Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research

Ian D. Gow
i-gow@kellogg.northwestern.edu
Kellogg School of Management
Northwestern University

Gaizka Ormazabal
gaizkao@stanford.edu
Graduate School of Business
Stanford University

Daniel J. Taylor*
djtaylor@stanford.edu
Graduate School of Business
Stanford University

August 4, 2009

*Corresponding author. We thank Bill Beaver, Mary Barth, John Core, Alan Jagolinzer, Dave Larcker, Xuesong Hu (the AAA discussant), Greg Sommers (the FARS discussant), two anonymous referees, Steven Kachelmeier, Mark Trombley (editor), and seminar participants at the 2008 AAA Annual Meeting, the 2009 FARS Mid-year Conference, and Stanford University, for comments and suggestions. We also thank Jennifer Francis and Per Olsson for providing their estimates of the monthly AQ factor, and Mary Barth and Steve Stubben for providing their estimates of the implied cost of capital. Daniel Taylor gratefully acknowledges funding from the Deloitte Foundation.

Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research

Abstract:

We review and evaluate the methods commonly used in the accounting literature to correct for cross-sectional and time-series dependence. While much of the accounting literature studies settings in which variables are cross-sectionally and serially correlated, we find that the extant methods are not robust to both forms of dependence. Contrary to claims in the literature, we find that the Z2 statistic and Newey-West corrected Fama-MacBeth standard errors do not correct for both cross-sectional and time-series dependence. We show that extant methods produce misspecified test statistics in common accounting research settings, and that correcting for both forms of dependence substantially alters inferences reported in the literature. Specifically, several findings in the implied cost of equity capital literature, the cost of debt literature, and the conservatism literature appear not to be robust to the use of well-specified test statistics.

JEL Classification: C12; C15; C23; M41

Keywords: cross-sectional dependence, time-series dependence, autocorrelation, Fama-MacBeth, Newey-West, cluster-robust standard errors, implied cost of capital, credit ratings, conservatism.

Data availability: Data are available from the sources indicated in the text.

I. INTRODUCTION

Much of the empirical accounting literature uses panel data sets, typically repeated observations on the same, or a substantially overlapping, set of firms over time. In these data sets the variables of interest are often both cross-sectionally and serially correlated (e.g., accounting items, audit fees, executive compensation, implied cost of capital, governance, and microstructure variables). As such, the common assumption of independence in regression errors is generally violated. While it is well known that residual dependence can result in misspecified test statistics (Bernard 1987), prior methodological work in the accounting literature focuses exclusively on cross-sectional dependence and does not examine the issues created by the presence of both cross-sectional and time-series dependence. This omission results in an important gap in the literature, as methods that correct for dependence in one direction typically assume independence in the other (e.g., Fama-MacBeth). As a result, we find that typical accounting studies either ignore one or both forms of dependence, or rely on methods developed within the accounting literature that have not been formally evaluated (e.g., aggregating firm- or industry-specific coefficients, Z2 statistic, and Newey-West corrected Fama-MacBeth standard errors).

In the 1980s, a significant stream of research examined the impact that cross-sectional dependence can have on test statistics and inferences in accounting research (Bernard 1987). However, since that time, accounting studies increasingly rely on panel data in which both cross-sectional *and* time-series dependence are present. To address these forms of dependence, a number of advances have been made in the econometrics literature (e.g., Newey-West and cluster-robust standard errors) and in the accounting literature (e.g., aggregating firm- or industry-specific coefficients, the Z2 statistic, and Newey-West corrected Fama-MacBeth standard errors). While the accounting literature now uses a number of methods to correct for cross-sectional and time-series dependence, several of these methods (e.g., the Z2 and Newey-West corrected Fama-MacBeth statistics) are not examined in prior research either within accounting or in neighboring fields. As a result, there is little guidance as to when each method is appropriate, such that the various approaches are inconsistently and often incorrectly applied. In this study, we review and evaluate the extant methods and show that they produce misspecified test statistics with significant consequences for reported inferences.

To understand the effect of cross-sectional and/or time-series dependence on inferences in accounting research, we examine four active streams of empirical accounting research in which panel data is frequently used and cross-sectional and/or time-series dependence is likely. In particular, we focus on: (1) asset pricing, where realized returns are the dependent variable, (2) cost of equity capital, where ex ante estimates of the cost of equity capital are the dependent variable, (3) cost of debt, where credit ratings are the dependent variable, and (4) conservatism, where the dependent and independent variables are aggregated over multiple years. We focus on broad research settings to show that our findings are not specific to individual studies.

We examine 121 studies that use cross-sectional and time-series data in their regression tests, including all such studies that appear in *The Accounting Review*, *Journal of Accounting and Economics*, or *Journal of Accounting Research* during the years 2002-2006. Notwithstanding the possibility of cross-sectional and/or time-series dependence, 25% (30) of these studies do not appear to address such dependence; in particular, tabulated test statistics are based on either "regular" OLS or White (1980) standard errors with no reported checks for robustness to cross-sectional or time-series dependence.

The remaining 75% (91) of these studies attempt to address cross-sectional and time-series dependence using a variety of approaches. Three approaches from econometrics and finance, Newey-West (N-W), Fama-MacBeth (FM-t) and one-way cluster-robust standard errors, are common in accounting research. These approaches address either cross-sectional or time-series dependence, but not both (see Petersen 2009). Three other methods have been developed by accounting researchers, but are not evaluated in prior research. The first, which we label FM-i, varies the FM-t approach by estimating a time-series regression for each cross-sectional unit (e.g., each firm) and basing inferences on the mean and standard deviation of the resulting coefficients.[1] The second approach, FM-NW, modifies the FM-t approach by adjusting the standard errors by applying the Newey-West procedure to the time-series of coefficient estimates. While intuitively appealing, the robustness of FM-NW to both cross-sectional and time-series dependence is not examined in prior research. The third approach involves calculating a statistic, typically labeled "Z2," based on the mean and standard deviation of *t*-statistics from either *T* cross-sectional regressions (Z2-t) or *N* firm- or industry-specific regressions (Z2-i). Like FM-NW, Z2 is frequently claimed to adjust for both forms of dependence, but its ability to do so is not formally evaluated in prior research.

In addition to methods applied in accounting research, we also consider two-way cluster-robust standard errors (CL-2). While the econometric literature shows that CL-2 is robust to *both* time-series and cross-sectional correlation (Thompson 2006; Cameron et al. 2009; Petersen 2009), we find only one study appearing in *The Accounting Review*, *Journal of Accounting and*

[1] In contrast, FM-t involves estimation of cross-sectional regressions for each period and basing inferences on the mean and standard deviation of those estimated coefficients.

Economics, and the *Journal of Accounting Research* that uses two-way cluster robust standard errors.[2] In contrast to Petersen's finding that "clustering standard errors by both firm and time appears unnecessary" in the finance applications he considers (2009, 473), we find that in a variety of accounting-specific applications (e.g., conservatism), CL-2 is necessary to produce valid inferences.

By evaluating the methods used to correct for cross-sectional and time-series dependence in the accounting literature, this study makes a number of contributions. First, we are the first to formally evaluate several common methods that are specific to the accounting literature, namely FM-i, Z2, and FM-NW. We find that 30% of all studies appearing in our survey use one of these three methods.

Second, we show that correcting for both cross-sectional and time-series dependence substantially affects inferences reported in the accounting literature. We examine how the use of these methods affects inferences in common accounting research settings including: (1) earnings quality, (2) cost of capital, (3) cost of debt, and (4) conservatism. We show that the accounting-specific methods, namely FM-i, Z2, and FM-NW, are not robust to both forms of dependence and produce misspecified test statistics in common applications. In particular, we find that FM-i produces substantially overstated *t*-statistics and rejects a true null hypothesis more than 80% of the time at the 1% level in the setting where it is most commonly used. We find that Z2 is highly correlated with Fama-MacBeth *t*-statistics, producing Type I error rates that are nearly identical to Fama-MacBeth, rejecting a true null hypothesis more than 30% of the time at the 1% level in

[2] Chen et al. (2008) cluster on both firm and year. While an additional ten studies in these journals cite Petersen (2008), none uses two-way cluster robust standard errors, consistent with Petersen's finding that two-way cluster-robust standard errors are not required in the finance settings he considers.

the setting where it is most commonly used. In contrast to claims in the literature, we find no evidence that FM-NW addresses time-series dependence.

We document that recent findings in the cost-of-capital literature relating to earnings quality, idiosyncratic risk, and governance are not robust to the use of well-specified test statistics. Additionally, we examine the determinants of credit ratings, a measure of the cost of debt, and provide evidence that recent findings in this literature related to earnings quality, research and development, and governance are also not robust to the use of well-specified test statistics. Finally, we examine studies in the conservatism literature that estimate conditional conservatism using the approach of Roychowdhury and Watts (2007), which aggregates returns and earnings over overlapping windows. We show that several findings in this literature are not robust to corrections for both the time-series and cross-sectional dependence that is induced by aggregating variables over overlapping windows. Further, and in contrast to claims in the literature, our analysis suggests that the increased *t*-statistics reported when using the Roychowdhury and Watts (2007) approach are not the result of increased power, but rather the result of biased standard errors.

Section II discusses the related literature. Section III discusses the various methods used in the accounting literature to control for residual dependence. Section IV presents our simulation evidence. Section V examines four applications to evaluate the robustness of results in the literature to controlling cross-sectional and time-series dependence. Section VI summarizes our recommendations for empirical researchers and concludes.

II. RELATED LITERATURE

Our study relates to streams of research in both accounting and econometrics. Prior methodological work in accounting research related to issues of residual dependence focuses exclusively on cross-sectional dependence, with particular emphasis on event studies. The general thrust of this literature is the use of feasible generalized least squares (FGLS) instead of OLS (e.g., Schipper and Thompson 1983; Collins and Dent 1984; Sefcik and Thompson 1986). However, as Bernard (1987) points out, FGLS is not feasible in many settings of interest to accounting researchers due to the limited length of the time-series relative to the number of covariance terms to be estimated. Consistent with this point, FGLS is rarely used in current research. Bernard (1987, 41) does not identify approaches to correct inferential statistics and concludes that "when OLS is used in such studies, inferential statistics should be interpreted cautiously." While cross-sectional and time-series dependence in panel data is common in many settings, to our knowledge no study in the accounting literature examines both forms of dependence in detail. This is an important gap in the literature, as methods that address dependence in one direction typically assume independence in the other. As a result, we find that prior studies either ignore this issue or rely on approaches that have not been formally evaluated—in the extreme relying on methods that we show produce misspecified test statistics.

We draw on a recent literature in econometrics that has produced variance estimators robust to both cross-sectional and time-series dependence. White (1984) proposes cluster-robust standard errors that are robust to heteroskedasticity and permit within-cluster correlation (e.g., within firm). These standard errors are common in the economics literature and are covered in standard econometrics textbooks (e.g., Wooldridge 2002; Greene 2003). Recent work by

Thompson (2006) and Cameron et al. (2009) shows that cluster-robust standard errors can be generalized to more than one dimension of clustering (e.g., within firm and within year), and Cameron and Trivedi (2006) show that cluster-robust standard errors are consistent for *m*-estimators, a wide class of estimators that includes logit and tobit. Petersen (2009) uses simulations to evaluate the performance of CL-2 relative to OLS, CL-i, and CL-t in the presence of both cross-sectional and time-series dependence and considers two finance applications, asset pricing and capital structure. He concludes that in these settings "clustering standard errors by both firm and time appears unnecessary" (2009, 473). In contrast, we find that, in a variety of accounting applications, two-way cluster-robust standard errors are required for valid inferences. The difference between our findings and those of Petersen is due to the fact that accounting variables (e.g., earnings and credit ratings) exhibit greater dependence both over time and in cross-section than finance variables (e.g., returns). Our study contributes to the literature by considering common methods specific to accounting research and not evaluated in prior work (e.g., FM-i, FM-NW, and Z2) and by examining the performance of all methods (including CL-2) in accounting research settings.

