm produce estimates which are unbiased. The presence of a fixed time effect, if it is controlled for with dummy variables, does not alter these results, except for accentuating the magnitude of the firm effect and thus making the bias in the OLS and Fama-MacBeth standard errors larger.

V) Estimating Standard Errors in the Presence of a Temporary Firm Effect

The analysis thus far has assumed that the firm effect is fixed. Although this is common in the literature, it may not always be true. The dependence between residuals may decay as the time between them increases (i.e. $\rho(\varepsilon_t , \varepsilon_{t-k})$ may decline with k). In a panel with a short time series, distinguishing between a permanent and a temporary firm effect may be impossible. However, as the number of years in the panel increases it may be feasible to empirically identify the permanence of the firm effect. In addition, if the performance of the different standard error estimators depends upon the permanence of the firm effect, researchers need to know this.

A) Temporary Firm Effects: Specifying the Data Structure.

To explore the performance of the different standard error estimates in a more general context, I simulated a data structure which includes both a permanent component (a fixed firm effect) and a temporary component (non-fixed firm effect) which I assume is a first order auto regressive process. This allows the firm effect to die away at a rate between a first order auto-regressive decay and zero. To construct the data, I assumed that non-firm effect portion of the residual (η_{it} from equation 4) is specified as

$$\begin{aligned} \eta_{it} &= \varsigma_{it} && \text{if } t = 1 \\ &= \varphi\, \eta_{it-1} + \sqrt{1 - \varphi^2}\, \varsigma_{it} && \text{if } t > 1 \end{aligned} \qquad \textbf{(16)}$$

Thus φ is the first order auto correlation between η_{it} and η_{it-1}, and the correlation between η_{it} and $\eta_{i,t-k}$ is φ^k.[15] Combining this term with the fixed firm effect (γ_i in equation 4), means the serial correlation of the residuals dies off over time, but more slowly than implied by a first order auto-regressive and asymptotes to ρ_ε (from equation 6). By choosing the relative magnitude of the fixed firm effect (ρ_ε) and the first order auto correlation (φ), I can alter the pattern of auto correlations in the residual. The correlation of lag length k is:

$$\begin{aligned} \text{Corr}(\varepsilon_{i,t}, \varepsilon_{i,t-k}) &= \frac{\text{Cov}(\gamma_i + \eta_{i,t}, \gamma_i + \eta_{i,t-k})}{\sqrt{\text{Var}(\gamma_i + \eta_{i,t})\,\text{Var}(\gamma_i + \eta_{i,t-k})}} \\ &= \frac{\sigma_\gamma^2 + \varphi^k \sigma_\eta^2}{\sigma_\gamma^2 + \sigma_\eta^2} \\ &= \rho_\varepsilon + (1 - \rho_\varepsilon)\,\varphi^k \end{aligned} \qquad \textbf{(17)}$$

An analogous data structure is specified for the independent variable. The correlation for lags one

[15] I multiply the ς term by $\sqrt{1 - \varphi^2}$ to make the residuals homoscedastic. From equation (16),

$$\begin{aligned} \text{Var}(\eta_{it}) &= \sigma_\varsigma^2 && \text{if } t = 1 \\ &= \varphi^2\, \sigma_\varsigma^2 + (1 - \varphi^2)\, \sigma_\varsigma^2 = \sigma_\varsigma^2 && \text{if } t > 1 \end{aligned}$$

where the last step is by recursion (if it is true for t=m, it is true for t=m+1). Assuming homoscedastic residuals is not necessary since the Fama-MacBeth and clustered standard errors are robust to heteroscedasticity (Jagannathan and Wang, 1998, & Rogers, 1993). However, assuming homoscedasticity makes the interpretation of the results simpler. If I assume the residuals are homoscedastic, then any difference in the standard errors I find is due to the dependence of observations within a cluster not heteroscedasticity.

through nine for the four data specifications I examine are graphed in Figure 6. They range from a fixed firm effect (ρ=0.25 and φ=0.00) to a standard AR1 process (ρ=0.00 and φ=0.75). I assume the same process for both the independent variable and the residual, since it is the interaction of the two correlations which cause the bias in the standard error estimates (see Section II).

B) Fixed Effects – Firm Dummies.

A significant minority of the papers in the survey, used fixed effects to control for dependence within a cluster. Using the simulations, I can compare the relative performance of OLS and clustered standard errors both with and without firm dummies. The results are reported in Table 7, Panel A. The fixed effect estimates are more efficient in this case (e.g. 0.0299 versus 0.0355), although this is not always true. The relative efficiency of the fixed effect estimates depends upon two offsetting effects. Including the firm dummies uses up N–1 additional degrees of freedom and this raises the standard deviation of the estimate. However, the firm dummies also eliminate the within cluster dependence of the residuals which reduces the standard deviation of the estimate. In this example, the second effect dominates and thus the fixed effect estimates are more efficient.

Once we include the firm effects, the OLS standard error are unbiased (see Table 7 - Panel A, column I).[16] The clustered standard errors are unbiased when we do not include the fixed effects and are slightly too large (5%) when we include the fixed effects (see Kezdi, 2004, for similar results). This conclusion, however, is sensitive to the firm effect being fixed. If the firm effect decays over time, the firm dummies no longer completely capture the within cluster dependence and

[16] I assumed the model is correctly specified [i.e. Corr($X_{i\,t}$, $\varepsilon_{i\,t}$) = 0]. In this case, the only purpose of including firm dummies is to correct the standard errors. In real applications, the model may not be correctly specified [i.e. Corr($X_{i\,t}$, $\varepsilon_{i\,t}$) ≠ 0], and so including fixed effects may be necessary to test the model's specification (see Hausman, 1978). Instead of including firm dummies, we could have first differenced the data within firm. However, it would still be necessary to use clustered opposed to OLS standard errors, since the residuals would be correlated.

OLS standard errors are still biased (see Table 7 - Panel A, columns II-IV). In these simulations, the firm effect decays over time (in column II, 92 percent of the firm effect dissipates after 9 years). Once the firm effect is temporary, the OLS standard errors again underestimate the true standard errors even when firm dummies are included in the regression (Wooldridge, 2004, Baker, Stein, and Wurgler, 2003). The magnitude of the underestimation depends upon the magnitude of the temporary component of the firm effect (i.e. φ). The bias rises from about 17% when φ is 50 percent (column IV) to about 33 percent when φ is 75% (columns II and III).

C) Adjusted Fama-MacBeth Standard Errors.

As I showed in Section II, the presence of a firm effect causes the Fama-MacBeth standard error to be biased downward. Several authors have acknowledged the bias and have suggested adjusting the standard errors for the estimated first order auto-correlation of the estimated slope coefficients (Pontiff, 1996; Graham, Lemmon, and Schallheim, 1998; Christopherson, Ferson, and Glassman, 1998; Chen, Hong, and Stein, 2001; Cochrane, 2001; Lakonishok, and Lee, 2001; Fama and French, 2002; Kemsley and Nissim, 2002; Bakshi, Kapadia, and Madan, 2003; Pastor and Veronesi, 2003; Chakravarty, Gulen, and Mayhew, 2004). The proposed adjustment is to estimate the correlation between the yearly coefficient estimates (i.e. $Corr[\beta_{t,} \beta_{t-1}] = \theta$), and then multiply the estimated variance by $(1+\theta)/(1-\theta)$ to account for serial correlation of the βs (see Chakravarty, Gulen, Mayhew, 2004 and Fama and French, 2002, footnote 1). This would seem to make intuitive sense since the presence of a firm effect causes the yearly coefficient estimates to be serially correlated.

To test the merits of this idea, I simulated data sets where the fixed firm effect accounts for 25 percent of the variance. For each simulated data set, ten slope coefficients are estimated, and the auto-correlation of the slope coefficients is calculated. I then calculate the original and adjusted

Fama-MacBeth standard errors, assuming both an infinite and a finite lag of T-1 periods (see Lakonishok and Lee, 2001).[17] The auto-correlation is estimated imprecisely as noted by Fama and French (2002). The 90th percentile confidence interval ranges from -0.60 to 0.41, but the mean is -0.1134 (see Table 7 - Panel B). Since the average first-order auto-correlation is negative, the adjusted Fama-MacBeth standard errors are even more biased than the unadjusted standard errors.

The problem is the correlation being estimated (the within sample auto-correlation of betas) is not the same as the one which is causing the bias in the standard errors (the population auto-correlation of betas). The co-variance which biases the standard errors and which I can estimate across the 5,000 simulations is:

$$Cov(\beta_t, \beta_{t-1}) = E[(\beta_t - \beta_{True})(\beta_{t-1} - \beta_{True})] \qquad \textbf{(18)}$$

To see how the presence of a fixed firm effect influences this covariance, consider the case where the realization for firm i is a positive value of $\mu_i \gamma_i$ (i.e. the realized firm effect in the independent variable and the residual). This positive realization will result in an above average estimate of the slope coefficient in year t, and because the firm effect is fixed it will also result in an above average estimate of the slope coefficient in year t-1 (see equation 8). The realized value of the firm effect (μ_i and γ_i) in a given simulation does not change the average β across samples. The average β across

[17] Thus, instead of multiplying the variance by the infinite period adjustment [$(1+\theta)/(1-\theta)$], I multiplied it by the 10 period adjustment

$$\text{Variance correction} = \left(1 + 2\sum_{k=1}^{10-1} (10 - k)\ \theta^k\right)$$

A third alternative is to estimate a k order auto-regressive model on the yearly βs, and then use the intercept and its standard error as an estimate of β and its standard error (see Pontiff, 1996). This bias in this method is similar to those reported in Table 7 (results available from the author).

data samples is the true β (one in the simulations). Thus when I estimate the true correlation between β_t and β_{t-1}, the firm effect causes this correlation to be positive and the Fama-MacBeth standard errors to be biased downward.[18]

Researchers, however, are given only one data set. They must calculate the serial correlation of the βs within the sample they are given. This co-variance is calculated as:

$$\text{Cov}(\beta_t, \beta_{t-1}) = E[(\beta_t - \overline{\beta}_{\text{Within sample}})(\beta_{t-1} - \overline{\beta}_{\text{Within sample}})] \quad \textbf{(19)}$$

The within sample serial correlation measures the tendency of β_t to be above the within sample mean when β_{t-1} is above the within sample mean. To see how the presence of a fixed firm effect influences this covariance, consider the same case as above. A positive realization of $\mu_i\gamma_i$ will raise the estimate of β_1 through β_T, as well as the average of the βs by the same amount. Thus a fixed firm effect will not influence the deviation of any β_t from the sample average β. The estimated serial correlation is asymptotically zero when the firm effect is constant and adjusting the standard errors based on this estimated serial correlation will still lead to biased standard error estimates.[19]

The adjusted Fama-MacBeth standard errors do better when there is an auto-regressive component in the residuals (i.e. $\varphi > 0$). In the three remaining simulations (Table 7 – Panel B), the estimated within sample auto correlation is positive in all cases, but the adjusted Fama-MacBeth standard errors are still biased downward. Adjusting the standard error estimates moves them closer

[18] In the simulation the correlation between β_t and β_s ranged from 0.0430 to 0.0916 and did not decline as the difference between t and s increased (the firm effect is fixed). The theoretical value of the correlation between β_t and β_s should be 0.0625 (according to equation 11) and would imply a true standard error of the Fama-MacBeth estimate of 0.0354 (according to equation 12). This matches the number I report in Table II.

