Evaluating Firm-Level Expected-Return Proxies: Implications for Estimating Treatment Effects

Charles M.C. Lee
Eric C. So
Charles C.Y. Wang

Evaluating Firm-Level Expected-Return Proxies: Implications for Estimating Treatment Effects

Charles M.C. Lee
Stanford University

Eric C. So
Massachusetts Institute of Technology

Charles C.Y. Wang
Harvard Business School

Working Paper 15-022

Evaluating Firm-Level Expected-Return Proxies: Implications for Estimating Treatment Effects[*]

Charles M.C. Lee
Stanford University
Graduate School of Business

Eric C. So
Massachusetts Institute of Technology
Sloan School of Management

Charles C.Y. Wang
Harvard University
Harvard Business School

April 2020

Abstract

We introduce a parsimonious framework for choosing among alternative expected-return proxies (ERPs) when estimating treatment effects. By comparing ERPs' measurement-error variances in the cross-section and in time series, we provide new evidence on the relative performance of firm-level ERPs nominated by recent studies. Generally, "implied-costs-of-capital" metrics perform best in time series; while "characteristic-based" proxies perform best in the cross-section. Factor-based ERPs, even the latest renditions, perform poorly. We revisit four prior studies that use ex-ante ERPs and illustrate how this framework can potentially alter either the sign or the magnitude of prior inferences.

I Introduction

Expected rates of return play a central role in the efficient allocation of scarce resources, and the literature on estimating expected returns is large. But no consensus exists on how firm-level expected-return estimates should be made, or even how the reliability of alternative proxies should be evaluated. As a result, the estimation techniques chosen by researchers vary widely across disciplines and across studies, often without justification or discussion of alternative approaches.

Disagreement over how to estimate firm-level expected equity returns has only increased with the recent proliferation of new expected return proxies (ERPs). One driver of this growth is the development of new asset-pricing models, each of which yields a different theoretical formulation for firm-level expected returns. With each new theoretical formulation, empirical researchers can further vary the input variables used in the estimation process, leading to a seemingly endless set of permutations (see Lewellen (2015) for a helpful discussion of this literature).[1] Without an objective evaluative framework, it is difficult to compare the relative merits of different ERPs, and perhaps even more difficult to assess the reliability of research that employs them.

This study offers a simple evaluation framework to guide researchers in choosing among expected-return proxies (ERPs) for estimating treatment effects. A large and important subset of the ERP literature examines how firms' expected returns are impacted by various "treatments," such as policy shocks and firm attributes. To illustrate, Table 1 summarizes 98 studies, published in top finance and accounting journals between 1997 and 2016, that

---

[1]Typical empirical innovations include different models that connect expected returns to firm attributes, techniques for forecasting earnings or growth, and inclusion of additional firm-level characteristics or asset-pricing factors. For example, while Fama and French (2002) uses a Gordon growth model and dividend-price ratios, Claus and Thomas (2001) and Gebhardt et al. (2001) use a residual-income model and analysts' earnings forecasts to derive expected returns. Subsequent studies have introduced alternative growth forecasts (Easton and Monahan, 2005) and corrections for bias in analysts' forecasts, for example by replacing analysts' forecasts with mechanical earnings forecasts (Hou et al., 2012). Others have derived alternative techniques, for example by cleaning up realized returns (Lewellen, 2015), or by leveraging the pricing of corporate bonds (Campello et al., 2008), dividend derivatives (Van Binsbergen et al., 2013), or stock options (Martin and Wagner, 2019).

feature different implementations of implied-cost-of-capital estimates (ICCs) as expected-return proxies. Most of these studies (77%) examine treatment effects on firms' expected returns, by exploiting either cross-sectional variation (e.g., comparing differences in ERPs across firms for each cross-section) or time-series variation (e.g., comparing differences in ERPs over time for each firm) in the data.

When ERPs are used to study treatment effects, the accuracy of the estimated effect can assume first-order importance. For example, when evaluating the impact of a regulatory shock on firms' expected returns, an estimated increase of 20% would likely have different policy implications than an estimated increase of 2%, even if both are statistically significant. We argue that, in this setting, an ERP's measurement-error variance (MEV) is the key distributional property of interest: minimizing MEV is both necessary and sufficient for identifying the best-performing ERPs. Optimizing ERP performance on alternative metrics—such as mean-squared errors, return-regression slope coefficients, or correlations with ex-post realized returns—is *either* not necessary *or* not sufficient, and can result in misleading inferences.

Our analyses result in a two-dimensional framework for evaluating the quality of ERPs in treatment-effect studies. First, when the research design depends primarily on cross-sectional variation in the data, such as Fama-MacBeth regressions or panel regressions with year-fixed effects, researchers should utilize ERPs with minimum cross-sectional MEV. These ERPs are preferable in that differences in the ERP *between* firms in each cross-section tracks the differences in expected returns. Second, when the research design depends primarily on time-series variation in the data, such as time-series regressions or panel regressions with firm-fixed effects, researchers should utilize ERPs with minimum time-series MEV. These ERPs are preferable in that variation in a given firm's ERP *over time* will track intertemporal variations in its expected returns. Our analyses show these two dimensions of ERP performance are not redundant. Therefore, given a set of alternative ERPs, the preferred candidate will depend, in part, on the relative importance a researcher assigns to each dimension.

We also introduce an empirical methodology for implementing this evaluative frame-

work. We show that, although measurement errors—the ERP's deviation from a firm's true expected returns—are unobservable, it is possible to derive empirical estimates that allow comparison of cross-section and time-series MEVs between ERPs. Our approach relies on the observation that news—the difference between realized returns and expected returns—is, by definition, not systematically predictable. This methodology is robust in that it makes no assumptions about the properties of ERPs' measurement errors other than finite variance.

We implement this framework by evaluating the relative performance of 17 ERPs, which we group into four families or types. These ERP families differ in conceptual origin (see Appendix I) and empirical construction (see Appendix II). The first family is derived from factor models (e.g., Fama and French, 2015; Hou et al., 2015); in general, a factor-based ERP is a linear combination of expected factor premiums, where the weights are based on a firm's factor sensitivities. The second family is firm-characteristic-based ERPs (e.g., Lewellen, 2015; Lyle and Wang, 2015). Characteristic-based ERPs are based on a linear combination of firm characteristics, where the weights derive from the historical cross-sectional relations between realized returns and characteristics. The third class of ERPs is the set of "implied-cost-of-capital" (ICC) measures, defined as the internal rates of return that equate firms' market values to the present value of their expected future earnings, based on a valuation model chosen by the researcher. Finally, we introduce a new ERP prototype—a hybrid of implied-cost-of-capital measures and characteristic-based ERPs—that we call "fitted" ICCs. Like characteristic-based ERPs, fitted ICCs are linear combinations of firm characteristics, with the difference that the weights are based on the historical cross-sectional relations between ICCs and characteristics.

Our empirical analyses begin by analyzing representative candidates from the four classes of ERPs. We denote the representative factor-based, characteristic-based, implied-cost-of-capital, and fitted implied-cost-of-capital ERPs as FER, CER, ICC, and FICC respectively. Our analyses of these candidates suggest that, in the cross-section, CER is the most reliable: CER exhibits significantly lower mean cross-sectional MEVs than other representative ERPs,

most of which perform no better than a trivial ERP (i.e., a fixed constant). In time-series, on the other hand, implied-cost-of-capital-based ERPs (ICC and FICC) are the most reliable. These measures exhibit the lowest mean time-series MEVs, while the other representative ERPs perform no better than a trivial ERP. In both cross-section and time series, FER is the least reliable, producing high MEV worse than that of a trivial ERP.

These broad conclusions are robust in subsamples based on market capitalization and in both halves of our sample period. However, the relative performance of individual ERPs vary across firm types and over time. For example, in the cross-section, FICC is the best-performing ERP among tiny firms—those whose market capitalization at the beginning of the year falls in the bottom 20th percentile of the NYSE distribution—edging out CER. In terms of time-series performance, CER performs better in the second half of our sample than in the first half (it outperforms the trivial estimator over recent years), but ICC and FICC remain the best performers in both subsample periods.

We also use the evaluation framework to analyze individual ERP variants within each family, in an effort to understand the sources of noise. For example, we examine the performance of seven factor-based ERPs, ranging from a single (market) factor model to multi-factor models consisting of up to six factors. Our findings attribute the poor performance of factor-based ERPs to noise from both estimating firm-level factor sensitivities and expected factor premiums using realized returns. These estimation errors compound with additional factors: generally, ERPs based on more factors exhibit greater MEVs. We tested several common noise-reduction techniques from prior studies, such as industry-level aggregations of factor sensitivities and rolling factor premium estimates. Although these approaches reduce MEVs to an economically significant degree, *none* of those we tested produced a factor-based ERP that outperforms a trivial ERP. A central insight from these analyses is that superior factor-based models—those that perform well at explaining in-sample variation in realized returns (e.g., Hou et al., 2019)—do not necessarily translate into ex-ante ERPs well suited to estimating treatment effects.

We also examine the relative performance of the most common ICC variants, which differ either in expected future earnings estimates (i.e., from analysts' forecasts or from a mechanical earnings-forecast model) or in the valuation model that links expected returns to market prices and expected future earnings (i.e., the residual-income or abnormal-earnings-growth valuation model). Although most ICCs are unreliable in the cross-section, there are standouts. In particular, the variant proposed by Gebhardt et al. (2001) produces average cross-sectional MEVs that outperform a trivial ERP whether we use analyst forecasts or mechanical earnings forecasts. ICCs perform markedly better in time series, with nearly every ICC variant outperforming the trivial ERP.

Our final analysis re-visits three lines of inquiry central to the risk-reward trade-off in financial economics, for which ICCs have been used in cross-sectional regression designs to resolve mixed evidence and debate in prior research. We re-evaluate the relation between expected returns and a) default risk (Chava and Purnanandam, 2010), b) idiosyncratic volatility (Fu, 2009), and c) information risk (Francis et al., 2004; Chen et al., 2015). In light of our findings that ICCs, in general, should be demoted in favor of characteristic-based ERPs in cross-sectional designs, we revisit these research question, replicate the key cross-sectional results that prior research establishes using common ICC variants, and examine the robustness of these findings to the use of CER.

Our re-evaluation of these prior studies underscores two key messages from our paper. First, choosing ERPs with the lowest MEV can sharpen inferences on the magnitude of the treatment effect of interest. For example, we find the positive relation between expected returns and distress risk is robust to the use of CER, but we document a significantly lower estimated premium (about 40% lower) compared to estimates based on ICCs. Second, choosing ERPs with the lowest MEV can overturn researchers' inferences. For example, when we re-examine the pricing of idiosyncratic volatility and information risks, we find the statistical significance of the premiums estimated in prior research fails to hold when using ERPs nominated by our framework.

In sum, prior studies of ERP performance focus almost exclusively on cross-sectional tests and report mixed results. We advance this literature by (a) providing a parsimonious methodology for ERP selection when estimating treatment effects, (b) introducing a time-series dimension to ERP performance evaluation, (c) presenting new evidence on the relative performance of common ERP variants, and (d) illustrating the importance of ERP selection (based on the MEV framework) for researchers' inferences.

In the concluding section of the paper, we discuss how ex-ante ERPs (such as ICCs) have been used in finance, as well as why, despite recent progress, significant opportunities for broader usage remain. We also provide guidance on the alternative applications our evaluative framework.

II   Theoretical Underpinnings

This section explains why minimizing MEV in ERPs, either in the cross-section or in time series, is desirable for accurately estimating treatment effects. We also explain how minimizing MEV differs from alternative criteria for assessing ERPs.

Definitions and Assumptions

We begin with a simple decomposition of realized returns:

$$r_{i,t+1} = er_{i,t} + \delta_{i,t+1}, \tag{1}$$

where $r_{i,t+1}$ is firm i's realized return in period $t + 1$, $er_{i,t}$ is the firm's true but unobserved expectation of future returns conditional on publicly available information at time t, and $\delta_{i,t+1}$ is the firm's unanticipated news or forecast error.[2] By the prediction property of conditional expectations (e.g., Angrist and Pischke, 2008), $er_{i,t}$ is an optimal forecast that captures all

---

[2]Campbell (1991), Campbell and Shiller (1988a,b), and Vuolteenaho (2002) analyze the relative importance of cash flow and the expected-return component of news; their work requires a log-linearization that is unnecessary for our purposes. Eq. (1) follows directly from the decomposition property of conditional expectations (e.g., Angrist and Pischke, 2008).

ex-ante predictability (on the basis of the information set) in returns; it is therefore not systematically correlated with its forecast errors $(\delta_{i,t+1})$ in time series or in the cross-section. Put differently, news cannot be predictable.

