

Petersen, Mitchell A.

Working Paper

Estimating standard errors in finance panel data sets: Comparing approaches

CSIO Working Paper, No. 0055

Provided in Cooperation with:

Department of Economics - Center for the Study of Industrial Organization (CSIO), Northwestern University

Suggested Citation: Petersen, Mitchell A. (2004) : Estimating standard errors in finance panel data sets: Comparing approaches, CSIO Working Paper, No. 0055, Northwestern University, Center for the Study of Industrial Organization (CSIO), Evanston, IL

This Version is available at:

<https://hdl.handle.net/10419/38640>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

THE CENTER FOR THE STUDY
OF INDUSTRIAL ORGANIZATION
AT NORTHWESTERN UNIVERSITY

Working Paper #0055

Estimating Standard Errors in Finance Panel
Data Sets: Comparing Approaches

By

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

November, 2004

* I thank the Financial Institutions and Markets Research Center at Northwestern University's Kellogg School for support. In writing this paper, I have benefitted greatly from discussions with Doug Staiger, Paola Sapienza, and Kent Daniel as well as the comments of seminar participants at Northwestern University and the Universities of Chicago and Iowa. The research assistance of Sungjoon Park is greatly appreciated.

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 Rogers standard errors, while asset pricing has used the Fama-MacBeth procedure to estimate standard errors. This paper will examine the different methods used in the literature and explain 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 thus give researchers guidance for their use.

I) Introduction

It is well known that OLS standard errors are correct 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 under estimate the true variability of the coefficient estimates. Although the use of panel data sets (e.g. data sets that contain observations on the same firm from multiple years) is common, the way that researchers have addressed possible biases in the standard errors varies widely. In recently published papers which include a regression on panel data, forty-five percent of the papers did not report adjusting the standard errors for possible dependence in the residuals.¹ Among the remaining papers, approaches for estimating the coefficients and standard errors in the presence of within cluster correlation varied. 31 percent of the papers included dummy variables for each cluster (e.g. for each firm). 34 percent of the papers estimated both the coefficients and the standard errors using the Fama-MacBeth procedure (Fama-MacBeth, 1973). The remaining two methods used OLS (or an analogous method) to estimate the coefficients but reported standard errors adjusted for correlation within a cluster. 7 percent of the papers adjusted the standard errors using the Newey-West procedure (Newey and West, 1987) modified for use in a panel data set, while 22 percent of the papers reported Rogers standard errors (Rogers, 1993, Williams, 2000, Moulton, 1990) which are White standard errors adjusted to account for possible correlation within a cluster. These are also called clustered standard errors.

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 to 2004 for a description of how the regression coefficients and standard errors were estimated in a panel data set. Panel data sets are data sets where observations can be grouped into clusters (e.g. multiple observations per firm, per year, or per country). I included only papers which reported at least five observations in each dimension (e.g. firms, years, or countries). Papers which did to report the method for estimating the standard errors, or reported correcting the standard errors only for heteroscedasticity (i.e. White standard errors), were coded as not having correcting the standard errors for within cluster dependence.

Although the literature has used a diversity of methods to estimate standard errors in panel data sets, it has provided little guidance to researchers as to when a given method is appropriate. Since the methods can sometimes produce different estimates it is important to understand how the methods compare, when they will produce different estimates of the standard errors, and when they differ how to choose among the estimates. This is the objective of the paper.

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 regression can be cross sectionally correlated (e.g. the observations of a firm in different years are correlated). I will call this a firm effect. Alternatively, the residuals of a given year may be correlated across firms. I will call this a time effect. I will simulate panel data set 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 performance. Section II contains the standard error estimates in the presence of a fixed firm effect. Both the OLS and the Fama-MacBeth standard errors are biased downward and the magnitude of this bias is increasing in the magnitude of the firm effect. The Rogers standard errors are unbiased as they account for the dependence created by the firm effect. The Newey-West standard errors, as modified for panel data, are also biased but their bias is small.

In section III, the same analysis is conducted with a time effect instead of a firm effect. Since Fama-MacBeth procedure is designed to address a time effect, not a firm effect, the Fama-MacBeth standard errors are unbiased and the coefficient estimates are more efficient than the OLS estimates. The intuition of these first two sections carries over to Section IV, were data with both a firm and a time effect is included. Thus far, the firm effect has been specified as a fixed effect (e.g. does not decay over time). In practice, the firm effect in the residual may decay over time such that the

correlation between residuals declines as the time between them grows. In Section V, I simulate data with a more general correlation structure. This not only allows me to compare OLS, clustered, and Fama-MacBeth standard errors in a more general setting, it also allows me to access the relative benefit of using fixed effects (firm dummies) to estimate the coefficients and whether this changes the way we should estimate standard errors.

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

A) Robust Standard Error Estimates.

To provide intuition on why the standard errors produced by OLS are incorrect and how Rogers robust estimates correct this problem, it is helpful to review the expression for the variance of the estimated coefficients. To simplify the formulas, I will assumed that all variables have been de-meaned. Thus, the intercepts can be ignored, and the variances can be directly calculated as sums of squares of a variable. The standard regression for a panel data set is:

$$Y_{it} = X_{it} \beta + \varepsilon_{it} \quad (1)$$

where we have observations on firms (i) across years (t). The estimated coefficient is:

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

and thus the estimated variance of the coefficient is:

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

This is the standard OLS formula and is based on the assumption that the errors are independent and identically distributed. The independence assumption is used to move from the first to the second line (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 It is the independence assumption which is often violated in panel data and which is the focus of the following section.

In relaxing the assumption of independent errors, I will initially assume the data has a fixed firm effect. Thus the residuals consist of a firm specific component as well as a component which is unique to each observation. 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.

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

Each of the components of X (μ and v) and ε (γ and η) are independent of each other. This is

² The Rogers standard errors are robust to heteroscedasticity residuals. However, since this is not my focus, I will assume that the errors are homoscedastic in the equations and simulations.

necessary for the coefficient estimates to be consistent.³ This is a typical panel data structure and implies a specific correlation among the observations of a given firm. Both the independent variable and the residual are correlated across two observations of the same firm, but are assumed to be independent across firms.

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

Given this data structure, I can calculate the true standard error of the OLS coefficient based on the data structure in equations (1), (4), and (5). Since the independent variable and residual are no longer independent within cluster, the square of the summed residuals is no longer equal to the sum of the residuals squared. The co-variances must be included as well.⁴ The variance of the OLS coefficient estimate is now:

³ Thus I am assuming that the model is correctly specified. I do this to focus on estimating the standard errors. Panel data sets often include a time effect as well as a firm effect. For the moment, I assume there is no time effect and return to the implications of a time effect in Section III.

⁴ When calculating the square of the sum of $X \varepsilon$ there are $(NT)^2$ terms (see Figure 1). There are NT variance terms $[\sigma^2(X) \sigma^2(\varepsilon)]$. These are the only ones included in the OLS standard errors. There are $NT(T-1)$ non-zero off diagonal terms $[(T(T-1)$ for each of N firms]. These are non-zero when there is a firm effect. The remaining $NT^2(N-1)$ diagonal terms are assumed to be zero. If there is a time effect, then $NT(N-1)$ of these would be non-zero as well $[N(N-1)$ for each of T years].