III. INFERENCE IN CURRENT ACCOUNTING RESEARCH

Accounting researchers use a variety of approaches to address issues of cross-sectional and time-series dependence that are widely understood to be present in common research settings. Our literature survey identified a number of common approaches: Fama-MacBeth, Newey-West, the Z2 statistic, and standard errors clustered by firm, industry, or time. In this section, we review each of these approaches and discuss the circumstances in which they produce valid inferences.

OLS and White standard errors

OLS standard errors assume that errors are both homoskedastic and uncorrelated across observations. While White (1980) standard errors are consistent in the presence of heteroskedasticity, it is well known that both OLS and White produce misspecified test statistics when either form of dependence is present.

Newey-West

Newey and West (1987) generalize the White (1980) approach to yield a covariance matrix estimator that is robust to both heteroskedasticity and serial correlation. While the Newey-West procedure was developed in the context of a single time series, it is frequently applied in panel data settings, and in such settings assumes cross-sectional independence.[3] Consistent with Andrews (1991) and Petersen (2009), we find that the Newey-West procedure produces slightly biased estimates of standard errors when time-series dependence alone is present. However, unlike Andrews (1991) and Petersen (2009), we assess the performance of Newey-West in the presence of both cross-sectional and time-series dependence, and find that it produces misspecified test statistics with even moderate levels of cross-sectional dependence.

Fama-MacBeth

The Fama-MacBeth approach (Fama and MacBeth 1973) is designed to address concerns about cross-sectional correlation. As originally formulated, the Fama-MacBeth approach (FM-t)

---

[3] The *newey* routine in Stata implements Newey-West standard errors in panel data. One observation we make is that the common approach of using the SAS MODEL procedure with the Bartlett kernel does *not* produce meaningful standard error estimates in panel data. We verify that this procedure: (1) treats the entire panel as a single time-series and (2) is sensitive to the ordering of the data. These limitations are to be expected, as the procedure does not allow the user to specify a time or cross-section index.

involves estimating *T* cross-sectional regressions (one for each period) and basing inferences on a *t*-statistic calculated as

$$t = \frac{\overline{\beta}}{se(\beta)}, \text{ where } \overline{\beta} = \frac{1}{T}\sum_{t=1}^{T}\hat{\beta}_t \qquad (1)$$

and *se(β)* is the standard error of the coefficients based on their empirical distribution. When there is no cross-regression (time-series) dependence, this approach yields consistent estimates of the standard error of the coefficients as *T* goes to infinity. However, cross-regression dependence in errors and regressors causes Fama-MacBeth standard errors to be understated (Schipper and Thompson 1983; Cochrane 2001).

Two common variants of the Fama-MacBeth approach appear in the accounting literature. The first variant, FM-i, involves estimating firm- or portfolio-specific time-series regressions with inferences based on the cross-sectional distribution of coefficients, and is used extensively in the recent accounting literature (e.g., Bartov and Mohanram 2004; Francis et al. 2004 2005; Larcker and Richardson 2004; Kallapur and Eldenburg 2005; Garfinkel and Sokobin 2006; Chen et al. 2007; Mohanram and Rajgopal 2009). This modification of the Fama-MacBeth approach is appropriate if there is time-series dependence but not cross-sectional dependence. However, FM-i is frequently used when cross-sectional dependence is likely, such as when returns are the dependent variable.[4]

The second common variant of the FM-t approach, FM-NW, is intended to correct for serial correlation in addition to cross-sectional correlation. FM-NW modifies FM-t by applying a Newey-West adjustment in an attempt to correct for serial correlation (e.g., Gebhardt et al. 2001; Doyle et al. 2003, 2006; Bushman and Piotroski 2006; Richardson et al. 2006). While some

[4] In such settings, we find that FM-i performs worse than OLS or White, which do not correct dependence at all.

articles suggest that FM-NW "generate[s] a conservative estimate of statistical significance" (Richardson et al. 2006. 733), FM-NW has not been formally evaluated in prior research. There are two reasons to believe that FM-NW may not correct for serial correlation. First, FM-NW involves applying Newey-West to a limited number of observations, a setting in which Newey-West is known to perform poorly (Andrews 1991). Second, FM-NW applies Newey-West to a time-series of *coefficients*, whereas the dependence is in the underlying *data*. For example, without using information in the underlying data, it seems difficult to determine whether a sequence such as {2.1, 1.8, 1.9, 2.2, 2.3} represents a highly autocorrelated draw from a mean-zero distribution or evidence of a population coefficient of around 2.[5]

Z2 statistic

The Z2 statistic first appears in Barth (1994) and appears in a number of subsequent studies in the accounting literature. The Z2-t (Z2-i) statistic is calculated using *t*-statistics from separate cross-sectional (time-series) regressions for each time period (cross-sectional unit) and is given by the expression:

$$Z2 = \frac{\bar{t}}{se(t)}, \text{ where } \bar{t} = \frac{1}{T}\sum_{t=1}^{T}\hat{t}_t\,, \tag{2}$$

se(t) is the standard error of the *t*-statistics based on their empirical distribution, and *T* is the number of time periods (cross-sectional units) in the sample. Some studies claim that Z2 adjusts for cross-sectional *and* serial correlation (e.g., Aboody and Lev 1998; Barth et al. 1998, 2001b; Davis 2002; Wang 2006). However, the basis for this claim is not specified and we could find no

[5] An alternative adjustment, essentially the Abarbanell and Bernard (2000) correction (FM-AB), is evaluated by Petersen (2009), who concludes that the adjustment only does well in correcting FM-t for time-series dependence when the serial correlation is of an autoregressive form and there are in excess of 50 years of data. We find in untabulated results that FM-AB performs similarly to FM-NW. This approach is common in finance, but rarely used in accounting, so we do not examine it in further detail.

formal analysis of the properties of the Z2 statistic. If anything, a comparison of (1) and (2) suggests that Z2 may suffer from cross-regression dependence in the same way as the Fama-MacBeth approach does.

One-way cluster-robust standard errors

A number of studies in our survey use cluster-robust standard errors, with clustering either along a cross-sectional dimension (e.g., analyst, firm, industry, or country) or along a time-series dimension (e.g., year); we refer to the former as CL-i and the latter as CL-t. Cluster-robust standard errors (also referred to as Huber-White or Rogers standard errors) were proposed by White (1984) as a generalization of the heteroskedasticity-robust standard errors of White (1980). With observations grouped into G clusters of N_g observations, for g in $\{1,\ldots,G\}$, the covariance matrix is estimated using the following expression,

$$\hat{V}(\hat{\beta}) = (X'X)^{-1}\hat{B}(X'X)^{-1}\ ,\ \hat{B} = \sum_{g=1}^{G} X'_g u_g u'_g X_g\ , \tag{3}$$

where X_g is the $N_g{\times}K$ matrix of regressors, and u_g is the N_g-vector of residuals for cluster g. If each cluster contains a single observation, equation (3) yields the White (1980) heteroskedasticity-consistent estimator.

While one-way cluster-robust standard errors allow for correlation of unknown form within cluster, it is assumed that errors are uncorrelated across clusters. For example, clustering by time (firm) allows observations to be cross-sectionally (serially) correlated, but assumes independence over time (across firms). While some studies consider both CL-i and CL-t separately (e.g., Bushee and Goodman 2007; Dhaliwal et al. 2007), separate consideration of CL-t and CL-i does not correct for both cross-sectional and time-series dependence. As we demonstrate, *t*-statistics for CL-t are inflated in the presence of time-series dependence and *t*-statistics for CL-i are inflated in the presence of cross-sectional dependence. Thus, when both forms of dependence are present, both CL-t and CL-i produce overstated *t*-statistics.[6]

It is also important to note that clustering by, say, industry-year does *not* produce standard errors robust to within-industry and within-year dependence. Instead, like all cluster-robust methods, this approach assumes independence across clusters. As such, clustering by industry-year assumes that each industry-year cluster is independent; that is, *neither* time-series *nor* cross-industry correlation is an issue (i.e., observations for industry *j* in year *t* are uncorrelated with those in industry *j* in year *t*+1 *and* those in industry *k* in year *t*). To see this point, recall that clustering by firm-year when observations are firm-years produces White standard errors.

Two-way cluster-robust standard errors

Thompson (2006) and Cameron et al. (2009) propose an extension of cluster-robust standard errors that allows for clustering along more than one dimension. In contrast to one-way clustering, two-way clustering (CL-2) allows for *both* time-series *and* cross-sectional dependence. For example, two-way clustering by firm and year allows for within-firm (time-series) dependence *and* within-year (cross-sectional) dependence (e.g., the observation for firm *j* in year *t* can be correlated with that for firm *j* in year *t+1* and that for firm *k* in year *t*). To estimate two-way cluster-robust standard errors, the expression in (3) is evaluated using clusters along each dimension (for example, clustered by industry and clustered by year) to yield V_1 and V_2. Then the same expression is calculated using the "intersection" clusters (in our example,

[6] This point applies to separate consideration of any two methods, each of which is robust to one form of dependence but not the other, such as using both FM-t and CL-i, or both Z2-i and Z2-t.

observations within an industry-year) to yield V_I. The two-way cluster-robust estimator V is calculated as $V = V_1 + V_2 - V_I$. Standard econometric software packages (e.g., Stata and SAS) contain routines for calculating one-way cluster-robust standard errors, making it relatively straightforward to implement two-way cluster-robust standard errors.[7]

One concern with cluster-robust methods is their finite sample properties (e.g., DeFond and Hung 2007). In particular, Cameron et al. (2008) document that cluster-robust methods will over-reject a true null when the number of clusters is small and attribute this issue to the use of limiting distributions (typically the normal distribution) in making inferences in small samples (e.g., using the standard 1.64, 1.96, and 2.58 critical values). However, it is important to note that most methods used in the literature (including White, N-W, FM-i, FM-t, Z2) have asymptotic foundations and thus are affected by this issue (e.g., FM-t with 10 years of data). Accordingly, researchers should exercise caution when applying any asymptotic methods in small-sample settings (e.g., FM-t, CL-t, or CL-2 when the number of time periods is small). This issue notwithstanding, we continue to find CL-2 produces unequivocally better inferences than any of these methods even with as few as 10 clusters.[8] Furthermore, as the econometrics literature has identified approaches to address this concern (e.g., Cameron et al. (2008) identify corrections based on bootstrapping methods), having few clusters does not seem to warrant relying on approaches that are not robust to cross-sectional and time-series dependence.

[7] A SAS macro and a Matlab routine that estimate two-way cluster-robust standard errors are posted at our website http://www.kellogg.northwestern.edu/faculty/gow/htm/GOT/ and Mitchell Petersen has posted Stata code at http://www.kellogg.northwestern.edu/faculty/petersen/htm/.

[8] With 10 time periods, type I rejection rates for CL-2 in the presence of cross-sectional and time-series dependence reach as high as 3.3%. However, these rejection rates are uniformly better than those for other methods we study, which range from 8.7% (FM-t) to 92% (FM-i). These results suggest that researchers using asymptotic methods in small samples may wish to consider bootstrap-based corrections (e.g., Cameron et al. 2008). Stata code that estimates two-way cluster-robust standard errors using the cluster bootstrapping technique discussed in Cameron et al. (2008) is posted on our website.