[19] The average within sample serial correlation I estimate is actually less than zero, but this is due to a small sample bias. With only ten years of data per firm, I have only nine observations to estimate the serial correlation. To verify that this is correct, I re-ran the simulation using 20 years of data per firm and the average estimated serial correlation moved closer to zero, rising from -0.1134 to -0.0556.

to the truth when the firm effect is not fixed (ρ=0). The standard errors based on the infinite period adjustment underestimate the true standard error by 23 percent (1-0.0374/0.0484, see Table 7 - Panel B, column II). As the magnitude of the firm effect increases (compare columns II to III and IV), the bias in the estimated standard errors increases. Thus the Fama-MacBeth standard errors adjusted for serial correlation do better than the unadjusted standard errors when the firm effect decays over time, but they still significantly underestimate the true standard errors.

VI) Empirical Applications.

The analysis in the previous sections has used simulated data. I had the advantage of knowing the data structure, which made choosing the method for estimating standard errors much easier. In real world applications, we may have priors about the data's structure (are firm effects or time effects more important and are they permanent or temporary), but we do not know the data structure for certain. Thus in this section, I will apply the different techniques for estimating standard errors to two real data sets. This way I can demonstrate how the different methods for estimating standard errors compare, confirm that the methods used by some published papers have produced biased results, and show what we can learn from the different standard errors estimates.

For both data sets, I first estimate the regression using OLS, and report White standard errors as well as standard errors clustered by firm and year (Tables 8 and 9, columns I-III). By using White standard errors as my baseline, difference across columns are attributable only to within cluster correlations, not to heteroscedasticity. If the standard errors clustered by firm are dramatically different than the White standard errors, then we know there is a significant firm effect in the data [e.g. Corr($X_{i\ t}\,\varepsilon_{i\ t}$, $X_{i\ t\text{-}k}\,\varepsilon_{i\ t\text{-}k}$) $\neq 0$]. I then estimate the slope coefficients and the standard errors using Fama-MacBeth (Tables 8 and 9, columns IV-V). Each of the OLS regressions include time dummies. This makes the efficiency of the OLS and Fama-MacBeth coefficients similar.[20]

A) Asset Pricing Application.

For the asset pricing example, I used the equity return regressions from Daniel and Titman (2004, "Market Reactions to Tangible and Intangible Information"). To demonstrate the effect of equity issues on future equity returns, they regress monthly returns on annual values of lagged book to market ratios, historic changes in book and market values, and a measure of the firm's equity issuance. The data is briefly described in the appendix and in detail in their paper. Each observation of the dependent variable is a monthly equity return. However, the independent variables are annual values (based on the prior year). Thus for the twelve observations in a year, the dependent variable (equity returns) changes each month, but the independent variable (e.g. past book value) does not, and is therefore highly persistent.

The White standard errors are essentially the same as standard errors clustered by firm (ranging from three percent larger to one percent smaller). This occurs because the auto-correlation in the residuals is effectively zero (see Figure 7). The auto-correlation in the independent variable is large and persistent, starting at 0.98 the first month and declining to 0.49 to 0.75 by the 24th month depending upon the independent variable. However, since the adjustment in the standard error, and the bias in White standard errors, is a function of the monthly auto-correlation in the Xs (a large number) times the auto-correlation in the residuals (zero), the standard errors clustered by firm are equal to the White standard errors.

---

[20] The reported R^2s do not include the explanatory power attributable to the time dummies. This is done to make the R^2 comparable between the OLS and the Fama-MacBeth results. Although the Fama-MacBeth procedure estimates a separate intercept for each year, the constant is calculated as the average of the yearly intercepts. Thus the Fama-MacBeth R^2 does not include the explanatory power of time dummies. Procedurally, I subtracted the yearly means off of each variable before running the OLS regressions.

The story is very different when I clustered by time (months). The standard errors clustered by month are two to four times larger than the White standard errors. For example, the t-statistic on the lagged book to market ratio is 7.2 if we use the White standard error and 1.9 if we cluster by month. This means there is a significant time effect in the data (see Figure 8), even after including time dummies. Thus any constant time effect (i.e. one which raises the monthly return for every firm in a given month by the same amount) has already been removed from the data and will not affect the standard errors. The remaining correlation in the time dimension must vary across observations (i.e. Corr[$\varepsilon_{it}\, \varepsilon_{jt}$] varies across i and j).

Understanding a temporary firm effect is straightforward. A firm effect is temporary (dies off over time) if the 1980 residual for firm A is more highly correlated to the 1981 residual for firm A than to the 1990 residual. This is how I simulated the data in Section V. Understanding a non-constant time effect is more difficult. For the time effect to be non-constant, it must be that a shock in 1980 has a large effect on firm A and B, but has a significantly smaller affect on firm Z. If the time effect influenced each firm in a given month by the same amount, the time dummies would absorb the effect and clustering by time would not change the reported standard errors. The fact that clustering by time does change the standard errors, means there must be a non-constant time effect.

If we know the data, we can use our economic intuition to determine how the data should be organized and predict the source of the dependence within a cluster. For example, since this data contains monthly equity returns we might consider how a shock to returns would effect firms differently. If the economy booms in a given month, firms in the durable goods industry may rise more than firms in the non-durable goods industry. This can create a situation where the residuals of firms in the same industry are correlated (within the month) with each other but less correlated with firms in another industry. When I sort the data by month, four-digit industry and then firm, I see evidence of this in the auto-correlation for the residuals and the independent variables within each month (see Figure 8). The auto-correlations of the residual is much larger than when I sorted by firm then month (compare Figure 7 and 8) and they die away as we consider firms in more distant industries.[21]

When calculating the standard errors clustered by time, we don't need to make an assumption about how to sort the data to obtain unbiased standard errors. However, if researchers are going to understand what the standard errors are telling them about the structure of the data, they need to consider the source of the dependence in the residuals. By examining how standard errors change when we cluster by firm or time (i.e. compare columns I to II and I to III), we can determine the nature of the dependence which remains in the residuals and this can guide us on how to improve our models.

According to results in Sections II and III, the Fama-MacBeth standard errors perform better in the presence of a time effect than a firm effect, and so given the above results should do well in this data set. The Fama-MacBeth coefficients and standard errors are reported in column IV (they are a replica of those reported by Daniel and Titman, Table 3, row 8). The coefficient estimates are similar to the OLS coefficients, and the standard errors are much larger than the White standard errors (2.0 to 3.4 times) as we would expect in the presence of a time effect. The Fama-MacBeth

[21] A non-constant time effect can be generated by a random coefficient model (Greene, 2004). For example, if the firm's return depends upon the firm's β times the market return, but only the market return or time dummies are included in the regression, then the residual will contain the term { [β_i -Average(β_i)] Market return$_t$ }. Firms which have similar βs will have highly correlated residuals within a month, and firms which have very different βs will have residuals whose correlation is small. This is a non-constant time effect. This logic suggests that I should instead sort by month, β, and then firm. When I sort this way, the auto-correlations are smaller and die away more slowly (declining from 0.030 at a lag of one to 0.028 at a lag of 24) than when I sorted by month, four-digit industry and firm (declining from 0.096 at a lag of one to 0.042 at a lag of 24).

standard errors are close to the standard errors when we cluster by time, as both methods are designed to account for dependence in the time dimension. The Fama-MacBeth standard errors are consistently smaller than the clustered standard errors, but the magnitude of the difference is small (twelve to eighteen percent, compare columns III and IV of Table 8).

Cross-sectional, time-series regressions on panel data sets treat each observation equally. In the Fama-MacBeth procedure, the monthly coefficient estimates are averaged using equal weights for each month, not each observation. In an unbalanced panel, Fama-MacBeth effectively weights each observation proportional to $1/N_t$ where N_t is the number of firms in month t. Since the number of firms in the sample grows from about 1,000 per month to almost 2,300 near the end of the sample, Fama-MacBeth will effectively under weight the later observations. To correct this, I took a weighted average of the monthly coefficient estimates where the weights were proportional to N_t. When I compare the weighted and un-weighted coefficient estimates and standard errors, they are very close (compare columns IV and V of Table 8). The weights are uncorrelated with the variables and thus do not effect the results, in this case. More sophisticated weighting schemes should be able to improve the efficiency of Fama-MacBeth estimates. However such estimators can be unstable in small samples (see Skoulakis, 2005).

B) Corporate Finance Application.

For the corporate finance illustration, I used a capital structure regression since this is an example of where the literature uses Fama-MacBeth standard errors. The independent variables are those which are common from the literature (firm size, firm age, asset tangibility, and firm profitability). The sample contains NYSE firms which pay a dividend in the previous year for the years 1965-2003. I lagged the independent variables one year relative to the dependent variable. The results are reported in Table 9.