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

I used the assumption that residuals are independent across firms [e.g. $i \neq j$, see equation (6)] in deriving the second line. The true standard errors will thus be greater than those reported by OLS if and only if both ρ_X or ρ_ε are non-zero.⁵ The magnitude of the error is also increasing in the number of years in the data set (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.

Given our 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. Thus when the data has a fixed firm effect, the OLS standard errors will always understate the true standard

⁵ If the firm effect is not fixed, the correlation of ε_t and ε_s will not be the same. In this case, the equation will be a sum of all the covariances between ε_t and ε_s times the covariance between X_t and X_s . We will return to this issue in Section V when we examine non-fixed firm effects. If the panel is unbalanced, the true standard error and the bias in the OLS standard errors will be even larger (Moulton, 1986).

error. In other contexts, it is possible for the true standard errors to be greater or less than the reported (OLS) standard errors depending upon the sign of $\rho_x \rho_\varepsilon$.

The correlation of the residuals within cluster is the problem the Rogers standard errors (White standard errors adjusted for clustering) are designed to correct.⁶ 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. We return to this issue in Section III.

B) Testing the Standard Error Estimates by Simulation.

To demonstrate the relative accuracy of the different standard error estimates and confirm our intuition, 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.⁷ In the first version of the simulation, I included a fixed firm effect but no time effect in both the independent variable as well as in the residual. Thus the data was simulated as described in equations (4) and (5). Across simulations I assumed that the standard

⁶ The exact formula for the Rogers standard error is:

$$S^2(\beta) = \frac{N (NT - 1) \sum_{i=1}^N \left(\sum_{t=1}^T X_{it} \varepsilon_{it} \right)^2}{(NT - k) (N - 1) \sum_{i=1}^N \sum_{t=1}^T X_{it}^2} \quad (8)$$

⁷ Each simulated data set contains 10 yearly observations on 500 firms, for a data set of 5,000 observations. The components of the independent variable and the residual are assumed to be 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 the 5,000 simulations.

deviation of the independent variable and the residual were both constant at one and two respectively. This will produce an R^2 of 20 percent which is not unusual for empirical finance regressions. Across different simulations, I altered the fraction of the variance in the independent variable which is due to the firm effect. This fraction ranged 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 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 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 correct on average, even if the residuals are 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 of them 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 under estimate the true standard errors, and the magnitude of the under estimation 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 ($0.557 = 0.0283/0.0508$).⁸ The standard errors estimated by OLS do not change as I increase the firm effect 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 Rogers (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. Thus as expected, the robust standard errors correctly account for the dependence in the data common in a panel data set (Rogers, 1993, Williams, 2000).⁹

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 periods doubles, OLS attributes a doubling in the number of observations. 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 2). The robust standard errors are consistently close to the true standard errors independent of the number of time periods (see Figure 2).

⁸ In addition to a slope coefficient, all of the regressions also contained a constant whose true value is zero. The results for the slope coefficient carry over to the intercept estimation as well. For example when $\rho_x = \rho_\varepsilon = 0.50$, the estimated slope coefficient averages -0.0003 with a standard deviation of 0.0669. The OLS standard errors are biased down (0.0283) and the Rogers standard errors are correct on average (0.0663).

⁹ The variability of the standard errors is small relative to their mean. For example, when $\rho_x = \rho_\varepsilon = 0.50$, the mean OLS standard error is 0.283 with a standard deviation of 0.001 and the mean clustered standard error is 0.0508 with a standard deviation of 0.003. Instead of reporting average standard errors, I could also report the number of significant t-statistics. Using the OLS standard error, 15.3 percent of the t-statistics are statistically significant at the one percent level (i.e. the 99 percentile confidence interval contains the true coefficient 84.7 percent of the time). Using the clustered standard errors, 0.8 percent of the t-statistics are statistically significant at the one percent level. Since the standard deviation of the standard error is usually small and the distribution symmetric, t-statistics and standard errors give the same intuition. I will thus report only standard errors except where the intuition can be misleading.

C) Fama-MacBeth Standard Errors: The Equations

An alternative way to estimate the regression coefficients and standard errors which has been used in the literature, and one often suggested when the residuals are not independent, is the Fama-MacBeth approach (Fama and MacBeth, 1973).¹⁰ In this approach, the researcher runs T cross sectional regressions. The average of the T estimates is the coefficient estimate

$$\begin{aligned}\hat{\beta}_{\text{FM}} &= \sum_{t=1}^T \frac{\hat{\beta}_t}{T} \\ &= \frac{1}{T} \sum_{t=1}^T \left(\frac{\sum_{i=t}^N X_{it} Y_{it}}{\sum_{i=t}^N X_{it}^2} \right) = \beta + \frac{1}{T} \sum_{t=1}^T \left(\frac{\sum_{i=t}^N X_{it} \varepsilon_{it}}{\sum_{i=t}^N X_{it}^2} \right)\end{aligned}\quad (9)$$

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

$$S^2(\hat{\beta}_{\text{FM}}) = \frac{1}{T} \sum_{t=1}^T \frac{(\hat{\beta}_t - \hat{\beta}_{\text{FM}})^2}{T - 1}\quad (10)$$

The variance formula, however, assumes that the yearly estimates of the coefficient (β_t) are independent of each other. As we can see from equation (8), this is only true if $X_{it} \varepsilon_{it}$ is uncorrelated with $X_{is} \varepsilon_{is}$ for $t \neq s$. As I 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 will be too small in the presence of a firm effect

¹⁰ There are other differences between OLS estimates from pooled cross-sectional time-series and Fama-MacBeth regressions. Fama-MacBeth traditionally weights each year of data equally even if there is a different number of observations per year. Thus in an unbalanced panel data set, Fama-MacBeth and OLS coefficient estimates can differ due to a different weighting of observations (Cohen, Gompers, and Vuolteenaho, 2002, Vuolteenaho, 2002). In addition, since Fama-MacBeth runs cross sectional regressions, any variable which does not vary across firms within a year (e.g. the stock market return, or the growth of GDP) can not be estimated by the Fama-MacBeth method (Vuolteenaho, 2002, Cochrane, 2001). Since these have been dealt with elsewhere, I will not discuss them here.