IV. SIMULATION

Description of simulation

To evaluate the performance of the above methods in the presence of cross-sectional and time-series dependence, we follow Thompson (2006) and Petersen (2009) and design a simulation that allows for varying levels of cross-sectional and time-series correlation in observed regressors and unobserved errors. We focus on the case of a single explanatory variable, x and homoskedastic errors, ε, where x and ε are independent.[9] We generate the dependent variable, y, using the equation, $y = \beta_0 + \beta_1 x + \varepsilon$, where $\beta_0 = 0$ and $\beta_1 = 1$, and values of x and ε using the following equations:

$x_{it} = x_t^T + x_{it}^F$, where $x_{it}^F = \rho x_{it-1}^F + \nu_{it}$, and

$\varepsilon_{it} = \varepsilon_t^T + \varepsilon_{it}^F$, where $\varepsilon_{it}^F = \rho \varepsilon_{it-1}^F + \eta_{it}$,

where ν_{it} and η_{it} are normally distributed with mean zero and independent of all other terms. In words, both the unobserved errors (ε_{it}) and the observed regressor (x_{it}) have two components: a year-specific fixed effect that drives cross-sectional correlation (x_t^T and ε_t^T), and an autocorrelated firm effect that drives time-series correlation (x_{it}^F and ε_{it}^F). We hold the variances of x and ε constant across our simulations, setting $\sigma_x^2 = 1$ and $\sigma_\varepsilon^2 = 2$.[10] To illustrate the effects of cross-sectional dependence, we vary cross-sectional correlation (driven by the variances of

[9] Including an additional explanatory variable, whether correlated or uncorrelated with the first regressor, does not affect our conclusions.

[10] In this regard we follow Petersen (2009). To maintain homoskedasticity, the variance of ν_{it} and η_{it} is implied by the values of ρ, σ_x, σ_ε, and the magnitude of the cross-sectional correlation (the variance of the time effects). We generate the time effects and the terms ν_{it} and η_{it} using normal distributions with zero means without loss of generality.

x_t^T and ε_t^T) from 0 to 0.75, in increments of 0.25. To illustrate the effects of the time-series dependence, we consider values of ρ common in the literature: 0, 0.5 and 0.8 (see Section V).

Using the data-generating process described above, we generate values of *y* and *x* for a balanced panel of 200 firms and 40 years.[11] We then estimate coefficients and standard errors using OLS, FM-t and FM-i. For the OLS coefficient we calculate standard errors using OLS, N-W, CL-i, CL-t, and CL-2. And for the FM-t coefficient, we calculate standard errors using FM-t and FM-NW. In calculating N-W standard errors, we set the lag length to *T*-1, the maximum lag length possible with our data, as the AR(1) structure of our data-generating process implies correlations in errors at all lags. This approach maximizes the ability of the N-W approach to correct for time-series dependence and is consistent with the practice of determining the lag length empirically. For FM-NW standard errors, we use one lag (e.g., Core et al. 2006a).[12]

Using the coefficient estimates and standard errors denoted above, we calculate *t*-statistics for each of the above methods, as well as the Z2 statistic. We then test the null hypothesis that $\beta_1 = 1$ at the 1% level using each method. We repeat this process 1,000 times, and for each method we tabulate (1) the frequency that the test rejects the true null hypothesis, (2) the actual standard error using the realized distribution of coefficient estimates, and (3) the average estimated standard error.[13] Consistent with common practice, for OLS, N-W, FM-t,

---

[11] In untabulated results we continue to find support for CL-2 relative to the other methods when vary the number of firms (100 to 500) and vary the number of time periods (from 10 to 100). We have posted all code necessary to run our simulation at http://www.kellogg.northwestern.edu/faculty/gow/htm/GOT/.

[12] Our results are unaffected when we consider alternative lag lengths for N-W and FM-NW. We tabulate results using lag lengths T-1 and 1 respectively as we find these to be the least biased.

[13] In our analysis, we do not evaluate the power of the various methods in the presence of *both* cross-sectional and time-series dependence. This reflects our finding that only CL-2 is well-specified in this setting and, as pointed out by Kothari and Wasley (1989, 238), "power is an issue only when a researcher is making a choice among a set of valid tests." However, our simulation can be used to examine the power of cluster-robust methods in settings where other methods are valid. For example, when cross-sectional dependence is present, but time-series dependence is absent, both CL-t and CL-2 standard error estimators are consistent. For a null that β_1=0 against the alternative that

FM-NW, FM-i, and Z2 we use large-sample critical values (e.g., Botosan and Plumlee 2005; Aboody et al. 2004a; Core et al. 2006a; Hail and Leuz 2006). For CL-i, CL-t, and CL-2 we use critical values based on the *t*-distribution with degrees-of-freedom equal to the smallest number of clusters along either dimension (Cameron et al. 2009).[14]

Simulation results

The results of the simulation are given in Tables 1 and 2. Table 1 reports the rejection rates of the various methods for the (true) null hypothesis that $\beta_1 = 1$. Table 2 reports the actual and estimated standard errors for each method. As the null hypothesis is true by construction, the rejection rate for a well-specified test at the 1% level should be close to 1%. For 1,000 simulations, the 95% confidence interval for the null of a well-specified test is (0.4%, 1.7%).[15]

[INSERT TABLE 1 & TABLE 2 ABOUT HERE]

Table 1, Panel A presents simulation results for the case of zero serial correlation (ρ=0) and varying levels of cross-sectional correlation, such as settings in which the dependent variable is either a raw or abnormal return (Bernard, 1987). The first row represents the one setting in our simulations in which neither cross-sectional nor time-series dependence is present and thus most methods provide rejection rates close to the 1% size of the test. However, FM-t and FM-NW depart from their nominal values. This result is attributable to following the common practice in the literature of using large-sample critical values. As cross-sectional correlation is added (rows

$\beta_1 \neq 0$ in the setting of row 4 of panel A of Table 1, CL-t rejects 92.7% of the time and CL-2 rejects 92.8% of the time (1000 iterations); suggesting no material loss of power from adjusting for both forms of dependence. We find similar results when comparing OLS to cluster-robust methods when no form of dependence is present.

[14] In untabulated results, we find that using large-sample critical values for cluster-robust methods does not affect our conclusions. Following Cameron et al. (2006b), we emphasize the use of small-sample critical values, as these produce better specified test statistics.

[15] The rejection rate will generally differ from 1% for a finite number of replications due to both random noise and differences between actual size and nominal size for asymptotic test statistics (Davidson and MacKinnon 2004).

2 through 4), OLS, N-W, CL-i, and FM-i produce standard errors biased downward anywhere from five- to more than ten-fold, yielding rejection rates that exceed 50%. These results are consistent with OLS, N-W, CL-i and FM-i not being robust to cross-sectional dependence. Panel A of Table 1 suggests that, in the case of FM-i, a researcher would reject the null hypothesis 86.5% of the time at the 1% level even when the null is true (row 4). This finding is not unrealistic, as we show later that similar rejection rates occur with real data.

Table 1 Panel B (Panel C) presents simulation results for the case of serial correlation of 0.5 (0.8).[16] These levels of serial correlation are what one might expect in regressions where variables are naturally serially correlated, such as accounting items, audit fees, executive compensation, and the implied cost of capital. Similar to Panel A, when both cross-sectional and serial correlation are present OLS, N-W, CL-i, and FM-i continue to produce rejection rates well in excess of nominal levels, 87% in the extreme (row 4 of Panel C). The addition of serial correlation also increases the rejection rates of FM-t, FM-NW, and Z2-t. For FM-t, rejection rates increase from 1.7% in Panel A to 4.9% in Panel B and 24.3% in Panel C. The rapid progression and magnitude of these rejection rates illustrate that when the time-series independence assumption of FM-t is not satisfied, misspecified test statistics result. Rejection rates for FM-NW and Z2-t exhibit a similar progression. As serial correlation increases, FM-NW (Z2-t) rejection rates increase from 1.9% (1.6%) in Panel A to 3.2% (4.9%) in Panel B and 16.0% (24.2%) in Panel C.[17] Both the magnitude of these rates and the fact that they are increasing in the degree of serial correlation confirm that FM-NW and Z2-t are *not* robust to both cross-sectional and time-series dependence. Confirming the intuition that FM and Z2 are closely

[16] While serial correlation of 0.8 may seem extreme, our fourth application examines such a setting.
[17] While the absolute performance of FM-NW improves slightly when we consider larger values of T, it still fails to correct for time-series dependence and produces misspecified test statistics.

related, in untabulated analysis we find that Z2-t (Z2-i) is highly correlated with FM-t (FM-i) *t*-statistics, with correlation coefficients in excess of 0.99.[18]

Overall, the results in Tables 1 and 2 suggest the following. First, methods that are not robust to cross-sectional dependence, such as FM-i and N-W, can produce severely misspecified test statistics. In the presence of cross-sectional dependence, such methods perform worse that OLS, and can inflate *t*-statistics by as much as ten-fold, causing the probability of rejecting a true null to exceed 50%. Second, FM-NW and Z2-t are not robust to cross-sectional and time-series dependence, and in the presence of time-series dependence FM-NW and Z2-t produce misspecified test statistics. Third, CL-2 is robust to both forms of dependence.

V. APPLICATIONS

In this section, we examine four settings that cover a broad sample of empirical accounting research for which panel data is frequently used and for which cross-sectional and/or time-series dependence is likely. For each application, we first replicate a representative regression from a class of studies and show that (1) the assumptions underlying methods common in the literature are violated, and (2) methods robust to such violations produce substantially different *t*-statistics. Next, to assess the sensitivity of our simulation results to the actual data generating processes and correlation structures that are encountered in common panel datasets, for our first two applications we include a randomly generated regressor and tabulate the frequency that the various methods reject the null hypothesis at a given significance level.

[18] This high correlation mirrors that observed in studies reporting both FM-t and Z2-t statistics (0.96 for Barth et al. 2001b and 0.99 for Wang 2006).

Application I: Asset Pricing

Our first application illustrates the effect of cross-sectional dependence on inferences when time-series dependence is not present. This setting parallels the one examined in Panel A of Tables 1 and 2 from our earlier simulation. We choose an asset pricing setting for our first application because market efficiency implies that excess stock returns are serially uncorrelated, allowing us to focus exclusively on the effects of cross-sectional correlation on inferences. While earlier work documents that cross-sectional dependence in return-based regressions can confound inferences (Bernard 1987), this application demonstrates that the recent literature continues to use methods not robust to cross-sectional dependence.

Beginning with Francis et al. (FLOS 2004, 2005), a significant recent literature investigates the role of earnings quality in asset pricing. Using FM-i, FLOS find that an accruals quality (AQ) factor is highly significant, with *t*-statistics in excess of 50.[19]

As discussed earlier, the FM-i procedure assumes that either the regressors or regression residuals are uncorrelated across the firm-specific regressions. But, since a common factor takes only one value at each point in time, regressors will be perfectly correlated in each period, and prior work suggests that errors (abnormal returns) are also cross-sectionally correlated. To evaluate the effects of such cross-sectional dependence on inference, we replicate the primary asset pricing regressions of FLOS and CGV using a variety of methods, including FM-i.

The sample used in this application consists of all firm-months not missing returns on the CRSP monthly file from April 1971 through March 2002, the period analyzed by FLOS and

[19] Core et al. (CGV 2007) dispute FLOS's conclusion, arguing that the one-stage tests used in FLOS do not provide evidence of a priced risk factor. We focus only on the robustness of FLOS's tests to cross-sectional dependence.