An ERP $(\widehat{er}_{i,t})$ is defined as the expected return $(er_{i,t})$ measured with error $(\omega_{i,t})$:

$$\widehat{er}_{i,t} = er_{i,t} + \omega_{i,t}. \tag{2}$$

In concept, $\widehat{er}_{i,t}$ can be any ex-ante expected-return measure, including such simple characteristics as a firm's beta, book-to-market ratio, market capitalization, or some linear combination of these characteristics. The key is that the "true" expected return, whatever it may be, is not observable. What we can observe are empirical proxies that contain measurement error, which can vary across firms and over time, and may be correlated with firm characteristics (i.e., non-classical). The only assumption we impose is that the measurement errors have finite variances.

Differences between alternative ERPs are therefore reflected in the distributional properties (time-series and cross-sectional) of their $\omega_{i,t}$ terms. Consequently, the relative desirability of one ERP over another is, in essence, an expression of preference with regard to their measurement-error properties. The most relevant distributional properties, in turn, depend on their intended use.

Minimizing Measurement-Error Variances

For the purposes of accurately estimating treatment effects on expected returns—the predominant use-case of ERPs in the literature (see Table 1)—we contend that the single most important distributional property in selecting ERPs is their measurement-error variance (MEV). The most preferable ERPs are those that exhibit minimal MEV. The mean of the measurement error (i.e., the bias in the ERP), on the other hand, is irrelevant.

To illustrate the importance of MEV in this research context, consider the following example. Suppose we wish to assess the treatment effect of a corporate action on cross-sectional expected returns. Without loss of generality, assume that these expected returns

are generated from a simple stochastic process,

$$er_i = \alpha + \beta x_i + \epsilon_i, \tag{3}$$

where x_i is an action of interest (e.g., a firm characteristic that can be chosen by the manager), and β is the treatment effect parameter of interest. The researcher assumes standard OLS assumptions hold for (x_i, ϵ_i), and obtains a sample of observations $\{\widehat{er}_i, x_i\}_{i=1}^N$ to estimate β.

If the researcher observes expected returns perfectly, the OLS estimate ($\hat{\beta}[er]$) is consistent for the true parameter and asymptotically normally distributed:

$$\sqrt{N}\left(\hat{\beta}[er] - \beta\right) \longrightarrow_d \mathbb{N}\left(0, \frac{\sigma_\epsilon^2}{\sigma_x^2}\right), \tag{4}$$

where $\longrightarrow_d$ denotes convergence in distribution and σ_ϵ^2 and σ_x^2 denote the variances in ϵ_i and x_i. However, researchers' use of an ERP implies that, in practice, the following specification is being estimated in lieu of Eq. (3):

$$\underbrace{er_i + \omega_i}_{\widehat{er}_i} = \alpha + \delta x_i + \epsilon_i. \tag{5}$$

Thus, treatment effects estimated using the ERP obtains the following distribution:

$$\sqrt{N}\left(\hat{\delta}[\widehat{er}] - \beta\right) \longrightarrow_d \mathbb{N}\left(\frac{\sigma_{x,\omega}}{\sigma_x^2}, \frac{\sigma_\epsilon^2 + \sigma_\omega^2}{\sigma_x^2}\right), \tag{6}$$

where $\sigma_{x,\omega}$ denotes the covariance between x_i and ω_i, and σ_ω^2 denotes the ERP's MEV.

Eq. (6) shows the accuracy of the treatment-effect estimate ($\hat{\beta}[erp]$)—its precision and bias—depends on the ERP's MEV. As MEV reaches its lower bound (i.e., zero), we ensure best-possible precision in the OLS estimate (because standard errors are minimized when $\sigma_\omega^2 = 0$) as well as consistency of the treatment-effect estimate (because $\sigma_{x,\omega} = 0$ when $\sigma_\omega^2 = 0$). Moreover, Eq. (6) shows any (constant) bias in the ERP (i.e., a non-zero mean

in ω) does not affect the distributional properties of the treatment-effect estimate and is therefore irrelevant. Intuitively, this is because any bias in the measurement error is picked up by the intercept term ($\hat{\alpha}$), which is typically unimportant to the researcher.

Accordingly, researchers interested in accurately estimating treatment effects on expected returns should focus on the MEV properties of ERPs. Specifically, depending on the research design, researchers should select ERPs with minimum cross-sectional or time-series MEV. An ERP that perfectly tracks true expected return would exhibit zero MEV both in time-series and in the cross-section.

The Two-Dimensional Evaluative Framework

We stress that an ERP's performance in cross-sectional and time-series MEVs need not be redundant. The same ERP that performs well in time series may perform poorly in the cross-section. For example, an ERP can have firm-specific measurement errors that are constant across time, resulting in zero time-series MEV, but these measurement errors can still obscure the cross-sectional ordering of expected returns across firms. Analogously, an ERP that performs well in the cross-section may perform quite poorly in time series. For example, an ERP can have time-specific measurement errors that are constant across firms but vary over time, resulting in zero cross-sectional MEV but substantial time-series MEV.[3]

In general, for research designs that seek to identify treatment effects by exploiting cross-sectional variation in the data, such as Fama-MacBeth regressions or panel regressions with

---

[3]Consider two stocks, A and B, with constant true expected returns of 10 percent and 2 percent respectively. Suppose that a particular ERP model produces expected-returns estimates of 2 percent and 10 percent for stocks A and B respectively. Such an ERP produces a zero time-series MEV for both stocks, since the measurement errors are constant across time for each firm, but such an ERP mis-orders the stocks' expected returns in the cross-section and results in cross-sectional MEV. Analogously, suppose again that the true expected returns of stocks A and B are always 10 and 2 percent respectively. Now consider an ERP model that estimates expected returns for A and B to be 13 and 5 percent in some years and 10 and 2 percent in others. This ERP always exhibits constant measurement errors for each firm in the cross-section, and thus accurately reflects the differences in expected returns between the two firms (i.e., zero MEV in each cross-section). However, the time-series variation in the ERPs does not reflect variations in true expected returns; instead it reflects variation in measurement errors.

year-fixed effects, the class of best-performing ERPs would have the following structure:

$$\widehat{er}_{i,t} = a_t + er_{i,t}. \tag{7}$$

Although ERPs satisfying such a structure contain error, the error term is common across all firms in each cross-section and is absorbed by the regression constant or year-fixed effects. These ERPs are preferable in that differences in the ERPs *between* firms in each cross-section track the differences in expected returns. Notably, ERPs satisfy such a structure if, and only if, they exhibit zero cross-sectional MEV.

Analogously, for research designs that seek to identify treatment effects of interest by exploiting time-series variation in the data, such as time-series regressions or panel regression designs with firm-fixed effects, the best-performing ERPs have the following structure:

$$\widehat{er}_{i,t} = a_i + er_{i,t}. \tag{8}$$

Here, the firm-specific measurement error is absorbed by the regression constant or firm-fixed effects. These ERPs are preferable in the sense that differences in a firm's ERP over time track the differences in its expected returns. Notably, ERPs satisfy such a structure if, and only if, they exhibit zero time-series MEV.

Relation to Alternative Evaluative Criteria

We argued above that achieving the best "score" in terms of MEV—the minimum of zero—is necessary and sufficient to identify the best-performing ERPs for treatment-effects studies. This section shows alternative criteria used in the literature are suboptimal in that achieving the best possible score is *either* not necessary *or* not sufficient to identify best-performing ERPs. Consider minimizing mean-squared errors (MSE), the standard criterion in statistics for assessing the quality of estimators. MSE incorporates two aspects of an

ERP's measurement-error properties: bias and variance.[4] Because MSE conflates variance and bias, which is irrelevant in treatment-effects applications, achieving a minimum MSE of zero is not a necessary condition for identifying best-performing ERPs (e.g., Eq. 7 and 8).

Another commonly used evaluation technique is the regression slope-coefficient test. As implemented in the literature (e.g., Greenwood and Shleifer, 2014; Lewellen, 2015), this test typically involves cross-sectional regressions of future realized returns on each ex-ante ERP. An ERP with an estimated slope coefficient that is closer to 1 is generally deemed superior to one whose estimated slope is more distant from 1.[5]

Although a perfect ERP exhibits a return-regression slope coefficient of 1, obtaining a slope coefficient of 1 is not sufficient to identify best-performing ERPs, because ERPs (e.g., with non-classical measurement errors) can still exhibit a slope of 1 even when they contain substantial MEV.[6] The intuition is that the magnitude of the regression slope coefficient is sensitive to re-scaling the ERP. That is, a re-scaled ERPs can yield slope coefficients of 1 even when substantial measurement errors remain.

Another conceptual alternative is to select the ERP that maximizes correlation with future realized returns. Here again, although the perfect ERP would achieve a correlation of 1 with expected returns, achieving a correlation of 1 does not ensure that an ERP is best-performing. This is because correlations are invariant to linear transformations; thus an ERP of the form $\widehat{er}_i = 100 \times er_i$ would be considered perfect using this criterion. From the perspective of accurately estimating treatment effects, such an ERP would clearly paint a misleading picture of the magnitude of the treatment effect. Maximum correlation with expected (or realized) returns is thus also insufficient to identify best-performing ERPs.

To circumvent the issues with the slope-coefficient or return correlation tests, researchers may of course supplement the tests with additional criteria, such as matching on additional

---

[4]Recall that $\mathbb{E}\left[(er_t - \widehat{er}_t)^2\right] = \mathbb{E}\left[(\omega_t)^2\right] = Var(\omega_t) + [\mathbb{E}(\omega_t)]^2$.

[5]Although we have never seen its rationale articulated, obtaining a slope coefficient higher than 1 or a slope coefficient that is equidistant to but lower than 1 are presumably equally harmful to an ERP's cause.

[6]The Technical Appendix illustrates this point by providing an example of a set of ERPs with slope coefficients of 1 that vary significantly in terms of MEV.

distributional moments of realized returns. However, it is unclear which moments to match and how to weight these additional distributional properties, particularly in research settings where the estimation of treatment effects is of first-order importance. The focus on MEVs, by contrast, is a single parsimonious criterion that, when achieving a lower bound of 0, is necessary and sufficient for identifying the best-performing ERPs in treatment-effect studies.

III   Methodology

This section shows that although ERPs' measurement errors are not directly observable, it is possible to derive an empirically estimable measure to compare alternative ERPs in terms of their average cross-sectional and time-series MEVs. For brevity, we summarize only the key results here. Most of the technical details appear in the Technical Appendix.

Comparing ERPs' Time-Series Measurement-Error Variance

We begin by comparing ERPs' time-series MEVs. The Technical Appendix shows that, under reasonable assumptions (i.e., that news is not predictable and thus uncorrelated with the level of expected returns), we can express firm i's time-series MEV as:

$$Var_i(\omega_{i,t}) = Var_i(\widehat{er}_{i,t}) - 2Cov_i(r_{i,t+1}, \widehat{er}_{i,t}) + Var_i(er_{i,t}), \tag{9}$$

where $Var_i(\widehat{er}_{i,t})$ is the time-series variance of a given ERP for firm i, $Var_i(er_{i,t})$ is the time-series variance of firm i's expected returns, and $Cov_i(r_{i,t+1}, \widehat{er}_{i,t})$ is the time-series covariance between a given ERP and realized returns for firm i in period $t + 1$.

The first term on the right-hand side indicates the (time-series) variance of a given ERP's measurement error is increasing in the variance of the ERP. This is intuitive: as the time-series variance of measurement errors of a given ERP for firm i increases, all else equal, so will the observed variance of the ERP.

The second term on the right-hand side indicates the MEV for a given ERP is decreasing in the covariance of the ERP and future returns. This is also intuitive: to the extent that

a given ERP consistently predicts variation in future returns for the same firm, time-series variation in that proxy is more likely to reflect variation in the firm's true expected returns than in measurement errors.

Finally, notice the third term is the time-series variance of firm i's true (but unobserved) expected returns. For a given firm, this variable is constant across alternative ERPs and thus we need only the first two terms of Eq. (9) to determine which ERP exhibits lower time-series variance in measurement errors. Accordingly, we define the sum of the first two terms of Eq. (9) as a firm's Scaled TS Variance:

$$SVar_i(\omega_{i,t}) = Var_i(\widehat{er}_{i,t}) - 2Cov_i(r_{i,t+1}, \widehat{er}_{i,t}). \tag{10}$$

Note that $SVar_i$ can be a negative number if the covariance between the ERP and future realized returns is positive and sufficiently large.

In our empirical tests, we compute for each ERP and each firm the scaled error variance measure using Eq. (10), and then assess the time-series performance of ERPs based on the average of $SVar_i$ across the N firms in our sample:

$$AvgSVar^{TS} = \frac{1}{N} \sum_i SVar_i(\omega_{i,t}). \tag{11}$$

All else equal, the time-series variation of ERPs with lower $AvgSVar^{TS}$ is deemed to be more informative about the time-series variation in expected returns. $AvgSVar^{TS}$ facilitates *relative* comparisons of ERPs' MEVs. If we make assumptions about the time-series variance in firm-level expected returns, such as by imposing additional structure about the time-series behavior of expected returns, it is possible to obtain an empirically estimable absolute measure of the time-series MEV.