(Cochrane, 2001). In the presence of a firm effect, the true variance of the Fama-MacBeth estimate is:

$$\begin{aligned}
\text{Var}(\hat{\beta}_{\text{FM}}) &= \frac{1}{T^2} \text{Var}\left(\sum_{t=1}^T \hat{\beta}_t\right) \\
&= \frac{\text{Var}(\hat{\beta}_t)}{T} + \frac{2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T \text{Cov}(\hat{\beta}_t, \hat{\beta}_s)}{T^2} \\
&= \frac{\text{Var}(\hat{\beta}_t)}{T} + \frac{T(T-1)}{T^2} \text{Cov}(\hat{\beta}_t, \hat{\beta}_s)
\end{aligned} \tag{11}$$

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 if $t \neq s$:

$$\begin{aligned}
\text{Cov}(\hat{\beta}_t, \hat{\beta}_s) &= E \left[\left(\sum_{i=1}^N X_{it}^2 \right)^{-1} \left(\sum_{i=1}^N X_{it} \varepsilon_{it} \right) \left(\sum_{i=1}^N X_{is} \varepsilon_{is} \right) \left(\sum_{i=1}^N X_{is}^2 \right)^{-1} \right] \\
&= (N \sigma_X^2)^{-2} E \left[\left(\sum_{i=1}^N X_{it} \varepsilon_{it} \right) \left(\sum_{i=1}^N X_{is} \varepsilon_{is} \right) \right] \\
&= (N \sigma_X^2)^{-2} E \left[\sum_{i=1}^N X_{it} X_{is} \varepsilon_{it} \varepsilon_{is} \right] \\
&= (N \sigma_X^2)^{-2} N \rho_X \sigma_X^2 \rho_\varepsilon \sigma_\varepsilon^2 \\
&= \frac{\rho_X \rho_\varepsilon \sigma_\varepsilon}{N \sigma_X^2}
\end{aligned} \tag{12}$$

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

$$\begin{aligned}
\text{Var}(\hat{\beta}_{\text{FM}}) &= \frac{\text{Var}(\hat{\beta}_t)}{T} + \frac{T(T-1)}{T^2} \text{Cov}(\hat{\beta}_t, \hat{\beta}_s) \\
&= \frac{1}{T} \left(\frac{\sigma_\varepsilon^2}{N \sigma_X^2} \right) + \frac{T(T-1)}{T^2} \left(\frac{\rho_X \rho_\varepsilon \sigma_\varepsilon^2}{N \sigma_X^2} \right) \\
&= \frac{\sigma_\varepsilon^2}{NT \sigma_X^2} (1 + (T-1) \rho_X \rho_\varepsilon)
\end{aligned} \tag{13}$$

This is same as our expression for the variance of the OLS coefficient (see equation 7). Thus the Fama-MacBeth estimated standard error will be too small in exactly the same cases as the OLS estimated standard error. In both cases, the magnitude of the under estimation will be a function of the 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 Error Estimates.

To document the bias of the Fama-MacBeth standard error estimates, I calculated the Fama-MacBeth estimate of the slope coefficient and the standard error in each of the 5,000 simulated data sets which were used above. 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 errors of the two estimates is the same estimates (compare the second entry in each cell of Table 1 and 2). Like the OLS standard error estimates, the Fama-MacBeth standard error estimates are biased downward (see Table 2).

The magnitude of the bias, however, is larger than implied by equation (12) and larger than the OLS bias. Moving down the diagonal of Table 2 from top left to bottom right, the true standard error increases but the standard error estimated by Fama-MacBeth falls. The standard error estimated by OLS equals 0.283 in all four cells. 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]. As the firm effect becomes larger ($\rho_x \rho_\varepsilon$ increases), the bias in the OLS standard error will grow, but the bias in the Fama-MacBeth standard error will grow even faster (see Figure 2). 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 10).

$$\begin{aligned} \text{Var}[\beta_{\text{FM}}] &= \frac{1}{T(T-1)} \sum_{t=1}^T \left[\frac{\sum_{i=t}^N X_{it} \varepsilon_{it}}{\sum_{i=t}^N X_{it}^2} - \frac{1}{T} \sum_{t=1}^T \left(\frac{\sum_{i=t}^N X_{it} \varepsilon_{it}}{\sum_{i=t}^N X_{it}^2} \right) \right]^2 \\ &= \frac{1}{T(T-1)} \sum_{t=1}^T \left[\frac{\sum_{i=t}^N (\mu_i + v_{it})(\gamma_i + \eta_{it})}{\sum_{i=t}^N (\mu_i + v_{it})^2} - \frac{1}{T} \sum_{t=1}^T \left(\frac{\sum_{i=t}^N (\mu_i + v_{it})(\gamma_i + \eta_{it})}{\sum_{i=t}^N (\mu_i + v_{it})^2} \right) \right]^2 \end{aligned} \quad (14)$$

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 affects both the yearly coefficients, and the average of the yearly coefficients, it does not appear in the estimated variance. Thus increases in the firm effect (increases in $\rho_x \rho_\varepsilon$) will 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. OLS will under estimate the standard errors by a factor of \sqrt{T} (the standard error estimated by OLS is $\sqrt{\sigma_\varepsilon^2 / NT \sigma_x^2}$ while the true standard error is $\sqrt{\sigma_\varepsilon^2 / N \sigma_x^2}$). The estimated Fama-MacBeth standard error will be zero. This additional source of bias will shrink as the number

of years increases since the estimate slope coefficient will converge to the true coefficient (see Figure 2).

The firm effect may be less important in regressions where the dependent variable is returns (and excess returns are serially uncorrelated) than in corporate finance applications where unobserved firm effects can be very important. The biases which I have highlighted will be less important in those applications. This isn't surprising since the Fama-MacBeth technique 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. In fact, Fama and MacBeth (1973) examine the serial correlation of the residuals in their results and find that it is close to zero. Its application in the literature, however, has not always been consistent with its roots. 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 is traditionally used to account for serial correlation of unknown form in the residuals of a single time series. It can be 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). This also simplifies the problem of choosing a lag length since the maximum lag length is one less than the maximum number of years per firm.¹¹ To examine the relative performance of the

¹¹ 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 (excluding 0). 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 and thus uses a maximum lag length (M) of $T-1$, would include $t-1$ lags

Newey-West and the robust/clustering approach to estimating standard errors. I simulated 5,000 data sets with 5,000 observations each. Each data set includes 500 firms and ten years of data per firm. The fixed firm effect was assumed to comprise twenty-five percent of the variability of both the independent variable and the residual.

The standard error estimated by the Newey-West will be an increasing function of the lag length in this 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. This is the same as the OLS standard error in my simulation, since I have assumed the errors are homoscedastic. Not surprisingly, this estimate significantly under estimates the true standard error (see Figure 3). As the lag length is increased from 0 to 9, the 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 3). In the presence of a fixed firm effect, an observation of a given firm is correlated with all other 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 (we return to non-permanent firm effects in Section V). However, even with the maximum lag length of 9, the Newey-West estimates have a small bias – underestimating the true standard error by 8% [$0.084 = 1 - 0.0328/0.0358$]. The robust standard errors under estimate the true standard error by less than 2%.