CGV. We require firms to have at least 18 months of such returns. Table 3 presents results from regressions of monthly excess returns on the three Fama-French factors and the AQ factor.[20]

[INSERT TABLE 3 ABOUT HERE]

Panel A presents *t*-statistics from pooled time-series cross-sectional regressions calculated using White standard errors. These standard errors are robust to heteroskedasticity but not cross-sectional correlation, and provide a benchmark for assessing the effect of cross-sectional correlation on inferences. Panel A shows a *t*-statistic on the AQ factor of 72.32. Panel B presents results from estimating the four-factor FM-i regressions of FLOS and CGV. Our results are within 5 basis points of CGV.[21] Panel B reports that the FM-i (Z2-i) *t*-statistic for the AQ factor is 50.41 (49.57), which is similar in magnitude to that reported by FLOS and CGV. However, as demonstrated above, FM-i (Z2-i) does *not* adjust for cross-sectional correlation. Panel C presents results from applying the traditional FM-t and Z2-t procedures, which produce test statistics that are robust to cross-sectional correlation. Panel C reports that the FM-t (Z2-t) *t*-statistic for the AQ factor is 13.62 (8.00), less than one-fifth of the uncorrected White *t*-statistic. Panel D presents the coefficient and *t*-statistics from OLS regression with standard errors clustered by month.[22] These *t*-statistics are generally one-tenth of the uncorrected White *t*-statistics, and the *t*-statistic on the AQ factor is around one-eighth (one-fifth) of the White (FM-i) *t*-statistic. A bias of this magnitude is consistent with the simulation results in Panel A of Table

[20] AQ factor data provided by Jennifer Francis and Per Olsson. All other factors are from Ken French's website.

[21] Our sample has just three more firms than the sample used in CGV and, in untabulated results for the three-factor model, our results are within 1 basis point of CGV. Differences in the four factor results are likely attributable to the fact that we use the monthly AQ factor provided by FLOS, whereas CGV presumably replicate the factor.

[22] OLS and Fama-MacBeth will generally produce different coefficient estimates in our setting because we have an unbalanced panel (different number of observations in each time interval); OLS equally weights the influence of each observation and FM-t equally weights the influence of each year. Absent reasons to under-weight observations drawn from a year with more firms, it is more efficient to base inferences on OLS coefficients adjusted for cross-sectional correlation than to use FM-t.

1, where, in the presence of cross-sectional correlation, OLS and FM-i produce standard errors that are downward biased anywhere from five- to ten-fold.

While adjusting for cross-sectional dependence does not change FLOS's conclusion that AQ is related to returns, it shrinks *t*-statistics by a factor of eight, implying that FM-i results in seriously misspecified test statistics when returns is the dependent variable. To investigate the effect on inferences, we replace the AQ factor with a randomly generated factor and use the methods above to test the significance of this factor. We implement this test in three steps. First, in each month we *randomly* assign all firms with non-missing returns on CRSP into five quintiles and subtract the equal-weighted return of the first quintile from that of the fifth quintile to get the random factor. Second, we regress stock returns in excess of the risk-free rate on the three Fama-French factors and this random factor, and test the significance of the random factor using each of the methods employed in Table 3. Finally, we repeat this procedure 1,000 times and tabulate the probability of rejecting the null hypothesis of a zero coefficient on the random factor at each of the 1%, 5%, and 10% significance levels. Because the factor we create is random, we expect to find no relationship between individual asset returns and the factor; a well-specified test should produce a rejection rate approximately equal to the level of the test.[23]

[INSERT TABLE 4 ABOUT HERE]

Results in Table 4 suggest that failure to adjust for cross-sectional dependence in return-based regressions leads to *highly* misspecified test statistics. Methods not robust to cross-

[23] Note that the factor is constructed as the difference in returns between two randomly constructed portfolios each month (i.e. in expectation these portfolios will have equivalent risk). So there is no reason to believe that there will be any association with a missing systematic factor. Moreover, the fact that methods known to be robust with respect to cross-sectional dependence produce rejection rates very close to 1% (FM-t, Z2-t, CL-t) where as non-robust methods (White, Z2-i, FM-i) produce rejection rates vastly exceeding 1% suggests that we have successfully created a random factor. If the factor were associated with a systematic factor, we would expect rejection rates for all methods to exceed 1%.

sectional dependence, such as White, FM-i, and Z2-i, produce standard errors around one-tenth their actual value and rejection rates in excess of 80% at the 1% significance level. In contrast, methods robust to cross-sectional dependence, such as FM-t, Z2-t, and CL-t, produce standard errors close to actual values and rejection rates close to nominal levels. However, these methods are not robust to time-series dependence. When both cross-sectional and serial dependence are present, these methods will produce misspecified test statistics and we explore the effects of such dependence in the remaining applications.

Application II: Implied Cost of Capital

Our second application illustrates the effect of both cross-sectional and time-series dependence on inferences. The implied cost of capital setting is a natural one to examine the effect of both forms of dependence, as commonly used dependent and independent variables are both cross-sectionally and serially correlated. Starting with Botosan (1997), a number of studies examine the association between a firm's implied cost of capital and its information environment. Four variables examined recently are earnings variability and predictability (e.g., Gebhardt et al. 2001; FLOS 2004, 2005), idiosyncratic volatility (e.g., CGV; Liu and Wysocki 2007; Ashbaugh-Skaife et al. 2009), and governance quality (e.g., Ashbaugh et al. 2004; Cheng et al. 2006; Hail and Leuz 2006). Studies in this area typically use the Fama-MacBeth procedure to correct for cross-sectional dependence and either do not adjust for time-series dependence or use FM-NW.

While our simulation evidence (Tables 1 and 2) suggests that in the presence of both cross-sectional and time-series dependence these methods produce misspecified test statistics, it is unclear what effect the use of these methods has on inferences reported in the literature. We investigate the robustness of inferences in the cost-of-capital literature examining the four variables above. The sample we use consists of all firms on Compustat with positive book value of equity as of the fiscal year-end, current and lagged net income (data #178), and beginning of period total assets (data #6) over the prior 10 years. We also require market value as of the fiscal year-end and monthly returns for 18 of the prior 60 months from CRSP. Consistent with the extant literature, we use *Beta*, *Size*, *BM, and Volatility* as control variables. *Beta* is computed using rolling 60-month firm-specific CAPM regressions, *Size* is the natural log of fiscal year end market value, *BM* is the fiscal year-end book-to-market ratio, and *Volatility* is the standard deviation of returns in excess of the risk free rate over the prior 60 months.

Similar to FLOS (2004), we compute earnings variability (*EarnsVol*) as the variance of a firm's ROA over the prior ten years, and earnings predictability as the variance of residuals from a rolling ten-year firm-specific AR(1) model with drift (*Predict*), where ROA is net income scaled by beginning-of-period total assets. As in FLOS (2004), higher values of *EarnVol* (*Predict*) correspond to more volatile (less predictable) earnings. Like CGV and Liu and Wysocki (2007), we compute idiosyncratic (systematic) volatility, *IdVol* (*SysVol*), as the variance of residuals (predicted values) from a rolling 60-month firm-specific CAPM regression. Following Ashbaugh et al. (2004) and Cheng et al. (2006), we use *G-Index* as our measure of governance quality, with higher values of *G-Index* corresponding to lower governance quality. Since *G-Index* is only available from IRRC for a limited period (1990–2004) and for a limited number of firms, we do not require *G-Index* for our analyses of the other variables. Finally, we require data to compute the implied cost of capital (*CoC*), which is constructed as in Hail and Leuz (2006) and Barth et al. (2008). The resulting sample comprises 19,125 firm-years from

1986–2004, with 11,977 firm-years from 1990–2004 with data on *G-Index*. Following the prior literature, we regress the implied cost of capital (*CoC*) on control variables (*Beta*, *Size*, *BM*, and *Volatility)*, year fixed effects, and measures of earnings quality (*EarnVol* and *Predict*), idiosyncratic volatility (*Idvol*), and governance quality (*G-Index*), and we tabulate *t*-statistics using FM-t, FM-NW, Z2-t, CL-t, and CL-2.[24] Results are presented in Table 5.

[INSERT TABLE 5 ABOUT HERE]

Panel A shows results from estimating the model using only the control variables. For both pooled OLS and FM-t, we find that residuals exhibit serial correlation in excess of 0.5, which is statistically significant at the 1% level, implying that the assumptions underlying FM-t, FM-NW, Z2-t, and CL-t do not hold. Therefore, by comparing *t*-statistics based on FM-t (CL-t) with those based on FM-NW (CL-2), we can gain insight into the extent to which FM-NW (CL-2) adjusts for time-series dependence. Panel A reports that the coefficient on *Volatility* is statistically significant at the 1% level using FM-t, FM-NW, and Z2-t. Using *t*-statistics based on CL-2 standard errors, the coefficient on *Volatility* is not significantly different from zero. The results in Panels B and C are similar.

In Panel B, the FM-t, FM-NW, and Z2-t methods suggest that the coefficient on *EarnVol* is significantly different from zero at the 1% level. In contrast, the CL-2 method suggests the coefficient on *EarnVol* is not significantly different from zero. Consistent with FLOS (2004), in Panel C, the FM-t, FM-NW, and Z2-t methods suggest that the coefficient on *Predict* is significantly different from zero at the 1% level. However, the CL-2 method suggests the coefficient on *Predict* is not significantly different from zero. Consistent with the findings of

[24] Because volatility is mechanically the sum of its idiosyncratic and systematic components, when *IdVol* is included in the regression we use *SysVol* instead of *Volatility*.

CGV, Liu and Wysocki (2007), and Ashbaugh-Skaife et al. (2009), Table 5, Panel D reports that *IdVol* is statistically significant at the 1% level using the FM-t, FM-NW, Z2-t, and CL-t methods. However, after adjusting standard errors for both forms of dependence (CL-2), the coefficient on *IdVol* is not statistically significant. Consistent with Ashbaugh et al. (2004) and Cheng et al. (2006), in Table 5, Panel E, the FM-t, FM-NW, Z2-t, and CL-t methods indicate a coefficient on *G-Index* that is significantly different from zero at the 1% level. However, after adjusting for both forms of dependence (CL-2), the coefficient on *G-Index* is not statistically significant.

Note that inferences from Table 5 do not vary across the FM-t, FM-NW, and Z2-t approaches, consistent with our finding in simulation analysis that FM-NW and Z2-t produce *t*-statistics similar to FM-t regardless of the level of serial correlation.

The evidence thus far suggests that inferences based on methods robust to cross-sectional dependence alone are frequently different from inferences based on methods robust to both forms of dependence. To examine this further, we regress the cost of capital on control variables and a serially correlated random variable (*RAND*). We repeat this procedure 1,000 times for three levels of serial correlation (0.2, 0.5, and 0.8) and calculate the actual and average estimated standard error of the coefficient on *RAND*, as well as the probability that the indicated estimation method rejects the (true) null hypothesis that the coefficient on *RAND* is zero.

[INSERT TABLE 6 ABOUT HERE]

Table 6 presents results for autocorrelation of 0.2, 0.5, and 0.8. Consistent with FM-NW and Z2-t not being robust to time-series dependence, across all levels of serial correlation their rejection rates and standard errors are very similar to those of FM-t and are well in excess of nominal significance levels. In the presence of serial correlation of 0.5 (0.8), methods not robust to serial dependence, FM-t, FM-NW, Z2-t, and CL-t, yield rejection rates up to 5 (15) times nominal levels and standard errors that underestimate the true standard error by around 20% (50%). In contrast, across all panels, CL-2 has rejection rates close to theoretical levels and unbiased standard errors, suggesting that only CL-2 is robust to both forms of dependence.