Comparing ERPs' Cross-Sectional Error Variance

It is also possible to derive an empirically estimable measure that allows us to compare the cross-sectional MEVs of alternative ERPs. The Technical Appendix derives the following measure for comparing ERPs' average cross-sectional MEV:

$$AvgSVar^{CS} = \frac{1}{T} \sum_t Var_t(\widehat{er}_{i,t}) - 2Cov_t(r_{i,t+1}, \widehat{er}_{i,t}). \tag{12}$$

Similar to the intuition for the time-series case, Eq. (12) indicates that, all else equal, an ERP's average cross-sectional MEV is increasing in the cross-sectional variance of the ERP and decreasing to the degree that ERPs predict future returns in the cross-section. Thus, the cross-sectional variation of ERPs with lower $AvgSVar^{CS}$ is deemed to be more informative about the cross-sectional variation in expected returns.[7]

Interpreting the magnitude of $AvgSVar^{CS}$ and $AvgSVar^{TS}$

Because the variance in expected returns is ignored in computing $AvgSVar^{CS}$ and $AvgSVar^{TS}$, these measures are bounded below not by zero but by $-Var(er)$. Nevertheless, zero remains a natural reference point in our framework because a trivial estimator—one that specifies the ERP as a constant for all firms in the cross-section and over time—produces $AvgSVar^{CS}$ and $AvgSVar^{TS}$ of 0. Thus, when examining a new ERP, a researcher can readily determine whether it constitutes an MEV improvement over a trivial estimator: those ERPs that perform better than a constant will have negative $AvgSVar^{CS}$ or $AvgSVar^{TS}$.

Additionally, because $AvgSVar^{CS}$ and $AvgSVar^{TS}$ measure average cross-sectional and time-series MEVs only up to a constant, these metrics will accurately reflect the *relative* performance of different ERPs in terms of MEVs. However, they do not tell us the magnitude of the improvement (in terms of percentage reduction in MEV) of one ERP over another without making further assumptions about the variance of true expected returns. In the next

---

[7]The cross-sectional MEV metric in this study is similar in spirit to that of Easton and Monahan (2005) ("EM"), but our approach differs from theirs in several ways. First, EM's framework is based on stricter assumptions, making it more difficult to apply their methodology to compare broad classes of ERPs. Specifically, the EM approach requires an ERP's measurement errors to be uncorrelated with true expected returns, disqualifying large classes of ERPs from consideration. Second, our approach circumvents the requirement of the EM framework to estimate multiple firm-specific and cross-sectional parameters (e.g., cash-flow news); ours is therefore much simpler to implement empirically. Finally, our evaluation focuses on cross-sectional *and* time-series MEVs, thus allowing researchers to more comprehensively assess the relative performance of ERPs and to identify the ERPs most suitable for a given research design. In sum, our study builds on EM by (a) extending the measurement-error approach to the time-series dimension, (b) simplifying the empirical estimation of ERPs' MEVs, and (c) broadening the applicability of this approach to broader classes of ERPs.

section, as part of our empirical analysis, we will impose conservative assumptions about the variance of expected returns to derive an estimate of the percentage improvement of one ERP over another (in terms of MEV reduction).

IV   Assessing the Performance of Representative ERP

We now implement the two-dimensional framework to evaluate the relative performance of four types of ERPs: ICC-based, characteristic-based, factor-based, and fitted-ICC-based. Collectively, these four types encompass the major prototype classes of ERPs nominated by the academic finance and accounting literature over the past 50 years. We will first define and compute a representative ERP from each type; then we will report their MEV properties to understand their reliability for different research designs.

Representative ERPs and Descriptive Statistics

We begin by computing "state-of-art" representatives of three of the classes of ERPs: ICC-based, characteristic-based, and factor-based. The representative ICC we consider, denoted ICC, is a commonly used composite that takes the equal-weighted average of four commonly used ICC variants: the residual-income-model-based ICCs proposed by Gebhardt et al. (2001) (GLS) and Claus and Thomas (2001) (CAT) and the abnormal-earnings-model-based ICCs proposed by Easton (2004) (PEG) and Ohlson and Juettner-Nauroth (2005) (AGR). We use as inputs the mechanical earnings forecasts from the cross-sectional forecast model of Hou et al. (2012), which facilitates the computation of ICCs for a large cross-section of firms (including those without analyst coverage).

The representative characteristic-based ERP we consider, denoted CER, is a composite that takes the equal-weighted average of two variants. Each variant is constructed as a linear combination of firm characteristics, where the weights are estimated by historical cross-sectional regressions of realized returns on those characteristics. We implement an

ERP from Lewellen (2015), denoted JLR, based on three firm characteristics: the log of market capitalization (size), the log of book-to-market (bm), and cumulative stock returns from twelve months to two months prior to the forecast date (momentum). The other ERP we implement, denoted LPV, is taken from Lyle and Wang (2015) and Chattopadhyay et al. (2020), and is also based on three firm characteristics: bm, the log of gross return on equity (roe), and the mean of prior-month daily squared returns (vol).

The representative factor-based ERP we consider, denoted FER, is a composite that takes the equal-weighted average of two variants that incorporate the latest innovations from this literature. The first factor-based ERP variant, q^5, is derived from the q-factor model (Hou et al., 2015), augmented with an expected growth factor (Hou et al., 2019); the second variant, FF6, is derived from the Fama and French (2015) five-factor model, augmented with a momentum factor (Fama and French, 2018). Each of these ERPs is constructed as a linear combination of estimated factor premiums, where the weights are a firm's factor sensitivities, estimated by historical time-series regressions of realized returns on factor returns.

Finally, we introduce a new ERP prototype by computing a "fitted" ICC measure, based on the systematic cross-sectional relation between ICCs and firm characteristics. These ERPs can be thought of as a cleaned-up version of ICCs, where the noise in ICC estimates is filtered through a linear projection onto firm characteristics. Our representative "fitted" ICC, denoted FICC, derives from our representative ICC measure and is based on a linear combination of three firm characteristics: size, bm, and momentum.

We compute these ERPs as of the last trading day of each month. To do so, we obtain market-related data on all U.S.-listed firms (excluding ADRs) from the Center for Research in Security Prices (CRSP) and annual accounting data from Compustat for the period 1977-2018. For each firm-month, we estimate ERPs using data from the CRSP Monthly Stock file and, when applicable, firms' most recently available annual financial statements.[8] To be included in our main sample, each firm-month observation must include information on stock

---

[8]We assume that financial statements are available four months after the fiscal year-end date.

price, shares outstanding, book values, earnings, dividends, and industry identification (SIC) codes. We also require each firm-month observation to include valid, non-missing values for each of our four representative ERPs. To exclude the influence of extreme observations, each ERP is winsorized at the 5th and 95th percentiles of the pooled-sample distribution.[9]

Table 2 reports annual means and standard deviations and their distributions for four monthly ERPs from 1977 to 2018. The annual mean ERP values vary across the ERP types. For example, the time-series average of the monthly mean ERP values range from 1.09 percent (for ICC) to 1.49 percent (for FER). There are also differences in variability across ERP types. For example, the time-series average of the annual standard deviation of monthly CER values is 0.37%, less than one-tenth that of FER (5.65%). In general, FER appears noisier than its non-factor-based counterparts, which likely reflects the instability of the estimated factor risk premiums or their corresponding sensitivities.

Table 3 reports the time-series average correlation coefficient among the four representative ERPs: above (below) the diagonal are the average monthly Pearson (Spearman) correlations. The table shows the ICC, CER, and FICC—the non-factor-based proxies— are significantly and positively correlated. CER and FICC exhibit the highest correlation coefficient (64.7%), followed by the correlation between ICC and FICC (62.7%) and the correlation between ICC and CER (45.7%). The relatively high correlation between CER and FICC is unsurprising given that each depends, at least in part, on a linear combination of size, bm, and momentum. Likewise the relatively high correlation between ICC and FICC is also unsurprising given that FICC is based on a linear projection of ICC. In contrast, ICC, CER, and FICC all exhibit negative and significant correlations with FER. Moreover, the correlation coefficients are relatively small, ranging from -3.7% to -6.9%.

---

[9]In our Online Appendix, we also conduct our main tests by trimming extreme values of ERPs based on the cross-sectional distribution each calendar month instead of winsorizing using the pooled distribution. Our main inferences are qualitatively unchanged.

Comparison of Representative ERPs' MEVs

We compare cross-sectional MEVs using this data on the four ERPs. Table 4, Panel A, reports the summary statistics of each ERP's monthly cross-sectional SVARs ($Var_t(\widehat{er}_{i,t}) - 2Cov_t(r_{i,t+1}, \widehat{er}_{i,t})$).

For ease of reporting, we multiply SVAR estimates by 100. Table values in this panel represent distributional statistics of SVAR computed across the 497 months or cross-sections in our 1977-2018 sample period. Table 4, Panel B, reports t-statistics based on Newey-West-corrected standard errors (using 12-month lags) corresponding to the pairwise comparisons of cross-sectional SVARs within the sample of 497 months used in Table 4, Panel A. The reported values are negative (positive) when the ERP displayed in the leftmost column has a larger (smaller) scaled MEV than the ERP displayed in the topmost row.

The two panels of Table 4 indicate the three non-factor-based proxies (ICC, CER, and FICC) generate significantly lower average cross-sectional MEV than the factor-based ERP (FER). Overall, CER outperforms the remaining ERPs, with lower MEVs than ICC, FICC, and FER. On the other hand, FER is the worst-performing ERP, with a mean SVAR that is positive and significantly larger than that of ICC, CER, or FICC.

Moreover, CER is the only ERP with an average cross-sectional SVAR that is negative and significant; thus it is the only ERP that outperforms a trivial ERP (i.e., a constant) in terms of cross-sectional MEV. Although ICC and FICC outperform FER, neither outperforms the trivial ERP in the cross-section: their mean cross-sectional SVARs are not statistically different from 0 (the SVAR of a constant). Our findings therefore suggest that, although many studies exploit the cross-sectional variation in ICCs to study treatment effects on expected returns, ICCs in general are unreliable in the cross-section, consistent with the conclusions of Easton and Monahan (2005).

The second dimension of our framework evaluates ERPs on the basis of their time-series MEVs. Table 5, Panel A, reports the summary statistics of each ERP's time-series SVARs ($Var_i(\widehat{er}_{i,t}) - 2Cov_i(r_{i,t+1}, \widehat{er}_{i,t})$) in Table 5; Panel B reports the pairwise comparisons. To

ensure the stability of our time-series SVAR estimates, we require firms to have a minimum of 20 (not necessarily consecutive) months of data over our sample period. Table values in this panel represent distributional statistics of time-series SVARs (multiplied by 100) across 11,620 unique firms.

The results in Table 5 show the relative performance of ERPs in time series differs from the cross-section. ICC and FICC exhibit the lowest MEV in time series, followed by CER and FER. Only ICC and FICC outperform the trivial ERP in time series, with average SVARs that are negative and statistically significant. The mean SVAR of CER is negative but not statistically significant. As for FER, its mean time-series SVAR is positive and statistically significant, suggesting that factor-based ERPs are noisy both in the cross-section and in time series.

In evaluating the relative performance of ERPs, we also explore mean squared errors (MSEs) and slope coefficients as alternative criteria (results not tabulated for parsimony). When employing MSEs, we find no significant difference in the performance of ICC, CER, and FICC. Similarly, although slope-coefficient tests also nominate CER as the best-performing ERP in cross-sectional tests, we find no difference in performance between ICC and FICC. Both of these results contrast with our findings in Table 4, thus underscoring that the implementation of our evaluative framework can change researchers' inferences about the relative performance of ERPs.

Subsample Analyses

Next we examine the robustness of these performance results in various subsamples in Table 6. Following Lewellen (2015), we define four subsamples based on beginning-of-the-calendar-year market capitalization: large (larger than the NYSE median), all-but-tiny (above the 20th percentile of the NYSE), small (below the NYSE median), and tiny (below the 20th percentile of the NYSE) firms. The distributional summary statistics of cross-sectional and time-series SVARs of the representative ERPs in these subsamples are reported

in Table 6, Panels A, B, C, and D respectively.

The cross-sectional performance of ERPs across size subsamples are similar to Table 4. In each subsample, CER exhibits a significantly negative average cross-sectional SVAR, and FER exhibits a significantly positive average cross-sectional SVAR. However, the cross-sectional performances of ICC and FICC vary across subsamples. Neither ICC nor FICC outperforms the trivial ERP among large firms and all-but-tiny firms; nevertheless, FICC performs relatively well among small firms, with a negative and significant average cross-sectional SVAR that is similar to the magnitude of that of CER. Among tiny firms, both ICC and FICC exhibit negative and significant mean cross-sectional SVARs. In fact, FICC exhibits the lowest (most negative) average cross-sectional SVAR in this subsample.