As the simulation demonstrates, the Newey-West approach to estimating standard errors, as applied to panel data, does not yield the same estimates as the Rogers standard errors. The difference between the two estimates is due to the weighting function used by Newey West. When estimating

and T-t leads for the t^{th} observation where t runs from 1 to T. Thus for the 4th year of data (t=4), we would include 3 lags and 6 leads [$\rho(\varepsilon_t \varepsilon_{t-3})$ to $\rho(\varepsilon_t \varepsilon_{t-6})$].

the standard errors, Newey-West multiplies the covariance term 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 Newey-West (when there is only one independent variable) is:

$$\begin{aligned}
\sum_{i=1}^N \left(\sum_{t=1}^T X_{it} \varepsilon_{it} \right)^2 &= \sum_{i=1}^N \left(\sum_{t=1}^T X_{it}^2 \varepsilon_{it}^2 + 2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T w(t-s) X_{it} X_{is} \varepsilon_{it} \varepsilon_{is} \right) \\
&= \sum_{i=1}^N \left(\sum_{t=1}^T X_{it}^2 \varepsilon_{it}^2 + 2 \sum_{t=1}^{T-1} \sum_{j=1}^{T-t} w(j) X_{it} X_{it-j} \varepsilon_{it} \varepsilon_{it-j} \right) \\
&= \sum_{i=1}^N \left(\sum_{t=1}^T X_{it}^2 \varepsilon_{it}^2 + 2 \sum_{t=1}^{T-1} \sum_{j=1}^{T-t} \left(1 - \frac{j}{T} \right) X_{it} X_{it-j} \varepsilon_{it} \varepsilon_{it-j} \right)
\end{aligned} \tag{15}$$

This is identical to the term in the Rogers standard error formula (see equation 7) except for the weighting function $[w(j)]$. The Rogers standard errors use a weighting function of one for all covariances. 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 with the sample size (T), then their estimate is consistent. However, in the panel data setting, the number of time periods is usually small. The consistency of the Rogers 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 are slightly smaller than the truth in a panel data setting.

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

To demonstrate how the two techniques work in the presence of a time effect I will generate data sets which contain only a time effect (observations on different firms with in the same year are

correlated). This is the data structure which the Fama-MacBeth approach was designed to address (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 have exchanged N and T.

A) Robust 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:

$$\begin{aligned} \varepsilon_{it} &= \delta_t + \eta_{it} \\ X_{it} &= \zeta_t + v_{it} \end{aligned} \tag{16}$$

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 affect to range from zero to seventy-five percent in twenty-five percent increments. The OLS coefficient and the true standard error as well as the OLS and robust standard error estimates 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 either 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 amount by which the OLS standard errors underestimate the true standard errors. When half of the variability in both comes from the time effect, the OLS standard errors underestimate the true standard errors by 91 percent [$0.909 = 1 - 0.0282/0.3105$, see Table 3].

The robust standard errors are much more accurate, but unlike our results with the firm effects, they also underestimate the true standard error. The magnitude of the underestimate is smaller, ranging from 13 percent [$1-0.1297/0.1490$] when the time effect comprises 25 percent of

the variability to 19 percent [$1-0.3986/0.4927$] when the time effect comprises 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) and ten years of data per cluster. When I estimated the standard errors in the presence of a time effect, I have only 10 years (clusters) and 500 firms per year. Since the robust standard errors method places no restriction on the correlation structure of the residuals within a cluster, its consistency depends upon having a sufficient number of clusters to estimate these standard errors. Based on these results, 10 clusters is too small and 500 is sufficient (see Kezdi, 2002, and Bertrand, Duflo, and Mullainathan, 2004 for similar results).

To explore this issue further, I simulated data sets of 5,000 observations but 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 are due to the time effect [i.e. $\sigma^2(\delta)/\sigma^2(\varepsilon) = \sigma^2(\zeta)/\sigma^2(X) = 0.25$]. The bias in the robust 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 4). Thus the bias in the robust standard errors estimates is a product of the small number of clusters. However, since panel data sets of 10 or 20 years are not uncommon in finance, this could be a problem in practice.

B) Fama-MacBeth Estimates

When there is only a time effect, the correlation of the estimated slope coefficients across the years will be zero and the standard errors estimated by Fama-MacBeth will thus be correct (see equation 9). This is exactly what we find in the simulation (see Table 4). The estimated standard errors are extremely close to the true standard errors and thus the confidence intervals will be the

correct size. In addition to producing unbiased standard error estimates, Fama-MacBeth also produces more efficient estimates than OLS.¹² For example, when 25 percent of the variability of both the independent variable and the residual is due to the time effect, the standard error of the Fama-MacBeth estimate is 81 percent $[1 - 0.0284/0.1490]$ smaller than the standard error 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 and is an estimate of the time effect. Since the variability due to the time effect is no longer in the residual, the residual variability in the Fama-MacBeth regressions is significantly smaller than in the OLS regression. 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.

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 robust standard errors are superior as they produce standard errors which are correct on average. If the data has only a time effect, the Fama-MacBeth estimates perform better than Rogers standard errors when there are few clusters (years). When the number of years is large, both the Rogers and Fama-MacBeth standard errors are correct. The Fama-MacBeth estimates are more efficient than the OLS coefficients, although as we will see below this advantage disappears if time dummies are included.

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

¹² The robust (White or Rogers) approach to estimating standard errors changes only the estimated standard errors. The coefficient estimates are numerically identical to OLS and thus have the same efficiency and variance as OLS.

Although the above results are instructive, they are unlikely to be completely descriptive of actual data confronted by empirical financial researchers. Most panel data sets will likely 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.¹³

The conceptual problem with using these techniques (Rogers 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 robust standard error approach allows us to be agnostic about the form of the correlation within a cluster. However, the price 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 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 Faulkendar, 2004, Sapienza, 2004, and Lamont and Polk, 2001). I will use this approach in our simulations.

A) Rogers Standard Error Estimates.

To test the relative performance of the two methods, 5,000 data sets were simulated with

¹³ Since we know that both OLS and Fama-MacBeth do well when the firm and time effect is zero in either the independent variable or the residual, I will not examine these cases.

¹⁴ It is possible to estimate robust 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, they could cluster by industry and year (i.e. each cluster would be a specific year and industry). 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 OLS.

both a firm and a time effect. Across the simulations, the fixed firm effect comprises 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 scenario where fifty percent of the variability is due to the firm effect and fifty percent is due to the time effect is excluded, as this would allow no remaining variability in the firm-year specific component. The same three scenarios were used for simulating the residual which generated nine different 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 under estimate the true standard errors whereas the Rogers (clustered by firm) standard errors are consistently accurate 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. The magnitude of the time effect does, however, appear to effect the magnitude of the bias in the OLS estimates, but this effect is subtle. To see this intuition, it is useful to examine a couple of examples. In Table 1, when the firm effect comprises 25 percent of the variability of both the independent variable and the residual, OLS under estimated the standard error by 20 percent [$1-0.0283/0.0353$, see Table 1]. In Table 5, there are two scenarios where the fixed firm effect is 25 percent for both the independent variable and the residual. When the magnitude of the time effect rises to 25 percent, the bias in OLS rises to 31 percent [$1-0.0283/0.0407$, see Table 5], and when the magnitude of the time effect rises to 50 percent, the bias in the OLS standard error is 45 percent [$1-0.0283/0.0515$, see Table 5]. The dummy variables, by absorbing variability from the residual and the independent variable which is due to the time effect, raise the fraction of the remaining variability which is due to the firm effect

(i.e. ρ_x or ρ_e rise). This increases the bias in the OLS standard errors.¹⁵

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, which are effectively included in the Fama-MacBeth estimates, 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. However, the standard errors estimated by Fama-MacBeth are again too small, as we found in the absence of a time effect (Table 2). As an example, when 25 percent of the variability in both the independent variable and the residual comes from the firm effect and 25 percent comes from a time effect, the Fama-MacBeth standard errors under estimate the true standard errors by 37 percent $[1-0.0258/0.0407]$.