Application III: Cost of Debt

Similar to the cost of equity capital literature, a significant recent literature has examined the determinants of firms' cost of debt capital. As in the prior application, the cost of debt setting is a natural one to examine the effects of both forms of dependence as research in this areas uses a common research design, and dependent and independent variables are known to be both cross-sectionally and serially correlated. Recent studies in this literature have examined the relationship between the cost of debt and measures of information risk (e.g., Mansi et al. 2009), earnings quality (e.g., FLOS 2005; Bharath et al. 2008; Francis et al. 2008), corporate governance (e.g., Bhojraj and Sengupta 2003; Anderson et al. 2004; Ashbaugh-Skaife et al. 2006; Cremers et al. 2007), the use of performance pricing (e.g., Asquith et al. 2005), flexibility in debt covenants (e.g., Beatty et al. 2002), research and development expenditure (e.g., Shi 2003; Eberhardt et al. 2008), pension disclosures (e.g., Hann et al. 2007), beating earnings benchmarks (e.g., Jiang 2008), employee stock options (e.g., Lee 2008), auditor choice (e.g., Pittman and Fortin 2004; Mansi et al., 2004), and conservatism (e.g., Ahmed et al. 2002). Studies in this literature use a variety of proxies for the cost of debt, including interest expense to debt, spread over LIBOR, yield to maturity, and credit ratings.

To illustrate the applicability of cluster-robust standard errors to non-linear estimation procedures, we follow those studies that use ordered logit and credit ratings as the dependent variable (e.g., Ahmed et al. 2002; Shi 2003; Ashbaugh-Skaife et al. 2006; Cremers et al. 2007; Eberhardt et al. 2008; Lee 2008; Hann et al. 2007; Jiang 2008; Mansi et al. 2004, 2009). Although almost all of these studies use panel data, they either do not adjust for either form of dependence, or address time-series but not cross-sectional dependence (Lee 2008 and Jiang 2008 report standard errors clustered by firm, and Cremers et al. 2007 use the Newey-West approach).

We estimate ordered logit regressions of credit ratings (*Rating*) on commonly used control variables (*Size*, *BM*, *ROA*, *Leverage*, *Beta*, *Volatility)*, year fixed effects, and measures of earnings quality (*EarnVol* and *Predict*), research and development (*R&D*), and governance quality (*G-Index*), tabulating *t*-statistics using uncorrected standard errors, CL-t, and CL-2.[25] *Rating* is a transformation of a firm's S&P's senior debt rating in year *t* (data #280) to range from 1 through 20, where *higher* values correspond to more favorable debt ratings (Ahmed et al. 2002; Ashbaugh-Skaife et al. 2006). *Size*, *BM*, *ROA, Beta, Volatility, EarnVol*, *Predict*, and *G-Index* are defined as above. *Leverage* is total debt (data #9 + data #34) scaled by total assets, and *R&D* is research and development expense (data #46) scaled by total assets.

[INSERT TABLE 7 ABOUT HERE]

Table 7 reports results. Panel A shows results using only the control variables. The results in Panels B through E confirm that we are able to replicate inferences in prior research for each of the four variables of interest. That is, without correcting for dependence, we find a statistically significant relation between credit ratings and earnings volatility, earnings predictability, research and development, and governance quality (uncorrected *t*-stats of –2.95, –2.45, –3.72,

[25] We do not tabulate results using FM-t, FM-NW, or Z2-t in this application, as researchers do not use these methods in this setting.

and 2.81, respectively). However, after accounting for both forms of dependence, none of these relations is statistically significant (*t*-stats of –1.28, –1.27, –1.29, and 1.01, respectively).

Application IV: Conservatism

Our final application examines the effects of cross-sectional and time-series dependence when dependent and independent variables are aggregated over overlapping periods (e.g., the dependent variable in year *t* is the sum of earnings from year *t* to *t–4*). While this kind of aggregation is common in accounting research (e.g., Doyle et al., 2003, 2006), many studies using overlapping observations do not correct for the induced time-series dependence. To examine the effect of using overlapping observations on inferences, we examine inferences in recent research on conservatism that follows Roychowdhury and Watts (2007, RW) in estimating Basu (1997) regressions with up to four years of overlap.

Basu (1997) estimates the asymmetric timeliness of earnings, or the level of conditional conservatism, as the coefficient on *D*R* in a regression of earnings on *R*, *D*, and *D*R*. *R* is buy-and-hold return over the period, *D* is an indicator variable equaling 1 if *R* is negative, and earnings is net income before extraordinary item (data #18) scaled by beginning of period price. While Basu (1997) estimates this regression using observations measured over a single year, RW argue that extending the horizon over which earnings and returns are measured will reduce the measurement error associated with the Basu (1997) specification. In each year they measure earnings and returns using aggregation windows of up to five years.

RW examine the relation between conditional conservatism and year-end market-to-book ratio. They find that the negative relation documented in Givoly et al. (2007) becomes positive once earnings and returns are aggregated over three-, four-, and five-year periods. RW also find a negative relation between goodwill and conditional conservatism, but only when the accumulation period is extended to three or more years. Recent studies use the RW approach to examine relations between conditional conservatism and director independence (e.g., Ahmed and Duellman 2007), the probability of informed trade (e.g., Lafond and Watts 2007), and managerial ownership (e.g., LaFond and Roychowdhury 2008). Studies in this literature rely almost exclusively on FM-t or FM-NW to compute standard errors, suggesting that inferences may not be robust to the time-series dependence induced by overlapping observations. When they consider accumulation periods of three years, LaFond and Roychowdhury (2008, 127) find that "the coefficient on [*R*D*OWN*] … is statistically more significant than the corresponding coefficient in [regressions without accumulation], consistent with the three-year specification yielding a more powerful test." An alternative explanation is that this test statistic does not account for the time-series dependence induced by using overlapping observations and its significance is overstated. For both pooled OLS and FM-t, we find that residuals exhibit serial correlation in excess of 0.8, which is statistically significant at the 0.01% level, implying that the assumptions underlying FM-t, FM-NW, Z2-t, and CL-t do not hold.

To assess the impact of correcting for both cross-sectional and time-series dependence on inferences in the conservatism literature, we use the RW approach, cumulating earnings and returns over five periods, and we examine four variables identified in the studies discussed above: *Leverage*, the market-to-book ratio (*MB*), the probability of informed trade (*PIN*), and managerial ownership (*OWN*). *Leverage* is as previously defined, *MB* is the end of period market to book ratio, *PIN* is the probability of informed trade, and *OWN* is the percentage of shares held by the top five highest paid officers at the firm.[26] Like RW, we use quintile ranks of all variables.

[INSERT TABLE 8 ABOUT HERE]

Table 8 shows results. In each case, we find that the coefficient on *D*R*X*, where *X* is the variable of interest, is statistically significant and of the sign documented in the original studies (FM-t *t*-stats of 3.96, 2.54, 3.03, and –2.17, respectively). However, in each case, when we use two-way cluster-robust standard errors, the coefficient is no longer statistically significant (CL-2 *t*-stats of 1.48, 1.60, 1.41, and –0.96, respectively). This is consistent with the apparent statistical significance of the coefficients being attributable to failure to correct for time-series dependence arising from overlapping observations, rather than increased power.

VI. CONCLUSIONS

Accounting researchers frequently encounter cross-sectional and time-series dependence in panel data sets. In such settings, researchers either rely on methods that assume either cross-sectional or time-series independence, or rely on methods developed in the accounting literature that have not been formally evaluated. We show that the assumptions underlying these methods are often violated in common research settings, producing misspecified test statistics and spurious inferences. Methods robust to such violations produce substantially different inferences. We provide guidance on the assumptions and appropriateness of each method in Table 9.

[INSERT TABLE 9 ABOUT HERE]

Despite growing use of FM-i, Z2, and FM-NW in the accounting literature, ours is the first study to evaluate the properties of these methods. Using both simulations and four

[26] PIN data were provided by Stephen Brown.

applications, we show that these estimators are robust to one but not both forms of dependence. In particular, we find that FM-i produces substantially overstated *t*-statistics and rejects a true null hypothesis more than 80% of the time at the 1% level in the setting in which it is most frequently used. In contrast to claims in the literature that Z2 and FM-NW are robust to *both* cross-sectional and time-series dependence, we find no evidence that either method corrects for time-series dependence. In the presence of time-series dependence we find that these methods can reject a true null hypothesis more than 24% of the time at the 1% level. In contrast, we find that CL-2 produces well-specified test statistics in all settings examined.

We examine the sensitivity of inferences reported in the literature to correcting for both cross-sectional and time-series dependence. We document that in return-based regressions, statistics aggregated from firm- or portfolio-specific regressions (FM-i, Z2-i) produce *t*-statistics biased upwards by as much as ten-fold. We document both cross-sectional and time-series dependence in popular cost of equity capital, cost of debt, and conservatism regressions, and we show that the Fama-MacBeth (FM-t) and Newey-West corrected Fama MacBeth (FM-NW) methods produce misspecified test statistics in these settings. Moreover, we show that several recent findings in literatures examining the cost of equity capital, the cost of debt, and conservatism are not robust to correcting for cross-sectional and time-series dependence. We argue that our applications provide evidence that properly adjusting for both forms of dependence can affect conclusions in actual research settings, and we caution researchers against using methods that are robust to only one form of dependence.

References

Abarbanell, J. S., and V. L. Bernard. 2000. Is the U.S. stock market myopic? *Journal of Accounting Research* 38 (2): 221–242.

Aboody, D., M. E. Barth, and R. Kasznik. 2004. Firms' voluntary recognition of stock-based compensation expense. *Journal of Accounting Research* 42 (2): 123–150.

Aboody, D., and B. Lev. 1998. The value relevance of intangibles: The case of software capitalization. *Journal of Accounting Research* 36: 161–191.

Ahmed, A. S., B. K. Billings, R. M. Morton, and M. Stanford-Harris. 2002. The role of accounting conservatism in mitigating bondholder-shareholder conflicts over dividend policy and in reducing debt costs. *The Accounting Review* 77 (4): 867–890.

Ahmed, A., and S. Duellman. 2007. Accounting conservatism and board of director characteristics: An empirical analysis. *Journal of Accounting and Economics* 43: 411–437.

Anderson, R. C., S. A. Mansi, and D. M. Reeb. 2004. Board characteristics, accounting report integrity, and the cost of debt. *Journal of Accounting and Economics* 37 (3): 315–342.

Andrews, D. W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59 (3): 817–858.

Ashbaugh-Skaife, H., D. W. Collins, and R. LaFond. 2006. The effects of corporate governance on firms' credit ratings. *Journal of Accounting and Economics* 42 (1–2): 203–243.

Ashbaugh-Skaife, H., D. W. Collins, W. R. Kinney, and R. LaFond. 2009. The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research* 47 (1): 1–43.

Ashbaugh, H., D. W. Collins, and R. LaFond. 2004. Corporate governance and the cost of equity capital. Working paper, Universisty of Wisconsin–Madison, University of Iowa, Massachusetts Institute of Technology

Asquith, P., A. L. Beatty, and J. P. Weber. 2005. Performance pricing in bank debt contracts. *Journal of Accounting and Economics* 40 (1–3): 101–128.

Barth, M. E. 1994. Fair value accounting: Evidence from investment securities and the market valuation of banks. *The Accounting Review* 69 (1): 1–25.

Barth, M. E., M. B. Clement, G. Foster, and R. Kasznik. 1998. Brand values and capital market valuation. *Review of Accounting Studies* 3 (1): 41–68.

Barth, M. E., L. D. Hodder, and S. R. Stubben. 2008. Fair value accounting for liabilities and own credit risk. *The Accounting Review* 83 (3): 629–664.

Barth, M. E., R. Kasznik, and M. F. McNichols. 2001. Analyst coverage and intangible assets. *Journal of Accounting Research* 39 (1): 1–34.

Bartov, E., and P. Mohanram. 2004. Private information, earnings manipulations, and executive stock-option exercises. *The Accounting Review* 79 (4): 889–920.

Basu, S. 1997. The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics* 24 (1): 3–37.

Beatty, A. L., K. Ramesh, and J. P. Weber. 2002. The importance of accounting changes in debt contracts: the cost of flexibility in covenant calculations. *Journal of Accounting and Economics* 33 (2): 205–227.