The relative importance of the firm effect and the time effect can be seen by comparing the standard errors across the first three columns. The standard errors clustered by firm are dramatically larger than the White standard errors (3.1 to 3.5 times larger, see columns I and II). For example, the t-statistic on the profit margin variable is -3.1 when I use the White standard errors and -0.9 when I use the standard errors clustered by firm. It is not surprising that the profit margin is highly persistent, and thus the auto-correlation for the profit margin is extremely high and persistent (see Figure 9). However, the auto-correlation in the residuals is also high and even after 10 years it remains above 40 percent.

The importance of the time effect (after including time dummies) is generally much smaller in this data set than in the previous one. The standard errors clustered by time are larger than the White standard errors but the magnitude of the difference is not large (except for the market to book ratio). This is due to a smaller auto-correlation in the residuals (see Figure 10) and the independent variables. When I sorted by year, industry, and then firm, the residual first-order auto-correlation is less than 12 percent.

The Fama-MacBeth standard errors (Table 9, column IV) are close to the standard errors clustered by year and the White standard errors. For example, the Fama-MacBeth t-statistic on the profit margin is -3.1, the same as the White t-statistic. The results are similar for firm size, firm age, asset tangibility (the ratio of property, plant, and equipment to assets), and R&D expenditure. The White and Fama-MacBeth t-statistics are significantly larger than the clustered t-statistics (when clustering by firm). This was the conclusion of Section II. In the presence of a firm effect, as in a capital structure regression, White and Fama-MacBeth standard errors are significantly biased.

VII) Conclusions.

It is well known from first-year econometrics classes that OLS and White standard errors are biased when the residuals are not independent. What has been less clear is how financial researchers should estimate standard errors when using panel data sets. The empirical finance literature has proposed and used a variety of methods for estimating standard errors when the residuals are correlated across firms or years. In this paper, I show that the performance of the different methods varies considerably and their relative accuracy depends upon the structure of the data. Simply put, estimates which are robust to the form of dependence in the data produce unbiased standard errors and correct confidence intervals; estimates which are not robust to the form of dependence in the data produce biased standard errors and often produce confidence intervals which are too small. The two illustrations in Section VI, demonstrate that the magnitude of the biases can be quite large.

Although it may seem obvious that choosing the correct method is important, the absence of good advice in the literature means the correct decision has not always been made. The purpose of this paper is to provide such guidance. In the presence of a firm effect [e.g. $Cov(X_{i,t}\,\varepsilon_{i,t}, X_{i,t-k}\,\varepsilon_{i,t-k}) \neq 0$], standard errors are biased when estimated by OLS, Newey-West (modified for panel data sets), Fama-MacBeth, or Fama-MacBeth corrected for first-order auto-correlation.[22] Despite this, these methods are often used in the literature when the regressions being estimated contain a firm effect. The standard errors clustered by firm are unbiased and produce correctly sized confidence intervals whether the firm effect is permanent or temporary. The fixed effect model also produces unbiased standard errors but only when the firm effect is permanent (and the model is correctly

[22] Skoulakis (2005) proposes applying the logic of Fama-MacBeth to each firm, instead of each year. He demonstrates that running N time series regressions and using the standard deviation of the N coefficients produces an estimate which is correct in the presence of a firm effect. I found only one paper in the literature which has used the Fama-MacBeth approach in this way (Coval and Shumway, 2005).

specified).

In the presence of a time effect [e.g. $\text{Cov}(X_{i,t}\,\varepsilon_{i,t}, X_{i,t-k}\,\varepsilon_{i,t-k}) \neq 0$], Fama-MacBeth produces unbiased standard errors and correctly sized confidence intervals. This is not surprising since it was designed for just such a setting. Standard errors clustered by time also produce unbiased standard errors and correctly sized confidence intervals, when there are a sufficient number of clusters. When there are too few clusters, clustered standard errors are biased even when clustered on the correct dimension (see Figure 5).

None of the methods for estimating standard errors which I discussed are designed to deal with a firm and a time effect simultaneously. When both a firm and a time effect are present in the data, researchers need to address one parametrically (e.g. by including time dummies) and then estimate standard errors clustered on the other dimension. This raises the question of how researchers can determine what form of dependence may exist in the data and is the second purpose of the two applications I examined in Section VI. By comparing the different standard errors, one can quickly observe the presence and general magnitude of a firm or a time effect. As we saw in Section VI, when the standard errors clustered by firm are much larger than the White standard errors (three to four times larger), this indicates the presence of a firm effect in the data (Table 9). When the standard errors clustered by time are much larger than the White standard errors, this indicates the presence of a time effect in the data (Table 8). Which dependence is most important will vary across data sets, and thus the researcher must consult their data. This information can provide researchers with the knowledge of which method for estimating standard errors is appropriate, intuition as to the deficiency of their models, and guidance for improving their models.

Appendix I: Data Set Constructions

A) Asset Pricing Application.

The data for the regressions in Table 8 are taken from Daniel and Titman's paper "Market Reactions to Tangible and Intangible Information" (2004). A more detailed description of the data can be found in their paper. The dependent variable is monthly returns on individual stocks from July, 1968 to December, 2001. The independent variables are:

Log(Lagged book to market) is the log of the total book value of the equity at the end of the firm's fiscal year ending anywhere in year t-6 divided by the total market equity on the last trading day of calendar year t-6.

Log(Book return) measures changes in the book value of the firm's equity over the previous five years. It is calculated as the log of one plus the percentage change in book value over the past five years. Thus if you purchased one percent of book value five years ago, and neither invested additional cash or nor took any cash out of the investment, book return is the current percentage ownership divided by the initial one percent.

Log(Market return) measures changes in the market value of the firm over the previous five years. It is calculated as the log of one plus the market return from the last day of year t-6 to the last day of year t-1.

Share issuance is a measure of net equity issuance. It is calculated as minus the log of the percentage ownership at the end of five years, assuming the investor started with 1 percent of the firm. Thus if investor purchases 1 percent of the firm and five years later they own 0.5 percent of the firm, then share issuance is equal to -log(0.5/1.0) = 0.693. Investors are assumed to neither take money out of their investment nor add additional money to their investment. Thus any cash flow which investors receive (e.g. dividends) would be reinvested. For transactions such as equity issues and repurchases, the investor is assumed not to participate and thus these will lower or raise the investor's fractional ownership.

To make sure that the accounting information is available to implement a trading strategy, the independent variables are lagged at least six months. Thus the independent variables for a fiscal year ending anytime during calendar year t-1, are used to predict future monthly returns from July of year t through June of year t+1. The independent variables are annual measures and are thus constant for each of the twelve monthly observations during the following year (July through June).

B) Corporate Finance Application

The data for the regressions in Table 9 are constructed from Compustat and include data from 1965 to 2003 (the dependent variable). The dependent variable, the market debt ratio, is defined as the book value of debt (data9 + data34) divided by the sum of the book value of assets (data6) minus the book value of equity (data60) plus the market value of equity (data25 * data199). The independent variables are lagged one year and I only include observations where the firm paid a dividend (data21) in the previous year (Fama and French, 2002). To reduce the influence of outliers, I capped ratio variables (e.g. profits to sales, tangible assets, advertising to sales, and R&D to sales) at the one and 99th percentile (Petersen and Faulkender, 2005, and Richardson and Sloan, 2003). The independent variables are:

Ln(Market Value of Assets) is the log of the sum of the book value of assets (data6) minus the book value of equity (data60) plus the market value of equity (data25 * data199).

Ln(1 + Firm Age). Firm age is calculated as the number of years the firm's stock has been listed. Firm age is calculated as the current year (fyenddt) minus the year the stock began trading (linkdt).

Profits / Sales is defined as operating profits before depreciation (data13) divided by sales revenue (data12).

Tangible assets is defined as property, plant, and equipment (data8) divide by the book value of total assets (data6).

Advertising / Sales is defined as advertising expense (data45) divided by sales (data12).

R&D / Sales is defined as R&D expenditure (data46) divided by sales (data12). If R&D is missing, it is coded as zero.

R&D > 0 is a dummy variable equal to one if R&D expenditure is positive, and zero otherwise.

References

Anderson, Ronald, and David Reeb, 2004, "Founding-Family Ownership and Firm Performance: Evidence from the S&P 500" *Journal of Finance* 58, 1301-1328.

Andrews, D. W. K, 1991, “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Econometrica 59, 817-858.

Arellano, Manuel, 1987, “Computing Robust Standard Errors for Within-Groups Estimators,”*Oxford Bulletin of Economics and Statistic* 49, 431-433.

Baker, Malcolm, Jeremy Stein, and Jeffrey Wurgler, 2003, “When Does the Market Matter? Stock Prices and the Investment of Equity-Dependent Firms” Quarterly Journal of Economics 118, 969-1005.

Baker, Malcolm and Jeffrey Wurgler, 2002, “Market Timing and Capital Structure” *Journal of Finance* 57, 1-32.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, “Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options” *Review of Financial Studies* 16, 101-143.

Bertrand, Marianne; Duflo, Esther and Mullainathan, Sendhil. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 2004, *119*(1), pp. 249-75.

Brockman, Paul, and Dennis Chung, 2001, “Managerial timing and corporate liquidity: evidence from actual share repurchases” Journal of Financial Economics 61, 417-448.

Chakravarty, Sugato, Huseyin Gulen, and Stewart Mayhew, 2004, “Informed Trading in Stock and Option Markets” Journal of Finance 59, 1235-1257.

Chen, Joseph, Harrison Hong, and Jeremy Stein, 2001, “Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices” *Journal of Financial Economics* 61, 345-381

Christopherson, Jon, Wayne Ferson, and Debra Glassman, 1998, "Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance” *Review of Financial Studies* 11, 111-142.

Cochrane, John, 2001, *Asset Pricing*, Princeton University Press, Princeton, NJ.

Coval, Joshua, and Tyler Shumway, 2005, “Do Behavioral Biases Affect Prices” *Journal of Finance* 60, 1-34.

Daniel, Kent and Sheridan Titman, 2004, “Market Reactions to Tangible and Intangible Information” Northwestern University working paper.

Denis, David, Diane Denis, and Keven Yost, 2002, "Global Diversification, Industrial Diversification, and Firm Value" *Journal of Finance* 57, 1951-1979.

Doidge, Craig, 2004, "U.S. cross-listings and the private benefits of control: evidence from dual-class firms" *Journal of Financial Economics* 72, 519-553 .