Thus most of the intuition from the earlier tables carry over. In the presence of a fixed firm effect both OLS and Fama-MacBeth standard error estimates are biased down significantly. Rogers standard errors which account for clustering by firm produce estimates which are correct on average. The presence of a 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 non-Fixed Firm Effect

¹⁵ The standard errors reported in Table 5 are very close to what is implied by the equations in Section II. Once I adjust the definition of ρ_x and ρ_e to equal the fraction of variability due to the firm effect after removing the time effect, the OLS standard errors are very close to those produced by equation (3) and the Rogers standard errors are very close to those produced by equation (7).

The analysis thus far has assumed that the firm effect is fixed. Although this is common in the literature, it may not always be accurate. 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 fixed effect, researchers need to know this.

A) Non-Fixed 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 which I assumed was 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) follows a first order auto regressive process:

$$\begin{aligned} \eta_{it} &= \zeta_{it} && \text{if } t = 1 \\ &= \varphi \eta_{it-1} + \sqrt{1 - \varphi^2} \zeta_{it} && \text{if } t > 1 \end{aligned} \quad (17)$$

Thus φ is the first order auto correlation between η_{it} and η_{it-1} , and the correlation between η_{it} and η_{it-k} is φ^k .¹⁶ Combining this term with the fixed firm effect (γ_i in equation 4), means the serial correlation

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

$$\begin{aligned} \text{Var}(\eta_{it}) &= \sigma_{\zeta}^2 && \text{if } t = 1 \\ &= \varphi^2 \sigma_{\zeta}^2 + (1 - \varphi^2) \sigma_{\zeta}^2 = \sigma_{\zeta}^2 && \text{if } t > 1 \end{aligned} \quad (18)$$

where the last step is by recursion (if it is true for $t=k$, it is true for $t=k+1$). Assuming homoscedastic residuals is not necessary since the Rogers standard errors are robust to heteroscedasticity. However, assuming homoscedasticity makes the interpretation of the results simpler. If I assume the residuals are homoscedastic, then any difference in the standard

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. An analogous data structure is specified for the independent variable. The correlation of lag length k is:

$$\begin{aligned}
\text{Corr}(\varepsilon_{i,t}, \varepsilon_{i,t-k}) &= \frac{\text{Cov}(\gamma_i + \eta_{i,t}, \gamma_i + \eta_{i,t-k})}{\sqrt{\text{Var}(\gamma_i + \eta_{i,t}) \text{Var}(\gamma_i + \eta_{i,t-k})}} \\
&= \frac{\sigma_\gamma^2 + \varphi^k \sigma_\eta^2}{\sigma_\gamma^2 + \sigma_\eta^2} \\
&= \rho_\varepsilon + (1 - \rho_\varepsilon) \varphi^k
\end{aligned} \tag{18}$$

The correlations for lags one through nine for the four data specifications I will examine are graphed in Figure 5. They range from the standard fixed firm effect ($\rho=0.25$ and $\varphi=0.00$) to a standard AR1 process ($\rho=0.00$ and $\varphi=0.75$). I have assumed the same process for both the independent variable and the residual, since as we know from Section II, if there is no within cluster dependence in the independent variables, OLS standard errors are correct.

B) Fixed Effects – Firm Dummies.

The one remaining approach used in the literature for addressing within cluster dependence in the residuals, but which I have not yet considered, is the use of fixed effects or firm dummies. A significant minority of the papers used fixed effects to control for dependence within a cluster. Using the simulations, I can compare the relative performance of OLS and Rogers standard errors both with and without firm dummies. The results are reported in Table 7, Panel A, column I.

The fixed effect estimates are more efficient than in this case (0.0299 versus 0.0355). This

errors I find is due to the dependence of observations within a cluster opposed to the presence and magnitude of the heteroscedasticity.

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 estimates. However, the firm dummies also eliminate the within cluster dependence of the independent variable and the residual (if the firm effect is fixed) 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 have included the firm effects, the OLS standard error are now correct and robust standard errors, such as the Rogers's, are not necessary (see Table 7 - Panel A, column I). The Rogers standard errors are correct when we do not include the fixed effects and are slightly too large (5%) when we include the fixed effects (see Kezdi (2002) 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. To show this, I ran three additional simulations (see Table 7 - Panel A, columns II-IV). In these simulations, the firm effect does decay over time (in column II, 92 percent of the firm effect dissipates after 9 years). Once the firm effect is not permanent, the OLS standard errors again under estimate the true standard errors, even when firm dummies are included in the regression. The magnitude of the under estimation 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). The Rogers standard errors are much closer to the truth, but consistently over estimate the true standard error by about 5 percent across the simulations.

C) Adjusted Fama-MacBeth Standard Errors.

As noted above, the presence of a firm effect cause the Fama-MacBeth yearly coefficient

estimates to be correlated and this causes the Fama-MacBeth standard error to be biased downward. Several authors who have used the Fama-MacBeth approach have acknowledged the downward bias and have suggested adjusting the standard errors for the estimated first order auto-correlation of the estimated slope coefficients (Chen, Hong, and Stein, 2001; Cochrane, 2001; Lakonishok, and Lee, 2001; Fama and French, 2002; Bakshi, Kapadia, and Madan, 2003; Chakravarty, Gulen, and Mayhew, 2004). The proposed adjustment is to estimate the correlation between the yearly coefficient estimates (i.e. $\text{Corr}[\beta_t, \beta_{t-1}] = \theta$), and then multiply the estimated variance by $(1+\theta)/(1-\theta)$ to account for serial correlation of the β s (see Chakravarty, Gulen, Mayhew, 2004 and Fama and French, 2002, especially footnote 1). This makes intuitive sense since the presence of a firm effect will cause the yearly coefficient estimates to be serially correlated.

To test the merits of this idea, I simulated data sets where the fixed firm effect comprised 25 percent of the variance. For each simulated data set, ten slope coefficients were estimated, and the auto correlation of the slope coefficients was calculated. I then calculated 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 estimated autocorrelation is imprecisely estimated as predicted by Fama and French (2002). The 90th percentile confidence interval ranges from -0.602 to 0.413, 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 smaller and more biased than the

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

$$\text{Variance correction} = \left(1 + 2 \sum_{k=1}^{10-1} (10 - k) \theta^k \right) \quad (19)$$

unadjusted standard errors.