Bernard, V. L. 1987. Cross-sectional dependence and problems in inference in market-based accounting research. *Journal of Accounting Research* 25 (1): 1–48.
Bharath, S. T., J. Sunder, and S. V. Sunder. 2008. Accounting quality and debt contracting. *The Accounting Review* 83(1): 1–28.
Bhojraj, S., and P. Sengupta. 2003. Effect of corporate governance on bond ratings and yields: The role of institutional investors and outside directors. *Journal of Business* 76 (3): 455–475.
Botosan, C. A. 1997. Disclosure level and the cost of equity capital. *The Accounting Review* 72 (3): 323–349.
Botosan, C. A., and M. A. Plumlee. 2005. Assessing alternative proxies for the expected risk premium. *The Accounting Review* 80 (1): 21–54.
Bushee, B. J., and T. H. Goodman. 2007. Which institutional investors trade based on private information about earnings and returns? *Journal of Accounting Research* 45 (2): 289–321.
Bushman, R. M., and J. D. Piotroski. 2006. Financial reporting incentives for conservative accounting: The influence of legal and political institutions. *Journal of Accounting and Economics* 42 (1–2): 107–148.
Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2008. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics* 90 (3): 414-427.
Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2009. Robust inference with multi-way clustering. *Journal of Business and Economics Statistics* (forthcoming).
Cameron, A., and P. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York, NY: Cambridge University Press.
Chen, S., T. Shevlin, and Y. H. Tong. 2007. Does the pricing of financial reporting quality change around dividend changes? *Journal of Accounting Research* 45 (1): 1–40.
Cheng, C., D. Collins, and H. Huang. 2006. Shareholder rights, financial disclosure and the cost of equity capital. *Review of Quantitative Finance and Accounting* 27 (2): 175–204.
Cochrane, J. H. 2001. *Asset Pricing*. Princeton, NJ: Princeton University Press.
Collins, D. W., and W. T. Dent. 1984. A comparison of alternative testing methodologies used in capital market research. *Journal of Accounting Research* 22 (1): 48–84.
Core, J. E., W. R. Guay, and T. Rusticus. 2006. Does weak governance cause weak stock returns? An examination of firm operating performance and investors' expectations. *Journal of Finance* 61 (2): 655–687.
Core, J. E., W. R. Guay, and R. Verdi. 2008. Is accruals quality a priced risk factor? *Journal of Accounting and Economics* 46 (1): 2–22.
Cremers, K. J. M., V. B. Nair, and C. Wei. 2007. Governance mechanisms and bond prices. *Review of Financial Studies* 20(5): 1359–1388.
Davidson, R., and J. G. MacKinnon. 2004. *Econometric Theory and Methods*. Oxford, U.K.: Oxford University Press.
Davis, A. K. 2002. The value relevance of revenue for internet firms: Does reporting grossed-up or barter revenue make a difference? *Journal of Accounting Research* 40 (2): 445–477.
DeFond, M. L., and M. Hung. 2007. Investor protection and analysts' cash flow forecasts around the world. *Review of Accounting Studies* 12 (2–3): 377–419.

Dhaliwal, D. S., L. Krull, and O. Z. Li. 2007. Did the 2003 Tax Act reduce the cost of equity capital? *Journal of Accounting and Economics* 43: 121–150.
Doyle, J. T., R. J. Lundholm, and M. T. Soliman. 2003. The predictive value of expenses excluded from pro forma earnings. *Review of Accounting Studies* 8 (2–3): 145–174.
Doyle, J. T., R. J. Lundholm, and M. T. Soliman. 2006. The extreme future stock returns following I/B/E/S earnings surprises. *Journal of Accounting Research* 44: 849–887.
Eberhardt, A., W. F. Maxwell, and A. Siddique. 2008. A reexamination of the tradeo between the future benefit and riskiness of rd increases. *Journal of Accounting Research* 46 (1): 27–52.
Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy* 81 (3): 607–636.
Francis, J., R. LaFond, P. M. Olsson, and K. Schipper. 2004. Costs of equity and earnings attributes. *The Accounting Review* 79 (4): 967–1010.
Francis, J., R. LaFond, P. M. Olsson, and K. Schipper. 2005. The market pricing of accruals quality. *Journal of Accounting and Economics* 39: 295–327.
Francis, J., D. J. Nanda, and P. Olsson. 2008. Voluntary disclosure, earnings quality, and cost of capital. *Journal of Accounting Research* 46 (1): 53–99.
Garfinkel, J. A., and J. Sokobin. 2006. Volume, opinion divergence, and returns: A study of post-earnings announcement drift. *Journal of Accounting Research* 44 (1): 85–112.
Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan. 2001. Toward an implied cost of capital. *Journal of Accounting Research* 39: 135–176.
Givoly, D., C. K. Hayn, and A. Natarajan. 2007. Measuring reporting conservatism. *The Accounting Review* 82 (1): 65–106.
Greene, W. 2003. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.
Hail, L., and C. Leuz. 2006. International differences in the cost of equity capital: Do legal institutions and securities regulation matter? *Journal of Accounting Research* 44: 485–531.
Hann, R. N., F. Heflin, and K. R. Subramanayam. 2007. Fair-value pension accounting. *Journal of Accounting and Economics* 44 (3): 328–358.
Jiang, J. X. 2008. Beating earnings benchmarks and the cost of debt. *The Accounting Review* 83 (2): 377–416.
Kallapur, S., and L. Eldenburg. 2005. Uncertainty, real options, and cost behavior: Evidence from Washington State hospitals. *Journal of Accounting Research* 43 (5): 735–752.
Kothari, S. P., and C. E. Wasley. 1989. Measuring security price performance in size-clustered samples. *The Accounting Review* 64 (2): 228–249.
LaFond, R., and S. Roychowdhury. 2008. Managerial ownership and accounting conservatism. *Journal of Accounting Research* 46 (1): 101–135.
LaFond, R., and R. L. Watts. 2008. The information role of conservatism. *The Accounting Review* 83 (2): 447–478.
Larcker, D. F., and S. A. Richardson. 2004. Fees paid to audit firms, accrual choices, and corporate governance. *Journal of Accounting Research* 42 (3): 625–658.
Lee, Y.-J. 2008. The effects of employee stock options on credit ratings. *The Accounting Review* 83 (5): 1273–1314.

Liu, M., and P. Wysocki. 2007. Operating risk, information risk, and cost of capital. Working paper, Penn State University and Massachusetts Institute of Technology.

Mansi, S. A., W. F. Maxwell, and D. P. Miller. 2004. Does auditor quality and tenure matter to investors? Evidence from the bond market. *Journal of Accounting Research* 42: 755–793.

Mansi, S. A., W. F. Maxwell, and D. P. Miller. 2009. Information risk and the cost of debt capital. *Journal of Accounting Research* (forthcoming).

Mohanram, P., and S. Rajgopal. 2009. Is PIN priced risk? *Journal of Accounting and Economics* 47 (3): 226–243.

Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (3): 703–708.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1): 435–480.

Pittman, J. A., and S. Fortin. 2004. Auditor choice and the cost of debt capital for newly public firms. *Journal of Accounting and Economics* 37 (1): 113–136.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna. 2006. The implications of accounting distortions and growth for accruals and profitability. *The Accounting Review* 81 (3): 713–748.

Roychowdhury, S., and R. L. Watts. 2007. Asymmetric timeliness of earnings, market-to-book and conservatism in financial reporting. *Journal of Accounting and Economics* 44 (1–2): 2–31.

Chen, W., C. Liu, and S. Ryan. 2008. Charateristics of securitizations that determine issuers' retention of the risks of the securitized assets. *The Accounting Review* 83 (5): 1181 – 1215.

Schipper, K., and R. Thompson. 1983. The impact of merger-related regulations on the shareholders of acquiring firms. *Journal of Accounting Research* 21 (1): 184–221.

Sefcik, S. E., and R. Thompson. 1986. An approach to statistical inference in cross-sectional models with security abnormal returns as dependent variable. *Journal of Accounting Research* 24 (2): 316–334.

Shi, C. 2003. On the trade-off between the future benefits and riskiness of R&D: A bondholders' perspective. *Journal of Accounting and Economics* 35 (2): 227–254.

Thompson, S. 2006. Simple formulas for standard errors that cluster by both firm and time. Working paper, Harvard University.

Wang, D. 2006. Founding family ownership and earnings quality. *Journal of Accounting Research* 44 (3): 619–656.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4): 817–838.

White, H. 1984. *Asymptotic Theory for Econometricians*. Orlando, FL: Academic Press, Harcourt Brace Jovanovich.

Wooldridge, J. M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

TABLE 1. Simulation results: Rejection rates

This table reports rejection rates of the true null (β_1=1) at the 1% level from the simulation analysis described in Section 4.0. The simulation uses 1,000 iterations with data for 200 firms and 40 years. Panels A, B, and C depict autocorrelation levels (ρ of 0, 0.5, and 0.8 respectively. Within each panel, cross-sectional correlation takes on values 0, 0.25, 0.50, and 0.75. Actual standard errors refer to the standard deviation of the 1,000 estimates and the numbers for each method refer to the mean of the estimated standard errors using that method. Methods are as defined in Table 9.

Panel A. No autocorrelation (ρ = 0)

Assumptions	Rejection rates (α = 1%)								
Cross-sectional corr.	OLS	N-W	CL-i	CL-t	CL-2	FM-t	FM-NW	FM-i	Z2-t
0.00	1.1%	1.3%	1.3%	0.9%	1.0%	1.7%	1.9%	1.0%	1.6%
0.25	48.2%	49.1%	49.0%	1.3%	1.2%	1.7%	1.9%	48.6%	1.6%
0.50	70.8%	73.1%	74.4%	1.1%	1.1%	1.7%	1.9%	75.8%	1.6%
0.75	80.2%	84.3%	86.5%	1.5%	1.5%	1.7%	1.9%	86.5%	1.6%

Panel B. Autocorrelation (ρ = 0.5)

Assumptions	Rejection rates (α = 1%)								
Cross-sectional corr.	OLS	N-W	CL-i	CL-t	CL-2	FM-t	FM-NW	FM-i	Z2-t
0.00	4.0%	1.2%	0.9%	4.1%	0.7%	4.9%	3.2%	1.1%	4.9%
0.25	47.6%	42.3%	41.9%	1.2%	1.2%	4.9%	3.2%	44.0%	4.9%
0.50	70.6%	70.5%	71.2%	1.1%	1.1%	4.9%	3.2%	72.6%	4.9%
0.75	80.1%	83.9%	85.8%	1.5%	1.5%	4.9%	3.2%	86.3%	4.9%

Panel C. Autocorrelation (ρ = 0.8)

Assumptions	Rejection rates (α = 1%)								
Cross-sectional corr.	OLS	N-W	CL-i	CL-t	CL-2	FM-t	FM-NW	FM-i	Z2-t
0.00	23.0%	1.4%	1.2%	23.2%	1.1%	24.3%	16.0%	0.8%	24.2%
0.25	50.8%	29.3%	28.2%	1.6%	1.3%	24.3%	16.0%	41.3%	24.2%
0.50	69.8%	65.3%	65.2%	1.2%	1.2%	24.3%	16.0%	71.3%	24.2%
0.75	80.4%	83.4%	84.8%	1.6%	1.6%	24.3%	16.0%	87.0%	24.2%

TABLE 2. Simulation results: Standard errors

This table reports actual and estimated standard errors from the simulation analysis described in Section 4.0. The simulation uses 1,000 iterations with data for 200 firms and 40 years. Panels A, B, and C depict autocorrelation levels (ρ of 0, 0.5, and 0.8 respectively. Within each panel, cross-sectional correlation takes on values 0, 0.25, 0.50, and 0.75. Actual standard errors refer to the standard deviation of the 1,000 estimates and the numbers for each method refer to the mean of the estimated standard errors using that method. Methods are as defined in Table 9.