Donald, Stephen and Kevin Lang, 2001, "Inference with Difference in Differences and Other Panel Data." University of Texas working paper.

Efron, B., Tibshirani, R. J., 1986. Bootstrap measures for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1, 54-77.

Fama, Eugene, and J. MacBeth, 1973, "Risk, return and equilibrium: Empirical tests," *Journal of Political Economy* 81, 607-636.

Fama, Eugene, and Ken French, 2002, "Testing Tradeoff and Pecking Order Predictions about Dividends and Debt" *Review of Financial Studies* 15, 1-33.

Fama, Gene and Kenneth French, 2001, "Disappearing dividends: changing firm characteristics or lower propensity to pay?" *Journal of Financial Economics* 60, 3-43.

Faulkender, Mike, and Mitchell Petersen, 2005, "Does the Source of Capital Affect Capital Structure?" *forthcoming Review of Financial Studies*.

Graham, John, Michael Lemmon, and James Schallheim, 1998, "Debt, Leases, Taxes and the Endogeneity of Corporate Tax Status," *Journal of Finance* 53, 131-162.

Green, William, 1990, *Econometric Analysis*, Macmillian Publishing Company, New York.

Gross, David and Nicholas Souleles, 2004, "An Empirical Analysis of Personal Bankruptcy and Delinquency" *Review of Financial Studies* 15, 319-347.

Hansen, Christina, 2005, "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data When T Is Large" University of Chicago working paper.

Hausman, Jerry, 1978, "Specification Tests in Econometrics, *Econometrica* 46, 1251-1271.

Jagannathan, Ravi, and Zhenyu Wang, 1998, "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression" *Journal of Finance* 53, 1285-1309.

Johnson, Shane, 2003, "Debt Maturity and the Effects of Growth Opportunites and Liquidity Risk on Leverage" *Review of Financial Studies* 16, 209-236.

Kayhan, Ayla and Sheridan Titman, 2004, "Firms' Histories and Their Capital Structures" University of Texas at Austin working paper.

Kemsley, Deen and Doron Nissim, 2002, "Valuation of the Debt Tax Shield" *Journal of Finance* 57, 2045-2073.

Kezdi, Gabor, 2004, "Robust Standard Error Estimation in Fixed-Effects Panel Models", *forthcoming. Hungarian Statistical Review*.

Lakonishok, Josef, and Inmoo Lee, 2001, "Are insider trades informative?" *Review of Financial Studies* 14, 79-111.

Lamont, Owen and Chris Polk, 2001, "Does diversification destroy value? Evidence from the industry shocks" *Journal of Financial Economics* 63, 51-77.

Lipson, Marc and Sandra Mortal, 2004, "Capital Structure Decisions and Equity Market Liquidity" University of Georgia working paper.

MacKay, Peter, 2003, "Real Flexibility and Financial Structure: An Empirical Analysis" *Review of Financial Studies* 16, 1131-1165.

Moulton, Brent, 1986, "Random Group Effects and the Precision of Regression Estimates" *Journal of Econometrics* 32, 385-397.

Moulton, Brent, 1990, "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units" *Review of Economics and Statistics* 72, 334-338.

Newey, Whitney, and Kenneth West, 1987, "A Simple, Positive Semi-Definite, Heteroscedastic and Autocorrelation Consistent Covariance Matrix," *Econometric* 55, 703-708.

Pontiff, Jeffrey, 1996, "Costly Arbitrage: Evidence from Closed-End Funds" *Quarterly Journal of Economics* 111, 1135-1151.

Pastor, Lubos and Pietro Veronesi, 2003, "Stock Valuation and Learning about Profitability" *Journal of Finance* 58, 1749-1789.

Richardson, Scott, and Richard Sloan, 2003, "External Financing, Capital Investment, and Future Stock Returns," University of Pennsylvania working paper.

Rogers, William, 1993, "Regression Standard Errors in Clustered Samples," *Stata Technical Bulletin* 13, 19-23.

Sapienza, Paola, 2004, "The effects of government ownership on bank lending" *Journal of Financial Economics* 72, 357-384.

Skoulakis, Georgios, 2005, "Panel Data inference in Finance: Least-Squares vs Fama-MacBeth" Northwestern working paper.

White, Hal, 1984, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity," *Econometrica* 48, 817-838.

Williams, Rick, 2000, "A Note on Robust Variance Estimation for Cluster-Correlated Data," *Biometrics* 56, 645-646.

Wooldridge, Jeffrey, 2002, *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.

Wooldridge, Jeffrey, 2004, "Cluster-sample Methods in Applied Econometrics" Michigan State University working paper.

Wu, YiLin, 2004, "The choice of equity-selling mechanisms" *Journal of Financial Economics* 74, 93-119.

Table 1: Estimating Standard Errors with a Firm Effect
OLS and Rogers Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) Avg(SE_R)		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	1.0004 0.0286 0.0283 0.0283	1.0006 0.0288 0.0283 0.0282	1.0002 0.0279 0.0283 0.0282	1.0001 0.0283 0.0283 0.0282
	25%	1.0004 0.0287 0.0283 0.0283	0.9997 0.0353 0.0283 0.0353	0.9999 0.0403 0.0283 0.0411	0.9997 0.0468 0.0283 0.0463
	50%	1.0001 0.0289 0.0283 0.0282	1.0002 0.0414 0.0283 0.0411	1.0007 0.0508 0.0283 0.0508	0.9993 0.0577 0.0283 0.0590
	75%	1.0000 0.0285 0.0283 0.0282	1.0004 0.0459 0.0283 0.0462	0.9995 0.0594 0.0283 0.0589	1.0016 0.0698 0.0283 0.0693

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set (500 firms and 10 years per firm). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. The fraction of the residual variance which is due to a firm specific component varies across the rows of the table from 0% (no firm effect) to 75%. The fraction of the independent variable's variance which is due to a firm specific component varies across the columns of the table from 0% (no firm effect) to 75%. Each cell contains the average slope coefficient estimated by OLS and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the average OLS estimated standard error of the coefficient. The fourth entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years). As an example, when fifty percent of the variability in both the residual and the independent variable is due to the fixed firm effect ($\rho_X = \rho_\varepsilon = 0.50$), the true standard error of the OLS coefficient is 0.0508. The OLS standard error estimate is 0.0283 and the clustered standard error estimate is 0.0508.

Table 2: Estimating Standard Errors with a Firm Effect
Fama-MacBeth Standard Errors

Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM})		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	1.0004 0.0287 0.0276	1.0006 0.0288 0.0276	1.0002 0.0280 0.0277	1.0001 0.0283 0.0275
	25%	1.0004 0.0288 0.0275	0.9997 0.0354 0.0268	0.9998 0.0403 0.0259	0.9997 0.0469 0.0250
	50%	1.0000 0.0289 0.0276	1.0002 0.0415 0.0259	1.0007 0.0509 0.0238	0.9993 0.0578 0.0219
	75%	1.0000 0.0286 0.0277	1.0004 0.0460 0.0248	0.9995 0.0595 0.0218	1.0016 0.0699 0.0183

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set (500 firms and 10 years per firm). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. The fraction of the residual variance which is due to a firm specific component varies across the rows of the table from 0% (no firm effect) to 75%. The fraction of the independent variable's variance which is due to a firm specific component varies across the columns of the table from 0% (no firm effect) to 75%. The first entry is the average estimated slope coefficient based on a Fama-MacBeth estimation (e.g. the regression is run for each of the 10 years and the estimate is the average of the 10 estimated slope coefficients). The second entry is the standard deviation of the coefficient estimated by Fama-MacBeth. This is the true standard error of the Fama-MacBeth coefficient. The third entry is the average standard error estimated by the Fama-MacBeth procedure (see equation 9). As an example, when fifty percent of the variability in both the residual and the independent variable is due to the fixed firm effect ($\rho_X = \rho_\varepsilon = 0.50$), the true standard error of the Fama-MacBeth coefficient is 0.0509. The Fama-MacBeth standard error estimate is 0.0238.

Table 3: Estimating Standard Errors with a Time Effect
OLS and Rogers Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) Avg(SE_R)		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	1.0004 0.0286 0.0283 0.0277	1.0002 0.0291 0.0288 0.0276	1.0006 0.0293 0.0295 0.0275	0.9994 0.0314 0.0306 0.0270
	25%	1.0006 0.0284 0.0279 0.0268	1.0043 0.1490 0.0284 0.1297	0.9962 0.2148 0.0289 0.1812	0.9996 0.2874 0.0300 0.2305
	50%	0.9996 0.0276 0.0274 0.0258	0.9997 0.2138 0.0278 0.1812	0.9919 0.3015 0.0282 0.2546	1.0079 0.3986 0.0292 0.3248
	75%	1.0002 0.0273 0.0267 0.0244	0.9963 0.2620 0.0271 0.2215	0.9970 0.3816 0.0276 0.3141	0.9908 0.4927 0.0284 0.3986

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set (500 firms and 10 years per firm). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. The fraction of the residual variance which is due to a year specific component varies across the rows of the table from 0 percent (no time effect) to 75 percent. The fraction of the independent variable's variance which is due to a year specific component varies across the columns of the table from 0 percent (no time effect) to 75 percent. Each cell contains the average estimated slope coefficient from OLS and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the average standard error estimated by OLS. The fourth entry is the average standard error clustered by year (i.e. accounts for the possible correlation between observations on different firms in the same year).

Table 4: Estimating Standard Errors with a Time Effect
Fama-MacBeth Standard Errors

Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM})		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	1.0004 0.0287 0.0278	1.0004 0.0331 0.0318	1.0007 0.0396 0.0390	0.9991 0.0573 0.0553
	25%	1.0005 0.0252 0.0239	1.0003 0.0284 0.0276	1.0006 0.0343 0.0338	0.9999 0.0496 0.0480
	50%	1.0000 0.0200 0.0195	1.0002 0.0231 0.0227	1.0006 0.0280 0.0276	1.0007 0.0394 0.0387
	75%	1.0001 0.0142 0.0138	0.9996 0.0161 0.0159	1.0000 0.0200 0.0196	0.9999 0.0285 0.0276

Notes

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set (500 firms and 10 years per firm). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. The fraction of the residual variance which is due to a year specific component varies across the rows of the table from 0 percent (no time effect) to 75 percent. The fraction of the independent variable's variance which is due to a year specific component varies across the columns of the table from 0 percent (no time effect) to 75 percent. The first entry is the average estimated slope coefficient based on a Fama-MacBeth estimation (e.g. the regression is run for each of the 10 years and the estimate is the average of the 10 estimated slope coefficients). The second entry is the standard deviation of the coefficient estimated by Fama-MacBeth. This is the true standard error of the Fama-MacBeth coefficient. The third entry is the average standard error estimated by the Fama-MacBeth procedure (e.g. equation 9).