The intuition for why the proposed adjustment does not work is subtle. The problem is the correlation which is being estimated (the within sample serial correlation of the yearly coefficient estimates) is not the same as the one which is causing the bias in the standard errors (the correlation of betas across data sets). The co-variance which biases the standard errors and which I estimate across the 5,000 simulations is

$$\text{Cov}(\beta_t, \beta_{t-1}) = E[(\beta_t - \beta_{\text{True}})(\beta_{t-1} - \beta_{\text{True}})] \quad (19)$$

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 both 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 9). 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 samples is the true β or one in the simulations. Thus when I estimate the 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 are given only one data set. Thus they must calculate the serial correlation of the β s within the sample they are given. This co-variance is calculated as:

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, because 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 is what we found in Table II.

$$\text{Cov}(\beta_t, \beta_{t-1}) = E[(\beta_t - \bar{\beta}_{\text{Within sample}})(\beta_{t-1} - \bar{\beta}_{\text{Within sample}})] \quad (20)$$

The within sample serial correlation measures the tendency of β_t to be above its within sample mean when β_{t-1} is above its within sample mean. To see how the presence of a fixed firm effect affects 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 β s (the Fama-MacBeth coefficient estimate) by the same amount. Thus a fixed firm effect will no influence the deviation of any β_t from the average β . Since this deviation is the source of the estimated within sample serial correlation, we should expect that the serial correlation calculated in sample would be zero on average.¹⁹ Since the within sample correlation is asymptotically zero, adjusting the standard errors based on this estimated serial correlation will still lead to biased standard error estimates.

The adjusted Fama-MacBeth standard errors do better when there is an auto-regressive component in the residuals (i.e. $\phi > 0$). In the three remaining simulations in 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 to the truth when the firm effect is not fixed ($\rho=0$). In this case, the standard errors based on the infinite period adjustment under estimate the true standard error by 23 percent ($1-0.0374/0.0484$). 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

¹⁹ The within sample serial correlation we 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 is closer to zero, rising from -0.1134 to -0.0556.

they still significantly under estimate the true standard errors when a portion of the firm effect is permanent (i.e. $\rho > 0$).

VI) Conclusions.

It is well known from first year econometrics classes that OLS standard errors are biased when the residuals are not independent. How financial researchers should estimate standard errors when using panel data sets has been less clear. The empirical literature has proposed and used a variety of methods for estimating standard errors when the residuals are correlated across firms or years in the data. In this paper, I find that the performance of the different methods varies and their relative accuracy depends upon the nature of the dependence in the data.

Since Fama-MacBeth estimation was designed for a setting where residuals were correlated within a year, but not across firms, it does well in this context. It produces estimates which are more efficient than OLS estimates, and standard errors which are as good as Rogers standard errors when the number of clusters is large, and better when the number of clusters is small.

The Rogers standard errors produce more accurate standard errors in the presence of a firm effect. In addition these estimates are very robust to different specifications of the dependence. The Rogers estimates produce correct standard errors and correctly sized confidence intervals in the presence of time effects (if time dummies are included) and in the presence of a firm effect which is not constant. It is in the later case, that the Rogers standard errors are superior to a fixed effect model. Since the precise form of the dependence in the residual and the independent variables is often not known, an estimate which is robust to different specifications is an advantage.

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.
- 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
- Cochrane, John, 2001, *Asset Pricing*, Princeton University Press, Princeton, NJ.
- Cohen, Randolph, Paul Gompers, and Tuomo Vuolteenaho, 2002, "Who underreacts to cash-flow news? evidence from trading between individuals and institutions" *Journal of Financial Economics* 66, 409-462.
- 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 .
- Fama, E., 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.
- Faulkender, Mike, and Mitchell Petersen, 2004, "Does the Source of Capital Affect Capital Structure?" Northwestern working paper.
- Gross, David and Nicholas Souleles, 2004, "An Empirical Analysis of Personal Bankruptcy and Delinquency" *Review of Financial Studies* 15, 319-347.
- Kezdi, Gabor, 2002, "Robust Standard Error Estimation in Fixed-Effects Panel Models, University of Michigan working paper.

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.

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.

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.

Vuolteenaho, Tuomo, 2002, "What Drives Firm-Level Stock Returns?" *Journal of Finance* 57, 233-264.

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

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	1.0006	1.0002	1.0001
		0.0286	0.0288	0.0279	0.0283
		0.0283	0.0283	0.0283	0.0283
		0.0283	0.0282	0.0282	0.0282
	25%	1.0004	0.9997	0.9999	0.9997
		0.0287	0.0353	0.0403	0.0468
		0.0283	0.0283	0.0283	0.0283
		0.0283	0.0353	0.0411	0.0463
	50%	1.0001	1.0002	1.0007	0.9993
		0.0289	0.0414	0.0508	0.0577
		0.0283	0.0283	0.0283	0.0283
		0.0282	0.0411	0.0508	0.0590
	75%	1.0000	1.0004	0.9995	1.0016
		0.0285	0.0459	0.0594	0.0698
		0.0283	0.0283	0.0283	0.0283
		0.0282	0.0462	0.0589	0.0693

Notes:

The table contains estimates of the coefficient and standard errors based on 5000 simulation of a panel data set (10 years per firm and 500 firms). 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 is varied across the rows of the table and varies from 0% (no firm effect) to 75%. The fraction of the independent variable's variance which is due to a firm specific component also varies across the columns of the table and varies 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 OLS estimated standard error of the coefficient. The fourth entry is Rogers' (clustered) standard error which accounts for possible clustering at the firm level (i.e. accounts for the possible correlation between observations of the same firm in different years).

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 5000 simulation of a panel data set (10 years per firm and 500 firms). 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 is varied across the rows of the table and varies from 0% (no firm effect) to 75%. The fraction of the independent variable's variance which is due to a firm specific component is varied across the columns of the table and varies from 0% (no firm effect) to 75%. The first entry is the average estimated slope coefficient based on a Fama-MacBeth estimation (e.g. we ran the regression for each of the 10 years and took the average). 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 standard error estimated by the Fama-MacBeth procedure.

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	1.0002	1.0006	0.9994
		0.0286	0.0291	0.0293	0.0314
		0.0283	0.0288	0.0295	0.0306
		0.0277	0.0276	0.0275	0.0270
	25%	1.0006	1.0043	0.9962	0.9996
		0.0284	0.1490	0.2148	0.2874
		0.0279	0.0284	0.0289	0.0300
		0.0268	0.1297	0.1812	0.2305
	50%	0.9996	0.9997	0.9919	1.0079
		0.0276	0.2138	0.3015	0.3986
		0.0274	0.0278	0.0282	0.0292
		0.0258	0.1812	0.2546	0.3248
	75%	1.0002	0.9963	0.9970	0.9908
		0.0273	0.2620	0.3816	0.4927
		0.0267	0.0271	0.0276	0.0284
		0.0244	0.2215	0.3141	0.3986

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set (10 years per firm and 500 firms). 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 standard error estimated by OLS. The fourth entry is Rogers' (clustered) standard error which accounts for possible clustering 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 (10 years per firm and 500 firms). 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 standard error estimated by the Fama-MacBeth procedure (e.g. equation 10).