Panel A. No autocorrelation (ρ = 0)

Assumptions	Standard errors										
Cross-sectional	OLS						FM-t			FM-i	
correlation	Actual	OLS	N-W	CL-i	CL-t	CL-2	Actual	FM-t	FM-NW	Actual	FM-i
0.00	0.023	0.022	0.022	0.022	0.022	0.022	0.023	0.022	0.022	0.023	0.023
0.25	0.080	0.022	0.022	0.022	0.078	0.078	0.023	0.022	0.022	0.082	0.023
0.50	0.154	0.022	0.021	0.020	0.152	0.152	0.023	0.022	0.022	0.158	0.020
0.75	0.232	0.022	0.018	0.015	0.229	0.229	0.023	0.022	0.022	0.236	0.016

Panel B. Autocorrelation (ρ = 0.5)

Assumptions	Standard errors										
Cross-sectional	OLS						FM-t			FM-i	
correlation	Actual	OLS	N-W	CL-i	CL-t	CL-2	Actual	FM-t	FM-NW	Actual	FM-i
0.00	0.029	0.022	0.028	0.029	0.022	0.028	0.029	0.022	0.024	0.029	0.029
0.25	0.081	0.022	0.025	0.025	0.079	0.080	0.029	0.022	0.024	0.087	0.026
0.50	0.155	0.022	0.022	0.021	0.153	0.153	0.029	0.022	0.024	0.164	0.022
0.75	0.233	0.022	0.018	0.016	0.230	0.230	0.029	0.022	0.024	0.240	0.016

Panel C. Autocorrelation (ρ = 0.8)

Assumptions	Standard errors										
Cross-sectional	OLS						FM-t			FM-i	
correlations	Actual	OLS	N-W	CL-i	CL-t	CL-2	Actual	FM-t	FM-NW	Actual	FM-i
0.00	0.047	0.022	0.045	0.046	0.021	0.046	0.046	0.021	0.026	0.045	0.045
0.25	0.088	0.022	0.036	0.037	0.081	0.086	0.046	0.021	0.026	0.106	0.034
0.50	0.159	0.022	0.028	0.028	0.156	0.156	0.046	0.021	0.026	0.183	0.025
0.75	0.236	0.022	0.020	0.017	0.232	0.232	0.046	0.021	0.026	0.252	0.016

TABLE 3. Application I: Asset pricing regressions with the *AQ* factor

Regressions of stock returns on hypothesized factors. Panel A presents OLS coefficient estimates and *t*-statistics using White (1980) standard errors. Panel B (Panel C) presents cross-sectional (time-series) Fama-MacBeth coefficients and corresponding *t*-statistics and Z2 statistics. Panel D presents OLS coefficients and *t*-statistics using robust standard errors clustered by month. Inferences are quantitatively similar if standard errors are clustered by year and also by firm. */** indicate statistical significance at the 5/1% level respectively.

Panel A. Pooled OLS with White standard errors.

	Intercept	MKTRF	SMB	HML	AQ
Coeff.	-0.13**	0.93**	0.52**	0.40**	0.27**
t-stat.	-11.03	305.32	90.97	76.01	72.32

Panel B. Firm-specific regressions. Inferences based on average coefficients from 21,107 firm-specific regressions. *t*-stat is the FM-i *t*-statistic, Z2 is the Z2-i statistic

	Intercept	MKTRF	SMB	HML	AQ
Coeff.	-0.36**	0.89**	0.56**	0.33**	0.37**
t-stat.	-15.43	137.02	50.19	29.78	50.41
Z2	-5.16	140.38	94.72	71.05	49.57

Panel C. Annual regressions. Inferences are based on coefficients from 30 annual regressions. *t*-stat is the FM-t *t*-statistic, Z2 is the Z2-t statistic

	Intercept	MKTRF	SMB	HML	AQ
Coeff.	-0.04	0.87**	0.62**	0.27**	0.27**
t-stat.	-0.57	49.83	16.54	7.71	13.62
Z2	-0.59	18.87	12.96	7.74	8.00

Panel D. Pooled OLS with cluster-robust standard errors. Inferences are based on standard errors clustered by month (CL-t).

	Intercept	MKTRF	SMB	HML	AQ
Coeff.	-0.13	0.93**	0.52**	0.40**	0.27**
t-stat.	-1.84	45.67	9.67	11.25	9.27

TABLE 4. Application I: Asset pricing regressions with a random factor

This table presents results from 1,000 iterations of a simulation procedure. Each iteration contains three steps. First, in each month we *randomly* assign all firms with non-missing returns on the CRSP file into quintiles and subtract the equal-weighted return of the 'low' quintile from that of the 'high' quintile. Second, we estimate the following regression

$$R_{i,t} - R^f_t = \alpha + \beta_1 MKTRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RAND_t + \varepsilon_{i,t}\,.$$

Third, we test the significance of *RAND* using each of the following methods: White, FM-i, Z2-i, FM-t, Z2-t, and CL-t. The actual standard error is the standard deviation of the estimate of β_4 and the average estimated standard error for each method is the mean of the 1,000 estimates using that method. Rejection rates for each method refer to the percentage of the 1,000 iterations in which the null of $\beta_4 = 0$ is rejected at each of the 1%, 5%, and 10% nominal levels. Methods are as defined in Table 9.

	Standard error, β_4		Rejection rates (%)		
Method:	Actual	Avg. estimated	α=1%	α=5%	α=10%
White	0.24	0.02	83.50	88.70	90.30
FM-i	0.33	0.04	80.30	85.20	87.00
Z2-i	.	.	83.10	87.40	90.50
FM-t	0.19	0.18	1.20	5.80	11.40
Z2-t	.	.	1.10	5.90	10.90
CL-t	0.24	0.22	0.80	5.30	13.30

TABLE 5. Application II: Cost of capital regressions

This table reports results from generic regressions characteristic of the cost of capital literature. Specifically, we regress cost of equity capital on control variables (*Beta*, *Size*, *BM*, *Volatility*) and measures of each of the following constructs: earnings volatility (*EarnVol*), earnings predictability (*Predict*), idiosyncratic volatility (*IdVol*), and governance quality (*G-Index*). For the Fama-MacBeth (FM-t) approach, we present *t*-statistics using FM-t and Newey-West-corrected (FM-NW) standard errors, as well as the Z2-t statistic. For OLS, we present *t*-statistics using standard errors clustered by year (CL-t) to control for cross-sectional correlation and clustered by firm and year (CL-2) to control for both cross-sectional and time-series correlation. */**/*** denotes significance at the 10/5/1% level. Methods are as defined in Table 9.

Panel A. Base model

Variable	Coeff.	FM-t *t*-statistic	FM-NW *t*-statistic	Z2-t statistic	Coeff.	CL-t *t*-statistic	CL-2 *t*-statistic
Beta	0.52	5.76***	4.68***	5.44***	0.43	3.11***	2.93***
Size	-0.55	-9.75***	-7.67***	-8.21***	-0.63	-10.98***	-9.76***
BM	0.02	1.02	0.88	-0.84	-0.004	-1.38	-1.18
Volatility	0.002	2.93***	2.43**	3.14***	-0.0001	-0.72	-0.62
N = 19,125	t = 1986-2004						

Panel B. Earnings volatility

Variable	Coeff.	FM-t *t*-statistic	FM-NW *t*-statistic	Z2-t statistic	Coeff.	CL-t *t*-statistic	CL-2 *t*-statistic
Beta	0.50	5.99***	4.86***	5.44***	0.43	3.11***	2.93***
Size	-0.55	-9.64***	-7.55***	-8.09***	-0.63	-10.96***	-9.74***
BM	0.02	0.98	0.85	-0.91	-0.004	-1.38	-1.18
Volatility	0.003	3.48***	2.71***	3.89***	-0.0001	-0.57	-0.50
EarnVol	-9.27	-3.65***	-3.44***	-3.72***	-0.21	-0.79	-0.74
N = 19,125	t = 1986-2004						

TABLE 5. (Cont'd)

Panel C. Earnings predictability

Variable	Coeff.	FM-t *t*-statistic	FM-NW *t*-statistic	Z2-t statistic	Coeff.	CL-t *t*-statistic	CL-2 *t*-statistic
Beta	0.50	5.95***	4.83***	5.39***	0.43	3.11***	2.93***
Size	-0.54	-9.63***	-7.54***	-8.08***	-0.63	-10.97***	-9.75***
BM	0.02	0.98	0.85	-0.91	-0.004	-1.38	-1.18
Volatility	0.003	3.43***	2.67***	3.81***	-0.0001	-0.60	-0.53
Predict	-11.15	-3.44***	-3.17***	-3.64***	-0.17	-0.66	-0.61
N = 19,125 t = 1986-2004							

Panel D. Idiosyncratic volatility

Variable	Coeff.	FM-t *t*-statistic	FM-NW *t*-statistic	Z2-t statistic	Coeff.	CL-t *t*-statistic	CL-2 *t*-statistic
Beta	1.36	5.76***	4.99***	8.01***	1.09	7.11***	6.27***
Size	-0.54	-9.82***	-7.65***	-8.21***	-0.63	-11.03***	-9.79***
BM	0.02	1.04	0.90	-0.80	-0.004	-1.38	-1.17
SysVol	-0.01	-2.82***	-2.61***	-3.59***	-0.01	-4.72***	-4.36***
IdVol	0.003	3.48***	2.86***	4.56***	0.0003	1.84*	1.60
N = 19,125 t = 1986-2004							

Panel E. Governance quality

Variable	Coeff.	FM-t *t*-statistic	FM-NW *t*-statistic	Z2-t statistic	Coeff.	CL-t *t*-statistic	CL-2 *t*-statistic
Beta	0.39	3.83***	3.16***	3.71***	0.48	2.67***	2.56**
Size	-0.23	-2.87***	-2.24**	-3.06***	-0.46	-3.91***	-3.72***
BM	3.12	11.70***	9.55***	15.57***	1.50	2.36**	2.36**
Volatility	0.004	4.02***	3.17***	3.85***	-0.0002	-1.45	-1.38
G-Index	0.03	2.87***	2.61***	3.15***	0.03	2.19**	1.44
N = 11,977 t = 1990-2004							

TABLE 6. Application II: Cost of capital regressions with a serially correlated random variable

This table presents results from 1,000 iterations of a simulation procedure. In each iteration, we first randomly generate values of *RAND* for years 1986-2004. Second, we estimate the following equation

$$CoC_{i,t} = \beta_0 + \beta_1\, Beta_{i,t} + \beta_2\, Size_{i,t} + \beta_3\, BM_{i,t} + \beta_4\, Volatility_{i,t} + \beta_5\, RAND_{i,t} + \varepsilon_{i,t}$$

using the full sample examined in Table 5. Third, we calculate standard errors using FM-t, FM-NW, CL-t, and CL-2 methods. Fourth, we test the null of β_5=0 using test statistics calculated using the FM-t, FM-NW, Z2-t, CL-t, and CL-2 methods. The actual standard error is the standard deviation of the estimate of β_3 and the average estimated standard error for each method is the mean of the 1,000 estimates using that method. Rejection rates refer to the percentage of the 1,000 iterations in which the null of $\beta_3 = 0$ is rejected at each of the 1%, 5%, and 10% nominal levels. In Panels A, B, and C, we set the autocorrelation coefficient for *RAND* at 0.2, 0.5, 0.8 respectively. Methods are as defined in Table 9.