The time-series performances of these ERPs across size subsamples are similar to those reported in Table 5. For example, in all four subsamples, ICC and FICC are the best performers in time series, with the most negative (and statistically significant) mean time-series SVARs. As is the case with cross-sectional SVARs, the performance of CER improves in smaller firms: for example, the mean time-series SVAR in the tiny-firms sample is negative and statistically significant.

Table 6, Panels E and D, show that our assessment of the relative performance of ERPs based on cross-sectional and time-series SVARs are generally unchanged for the first and second half of our sample period. Interestingly, we find that the time-series MEV of CER improves in the second half of the sample period: CER's mean time-series SVAR goes from being positive and significant in the first half to negative and significant in the second half. Nevertheless, in the second period, the mean time-series SVARs for ICC and FICC are still significantly more negative than that of CER.

Discussion

It is also possible to estimate the economic magnitude of performance gains (in terms of the percent reduction in mean MEV) attained by one ERP over another. Doing so requires

an estimate of the expected variance in expected returns, or the difference between average MEV and SVAR (see, e.g., Eq. (9)). Assuming conservatively that 1% of realized variance is represented by the variance in expected returns, and applying this ratio to the average variance in monthly realized returns, we estimate an average cross-sectional variance in expected returns of 0.000168 and an average time-series variance of 0.000199.[10]

Based on these estimates, we find CER's performance in the cross-section represents a 12%, 8%, and 91% improvement over ICC, FICC, and FER (in terms of absolute MEV reduction). Similarly, ICC's performance in time series represents a 44%, 13%, and 96% improvement over CER, FICC, and FFF. It is also instructive to assess these ERPs' improvements over the trivial estimator, whose MEV is simply the variance of expected returns. The cross-sectional performance of CER represents a 10% improvement, while the time-series performance of ICC represents a 45% improvement over the trivial estimator.

Overall, the findings reported in Tables 4, 5, and 6 suggest the choice of ERPs in treatment-effects studies depends on the research design. In general, designs that exploit time-series variation in the data should select ICC-based ERPs; designs that exploit cross-sectional variation should select characteristic-based ERPs.

These findings also provide a more sanguine assessment of ICC-based ERPs than does the prior literature. Our results are consistent with the conclusions of Easton and Monahan (2005) that ICCs are unreliable in the cross-section: our representative ICC does not outperform a trivial estimator. However, our findings demonstrate the usefulness of ICCs in time series, a dimension of performance previously overlooked by the literature, and support the credibility of recent research that leverages time-series variation in ICCs (e.g., Pástor et al., 2008; Li et al., 2013a; Huang and Kilic, 2019). We conjecture that the earnings and payout forecasts—key inputs into ICCs—contain measurement errors that distort the cross-sectional ordering of firms in terms of their expected returns and that these measurement errors are

---

[10]For example, Lewellen (2015)'s results suggest the variance of expected returns accounting for roughly 0.3% of the variance in realized returns. Our assumption of 1% is conservative in that larger expected-return variance estimates imply lower percentage improvement in MEV.

likely to be relatively stable over time. Consequently, ICCs' time-series variation can be quite informative despite their noisy cross-sectional variation.

We hasten to point out these conclusions are based on the sample used in the analysis and on the types of ERPs tested. The next section will perform a deep dive into each class of ERPs to better understand variation in the cross-sectional and time-series MEVs of different factor-based, ICC-based, and characteristic-based ERP variants.

V   Detailed Examination of Individual ERP Variants

This section digs more deeply by examining the MEV performance of 13 variants of the ERPs examined above to better understand how perturbations in construction methodology within each ERP family can affect performance.

Factor-Based ERPs

In our main results, our representative factor-based ERP (FER) is consistently the worst-performing in terms of MEV. However, FER is based on two specific factor models and on particular estimation choices. To better understand the performance of factor-based ERPs, we provide an in-depth evaluation of factor-based ERPs that vary due either to the underlying model or to the estimation approaches taken.

In Table 7 we evaluate the MEV properties of seven factor-based ERP variants from the last 50 years of the empirical asset-pricing literature. We begin with the canonical single-factor model: the Capital Asset Pricing Model (Sharpe, 1964; Lintner, 1965). We then examine four multi-factor models drawn from the empirical asset-pricing research of Eugene Fama and Ken French: a three-factor model consisting of a market factor, a size factor, and a value factor (e.g., Fama and French, 1993, 1996); a four-factor model that adds a momentum factor to the three-factor model; a five-factor model that adds a profitability factor and an investment factor (Fama and French, 2015) to the three-factor model; and a six-factor model

that augments the five-factor model with a momentum factor.[11] We also examine two models motivated by the neoclassical q-theory of investment: a four-factor model that consists of a market factor, a size factor, an investment factor, and a profitability factor (Hou et al., 2015); and a five-factor model that augments the four-factor model with an expected growth factor (Hou et al., 2019).[12] Recent evidence shows the latter models efficiently summarize the cross-section of average stock returns as well as a large set of apparent anomalies.

Table 7, Panels A-F, report the MEV performance of these factor-based ERP variants, where each panel considers a different approach to estimating factor sensitivities (βs) or expected factor premiums (γs). The ERPs constructed using the Capital Asset Pricing Model are denoted CAPM; those using the Fama and French three-, four-, five-, and six-factor models are denoted FF3, FF4, FF5, and FF6; those using the four- and five-factor q-theory models are denoted $q4$ and $q5$. In each panel we also examine a variant, denoted FER, based on the equal-weighted average of FF6 and $q5$, as in prior tables.

Panel A estimates factor-based ERPs using the same approach as in prior tables: each firm's βs are estimated on a rolling basis, using available data from the most recent 60 months; the most recent month's premium for a given factor is the estimate for that factor's γ. Interestingly, we find the single-factor CAPM exhibits significantly lower MEV on average, both in the cross-section and in time series. In fact, MEVs appear to increase with the number of factors, suggesting a propagation of errors resulting from estimating a greater number of factor sensitivities or premiums.

Table 7, Panels B and C, examine the impact of different approaches to reducing the noise in estimated βs. One approach is to utilize the full sample of each firm's stock returns to estimate the full-sample-period βs. Understanding the degree to which this approach improves the performance of factor-based ERPs can be instructive about the source of measurement errors, even if the approach is potentially subject to look-ahead bias. Another approach is to use the industry-average βs as a means for filtering out noise in the estima-

---

[11]The momentum factor we use is the UMD factor, obtained from French's data library.

[12]We thank Lu Zhang for graciously sharing the data on q-factors' premiums.

tion process. Panels B and C report the MEV properties of ERPs that use full-sample βs and industry-average βs, estimated monthly as an average of firm-level rolling βs within each Fama-French 48 industry.

Panel B shows the use of full-sample-estimated βs lowers the mean SVAR for all factor-based ERP variants, and Panel C shows the use of industry-average βs has a dramatic impact on reducing the MEVs of factor-based ERPs. For example, assuming the cross-sectional variance in expected returns is 0.000168, the percentage improvement in mean cross-sectional MEV gained from the use of industry-average βs (relative to Panel A) is 72% for CAPM and ranges from 83% (for FF3) to 89% (for FF6) among the remaining variants. Similarly, assuming the time-series variance in expected returns is 0.000199, the percentage improvement in mean time-series MEV is 40% for CAPM and ranges from 61% (for FF3) to 73% (for FF6) among the remaining variants. Despite these economically significant improvements in MEV, none of the improvements produced a factor-based ERP in Panel B or C that outperforms the trivial estimator either in the cross-section or in time series.

Table 7, Panels D-F, examine the impact of different approaches to estimating γs. In each panel, ERPs are constructed with rolling βs, estimated using available data from the most recent 60 months. As estimates of γs, Panel D uses the prior-60-month average factor premiums; Panel E uses the cumulative average factor premiums (with a minimum requirement of 60 months). Relative to the performance summarized in Panel A, the use of smoothed factor premiums significantly reduces MEVs. The percentage improvement in mean cross-sectional MEV ranges from 52% to 84%; that of mean time-series MEV ranges from 65% to 86%.

Panel F estimates γs using future realized factor premium (i.e., measured contemporaneously with realized returns) in lieu of the prior period's factor premium. Despite potential look-ahead bias, we find that these factor-based ERPs continue to exhibit high MEVs—higher in fact than those of factor-based ERPs that use prior-month factor premiums. The intuition is that future realized factor premiums are noisy proxies for ex-ante *expected* fac-

tor returns, just as future realized returns are noisy proxies for ex-ante expected returns.[13] Therefore, although the realized factor premiums can significantly explain realized returns, they are noisy inputs for ex-ante expected returns.

Averaging across factor-based estimates is also effective at reducing MEVs. In each panel, the mean SVAR of FER is lower than that of its constituents, FF6 and $q5$. However, averaging does not produce as large an impact on lowering MEVs as does reducing the noise in factor sensitivities or expected factor premiums.

Together, these findings suggest the noisiness of factor-based ERPs is due to the noise both from estimated factor sensitivities and from estimated expected factor premiums. Additional factors generally compound these estimation errors. While common approaches for improving the precision of factor-based ERPs are effective in reduce MEVs, *none* of the approaches we consider results in a factor-based ERP that outperforms a trivial estimate. A central insight from these analyses is that superior factor models—those that perform well in explaining the in-sample variation in realized returns (e.g., Hou et al., 2019)—do not necessarily translate into ex-ante ERPs that are best suited for estimating treatment effects.

Implied-Cost-of-Capital Estimates

Next we provide an in-depth evaluation of ICCs—perhaps the most common ERP used by researchers in treatment-effects studies since the early 2000s. ICCs differ based on the earnings and earnings-growth forecasts used as inputs and on the underlying valuation model that links market prices and expected future earnings to expected returns.

Table 8 examines the relative performance of ICCs that use as inputs the mechanical

---

[13]As a formal illustration, consider a single-factor model of returns ($r_{t+1} = \alpha + \beta f_{t+1} + \epsilon_{t+1}$ with $\epsilon \sim_{iid} (0, \sigma_\epsilon^2)$) where the factor premiums follow an AR(1) process: $f_{t+1} = a + \phi f_t + \eta_t$ for $\phi \in (0,1)$ and $\eta \sim_{iid} (0, \sigma_\eta^2)$. To the degree that factor premiums are persistent, the last period's factor premium will be a less-noisy proxy for expected factor returns than contemporaneous factor returns:

$$Var[f_t - \mathbb{E}_t(f_{t+1})] = \frac{(1-\phi)^2}{1-\phi^2}\sigma_\eta^2 < \sigma_\eta^2 = Var[f_{t+1} - \mathbb{E}_t(f_{t+1})].$$

Note that the average (i.e., across the factors included in the seven factor-based models we examine) factor premium persistence in our sample is 0.10.

earnings forecasts derived from the cross-sectional earnings-forecast model of Hou et al. (2012) (Panel A) and those that use the consensus analyst earnings forecasts (Panel B), measured as the median of analysts' earnings forecasts for a given firm. In each panel, we report the distributional properties of cross-sectional and time-series SVAR for five ICC variants: two variants—proposed by Gebhardt et al. (2001) (GLS) and Claus and Thomas (2001) (CAT)—utilize the residual-income valuation model; two variants—proposed by Ohlson and Juettner-Nauroth (2005) (AGR) and Easton and Monahan (2005) (PEG)— utilize the abnormal-earnings-growth valuation model; and a composite measure (ICC) that is the equal-weighted average of the aforementioned four variants, as in prior tables. GLS, CAT, AGR, and PEG are the ICC variants most commonly used in the treatment-effects literature, and many studies rely on a composite measure like ICC to eliminate model-specific noise. Table 8, Panel A, reports the performance of mechanical-forecast-based ICCs. In the cross-section, GLS is the only ICC variant that exhibits an average MEV significantly lower than that of the trivial ERP. CAT, AGR, and PEG all produce a positive and significant average cross-sectional SVAR; ICC produces a positive but statistically insignificant average cross-sectional SVAR. In time series, on the other hand, nearly all ICC variants perform well relative to a constant. Except for AGR, each variant produces a negative and significant average time-series SVAR. The composite ICC has the lowest mean time-series SVAR, suggesting averaging ICC variants is effective at removing model-specific noise.

Table 8, Panel B, reports on the performance of analyst-forecast-based ICCs. Because analyst coverage is required, these ICCs are available for a smaller sample than that of Panel A. In the cross-section, the ICC variants based on the residual-income valuation model—GLS and CAT—both exhibit negative and significant mean SVARs within the sample of firms with analyst coverage; the remaining variants produce average SVARs that are not statistically different from zero. In time series, all of the analyst-based ICCs produce average SVARs that are negative and statistically significant. Like the results with mechanical-forecast-based ICCs, the composite ICC exhibits the lowest mean time-series MEV.