Table 5: Estimating Standard Errors with a Firm and Time Effect
OLS and Rogers Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) Avg(SE_R)		Independent Variable Volatility from Firm Effect		
		25%	25%	50%
		Independent Variable Volatility from Time Effect		
Residual Volatility from Firm Effect	Residual Volatility from Time Effect	25%	50%	25%
25%	25%	0.9997 0.0407 0.0283 0.0400	1.0004 0.0547 0.0347 0.0548	1.0004 0.0489 0.0283 0.0489
25%	50%	1.0005 0.0362 0.0231 0.0364	1.0015 0.0515 0.0283 0.0508	0.9993 0.0468 0.0231 0.0461
50%	25%	1.0002 0.0493 0.0283 0.0490	1.0008 0.0690 0.0347 0.0692	0.9994 0.0631 0.0283 0.0630

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set with 5,000 observations (500 firms and 10 years per firm). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. In these simulations, the proportion of the variance of the independent variable and the residual which is due to the firm effect is either 25 or 50 percent. The proportion which is due to the time effect is also 25 or 50 percent. For example, in the bottom left cell 25 percent of the variability in the independent variable is from the firm effect and 25 percent is from the time effect. 50 percent of the variability of the residual is from the firm effect and 25 percent is from the time effect. Each cell contains the average estimated slope coefficient from OLS and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the average standard error estimated from OLS. The fourth entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years). Each regression includes nine year dummies.

Table 6: Estimating Standard Errors with a Firm and Time Effect
Fama-MacBeth Standard Errors

Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM})				Independent Variable Volatility from Firm Effect		
				25%	25%	50%
				Independent Variable Volatility from Time Effect		
				25%	50%	25%
Residual Volatility from Firm Effect	25%	Residual Volatility from Time Effect	25%	0.9997 0.0407 0.0258	1.0004 0.0547 0.0309	1.0004 0.0489 0.0243
	25%		50%	1.0005 0.0362 0.0206	1.0015 0.0515 0.0239	0.9993 0.0469 0.0185
	50%		25%	1.0002 0.0493 0.0244	1.0008 0.0691 0.0275	0.9994 0.0632 0.0206

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set with 5,000 observations (500 firms and 10 years per firm). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. In these simulations, the proportion of the variance of the independent variable and the residual which is due to the firm effect is either 25 or 50 percent. The proportion which is due to the time effect is also 25 or 50%. For example, in the bottom left cell 25 percent of the variability in the independent variable is from the firm effect and 25 percent is from the time effect. 50 percent of the variability of the residual is from the firm effect and 25 percent is from the time effect. The first entry in each cell is the average estimated slope coefficient based on a Fama-MacBeth estimation (e.g. the regression is run for each of the 10 years and the estimate is the average of the 10 estimated slope coefficients). The second entry is the standard deviation of this coefficient. This is the true standard error of the Fama-MacBeth coefficient. The third entry is the average standard error estimated by the Fama-MacBeth procedure (e.g. equation 9).

Table 7: Estimated Standard Errors with a Non-Fixed Firm Effect
Panel A: OLS and Rogers Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) Avg(SE_R)	I	II	III	IV
ρ_X / ρ_ε	0.25 / 0.25	0.00 / 0.00	0.25 / 0.25	0.50 / 0.50
φ_X / φ_ε	0.00 / 0.00	0.75 / 0.75	0.75 / 0.75	0.50 / 0.50
OLS	1.0001 0.0355 0.0283 0.0352	1.0001 0.0483 0.0283 0.0488	1.0009 0.0566 0.0283 0.0569	1.0007 0.0587 0.0283 0.0578
OLS with firm dummies	1.0007 0.0299 0.0298 0.0314	1.0008 0.0443 0.0298 0.0466	1.0013 0.0442 0.0298 0.0465	1.0000 0.0357 0.0298 0.0377

Panel B: Fama-MacBeth Standard Errors

Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM}) Avg($SE_{FM\text{-}AR1}$)	I	II	III	IV
ρ_X / ρ_ε	0.25 / 0.25	0.00 / 0.00	0.25 / 0.25	0.50 / 0.50
φ_X / φ_ε	0.00 / 0.00	0.75 / 0.75	0.75 / 0.75	0.50 / 0.50
Fama-MacBeth	1.0001 0.0357 0.0267 0.0250 0.0250	1.0001 0.0484 0.0240 0.0374 0.0344	1.0008 0.0567 0.0221 0.0376 0.0336	1.0007 0.0588 0.0220 0.0296 0.0281
Avg(1st order auto-correlation)	-0.1134	0.2793	0.3250	0.1759

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set with 5,000 observations (500 firms and 10 years per firm). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. Across the columns the magnitude of the fixed firm effect (ρ) and the first order auto-correlation (φ) is changed. ρ_X (ρ_ε) is the fraction of the independent variable's (residual's) variance which is due to the fixed firm effect. φ_x (φ_ε) is the first order auto-correlation of the non-fixed portion of the firm effect of the independent variable (residual). Combining equations (4) with equation (16), the residual is specified as:

$$\begin{aligned} \varepsilon_{it} &= \mu_{it} + \eta_{it} = \mu_{it} + \varsigma_{it} && \text{if } t = 1 \\ &= \mu_{it} + \eta_{it} = \mu_{it} + \varphi_\varepsilon \eta_{it-1} + \sqrt{1 - \varphi^2}\, \varsigma_{it} && \text{if } t > 1 \end{aligned} \qquad \mathbf{(1)}$$

The independent variable is specified in a similar manner.

Panel A contains coefficients estimated by OLS. In the first row only the independent variable (X) is included; in the second row 499 firm dummies (for 500 firms) are also included in the regression. The first two entries in each cell contain the average slope estimated by OLS and the standard deviation of the coefficient (i.e. the true standard error). The third entry is the average standard error estimated from OLS. The fourth entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years).

Panel B contains coefficients and standard errors estimated by Fama-MacBeth. The first two entries in each cell contain the average slope estimated by Fama-MacBeth and the standard deviation of the coefficient (i.e. the true standard error). The third entry in these cells is the average standard error estimated by the Fama-MacBeth procedure (see equation 9). The last two entries are the Fama-MacBeth standard error estimates corrected for first order auto-correlation. The fourth entry assumes an infinite lag (i.e. multiplied by the square root of (1+φ)/(1-φ)), and the fifth entry assumes a finite lag of 9 periods (i.e. multiply by the square root of sum from k=1 to T of (T-k) φ^k). The bottom row contains the average first order autocorrelation of β_t and β_{t-1}.

Table 8: Asset Pricing Application
Equity Returns and Asset Tangibility

	I	II	III	IV	V
Log(B/M_{t-5})	0.1883[1]	0.1883[1]	0.1883[10]	0.1728[5]	0.1504[10]
	(0.0261)	(0.0270)	(0.1007)	(0.0824)	(0.0831)
Log(Book Return)	0.1946[1]	0.1946[1]	0.1946[5]	0.1691[5]	0.1544[10]
(last 5 years)	(0.0421)	(0.0433)	(0.0973)	(0.0848)	(0.0808)
Market Return	-0.3177[1]	-0.3177[1]	-0.3177[1]	-0.3002[1]	-0.2536[1]
(last 5 years)	(0.0283)	(0.0292)	(0.1092)	(0.0957)	(0.0910)
Share issuance	-0.5012[1]	-0.5012[1]	-0.5012[1]	-0.5172[1]	-0.5104[1]
	(0.0471)	(0.0466)	(0.1529)	(0.1275)	(0.1213)
R^2	0.0006	0.0006	0.0006	0.0006	0.0006
Coefficient Estimates	OLS	OLS	OLS	FM	WFM
Standard Errors	White	CL - F	CL - T	FM	FM

Notes:

The table contains coefficient and standard error estimates of the equity return regressions from Daniel and Titman (2004). The data is briefly described in the appendix and in detail in their paper. The sample runs from July, 1968 to December, 2001 and contains 699,707 firm-month observations. The estimates in columns I-III are OLS coefficients and the regressions contain time (monthly) dummies. Standard errors are reported in parenthesis. White standard errors are reported in column I, standard errors clustered by firm in column II and standard errors clustered by month in column III. Column IV and V contain coefficients and standard errors estimated by Fama-MacBeth. In column V, I weighted each of the monthly coefficient estimates by the number of observations in the given month.