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		
				25%	50%	25%
Residual Volatility from Firm Effect	25%	Residual Volatility from Time Effect	25%	0.9997	1.0004	1.0004
			0.0407	0.0547	0.0489	
	0.0283		0.0347	0.0283		
	0.0400		0.0548	0.0489		
	25%	50%	1.0005	1.0015	0.9993	
			0.0362	0.0515	0.0468	
50%	25%	0.0231	0.0283	0.0231		
		0.0364	0.0508	0.0461		
	50%	1.0002	1.0008	0.9994		
		0.0493	0.0690	0.0631		
0.0283	0.0347	0.0283				
0.0490	0.0692	0.0630				

Notes:

The table contains estimates of the coefficient and standard errors based on 1,000 simulation of a panel data set with 5,000 observations (10 years per firm and 500 firms). 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. 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 standard error estimated from OLS. The fourth entry is Rogers' (clustered) standard error which accounts for possible clustering at the firm level (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 (10 years per firm and 500 firms). 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. I ran the regression for each of the 10 years and took the average). 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 standard error estimated by the Fama-MacBeth procedure (e.g. equation 10).

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
$\rho_X / \rho_\varepsilon$	0.25 / 0.25	0.00 / 0.00	0.25 / 0.25	0.50 / 0.50
$\varphi_X / \varphi_\varepsilon$	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-AR1})	I	II	III	IV
$\rho_X / \rho_\varepsilon$	0.25 / 0.25	0.00 / 0.00	0.25 / 0.25	0.50 / 0.50
$\varphi_X / \varphi_\varepsilon$	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(1 st order auto-correlation)	-0.1134	0.2793	0.3250	0.1759

Notes:

The table contains estimates of the coefficient and standard errors based on 1,000 simulation of a panel data set with 5,000 observations (10 years per firm and 500 firms). 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 (see equation 6). φ_x (φ_ε) is the first order auto correlation of the non-fixed portion of the firm effect of the independent variable (the residual). Combining equations (6) with equations (15?) and (16), the residual is specified as:

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

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) was included; in the second row 499 firm dummies (for 500 firms) were 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 standard error estimated from OLS. The fourth entry is Rogers' (clustered) standard error which accounts for possible clustering at the firm level (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 standard error estimated by the Fama-MacBeth procedure, assuming the yearly estimates are independent. The last two entries are the Fama-MacBeth standard error estimate corrected for first order auto-correlation. The fourth entry assumes an infinite lag (i.e. multiplied by the square root of $(1+\varphi)/(1-\varphi)$), 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 $(T-k) \varphi^k$). The bottom row contains the average across the 5,000 simulation of the first order autocorrelation of β_t and β_{t-1} estimated within each of the 5,000 samples.

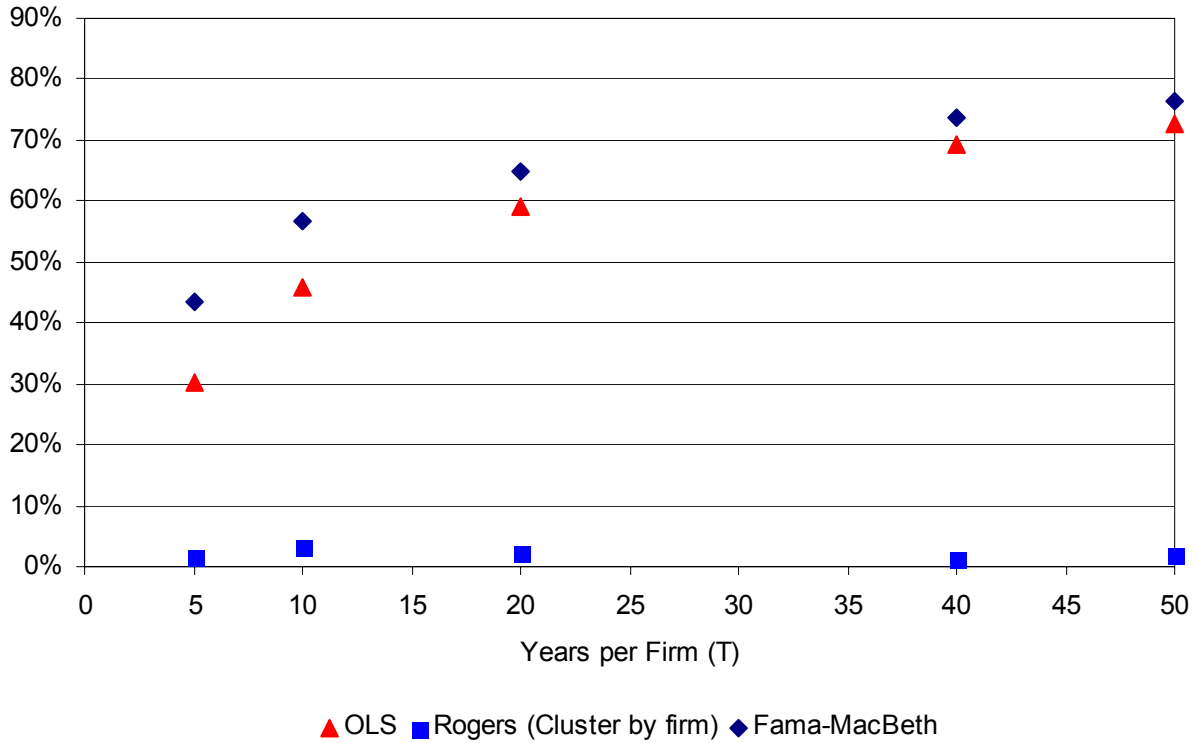
Figure 1: Residual Cross Product Matrix
Assumptions About Zero Co-variances

	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 only the NT diagonal terms are non-zero. The cluster assumption assumes that the correlation of the residuals within the cluster may be non-zero (these elements are shaded). Thus there are T^2 unique variances and co-variances to estimate and N observations of each variance or covariance. The cluster assumption assumes that residuals across clusters are uncorrelated. These are recorded as zero in the above matrix.