Finally, Panel C evaluates the performance of mechanical-forecast-based composite ICC and analyst-forecast-based composite ICC on a common sample of observations with non-missing values for both. Neither composite ICC variant outperforms the trivial ERP in the cross-section; on the other hand, both composite ICCs exhibit negative and statistically significant mean time-series SVARs, with mechanical-based ICC exhibiting lower time-series MEV.

The analyses in Table 8 provides some clues about the relative importance of potential sources of measurement errors. Wang (2015) suggests that, as proxies of for expected returns, ICCs' measurement errors arise from two main sources. One source arises from errors in forecasts of future fundamentals, such as estimates of expected future earnings; analysts exhibit systematic biases in their earnings and long-term growth forecasts (e.g., La Porta, 1996; Dechow and Sloan, 1997; Frankel and Lee, 1998; Guay et al., 2011). Another source arises from model mis-specification, resulting from erroneous assumptions embodied in the functional form that maps current stock price and expected future earnings to expected returns. To the extent that mechanical forecasts mitigate measurement errors stemming from analysts' earnings forecasts, our findings suggest that functional-form mis-specification is the primary driver of ICC MEVs. In the cross-section, the use of mechanical forecasts of earnings does not systematically lead to ICCs that outperform the trivial estimator; on the other hand, we find that ICCs based on the residual-income valuation model, particularly GLS, systematically outperform the trivial ERP. Our findings are also consistent with ICCs' measurement errors being highly persistent: in nearly all cases, irrespective of the earnings-forecast type or the valuation model used, ICCs outperform a trivial estimator in time series.

Together, these findings suggest that ICCs should generally be avoided by researchers utilizing cross-sectional designs, and even by those using composite ICCs in hopes of miti-gating the influence of measurement errors. Our conclusions therefore raise significant doubt about the inferences of a large percentage of the treatment-effects literature, a topic that we will address in Section VI. Nevertheless, our analyses suggest that researchers can rely on

ICCs in research designs that utilize time-series variation in the data.

Characteristic-Based ERPs

We conclude this section with an analysis of variants of characteristic-based ERPs. In general, these ERPs are linear combinations of firm characteristics, where weights derive from historical cross-sectional regressions of realized returns on those characteristics. Characteristic-based ERPs thus vary due either to the underlying model (i.e., the characteristics used) or the method for estimating weights (i.e., the regression technique or sample).

Table 9 compares the MEV properties of CER with its constituents, JLR and LPV. Panel A examines the variants constructed using weights estimated on a rolling-10-year basis, as we have done in previous results. In the cross-section, we find that both JLR and LPV produce negative and statistically significant mean SVARs; nevertheless, JLR is the better performer in the cross-section, with substantially lower average cross-sectional SVAR and larger t-statistics. Assuming again a cross-sectional variance in expected returns of 0.000168, LPV's cross-sectional MEV is 8% higher than that of JLR. However, LPV is the better performer in time series, with a mean SVAR that is negative and significant, whereas JLR's mean time-series SVAR is positive and statistically insignificant. Assuming that time-series variance in expected returns is 0.000199, LPV's time-series MEV is 7% lower than that of JLR. CER, which averages the JLR and LPV estimates, appears to embody the best attributes of JLR and LPV, in that it produces negative and significant MEVs in both the cross-section and time series. Its mean cross-sectional MEV is only 1% higher than that of CER, whereas its mean time-series MEV is only 3% higher than that of LPV.

Table 9, Panel B, reports on the performance of characteristic-based ERP variants where we recursively estimate weights on firm characteristics via Fama-MacBeth regressions using all available historical data, with a 10-year-minimum requirement. Ex ante, it is unclear which window to use to estimate these weights; it depends on the nature of the relationship between characteristics and expected returns. To the extent that a firm attribute is stably

associated with expected returns, incorporating more historical data in estimating weights is desirable. However, the opposite is likely true for characteristics that exhibit relationships that can vary substantially over time.

The results in Panel B suggest the use of recursively estimated weights has limited impact on cross-sectional MEV, but can be more important for time-series MEV. In the cross-section, the MEV of each variant changes by no more than 1.5%. In time series, on the other hand, the use of recursively estimated weights appears to impact JLR differently from LPV. JLR's mean MEV increases by 10%; LPV's mean MEV declines by 8%. The net effect is a moderate 2% improvement in CER's mean time-series MEV. One possible explanation for these findings is JLR's inclusion of momentum, whose relation with returns varies over time (e.g., Lewellen, 2015), with a greater impact on the ERP's time-series volatility. A comprehensive analysis of the payoffs for individual characteristics is clearly beyond the scope of this study. However, our findings suggest that more work is needed to better understand the determinants of the cross-sectional and time-series MEVs of characteristic-based ERPs.

VI   Re-Examining Prior Literature

Our results suggest researchers utilizing cross-sectional designs in treatment-effect studies should rely on characteristic-based ERPs rather than ICCs. However, most of the prior literature that utilize ex-ante measures of expected returns have drawn their inference from cross-sectional variations in ICCs. In this section, we re-evaluate three significant lines of inquiry central to the risk-reward trade-off in financial economics but for which there is mixed empirical findings in prior research. For each line of inquiry, we first replicate findings from prior research and then illustrate how the application of our ERP evaluative framework can either sharpen or alter inferences.

VI.1 Pricing of Default Risk

We begin by re-examining the link between default risk and the cross-section of expected returns hypothesized in prior research (e.g., Chan and Chen, 1991; Fama and French, 1996). Empirical tests of this link have produced mixed evidence. For example, Campbell et al. (2008) finds firms' default risk is negatively associated with ex-post measures of expected returns. However, Chava and Purnanandam (2010) shows that, by using an ex-ante measure of expected returns (an ICC measure derived from Gebhardt et al., 2001), there is a strong and positive empirical relation between default risk and expected returns.

We re-examine the findings of Chava and Purnanandam (2010) in light of the evidence on ERP performance presented in this paper. In particular, we examine whether the findings are robust to the use of CER, the ERP nominated by our evaluation framework to exhibit the lowest cross-sectional measurement-error variance.[14] Table 10, Panel A, reports the results of pooled cross-sectional OLS regressions (i.e., with year-fixed effects) of ERPs on measures of default risk and firm-level controls used by Chava and Purnanandam (2010).[15] The main variables of interest, DRP_{hazard} (columns 1 and 2) and DRP_{EDF} (columns 3 and 4), are the annual percentile rankings of firms' default risk based on their hazard rate and the expected distance to default, respectively, and range from 0 to 100, where higher values represent greater distress risk.[16] Columns 1 and 3 replicates the main results of Chava and Purnanandam (2010) using analyst-based GLS (expressed in percentage points). The estimated coefficients on DRP_{hazard} (column 1) and DRP_{EDF} (column 3) are positive and significant (at the 1% level), and very similar in magnitude to those of the original paper.

---

[14]For this exercise, as well as in our re-examination of the pricing of information risk, we use proxies of expected annual returns. To construct characteristic-based ERPs, we use historical regressions of annual returns on firm characteristics. Our evaluative framework provides similar relative rankings of annual ERP variants: CER achieves the lowest cross-sectional MEV and outperforms a trivial estimate; ICCs perform worse and do not outperform a constant. Moreover, CER's MEV is lower than that of not only ICC but also each of its components, including GLS.

[15]In our replications, we run pooled regressions with cross-sectional fixed effects with two-way cluster-robust standard errors, rather than Fama-MacBeth regressions and Newey-West-corrected standard errors, because Gow et al. (2010) shows the latter understates the true variability of treatment-effect estimates.

[16]We thank the authors of Chava and Purnanandam (2010) for providing their data.

When estimating the same regression specifications using CER as the dependent variable of interest, we also obtain positive and statistically significant (at the 1% level) coefficients on DRP_{hazard} (column 2) and DRP_{EDF} (column 4). However, CER produces effect magnitudes that are significantly lower: the coefficient on DRP_{hazard} obtained using CER is 40% lower than that obtained using GLS; the coefficient on DRP_{EDF} obtained using CER is 37% lower than that obtained using GLS. These differences suggest the use of ERPs with lower measurement error variances are useful in calibrating the economic magnitudes of estimated treatment effects.

VI.2   Pricing of Idiosyncratic Volatility

Next, we re-examine the relation between idiosyncratic volatility and the cross-section of expected returns. Modern portfolio theory (e.g., the Capital Asset Pricing Model) suggests only systematic risk should be priced and thus idiosyncratic risk should not be related to expected returns. Other theories (e.g., Levy, 1978; Merton, 1987; Malkiel and Xu, 2002) suggest under-diversified investors demand a return from bearing idiosyncratic risk. In empirical tests of these theories, Ang et al. (2006) finds a negative relation, whereas subsequent work demonstrates these inferences are sensitive to the measurement of idiosyncratic risk (Huang et al., 2010) as well as the measurement of expected returns (Fu, 2009).

A central finding in Fu (2009) is that traditional measures of idiosyncratic volatility ($IVOL$) are associated with a risk premium when using ICCs derived from Gebhardt et al. (2001) as an ex-ante measure of expected returns. We re-evaluate the relation between $IVOL$ and ex-ante ERPs nominated by our evaluative framework.[17] Table 10, Panel B, reports the results of pooled cross-sectional regressions (i.e., with year-month-fixed effects)

---

[17]It is worthwhile noting that one of the main innovations of Fu (2009) is in the measurement of idiosyncratic volatility (through the estimation of an EGARCH model). Fu (2009) examines the relation between ICCs and traditional measures of idiosyncratic volatility to argue that better measures of idiosyncratic volatility are unnecessary given cleaner measures of expected returns. The evidence using ICCs documented in Fu (2009) is particularly important in light of the subsequent work that has challenged validity of the paper's idiosyncratic volatility measurement innovation (e.g., Guo et al., 2014), heightening the relevance of our re-examination of Fu (2009)'s ICC analysis.

of month t ERPs on two measures of realized idiosyncratic volatility considered in Fu (2009): contemporaneous $IVOL_t$ (columns 1 and 2) and one-month lagged $IVOL_{t-1}$ (columns 3 and 4).

Columns 1 and 3 replicate the results of Fu (2009) using analyst-based GLS (expressed in percentage points); the coefficients on $IVOL_t$ and $IVOL_{t-1}$ are positive and significant at the 1% level. In columns 2 and 4, however, we find these positive relations are not robust to the use of CER. In each case, the main coefficient obtained using CER is less than one-third that obtained using GLS and no longer statistically significant at the 10% level. Overall, our findings cast doubt on the positive relation between expected returns and $IVOL$.

VI.3   Pricing of Information Risk

Finally, we revisit the link between a firm's information risk and expected returns. Whether the information risk of firms is diversifiable or priced is theoretically unclear (e.g., Diamond and Verrecchia, 1991; Easley and O'Hara, 2004; Lambert et al., 2007), and the empirical evidence has been mixed. The most enduring evidence on the relation between expected returns and information risk stems from the relation between measures of ICCs and measures of accrual quality. Core et al. (2008) asserts that "with respect to the evidence that [Francis et al. (2004)] and others present that accrual quality is priced, only the implied cost of capital results appear robust," but cautions that the interpretation of these findings depends largely on the merits of the proxies of expected returns.[18]

We re-examine this link in Column 1 and 2 of Table 10, Panel C, which reports the results of pooled cross-sectional OLS regressions (i.e., with year-fixed effects) of annual ERPs on accrual quality (AQ) and firm-level controls used by Francis et al. (2004) using a sample from 1975 to 2001. We calculate AQ using the approach detailed in Francis et al. (2004), which measures the extent to which the accrual portion of a firm's earnings is unexplained by their cash flow from operations. Francis et al. (2004) hypothesizes that a greater share

---

[18]Although the pricing of information risk has been largely examined in the accounting literature, it has also stimulated recent research in the finance literature (e.g., Yang et al., 2020).

of unexplained variation indicates greater information risk due to less informative earnings reports and therefore commands a expected return premium.

Column 1 replicates the results of Francis et al. (2004) using analyst-based PEG (expressed in percentage points).[19] We find a positive coefficient on AQ that is similar to Francis et al. (2004) and significant at the 1% level. In column 2, however, we find this result is not robust to the use of CER. The coefficient estimate on AQ is now negative, its magnitude is significantly smaller (less than one-tenth of that obtained using PEG), and is no longer statistically significant at the 10% level.

In columns 3 and 4, Table 10, Panel C, we examine the relation between information risk and expected returns using an alternative measure of information risk. Chen et al. (2015) introduces a measure of disclosure quality (DQ) that captures the level of disaggregation in financial-data items in firms' annual reports, where higher values of DQ (which is scaled to range from 0 to 1) reflect finer information and are deemed to be higher-quality disclosures and reflect lower information risk. Chen et al. (2015) shows higher DQ is cross-sectionally associated with lower expected returns, using a composite mechanical-forecast-based ICC.