[10] significant at 10%; [5] significant at 5%; [1] significant at 1%

Table 9: Corporate Finance Application
Capital Structure Regressions (1965-2003)

	I	II	III	IV	V
Ln(MV Assets)	0.0460[1]	0.0460[5]	0.0460[1]	0.0394[1]	0.0444[1]
	(0.0055)	(0.0184)	(0.0074)	(0.0076)	(0.0074)
Ln(1+Firm Age)	-0.0432[1]	-0.0432	-0.0432[1]	-0.0479[1]	-0.0461[1]
	(0.0084)	(0.0297)	(0.0067)	(0.0077)	(0.0076)
Profits / Sales	-0.0330[1]	-0.0330	-0.0330[1]	-0.0299[1]	-0.0299[1]
	(0.0107)	(0.0359)	(0.0098)	(0.0097)	(0.0099)
Tangible assets	0.1043[1]	0.1043[1]	0.1043[1]	0.1158[1]	0.1067[1]
	(0.0057)	(0.0197)	(0.0083)	(0.0096)	(0.0089)
Market to book	-0.0251[1]	-0.0251[1]	-0.0251[1]	-0.0272[1]	-0.0258[1]
(Assets)	(0.0006)	(0.0020)	(0.0013)	(0.0016)	(0.0014)
Advertising / Sales	-0.3245[1]	-0.3245	-0.3245[1]	-0.3965[5]	-0.4309[1]
	(0.0841)	(0.2617)	(0.0814)	(0.1712)	(0.1405)
R&D / Sales	-0.3513[1]	-0.3513[5]	-0.3513[1]	-0.3359[1]	-0.3532[1]
	(0.0469)	(0.1544)	(0.0504)	(0.0501)	(0.0481)
R&D > 0	0.0177[1]	0.0177[5]	0.0177[1]	0.0126[1]	0.0155[1]
(=1 if yes)	(0.0024)	(0.0076)	(0.0025)	(0.0034)	(0.0031)
R-squared	0.1360	0.1360	0.1360	0.1300	0.1238
Coefficient Estimates	OLS	OLS	OLS	FM	WFM
Standard Errors	White	CL - F	CL - T	FM	FM

Notes:

The table contains coefficient and standard error estimates of a capital structure regressions. The dependent variable is the market debt ratio (book value of debt divided by the sum of the book value of assets minus the book value of equity plus the market value of equity). The data is annual observations between 1965 and 2003. The sample contains NYSE firms which pay a dividend in the previous year. There are 24,286 firm years in the sample. The independent variables are lagged one year and are defined in the appendix. The estimates in columns I-III are OLS coefficients and the regressions contain time (year) dummies. Standard errors are reported in parenthesis. White standard errors are reported in column I, standard errors clustered by firm in column II, and standard errors clustered by year in column III. Column IV and V contain coefficients and standard errors estimated by Fama-MacBeth. In column V, I weighted each of the yearly coefficient estimates by the number of observations in the given year.

[10] significant at 10%; [5] significant at 5%; [1] significant at 1%

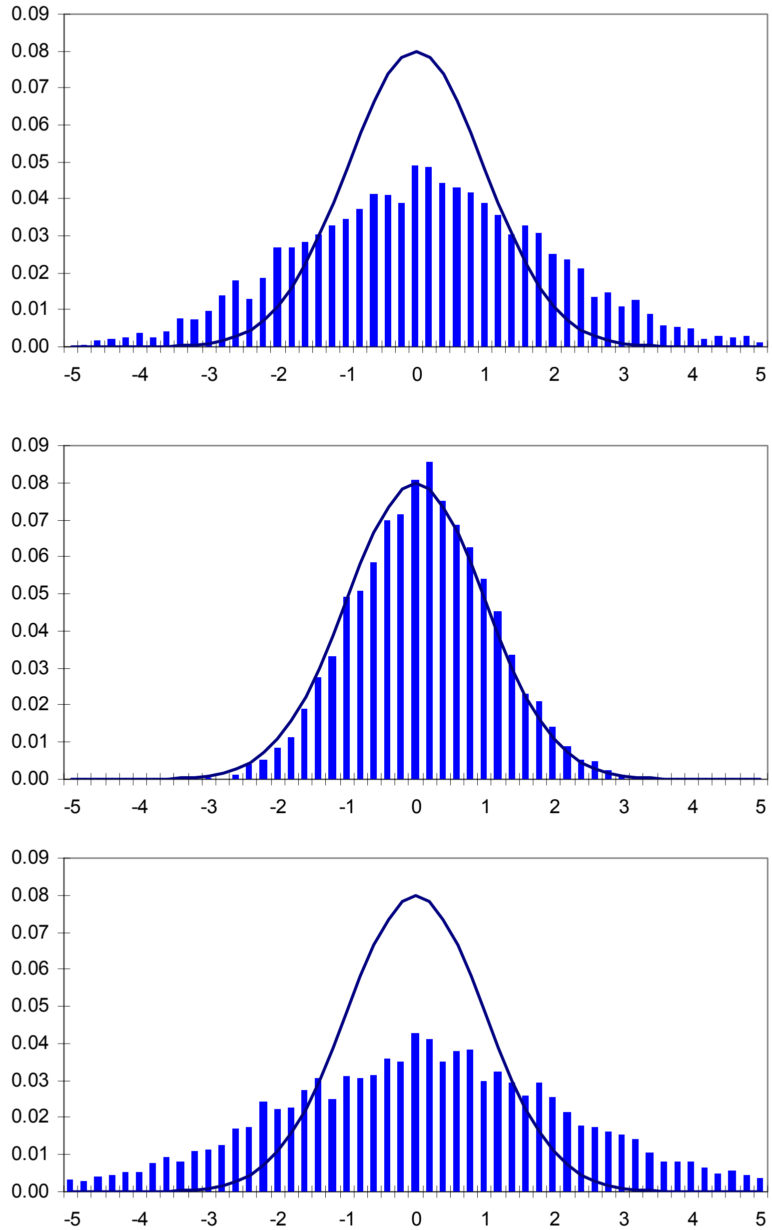
Figure 1: Residual Cross Product Matrix
Assumptions About Zero Co-variances

	Firm 1			Firm 2			Firm 3		
Firm 1	ε_{11}^2	$\varepsilon_{11}\,\varepsilon_{12}$	$\varepsilon_{11}\,\varepsilon_{13}$	0	0	0	0	0	0
	$\varepsilon_{12}\,\varepsilon_{11}$	ε_{12}^2	$\varepsilon_{12}\,\varepsilon_{13}$	0	0	0	0	0	0
	$\varepsilon_{13}\,\varepsilon_{11}$	$\varepsilon_{13}\,\varepsilon_{12}$	ε_{13}^2	0	0	0	0	0	0
Firm 2	0	0	0	ε_{21}^2	$\varepsilon_{21}\,\varepsilon_{22}$	$\varepsilon_{21}\,\varepsilon_{23}$	0	0	0
	0	0	0	$\varepsilon_{22}\,\varepsilon_{21}$	ε_{22}^2	$\varepsilon_{22}\,\varepsilon_{23}$	0	0	0
	0	0	0	$\varepsilon_{23}\,\varepsilon_{21}$	$\varepsilon_{23}\,\varepsilon_{22}$	ε_{23}^2	0	0	0
Firm 3	0	0	0	0	0	0	ε_{31}^2	$\varepsilon_{31}\,\varepsilon_{32}$	$\varepsilon_{31}\,\varepsilon_{33}$
	0	0	0	0	0	0	$\varepsilon_{32}\,\varepsilon_{31}$	ε_{32}^2	$\varepsilon_{32}\,\varepsilon_{33}$
	0	0	0	0	0	0	$\varepsilon_{33}\,\varepsilon_{31}$	$\varepsilon_{33}\,\varepsilon_{32}$	ε_{33}^2

Notes:

This figure shows a sample covariance matrix of the residuals. Assumptions about elements of this matrix and which are zero is the source of difference in the various standard error estimates. The covariance of the matrix of the residuals has $(NT)^2$ elements where N is the number of firms and T is the number of years. Both are three in this illustration. The standard OLS assumption is that only the NT diagonal terms are non-zero. Standard errors clustered by firm assumes that the correlation of the residuals within the cluster may be non-zero (these elements are shaded). The cluster assumption assumes that residuals across clusters are uncorrelated. These are recorded as zero in the above matrix.

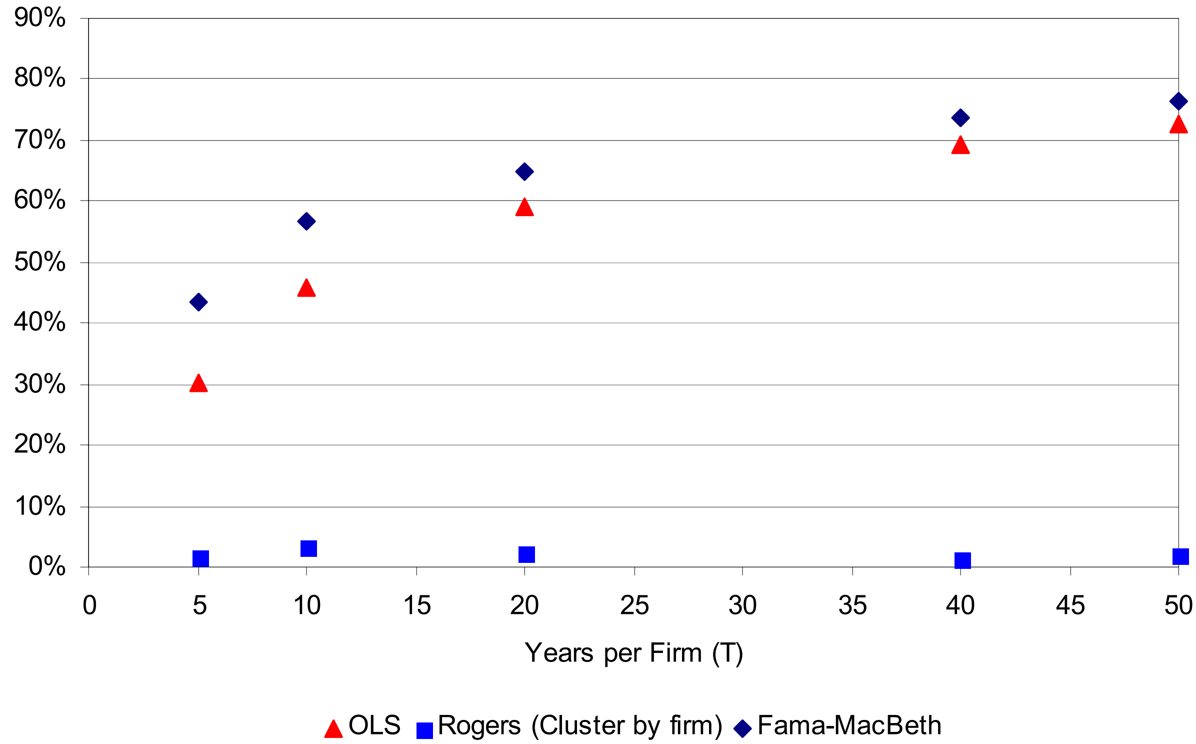
Figure 2: Distribution of Simulated T-Statistics



Notes:

The figures contain the theoretical t-distribution (the line), and the distribution of t-statistics produced by the simulation (the bars) when fifty percent of the variability in the independent variable and the residual is due to the firm effect. The top figure is the distribution of the t-statistics based on the OLS standard errors, the middle figure is the distribution of t-statistics based on the standard errors clustered by firm, and the bottom figure is the distribution of t-statistics based on Fama-MacBeth standard errors. When the standard errors estimates are too small (as with OLS and Fama-MacBeth) there are too many t-statistics which are large in absolute value.