Panel A. Autocorrelation (ρ = 0.2)

	Standard error of *RAND*		Rejection rates (%)		
Method:	Actual	Avg. estimated	α=1%	α=5%	α=10%
FM-t	0.022	0.020	2.5	8.8	15.4
FM-NW	0.022	0.020	2.9	9.5	15.7
Z2-t	.	.	2.9	8.6	15.5
CL-t	0.023	0.021	1.3	6.6	13.1
CL-2	0.023	0.022	2.8	7.8	12.5

Panel B. Autocorrelation (ρ = 0.5)

	Standard error of *RAND*		Rejection rates (%)		
Method:	Actual	Avg. estimated	α=1%	α=5%	α=10%
FM-t	0.023	0.017	5.5	15.6	23.2
FM-NW	0.023	0.018	5.3	13.3	21.3
Z2-t	.	.	5.3	15.4	23.4
CL-t	0.024	0.018	4.0	12.3	20.1
CL-2	0.024	0.024	1.5	6.3	9.0

Panel C. Autocorrelation (ρ = 0.8)

	Standard error of *RAND*		Rejection rates (%)		
Method:	Actual	Avg. estimated	α=1%	α=5%	α=10%
FM-t	0.022	0.012	16.0	27.7	35.7
FM-NW	0.022	0.013	13.5	24.6	32.4
Z2-t	.	.	16.9	28.6	36.6
CL-t	0.023	0.013	13.4	25.2	35.2
CL-2	0.023	0.022	1.7	6.7	12.7

TABLE 7. Application III: Cost of debt

This table reports results from ordered logit regressions of S&P credit ratings on control variables (*Size, BM, ROA, Leverage, Beta, Volatility*) and measures of constructs examined in prior research: earnings volatility (*EarnVol*), earnings predictability (*Predict*), research & development (*R&D*), and governance quality (*G-Index*). Following prior research we present unadjusted *t*-statistics (Uncorrected). We also present t-statistics using standard errors clustered by year (CL-t) to control for cross-sectional correlation and clustered by firm and year (CL-2) to control for both cross-sectional and time-series correlation. */**/*** denotes significance at the 10/5/1% level. Methods are as defined in the Table 9.

Panel A. Base Model

Variable	Coef	Uncorrected t-stat	CL-t t-stat	CL-2 t-stat
Size	0.74	86.40***	28.24***	18.84***
BM	0.01	18.58***	3.53***	1.79*
ROA	1.67	11.78***	2.67***	2.52**
Leverage	-2.80	-44.99***	-25.43***	-14.52***
Beta	-0.38	-15.99***	-5.84***	-4.47***
Volatility	-49.06	-44.06***	-15.16***	-10.93***
N = 25,764	t = 1985-2005			

Panel B. Earnings Volatility

Variable	Coef	Uncorrected t-stat	CL-t t-stat	CL-2 t-stat
Size	0.81	76.15***	43.10***	21.19***
BM	0.02	15.71***	4.46***	1.96**
ROA	1.98	9.93***	3.68***	3.28***
Leverage	-2.53	-32.23***	-12.26***	-8.82***
Beta	-0.58	-18.67***	-7.86***	-5.62***
Volatility	-41.18	-26.39***	-6.50***	-5.40**
EarnVol	-1.62	-2.95***	-1.74*	-1.28
N = 18,058	t = 1985-2005			

TABLE 7. (Cont'd)

Panel C. Earnings predictability

Variable	Coef	Uncorrected t-stat	CL-t t-stat	CL-2 t-stat
Size	0.81	76.11***	43.16***	21.19***
BM	0.02	15.71***	4.46***	1.96**
ROA	1.99	9.95***	3.66***	3.27***
Leverage	-2.54	-32.24***	-12.27***	-8.83***
Beta	-0.58	-18.62***	-7.85***	-5.61***
Volatility	-41.51	-26.63***	-6.59***	-5.46***
Predict	-1.24	-2.45**	-1.70*	-1.27
N = 18,058	t = 1985-2005			

Panel D. Research and Development

Variable	Coef	Uncorrected t-stat	CL-t t-stat	CL-2 t-stat
Size	0.77	57.94***	17.86***	12.54***
BM	0.01	13.45***	2.38**	1.34
ROA	1.33	6.32***	1.57*	1.46
Leverage	-2.89	-27.11***	-12.90***	-8.69***
Beta	-0.36	-9.62***	-4.97***	-3.58***
Volatility	-44.21	-27.21***	-10.30***	-7.91***
R&D	-1.47	-3.72***	-2.11**	-1.29
N = 10,819	t = 1985-2005			

Panel E. Governance Quality

Variable	Coef	Uncorrected t-stat	CL-t t-stat	CL-2 t-stat
Size	0.97	69.01***	29.71***	19.89***
BM	-0.02	-1.76*	-1.24	-0.88
ROA	3.02	11.90***	3.84***	3.34***
Leverage	-2.49	-25.81***	-20.22***	-8.96***
Beta	-0.36	-9.47***	-6.93***	-3.67***
Volatility	-71.33	-34.50***	-7.42***	-6.53***
G-Index	0.02	2.81***	2.20**	1.01
N = 12,870	t = 1990-2005			

TABLE 8. Application IV: Conservatism

This table reports results from regressions of earnings scaled by beginning-of-period price on returns, an indicator variable for negative returns, and an interaction term (base model, Panel A). In Panels B-E, we include each of the following variables and interact them with all variables in the base model: market-to-book ratio (*MB*), leverage (*Lev*), probability of informed trade (*PIN*), and managerial ownership (*Mown*). Following Roychowdhury and Watts (2007), we cumulate returns and earnings over current and prior four years and all variables except *D*, *R* and *D*R* are measured using quintile ranks. Following prior research we present *t*-statistics using FM-t and FM-NW, as well as the Z2-t statistic. We also present *t*-statistics using standard errors clustered by year (CL-t) to control for cross-sectional correlation and clustered by industry and year (CL-2) to control for both cross-sectional and time-series correlation. */**/*** denotes significance at the 10/5/1% level. Methods are as defined in the Table 9.

Panel A. Base Model

Variable	Coef	FM-t *t*-stat	FM-NW *t*-stat	Z2-t *t*-stat	Coef	CL-t *t*-stat	CL-2 *t*-stat
D	-0.28	-5.70***	-5.22***	-6.65***	-0.32	-5.36***	-5.87***
R	0.16	11.80***	9.50***	8.55***	0.11	4.10***	3.33***
D*R	0.65	10.60***	8.64***	7.46***	0.61	9.18***	7.72***
N = 92,511 t = 1976-2005							

Panel B. Leverage (N=92,511; t=1976-2005)

Variable	Coef	FM-t *t*-stat	FM-NW *t*-stat	Z2-t *t*-stat	Coef	CL-t *t*-stat	CL-2 *t*-stat
D	-0.27	-3.45***	-3.59***	-5.54***	-0.43	-5.08***	-6.74***
R	0.14	6.23***	5.19***	6.51***	0.03	1.16	1.04
D*R	0.35	3.89***	3.80***	4.30***	0.47	6.19***	4.11***
Lev	0.01	0.57	0.45	-0.75	-0.03	-1.57	-1.08
D*Lev	0.01	0.12	0.11	-0.09	0.05	2.80***	3.22***
R*Lev	0.01	2.31**	2.43**	3.05***	0.03	5.38***	3.97***
D*R* Lev	0.08	3.96***	4.14***	3.28***	0.04	2.09**	1.48
N = 92,511 t = 1976-2005							

TABLE 8. (Cont'd)

Panel C. Market-to-Book Ratio

Variable	Coef	FM-t *t*-stat	FM-NW *t*-stat	Z2-t *t*-stat	Coef	CL-t *t*-stat	CL-2 *t*-stat
D	0.09	0.66	0.67	4.23***	0.13	0.98	0.80
R	0.88	19.27***	17.84***	9.26***	0.80	12.36***	6.60***
D*R	0.10	0.93	0.87	2.26**	0.13	0.96	0.71
MB	-0.04	-2.74***	-2.29**	-2.47**	-0.02	-1.16	-0.96
D*MB	-0.07	-1.76*	-1.71*	-5.82***	-0.09	-2.35**	-1.74*
R*MB	-0.15	-17.03***	-16.75***	-9.04***	-0.15	-13.30***	-6.56***
D*R* MB	0.09	2.54**	2.41**	2.05**	0.08	2.30**	1.60
N = 92,511 t = 1976-2005							

Panel D. Probability of Informed Trade

Variable	Coef	FM-t *t*-stat	FM-NW *t*-stat	Z2-t *t*-stat	Coef	CL-t *t*-stat	CL-2 *t*-stat
D	-0.50	-0.43	-0.45	-3.00***	-0.46	-0.30	-0.47
R	0.06	0.35	0.32	1.05	0.05	0.51	0.47
D*R	-0.39	-0.18	-0.18	-1.85*	1.28	0.37	0.36
PIN	1.57	2.66***	2.92***	3.72***	1.01	2.67***	1.79*
D*PIN	-0.76	-0.90	-1.02	1.21	-0.44	-0.48	-0.48
R*PIN	-0.18	-1.44	-1.73*	-0.32	-0.06	-0.77	-0.71
D*R*PIN	1.56	3.03***	3.11***	4.23***	1.31	1.74*	1.41
N = 41,029 t = 1993-2005							

Panel E. Managerial Ownership

Variable	Coef	FM-t *t*-stat	FM-NW *t*-stat	Z2-t *t*-stat	Coef	CL-t *t*-stat	CL-2 *t*-stat
D	-0.20	-3.02***	-2.81***	-4.06***	-0.24	-5.67***	-5.38***
R	0.02	0.58	0.46	0.57	-0.04	-1.50	-1.06
D*R	0.71	4.99***	5.29***	5.27***	0.68	5.08***	3.56***
MOwn	-0.01	-0.85	-0.76	0.34	0.01	0.02	0.02
D*MOwn	0.02	0.84	0.79	-0.19	-0.01	-0.03	-0.04
R*MOwn	0.03	2.34**	1.86*	2.29**	0.03	4.09***	2.88***
D*R*MOwn	-0.07	-2.17**	-2.33**	-2.85***	-0.06	-1.38	-0.96
N = 19,658 t = 1992-2005							

TABLE 9. Robustness of methods to cross-sectional and time-series dependence

This table summarizes the robustness to cross-sectional and/or time-series dependence of methods commonly used in accounting research to calculate standard errors. A ✓ indicates the method is robust to the indicated form of dependence.

Key to methods: *OLS*—OLS standard errors; *White*—White (1980) standard errors; *N-W*—Newey-West (1987) standard errors; *FM-i*—Fama-MacBeth *t*-statistic based on mean and standard error of cross-section of coefficients from time-series regressions; *FM-t*—Fama-MacBeth *t*-statistic based on mean and standard error of time-series of coefficients from cross-sectional regressions; *FM-NW*—FM-t statistic with Newey-West (1987) correction; *Z2-i*—Z2 statistic based on mean and standard error of cross-section of *t*-statistics from time-series regressions; *Z2-t*—Z2 statistic based on mean and standard error of time-series of *t*-statistics from cross-sectional regressions; *CL-i*—robust standard errors clustered by firm (or industry); *CL-t*—robust standard errors clustered by time; *CL-2*—robust standard errors clustered by firm and time.

	Form of dependence			
Method	None	Cross-sectional	Time-series	Cross-sectional & time-series
OLS	☑	☐	☐	☐
White	☑	☐	☐	☐
N-W	☑	☐	☑	☐
FM-i	☑	☐	☑	☐
FM-t	☑	☑	☐	☐
FM-NW	☑	☑	☐	☐
Z2-i	☑	☐	☑	☐
Z2-t	☑	☑	☐	☐
CL-i	☑	☐	☑	☐
CL-t	☑	☑	☐	☐
CL-2	☑	☑	☑	☑