Column 3 and 4 of Table 10, Panel C, reports the results of pooled cross-sectional OLS regressions (i.e., with year-fixed effects and industry-fixed effects) of annual ERPs on DQ and firm-level controls used by Chen et al. (2015) and using a sample from 1976 to 2011.[20] Column 3 replicates the results of Chen et al. (2015) using the composite ICC (expressed in percentage points) that we evaluated in earlier sections. We find a negative coefficient on DQ that is significant at the 5% level and closely mirrors the effects estimated in Chen et al. (2015). In column 4, however, we find this negative and significant relation is not robust to the use of CER. The coefficient estimate on DQ is now positive, its magnitude is significantly smaller (about one-tenth of that obtained using the composite ICC), and is no

---

[19]The main results of Francis et al. (2004) use a measure of implied cost of capital derived from analysts' target prices and dividend forecasts, as compiled by Value Line. We did not construct this measure because it is an uncommon proxy in the literature, and because Francis et al. (2004) reports that all results remain very similar using an analyst-based PEG measure. Having previously computed and analyzed the PEG measure using IBES data, we use it as the main dependent variable in our replication of Francis et al. (2004).

[20]We thank the authors of Chen et al. (2015) for sharing their data.

longer statistically significant at the 10% level. Our findings cast doubt on the cross-sectional relation between expected returns and information risk, whether measured by *AQ* or *DQ*.

Overall, our examination of the prior literature illustrates the importance of ERP selection for researchers' inferences. Selecting ERPs that exhibit the lowest cross-sectional MEVs can sharpen researchers' inferences (by influencing the magnitude of estimated effects) or alter inferences altogether (by influencing the statistical significance of estimated effects).

VII   Implications for Financial Economists and Conclusion

Estimates of expected returns play a central role in many managerial and investment decisions that affect the allocation of scarce resources. Understanding the drivers of variation in expected returns is "*the* central organizing question in current asset-pricing research" (Cochrane, 2011). In pursuing this agenda, financial economists have called for innovations in the measurement of expected returns (Elton, 1999).

In recent years, an increasing number of studies published in top finance journals have turned to ex-ante ERPs (such as ICCs). For example, scholars have used them to address several long-standing financial puzzles, such as the magnitude of the equity risk-premium (Claus and Thomas, 2001; Fama and French, 2002), the degree of fragmentation in international equity markets (Lee et al., 2009), and the inter-temporal relation between market-wide volatility and expected returns (Pástor et al., 2008). Ex-ante ERPs have also been used to shed new light on the cross-sectional relation between expected returns and firm-level distress risk (Chava and Purnanandam, 2010), idiosyncratic volatility (Fu, 2009), and cross-listings (Hail and Leuz, 2009), to name a few. Other studies, such as Li et al. (2013b), Jones and Tuzel (2013), and Huang and Kilic (2019), have used these measures as key explanatory variables in various settings. In each case, researchers used ex-ante ERPs to obtain more precise estimates than what is possible with ex-post ERPs. Indeed, data suggests an increasing adoption of these measures in finance in recent years: for example, about 40% of the ICC studies in 2007-2016 period in Exhibit 1 were published in finance compared to 18% in

1997-2006.

Despite the recent uptick, significant opportunities for broader usage of ex-ante ERPs remain. Without prejudice, we can point to many illustrative studies where the use of ex-ante ERPs would have been useful: for example, in explaining the relation between book-to-market and future returns (e.g., Chen, 2011); in sharpening the relation between firm-level investments and Tobin's Q (e.g., Xing, 2008); or in estimating the impact of economic shocks on expected-returns (e.g., Kelly and Ljungqvist, 2012). These studies in finance would not show up in Exhibit 1, but our findings are nevertheless important to them. Furthermore, although this paper emphasizes the use of ERPs as dependent variables of interest, researchers interested in using ex-ante ERPs as explanatory variables could also adopt our MEV framework for selecting suitable measures. In the latter setting, minimizing MEVs helps to mitigate the potential bias from measurement errors in regressors.

Finally, we note financial economists are continuing to develop new approaches to estimate expected returns, such as by cleaning up historical realized returns (e.g., Lewellen, 2015), or by leveraging the pricing of corporate bonds (Campello et al., 2008), dividend derivatives (Van Binsbergen et al., 2013), and stock options (Martin and Wagner, 2019). Our paper facilitates the use and adoption of ex-ante ERPs by providing a framework that: (a) allows potential users of ERPs to select among alternatives the most suitable for a given research design, and (b) allows developers of new ERPs to assess when progress is being made by establishing a minimum bar for what should be demanded from new entrants.

References

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. 2006. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Angrist, J. D. and Pischke, J. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Botosan, C. 1997. Disclosure level and the cost of equity capital. *The Accounting Review*, 72(3):323–349.

Campbell, J. 1991. A variance decomposition for stock returns. *The Economic Journal*, 101(405):157–179.

Campbell, J. and Shiller, R. 1988a. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. and Shiller, R. 1988b. Stock prices, earnings, and expected dividends. *Journal of Finance*, 43(3):661–676.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. 2008. In search of distress risk. *The Journal of Finance*, 63(6):2899–2939.

Campello, M., Chen, L., and Zhang, L. 2008. Expected returns, yield spreads, and asset pricing tests. *The Review of Financial Studies*, 21(3):1297–1338.

Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.

Chan, K. and Chen, N.-F. 1991. Structural and return characteristics of small and large firms. *The Journal of Finance*, 46(4):1467–1484.

Chattopadhyay, A., Lyle, M. R., and Wang, C. C. 2020. Expected stock returns worldwide: A log-linear present-value approach. *Working Paper*.

Chava, S. and Purnanandam, A. 2010. Is default risk negatively related to stock returns? *Review of Financial Studies*, 23(6):2523–2559.

Chen, H. J. 2011. Firm life expectancy and the heterogeneity of the book-to-market effect. *Journal of Financial Economics*, 100(2):402–423.

Chen, S., Miao, B., and Shevlin, T. 2015. A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports. *Journal of Accounting Research*, 53(5):1017–1054.

Claus, J. and Thomas, J. 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts from domestic and international stock markets. *Journal of Finance*, 56(5):1629–1666.

Cochrane, J. 2011. Presidential address: Discount rates. *Journal of Finance*, 66(4):1047–1108.

Core, J. E., Guay, W. R., and Verdi, R. 2008. Is accruals quality a priced risk factor? *Journal of Accounting and Economics*, 46(1):2–22.

Dechow, P. and Sloan, R. 1997. Returns to contrarian investment strategies: Tests of naive expectations hypotheses. *Journal of Financial Economics*, 43(1):3–27.

Diamond, D. W. and Verrecchia, R. E. 1991. Disclosure, liquidity, and the cost of capital. *Journal of Finance*, 46(4):1325–1359.

Easley, D. and O'Hara, M. 2004. Information and the cost of capital. *Journal of Finance*, 59(4):1553–1583.

Easton, P. and Monahan, S. 2005. An evaluation of accounting-based measures of expected returns. *The Accounting Review*, 80:501–538.

Easton, P. D. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review*, 79:73–96.

Elton, E. J. 1999. Presidential address: expected return, realized return, and asset pricing tests. *The Journal of Finance*, 54(4):1199–1220.

Fama, E. and French, K. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. and French, K. 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51:55–84.

Fama, E. F. and French, K. R. 2002. The equity premium. *Journal of Finance*, 57(2):637–659.

Fama, E. F. and French, K. R. 2015. A five-factor asset pricing model. *Journal of Financial Economics*, 116:1–22.

Fama, E. F. and French, K. R. 2018. Choosing factors. *Journal of Financial Economics*, 128(2):234–252.

Francis, J., LaFond, R., Olsson, P. M., and Schipper, K. 2004. Costs of equity and earnings attributes. *The Accounting Review*, 79(4):967–1010.

Frankel, R. and Lee, C. 1998. Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics*, 25(3):283–319.

Fu, F. 2009. Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics*, 91:24–37.

Gebhardt, W., Lee, C., and Swaminathan, B. 2001. Toward an implied cost of capital. *Journal of Accounting Research*, 39(1):135–176.

Gode, D. and Mohanram, P. 2003. Inferring the cost of capital using the Ohlson-Juettner model. *Review of Accounting Studies*, 8(4):399–431.

Gow, I. D., Ormazabal, G., and Taylor, D. J. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review*, 85(2):483–512.

Greenwood, R. and Shleifer, A. 2014. Expectations of returns and expected returns. *The Review of Financial Studies*, 27(3):714–746.

Guay, W., Kothari, S., and Shu, S. 2011. Properties of implied cost of capital using analysts' forecasts. *Australian Journal of Management*, 36(2):125–149.

Guo, H., Kassa, H., and Ferguson, M. F. 2014. On the relation between egarch idiosyncratic volatility and expected stock returns. *Journal of Financial and Quantitative Analysis*, 49(1):271–296.

Hail, L. and Leuz, C. 2009. Cost of capital effects and changes in growth expectations around U.S. cross-listings. *Journal of Financial Economics*, 93(3):428–454.

Hou, K., Mo, H., Xue, C., and Zhang, L. 2019. q^5. *Working Paper*.

Hou, K., Van Dijk, M. A., and Zhang, Y. 2012. The implied cost of capital: A new approach. *Journal of Accounting and Economics*, 53:504–526.

Hou, K., Xue, C., and Zhang, L. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705.

Huang, D. and Kilic, M. 2019. Gold, platinum, and expected stock returns. *Journal of Financial Economics*, 132(3):50–75.

Huang, W., Liu, Q., Rhee, S. G., and Zhang, L. 2010. Return reversals, idiosyncratic risk, and expected returns. *The Review of Financial Studies*, 23(1):147–168.

Jones, C. S. and Tuzel, S. 2013. Inventory investment and the cost of capital. *Journal of Financial Economics*, 107(3):557–579.

Kelly, B. and Ljungqvist, A. 2012. Testing asymmetric-information asset pricing models. *The Review of Financial Studies*, 25(5):1366–1413.

La Porta, R. 1996. Expectations and the cross-section of stock returns. *Journal of Finance*, 51(5):1715–1742.

Lambert, R., Leuz, C., and Verrecchia, R. E. 2007. Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research*, 45(2):385–420.

Lee, C., Ng, D., and Swaminathan, B. 2009. Testing international asset pricing models using implied costs of capital. *Journal of Financial and Quantitative Analysis*, 44(2):307–335.

Levy, H. 1978. Equilibrium in an imperfect market: A constraint on the number of securities in the portfolio. *The American Economic Review*, 68(4):643–658.

Lewellen, J. 2015. The cross section of expected stock returns. *Critical Finance Review*, (4):1–44.

Li, Y., Ng, D. T., and Swaminathan, B. 2013a. Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics*, 110(2):419–436.

Li, Y., Ng, D. T., and Swaminathan, B. 2013b. Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics*, 110(2):419–436.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 73(1):13–37.

Lyle, M. R. and Wang, C. C. 2015. The cross section of expected holding period returns and their dynamics: A present value approach. *Journal of Financial Economics*, 116(3):505–525.

Malkiel, B. G. and Xu, Y. 2002. Idiosyncratic risk and security returns. *University of Texas at Dallas (November 2002)*.

Martin, I. W. and Wagner, C. 2019. What is the expected return on a stock? *The Journal of Finance*, 74(4):1887–1929.

Merton, R. C. 1987. A simple model of capital market equilibrium with incomplete information. *The journal of finance*, 42(3):483–510.

Ohlson, J. and Juettner-Nauroth, B. E. 2005. Expected EPS and EPS growth as determinants of value. *Review of Accounting Studies*, 10(2):349–365.

Pástor, L., Sinha, M., and Swaminathan, B. 2008. Estimating the intertemporal risk-return tradeoff using the implied cost of capital. *Journal of Finance*, 63(6):2859–2897.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Van Binsbergen, J., Hueskes, W., Koijen, R., and Vrugt, E. 2013. Equity yields. *Journal of Financial Economics*, 110(3):503–519.

Vuolteenaho, T. 2002. What drives firm-level stock returns? *Journal of Finance*, 57(1):233–264.

Wang, C. C. 2015. Measurement errors of expected return proxies and the implied cost of capital. *Working Paper, Harvard Business School*.

Xing, Y. 2008. Interpreting the value effect through the q-theory: An empirical investigation. *The Review of Financial Studies*, 21(4):1767–1795.

Yang, Y. C., Zhang, B., and Zhang, C. 2020. Is information risk priced? evidence from abnormal idiosyncratic volatility. *Journal of Financial Economics*, 135(2):528–554.