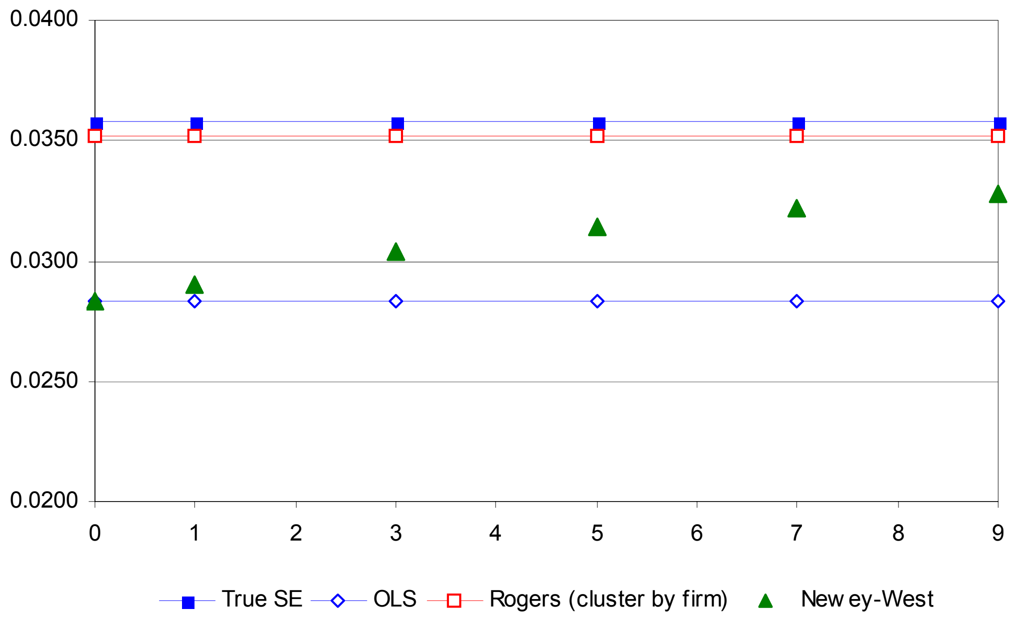
Figure 3: Bias in Estimated Standard Errors
as a function of observations per cluster



Notes:

The figure graphs the percentage by which the three methods underestimate the true standard error in the presence of a firm effect. The results are based on 5,000 simulations of a data set with 5,000 observations. The number of years per firm ranges from five to fifty. The firm effect was assumed to comprise fifty percent of the variability in both the independent variable and the residual. The underestimates are calculated as one minus the average estimated standard error divided by the true standard deviation of the coefficient estimate. For example, the standard deviation of the coefficient estimate was 0.0406 in the simulation with five years of data (T=5). The average of the OLS estimated standard errors is 0.0283. Thus the OLS underestimated the true standard error by 30% (1 - 0.0283/0.0406).

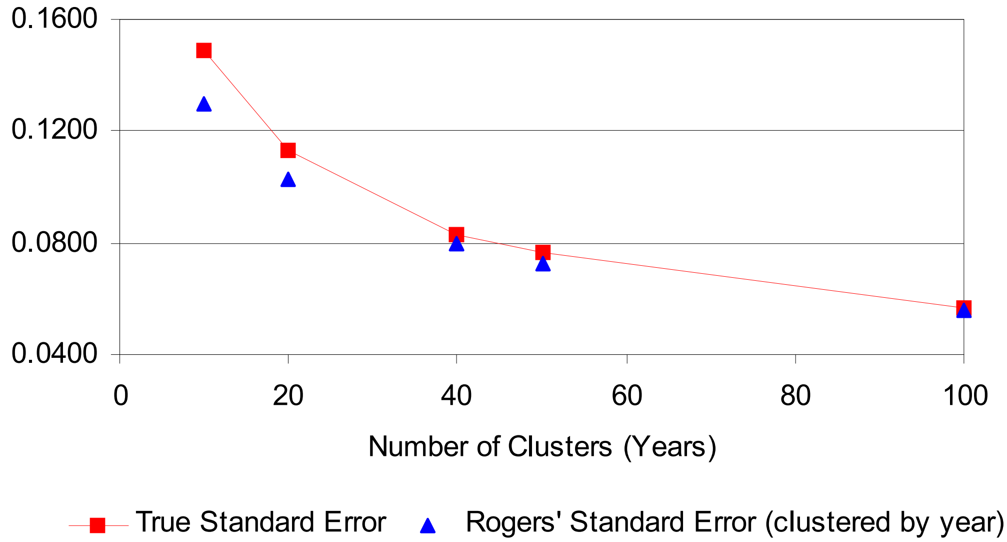
Figure 4: Relative Performance of OLS, Rogers, and Newey-West Standard Errors



Notes:

The figure contains OLS, standard errors clustered by firm, and Newey-West standard error estimates, as well as the true standard error. Estimates are based on 5,000 simulated data sets. Each data set contains 5,000 observations (500 firms and 10 years per firm). In each simulation, twenty-five percent of the variability in both the independent variable and the residual is due to a firm effect [i.e. $\rho_X = \rho_\varepsilon = 0.25$]. The true standard error (shaded squares), the OLS standard error estimates (empty diamonds), and the clustered standard errors (empty squares) are plotted as straight lines as they do not depend upon the assumed lag length. The Newey-West standard error estimates, which rise with the assumed lag length, are plotted as triangles.

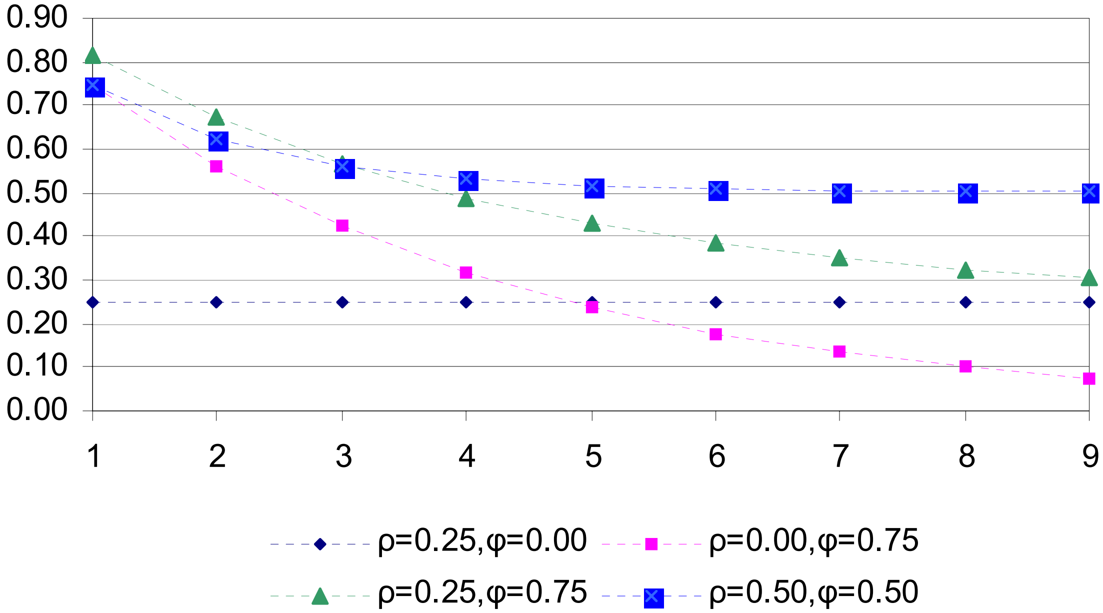
Figure 5: True Standard Errors and Robust Estimates
as a function of cluster size (T)



Notes:

The true standard errors (squares) and the standard error clustered by year (triangles) are graphed against the number of years (clusters) used in each simulation. The standard errors are the average across 5,000 simulations. Each simulated data set has 5,000 observations. In each simulation, twenty-five percent of the variability in both the independent variable and the residual is due to the time effect [i.e. $\rho_X = \rho_\varepsilon = 0.25$]. The clustered standard errors underestimate the true standard errors, but this underestimate declines with the number of years (clusters). In these simulations, the underestimation ranges from 15 percent when there were 10 years in the simulated data set to 1 percent when there were 100 years in the simulated data set.

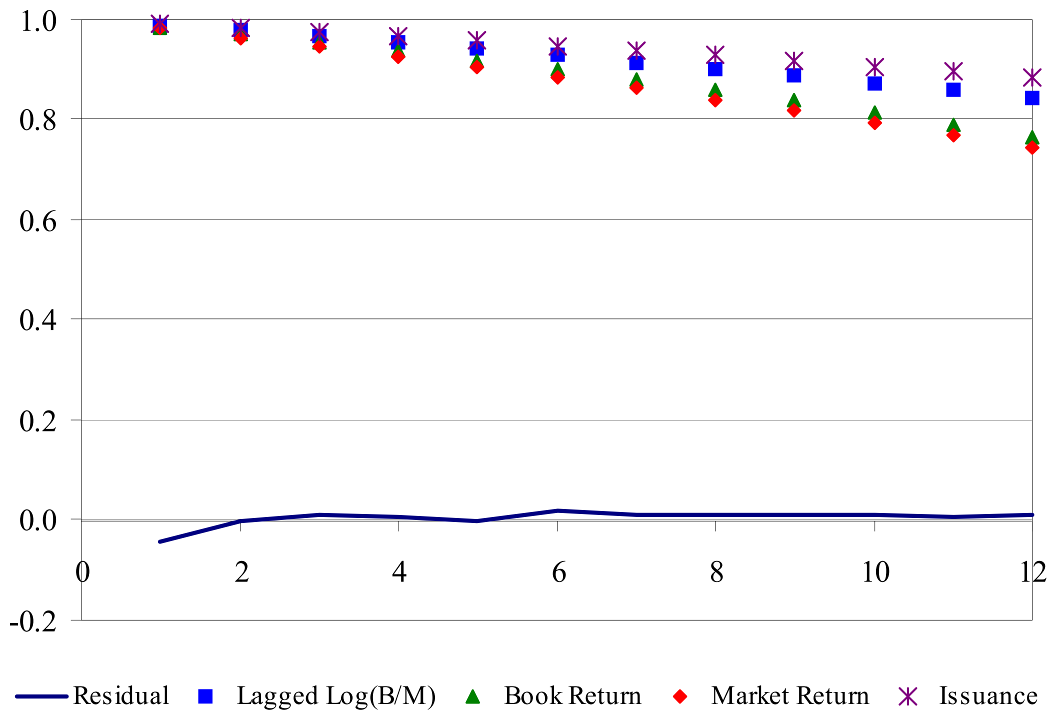
Figure 6: Auto-correlation Patterns of Non-Fixed Firm Effects



Notes:

This figure contains the auto-correlations of the residuals and the independent variable for lags one through nine for the data structures used in Table 7. The specifications contain a fixed and a temporary firm component. ρ is the fraction of the variance which is due to the fixed firm effect and φ is the first order auto-correlation of the non-fixed firm effect (see equation 17).