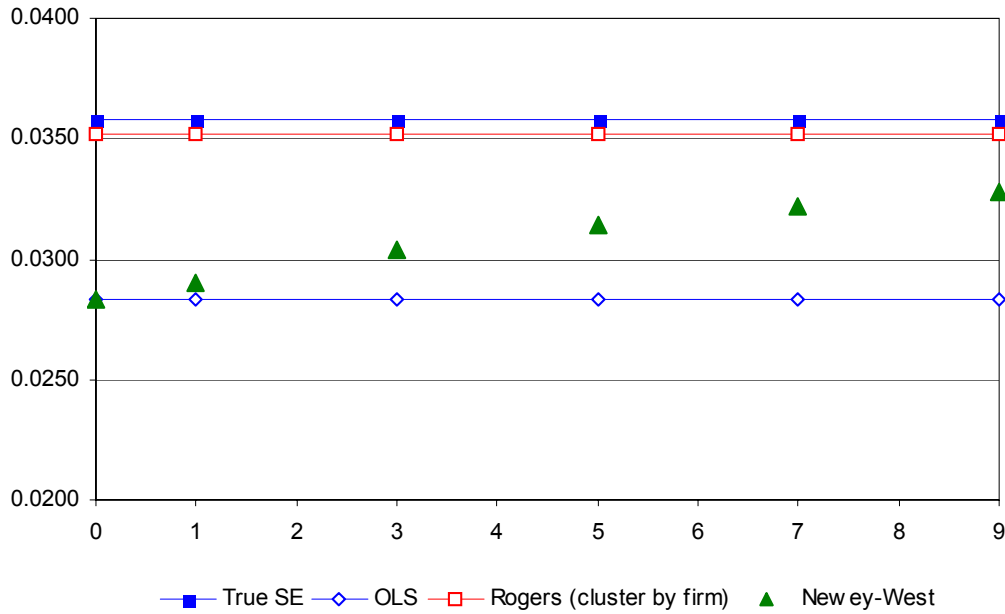
Figure 2: Bias in Estimated Standard Errors
As a function of observations per cluster



Notes:

The figure graphs the percentage by which the three methods under estimate 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 under estimates are calculated as one minus the average standard error estimated by the method 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 under estimated the true standard error by 30% ($1 - 0.0283/0.0406$). The under estimates are graphed as squares for the Rogers' (clustered by firm) standard errors, triangles for the OLS standard errors, and diamonds for the Fama-MacBeth standard errors.

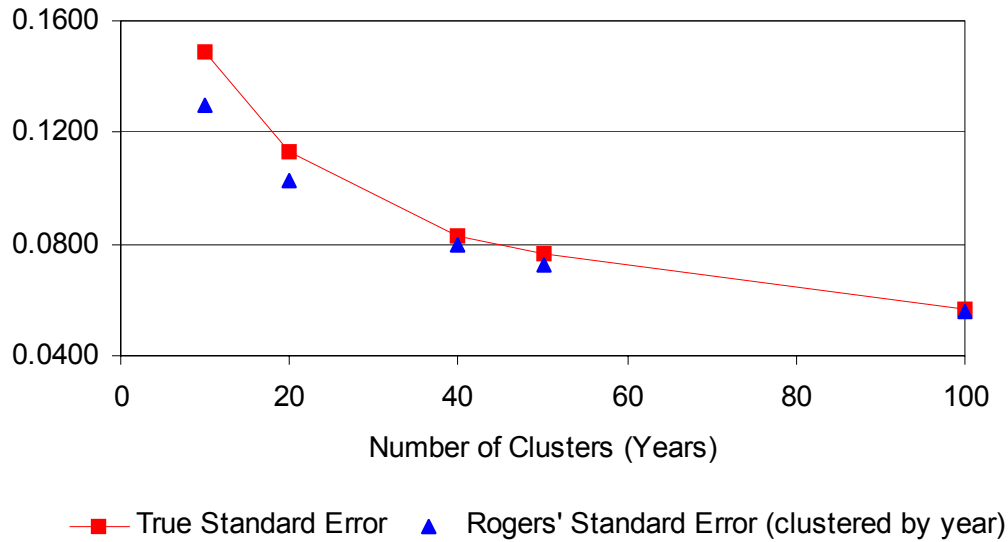
Figure 3: Relative Performance of OLS, Rogers, and Newey-West Standard Errors



Notes:

The figure contains OLS, Rogers (clustered by firm), and Newey-West standard error estimates. The estimates are based on 5,000 simulated data sets. Each data set contains 5,000 observations (500 firms and 10 years for each 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. $\sigma^2(\gamma)/\sigma^2(\varepsilon) = \sigma^2(\mu)/\sigma^2(X) = 0.25$]. The true standard error (filled in squares), the OLS standard error (empty diamonds), and the Rogers' standard error (empty squares) are plotted as straight lines since they do not depend upon the assumed lag length. The Newey-West standard errors, which rise with the assumed lag length, are plotted as triangles.

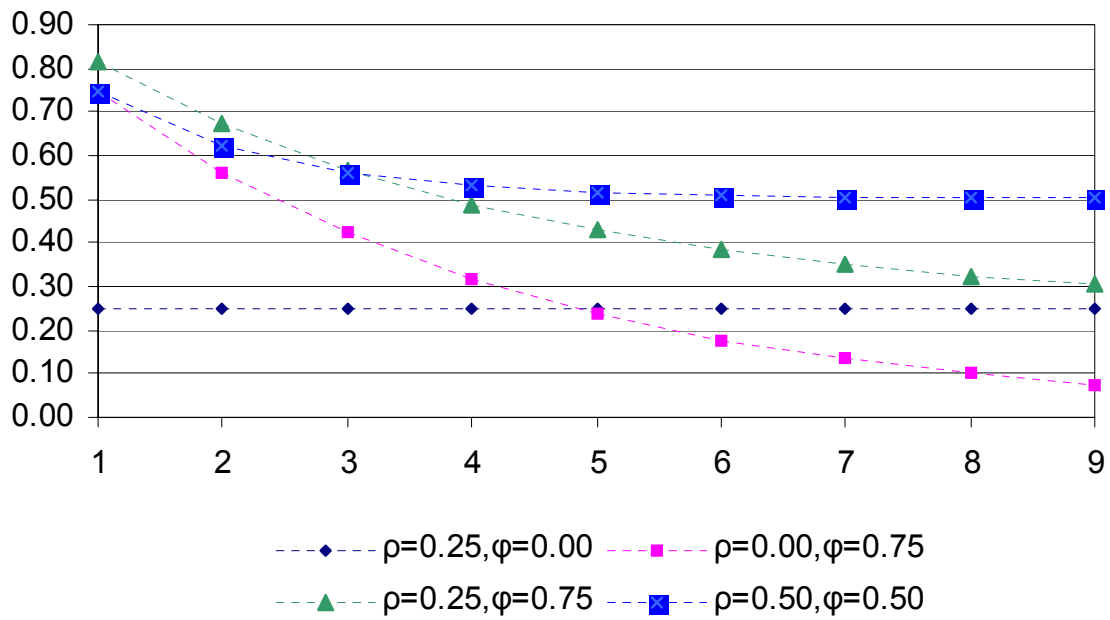
Figure 4: True Standard Errors and Robust Estimates as a function of cluster size (T)



Notes:

The true standard errors (squares) and the Roger’s (clustered by year) standard errors (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. $\sigma^2(\delta)/\sigma^2(\varepsilon) = \sigma^2(\zeta)/\sigma^2(X) = 0.25$]. The robust estimates 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 5: 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. ϕ is the fraction of the variance which is due to the fixed firm effect and ρ is the first order auto-correlation of the non-fixed firm effect.