Technical Appendix

Part A. Estimating Firm-Specific ERP Measurement-Error Variance

Part A of this appendix derives a measure to evaluate ERPs on the basis of average time-series MEV, a measure that is ERP-specific and empirically estimable. We call this measure Average Scaled TS Variance ($AvgSVar^{TS}$). The derivation proceeds in three steps. In Step 1 we decompose a firm's time-series ERP MEV and define a firm-specific Scaled TS Variance measure. In Step 2 we decompose realized returns and derive an expression for the time-series return-ERP covariance. In Step 3 we show how to estimate Scaled TS Variance using the time-series return-ERP covariance and define the Average Scaled TS Variance.

We make the following assumptions throughout:

A1 Expected returns ($er_{i,t+1}$), ERP measurement error ($\omega_{i,t+1}$), and realized returns ($r_{i,t+1}$) are jointly covariance stationary.

A2 Unexpected returns (or news, i.e., $\delta_{i,t+1} = r_{i,t+1} - er_{i,t}$) is not ex-ante forecastable, and is not systematically correlated with expected returns (in time series or cross-section).

Step 1. Decomposing a Firm's Time-Series Variance in ERP Measurement Errors and Defining $SVar_i(\omega_{i,t})$

We define an ERP as the sum of the true expected return and its measurement error:

$$\widehat{er}_{i,t+1} = er_{i,t+1} + \omega_{i,t+1}. \tag{T1}$$

Taking the time-series variance on both sides of Eq. (T1) and reorganizing terms, firm i's time-series variance in ERP measurement errors can be written as

$$Var_i(\omega_{i,t}) = Var_i(\widehat{er}_{i,t}) + Var_i(er_{i,t}) - 2Cov_i(er_{i,t}, \widehat{er}_{i,t}), \tag{T2}$$

which can be re-expressed as

$$Var_i(\omega_{i,t}) = Var_i(\widehat{er}_{i,t}) - 2[Var_i(er_{i,t}) + Cov_i(er_{i,t}, \omega_{i,t})] + Var_i(er_{i,t}). \tag{T3}$$

The last right-hand-side term, firm i's time-series variance in expected returns, does not depend on the choice of ERP model. Therefore, in comparing the time-series variance of ERP measurement errors for firm i, one need only compare the first two terms of Eq. (T3), which we refer to collectively as the Scaled Time-Series Variance of an ERP's measurement errors of firm i's expected returns $[SVar_i(\omega_{i,t})]$:

$$SVar_i(\omega_{i,t}) = Var_i(\widehat{er}_{i,t}) - 2[Var_i(er_{i,t}) + Cov_i(er_{i,t}, \omega_{i,t})]. \tag{T4}$$

Notice that the first right-hand-side term is firm i's time-series variance in the ERP, which can be empirically observed. The second right-hand-side term involves unobservables: specifically, firm i's variance in expected returns $[Var_i(er_{i,t})]$ and the time-series covariance

between the firm's expected returns and the ERP measurement errors $[Cov_i(er_{i,t}, \omega_{i,t})]$. In what follows, we re-express the second term on the right-hand side in terms of variables that can be empirically observed.

Step 2. Decomposing Realized Returns and Time-Series Return-ERP Covariance

In this step we show that $Cov_i(r_{i,t+1}, \widehat{er}_{i,t}) = Var_i(er_{i,t}) + Cov_i(er_{i,t}, \omega_{i,t})$. To obtain this result, note that realized returns is the sum of the expected returns and news:

$$r_{i,t+1} = er_{i,t} + \delta_{i,t+1}. \tag{T5}$$

We define $er_{i,t}$ as firm i's true but unobserved expected returns conditional on publicly available information at time t, capturing all ex-ante predictability (with respect to the information set) in returns. By the decomposition property of conditional expectations (Angrist and Pischke, 2008), $er_{i,t}$ is uncorrelated with its forecast errors ($\delta_{i,t+1}$). Intuitively, if expected returns were correlated with subsequent forecast errors, one could always improve on the expected-return measure by taking into account such systematic predictability, thus violating the efficiency or the prediction property of conditional expectations (Angrist and Pischke, 2008). This justifies assumption **A2**.

We can thus write the time-series covariance between returns and ERPs as:

$$\begin{aligned}
Cov_i(r_{i,t+1}, \widehat{er}_{i,t}) &= Cov_i(er_{i,t} + \delta_{i,t+1}, er_{i,t} + \omega_{i,t}) \\
&= Var_i(er_{i,t}) + Cov_i(er_{i,t}, \omega_{i,t}),
\end{aligned} \tag{T6}$$

where the first equality follows from the return decomposition of Eq. (T5) and the definition of ERP (Eq. (T1)), and the last equality follows from assumption **A2**, which implies that $Cov_i(\delta_{i,t+1}, er_{i,t}) = Cov_i(\omega_{i,t+1}, \delta_{i,t+1}) = 0$.

Step 3. Estimating *AvgSVar^{TS}*

Substituting Eq. (T6) into Eqs. (T3) and (T4), we obtain:

$$Var_i(\omega_{i,t}) = Var_i(\widehat{er}_{i,t}) - 2Cov_i(r_{i,t+1}, \widehat{er}_{i,t}) + Var_i(er_{i,t}), \tag{T7}$$

so that

$$SVar_i(\omega_{i,t}) = Var_i(\widehat{er}_{i,t}) - 2Cov_i(r_{i,t+1}, \widehat{er}_{i,t}). \tag{T8}$$

The first term of $SVar_i$ shows that, all else equal, an ERP's MEV is increasing in the variance of the ERP. The second term of $SVar_i$ shows that, all else equal, an ERP's MEV is decreasing in the degree to which ERPs predict future returns in time series.

Notice that Eq. (T8) expresses $SVar_i(\omega_{i,t})$ in terms of two empirically observable variables $\{\widehat{er}_{i,t}, r_{i,t+1}\}$. These variables can be computed empirically, with consistency achieved under standard regularity conditions. Our empirical tests compute, for each ERP and each firm, the relative error-variance measure using Eq. (T8), and assess the time-series performance

of ERPs based on the average of $SVar_i$ across the N firms in our sample:

$$AvgSVar^{TS} = \frac{1}{N} \sum_i SVar_i(\omega_{i,t}). \tag{T9}$$

Part B. Estimating Cross-Sectional ERP Measurement-Error Variance

Here we derive a measure to evaluate ERPs on the basis of their average cross-sectional MEV. We call this measure Average Scaled CS Variance ($AvgSVar^{CS}$). Our derivation proceeds in two steps. In Step 1, we decompose a firm's cross-sectional ERP MEV and define our cross-section-specific Scaled CS Variance measure. In Step 2, we show how to estimate Average Scaled CS Variance using the average cross-sectional return-ERP covariance. We make the same assumptions as in Part A.

Step 1. Decomposing an ERP's Cross-Sectional Measurement-Error Variance and Defining $SVar_t(\omega_{i,t})$

As in the case of time series, the cross-sectional variance in ERP measurement errors can be written as

$$Var_t(\omega_{i,t}) = Var_t(\widehat{er}_{i,t}) + Var_t(er_{i,t}) - 2Cov_t(er_{i,t}, \widehat{er}_{i,t}), \tag{C1}$$

which can be re-expressed as

$$Var_t(\omega_{i,t}) = Var_t(\widehat{er}_{i,t}) - 2[Var_t(er_{i,t}) + Cov_t(er_{i,t}, \omega_{i,t})] + Var_t(er_{i,t}). \tag{C2}$$

The final right-hand-side term, the cross-sectional variance in expected returns at time t, does not depend on the choice of ERP model. Therefore, in comparing the cross-sectional variance of ERP measurement errors at time t, one need only compare the first two terms of Eq. (C2), which we refer to collectively as the Scaled CS Variance of an ERP's measurement errors [$SVar_t(\omega_{i,t})$]:

$$SVar_t(\omega_{i,t}) = Var_t(\widehat{er}_{i,t}) - 2[Var_t(er_{i,t}) + Cov_t(er_{i,t}, \omega_{i,t})]. \tag{C3}$$

The first right-hand-side term is the cross-sectional variance in the ERP, which can be empirically estimated. The second right-hand-side term involves unobservables—specifically, the cross-sectional variance in expected returns [$Var_t(er_{i,t})$] and the cross-sectional covariance between the firm's expected returns and the ERP measurement errors [$Cov_t(er_{i,t}, \omega_{i,t})$].

Step 2. Defining and Estimating $AvgSVar^{CS}$

In our empirical tests, we assess the cross-sectional performance of ERPs based on the average of $SVar_t$ across the T cross-sections in our sample:

$$AvgSVar^{CS} = \frac{1}{T} \sum_t Var_t(\widehat{er}_{i,t}) - 2[Var_t(er_{i,t}) + Cov_t(er_{i,t}, \omega_{i,t})]. \tag{C4}$$

To estimate $AvgSVar^{TS}$, we note that the *average* cross-sectional covariance between returns and ERPs can be expressed as:

$$\frac{1}{T}\sum_t Cov_t(r_{i,t+1}, \widehat{er}_{i,t}) = \frac{1}{T}\sum_t Cov_t(er_{i,t} + \delta_{i,t+1}, er_{i,t} + \omega_{i,t})$$

$$= \frac{1}{T}\sum_t [Var_t(er_{i,t}) + Cov_t(er_{i,t}, \omega_{i,t})], \qquad (C5)$$

where the first equality follows from the realized returns decomposition (Eq. (T5)) and the definition of expected-returns proxy (Eq. (T1)), and the last equality follows from the assumption (**A2**) that news cannot exhibit systematic forecastability.[21] Combining Eqs. C4 and C5 yields the following expression for $AvgSVar^{CS}$ that can be estimated by data:

$$AvgSVar^{CS} = \frac{1}{T}\sum_t Var_t(\widehat{er}_{i,t}) - 2Cov_t(r_{i,t+1}, \widehat{er}_{i,t}). \qquad (C6)$$

Part C. Return-Regression Slope-Coefficient Test and MEV

Here we show that achieving a return-regression slope of 1 does not imply that an ERP has zero MEV. To illustrate this point, we identify a set of ERPs with slope coefficients of 1 but that vary significantly in terms of MEV. To construct such a set, let $ERP'_t = C \times ERP_t$ where $ERP_t = er_t + o_t$, o_t is a white-noise process, and $C = 1/\left[1 + \frac{Var(o_t)}{Var(erp_t)}\right]$. A regression of future returns on ERP'_t yields a slope of 1:

$$\hat{\beta}' = \frac{Cov(r_{t+1}, C \times ERP_t)}{Var(C \times ERP_t)} = \frac{1}{C}\frac{Cov(er_t + \delta_{t+1}, er_t + o_t)}{Var(er_t + o_t)} = \frac{1}{C}\frac{Var(er_t)}{Var(er_t) + Var(o_t)} = 1. \qquad (S1)$$

However, these ERPs exhibit MEV,

$$Var(\omega'_t) = Var(ERP'_t - er_t) = [Var(C \times (er_t + o_t) - er_t)] = Var(er_t)\left[\frac{1}{1 + \frac{Var(er_t)}{Var(o_t)}}\right], \qquad (S2)$$

that ranges from zero (i.e., when $Var(o_t)$ is arbitrarily small) to the variance of expected returns (i.e., when $Var(o_t)$ is arbitrarily large).

---

[21]Note that, for any given cross-section, it might be possible for realized news and measurement errors to be correlated, but that this cannot be true systematically (i.e., across many cross-sections) by the definition of news.

Appendix I. A Family Tree of Expected-Return Proxies based on Initial Conceptual Origin

```
┌────────────────────────────────────────┐
│ Firm-Level Expected-Return Proxy (ERP) │
└────────────────────────────────────────┘
```

Empirical Asset-Pricing Approach

Implied-Cost-of-Capital Approach

Basic Premise

- Equilibrium Pricing: only non-diversifiable risks are priced
- A firm's ERP is a linear function of its sensitivity to each factor (β's) and the price of each factor (γ's)

- The share price reflects the PV of the expected payoff to shareholders
- A firm's ERP is the IRR that equates expected future cash flow (CF) with current price
- Agnostic with respect to the source of risk

Estimation Challenges

- Factor Identification: which factors matter?
- Risk Premium for each factor (the γ's)
- Firm factor loadings (the β's)

- CF forecasting assumptions: future earnings/FCF/dividends? terminal value estimation?
- Expected return dynamics assumptions: constant expected returns? intertemporal variation expected returns?