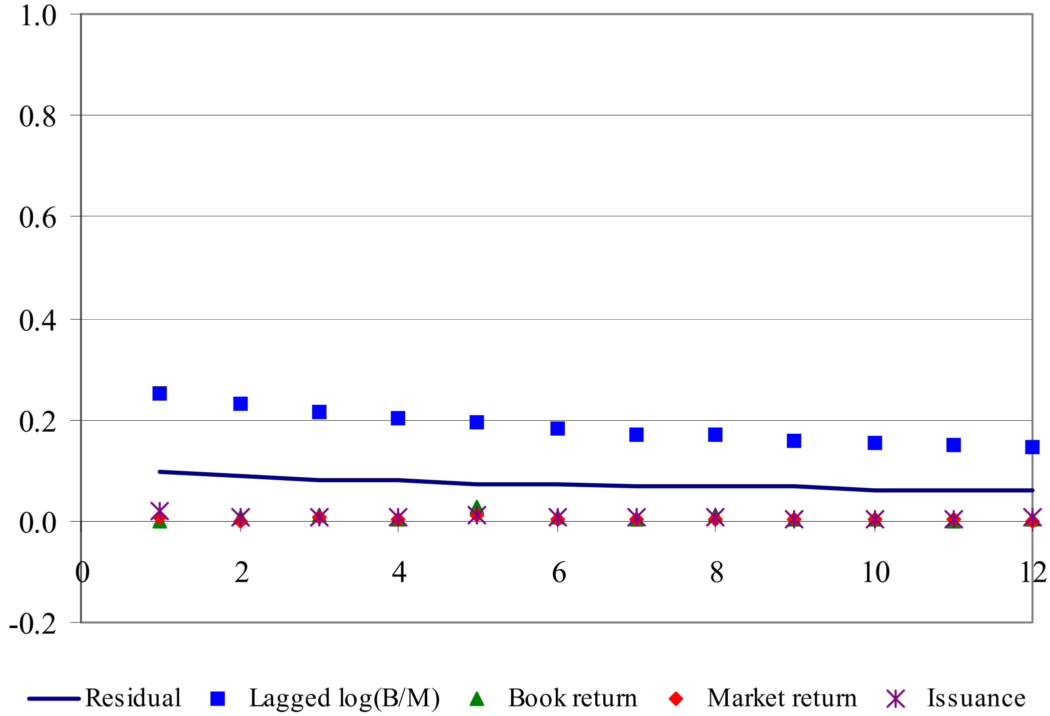
Figure 7: Within Firm Auto-Correlation in the Residuals and the Independent Variables
Asset Pricing Example



Notes:

The auto-correlations of the residual and the four independent variables are graphed for lags of one to twelve months. Correlations are calculated only for observations of the same firm [i.e. Corr ($\varepsilon_{i\,t}\, \varepsilon_{i\,t\text{-}k}$) for k equal one to twelve]. The independent variables are described in the appendix and in Daniel and Titman (2004).

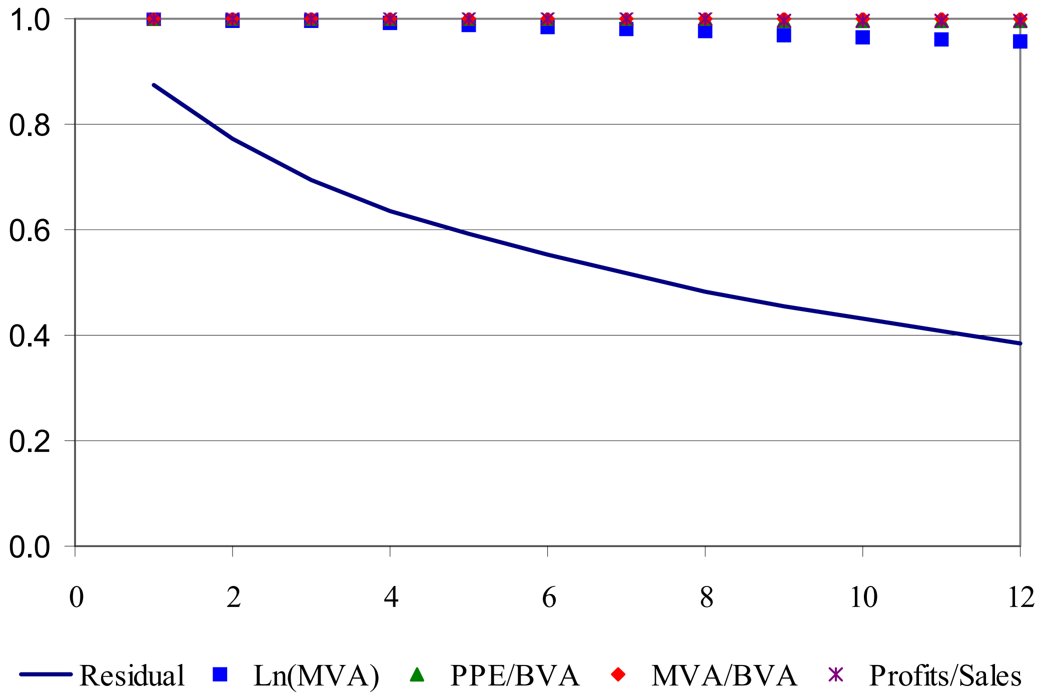
Figure 8: Within Month Auto-Correlation in the Residuals and the Independent Variables
Asset Pricing Example



Notes:

The auto-correlations of the residuals and the four independent variables are graphed for lags of one to twelve firms. Correlations are calculated only for observations of the same year [i.e. Corr ($\varepsilon_{i\,t}\, \varepsilon_{i\text{-}k\,t}$) for k equal one to twelve]. The data was sorted by month, then by four digit industry, and then by firm identifier (permco). The independent variables are described in the appendix and in Daniel and Titman (2004).

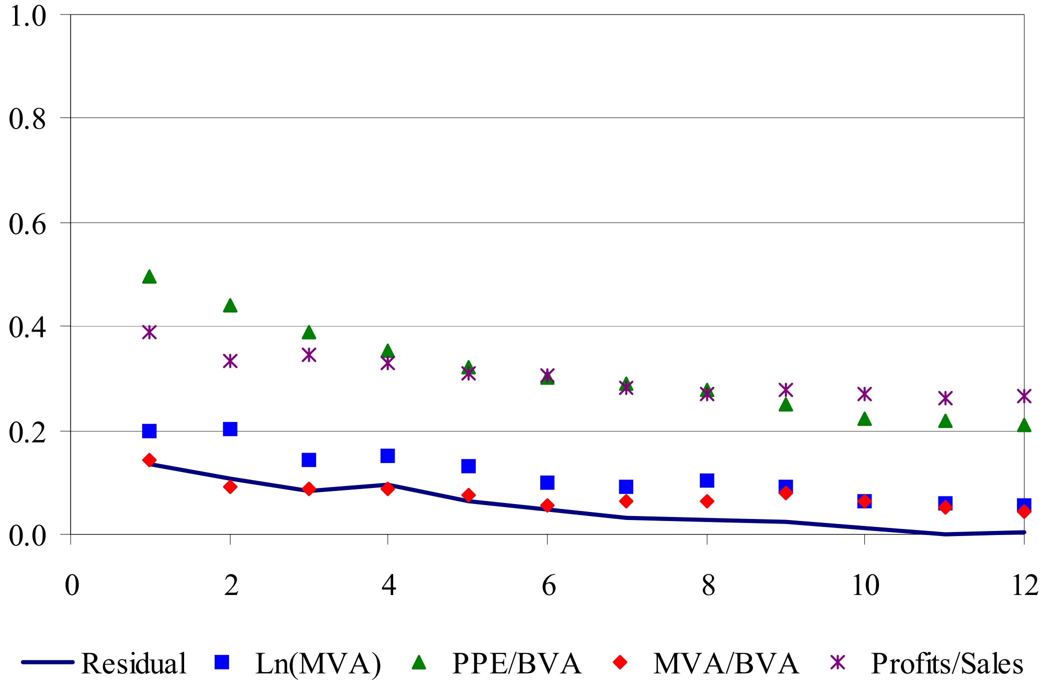
Figure 9: Within Firm Auto-Correlation in the Residuals and the Independent Variables
Corporate Finance Example



Notes:

The auto-correlations of the residual and four of the eight independent variables are graphed for lags of one to twelve years. Correlations are calculated only for observations of the same firm [i.e. Corr ($\varepsilon_{i\,t}\ \varepsilon_{i\,t-k}$) for k equal one to twelve]. The independent variables are described in the appendix. The graph for the remaining four variables are similar and are available from the author.

Figure 10: Within Month Auto-Correlation in the Residuals and the Independent Variables
Corporate Finance Example



Notes:

The auto-correlations of the residuals and four of the eight independent variables are graphed for lags of one to twelve firms. Correlations are calculated only for observations of the same year [i.e. Corr ($\varepsilon_{i\,t}\,\varepsilon_{i\text{-}k\,t}$) for k equal one to twelve]. The data was sorted by month, then by four digit industry, and then by firm identifier (gvkey). The independent variables are described in the appendix. The graph for the remaining four variables are similar and are available from the author.