Examples of ERP Variants

- CAPM (Sharpe, 1964; Lintner, 1965)
- FF4 (Fama and French, 1993; Fama and French, 1996)
- JLR (Lewellen, 2015)

- ICC (Gebhardt et al., 2001; Hou et al., 2012)
- LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2020)
- FICC (Fitted ICC, a new proxy)

We note that there has been considerable convergence between the Empirical Asset-Pricing approach and the Implied-Cost-of-Capital approach in recent years. Specifically, a recent Empirical Asset-Pricing ERP estimate using characteristic-based factor loadings (i.e., JLR from Lewellen, 2015) closely resembles present-value-model-based ERPs (such as LPV from Lyle and Wang, 2015; Chattopadhyay et al., 2020). In Appendix II, we regroup 17 ERPs from the literature into four ERP families, based on their empirical construction methodology rather than their initial conceptual origins.

Appendix II. Grouping Expected-Return Proxies by Empirical Construction Methodology

This table groups the 17 expected-return proxy (ERP) variants examined in this study into four families, based on similarities in their empirical construction. For each ERP family, we create a composite variable (named in the first column) by equal-weighting some or all of the individual family-member ERPs. The table then briefly describes each family (in the second column), explains how the factor risk-premium is estimated (in the third column), explains how the factor loadings on risk factors are computed (in the fourth column), and lists each family-member ERP and related studies (in the last column). Formulas for ERP construction follow the table.

Label	Family Description	Factor risk-premium (Gamma) estimation	Factor-loading (Beta) estimation	Individual family-member ERPs and related prior studies
FER	A factor-based ERP based on each firm's estimated sensitivity to various common factors (e.g., MKT, SMB, HML, and UMD) and estimated expected premiums for each factor	Based on each factor's realized risk premium over the preceding month	Based on time-series regressions using realized returns over the preceding 60 months	CAPM (Sharpe, 1964) FF3 (Fama and French, 1993) FF4 (Carhart, 1997) FF5 (Fama and French, 2015) FF6 (Fama and French, 2018) $q4$ (Hou et al., 2015) $q5$ (Hou et al., 2019)
ICC	The internal rate of return that equates a firm's forecasted cash-flows to its current market price	Not applicable	Not applicable	GLS (Gebhardt et al., 2001) CAT (Claus and Thomas, 2001) PEG (Easton, 2004) AGR (Ohlson and Juettner-Nauroth, 2005)
CER	A characteristic-based ERP whereby each firm's one-month-ahead is "fitted" to a set of firm characteristics (e.g., size, bm, momentum)	Based on ten-year-average Fama-MacBeth coefficients for each characteristic	Based on current year firm characteristics	JLR (Lewellen, 2015) LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2020)
FICC	A characteristic-based ERP whereby each firm's ICC is "fitted" to a set of firm characteristics	Based on ten-year-average Fama-MacBeth coefficients for each characteristic	Based on current-year firm characteristics	FICC is a new proxy, derived by regressing our composite ICC measure on three firm characteristics: size, B/M ratio, and momentum.

Factor-Based Expected-Return Proxies

Factor-based ERPs are constructed at the end of each calendar month t as

$$\widehat{\mathbb{E}}_t[r_{i,t+1}] = rf_{t+1} + \sum_{j=1}^{J} \hat{\beta}_i \widehat{\mathbb{E}}_t[f_{j,t}],$$

for a factor model with J factors where rf_{t+1} is the risk-free rate in period $t+1$, $\hat{\beta}_i$ are the estimated factor sensitivities, and $f_{j,t}$ are the corresponding factors in period t. In our baseline results, factor sensitivities are estimated in time series for each firm using monthly stock and factors' returns over the 60 months prior to the forecast date, and expected factor premiums are the most recent period's realized factor premium.

Implied Cost of Capital

Implied-cost-of-capital measures are computed at the end of each calendar month t as the internal rate of returns that equates the stock price to expected future cash flows, based on a particular valuation-model formulation. We describe two types of valuation models and four ICC variants below.

Two of the ICCs we examine are based on variants of the residual-income model:

$$P_{i,t} = B_{i,t} + \sum_{\tau=1}^{\infty} \frac{\mathbb{E}_t[EPS_{i,t+\tau}] - r_e \times \mathbb{E}_t[B_{i,t+\tau-1}]}{(1+r_e)^\tau},$$

where EPS_t is earnings per share, B_t is book value per share, P_t denotes the stock price, and r_e denotes the discount rate. We consider two of the most popular ICCs based on this model, proposed by Gebhardt et al. (2001) (GLS) and Claus and Thomas (2001) (CAT), which differ in terms of their forecasting horizon and terminal value estimation.

The other two ICC variants we examine are based on the abnormal-growth-in-earnings model:

$$P_t = \frac{EPS_{t+1}}{r_e} + \sum_{i=2}^{T} \left(\frac{agr_{t+i}}{r_e(1+r_e)^{i-1}} + \frac{agr_{t+T}(t+g)}{r_e(r_e-g)(1+r_e)} \right)$$

where $agr_{i,t} = EPS_t + r_e DPS_{t-1} - (1+r_e)EPS_{t-1}$, $DPS_{i,t}$ denotes dividends per share, and g denotes the perpetual growth rate in $agr_{i,t}$. We consider two of the most popular ICCs based on a two-period version of this model, proposed by Easton (2004) ("PEG") and Ohlson and Juettner-Nauroth (2005) ("AGR").

GLS: In this formulation, earnings are forecasted explicitly for the first three years using the Hou et al. (2012) methodology. For years 4 through 12, each firm's ROE forecast is obtained from a linear interpolation from the 3-year-ahead ROE forecast to the industry-median ROE. We compute each industry's median ROE each year using available data over the preceding ten years. The terminal value beyond year 12 is computed as the present value of capitalized period 12 residual income. Among the models we test, GLS alone uses industry-based profitability estimates. GLS is the r_e satisfying the following equation:

$$P_{i,t} = B_{i,t} + \sum_{n=1}^{11} \frac{\frac{\mathbb{E}_t[NI_{i,t+n}]}{\mathbb{E}_t[B_{i,t+n-1}]} - r_e}{(1+r_e)^n} \mathbb{E}_t[B_{i,t+n-1}] + \frac{\frac{\mathbb{E}_t[NI_{i,t+12}]}{\mathbb{E}_t[B_{i,t+11}]} - r_e}{r_e(1+r_e)^{11}} \mathbb{E}_t[B_{i,t+11}],$$

CAT: In this formulation, we use mechanical earnings forecasts up to five-years-ahead, an assumed growth rate g equal to the risk-free rate minus 3 percent; future book values are derived from the clean surplus relation. Thus, CAT is the r_e satisfying the following equation:

$$P_{i,t} = B_{i,t} + \sum_{n=1}^{5} \frac{\frac{\mathbb{E}_t[NI_{i,t+n}]}{\mathbb{E}_t[B_{i,t+n-1}]} - r_e}{(1+r_e)^n} \mathbb{E}_t[B_{i,t+n-1}] + \frac{\frac{\mathbb{E}_t[NI_{i,t+5}]}{\mathbb{E}_t[B_{i,t+4}]} - r_e}{(r_e-g) \times (1+r_e)^5} \mathbb{E}_t[B_{i,t+4}](1+g)$$

PEG: This formulation is a special case of the abnormal-earnings-growth model in which $T = 2$, and $g = 0$, with the additional assumption that $DPS_{t+1} = 0$. Here, the ICC estimate obtains a closed-form solution:

$$r_e = \sqrt{(EPS_{t+2} - EPS_{t+1})/P_t},$$

where the estimate requires that $EPS_{t+2} \geq EPS_{t+1}$. Note that a variant known as the modified PEG (MPEG), which we implement in Section VI, allows for non-zero DPS_{t+1}, so that $r_e = \sqrt{(EPS_{t+2} + DPS_{t+1} - EPS_{t+1})/P_t}$.

AGR: This formulation is a special case of the abnormal-earnings-growth model in which $T = 2$, using a specific computation for long-term growth in abnormal earnings. Here, the ICC estimate is the r_e that solves the following equation:

$$P_t = \frac{EPS_{t+1}}{r_e} + \frac{EPS_{t+1}\{EPS_{t+2} + r_e DPS_{t+1} - (1+r_e)EPS_{t+1}\}}{r_e\{r_e - \frac{EPS_{t+3} + r_e DPS_{t+2} - (1+r_e)EPS_{t+2}}{EPS_{t+2} + r_e DPS_{t+1} - (1+r_e)EPS_{t+1}}\}}.$$

In our main results, ICCs are based on earnings forecasts derived from the cross-sectional mechanical forecast model of Hou et al. (2012). This model forecasts future income based on historical regressions of the following model:

$$\mathbb{E}[E_{i,t+\tau}] = \beta_0 + \beta_1 TA_{i,t} + \beta_2 DIV_{i,t} + \beta_3 DD_{i,t} + \beta_4 E_{i,t} + \beta_5 NEGE_{i,t} + \beta_6 AC_{i,t} + \epsilon_{i,t},$$

where $E_{i,t+\tau}$ is income (Compustat data item "ib") in year $t + \tau$ ($\tau = 1$ to 5); $TA_{i,t}$ is total assets (Compustat item "at"); $DIV_{i,t}$ is dividends (Compustat item "dvt"); $DD_{i,t}$ is an indicator for firms with non-zero dividends; $E_{i,t}$ is income in year t; $NEGE_{i,t}$ is an indicator for loss-making firms; and $AC_{i,t}$ is working-capital accruals. For each forecasting horizon, we estimate the regression by pooling the most recent ten years of data available as of the forecasting date. For example, to generate one-year-ahead earnings forecasts in year t, we use data from years $t - 10$ to $t - 1$ as the regressors and income data from years $t - 9$ to t as the outcome variable. We then multiply the estimated coefficients by the values of the inputs in year t for each firm to arrive at the forecast of income for year $t + 1$. For two-year-ahead earnings forecasts in year t, on the other hand, we use data from years $t - 11$ to $t - 2$ as the regressors and income data from $t - 10$ to t as the outcome variable. This model produces forecasts of total earnings, which we scale by shares outstanding to derive EPS forecasts that are then used as inputs in ICC computations. As a summary, we report in the following table the Fama-Macbeth coefficients, their t-statistics (in parentheses), and the R^2 from our estimation of this model.

τ-Years-Ahead Earnings	Intercept	V	TA	DIV	DD	E	NEGE	AC	R^2
$\tau = 1$	2.670	0.009	-0.006	0.282	-2.706	0.729	-0.070	-0.025	0.852
	(6.54)	(45.80)	-(26.63)	(30.21)	-(4.06)	(160.78)	(1.33)	-(8.76)	
$\tau = 2$	4.636	0.011	-0.006	0.402	-3.280	0.663	0.136	-0.043	0.790
	(8.47)	(38.18)	-(17.01)	(31.86)	-(4.12)	(99.49)	(1.61)	-(9.83)	
$\tau = 3$	12.550	0.004	0.000	0.476	-7.128	0.362	-0.549	-0.025	0.519
	(24.41)	(18.44)	-(2.14)	(33.56)	-(9.35)	(54.26)	-(0.21)	-(5.05)	
$\tau = 4$	17.229	0.002	0.002	0.443	-9.550	0.286	-0.913	-0.009	0.413
	(30.74)	(10.64)	(3.07)	(30.51)	-(11.26)	(40.18)	-(0.82)	-(1.16)	
$\tau = 5$	20.921	0.001	0.003	0.410	-11.215	0.250	-1.055	0.002	0.345
	(34.81)	(5.23)	(6.49)	(27.61)	-(12.43)	(34.25)	-(1.20)	(1.02)	

Characteristic-Based Expected-Return Proxies and Fitted ICCs

Characteristic-based ERPs are constructed at the end of each calendar month t as

$$\widehat{\mathbb{E}}_t[r_{i,t+1}] = \hat{\delta}_0 + \sum_{j=1}^{K} \hat{\delta}_j X_{i,t},$$

for a model with K characteristics where $(\hat{\delta}_0, \hat{\delta}_1, ..., \hat{\delta}_K)$ are obtained from historical regressions of realized monthly returns on firm characteristics. Our baseline results construct characteristic-based ERPs using rolling 10-year Fama-MacBeth coefficients.

FICC is also a characteristic-based ERP. However, the weights on characteristics are obtained using historical regressions of ICCs on firm characteristics. In this paper, we construct FICC using rolling 10-year Fama-MacBeth coefficients estimated from the composite ICC.

Table 1. Use of ICCs in Academic Literature

This table reports the total number of papers that examine magnitudes and variations in ICCs. The papers tabulated have appeared between 1997 and 2016 in the following five finance journals—*Journal of Corporate Finance*, *Journal of Finance*, *Journal of Financial Economics*, *Journal of Financial and Quantitative Analysis*, and *Review of Financial Studies*—and five accounting journals— *Contemporary Accounting Research*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Review of Accounting Studies*, and *The Accounting Review*. These articles were identified by searching for keywords and citations. Keyword searching involves searching abstracts and full texts in ABI-Proquest, Business Source Complete, and the historical archives of each journal for combinations of the terms *implied* and *ex ante* with variations on the term *cost of capital* (e.g., *cost of equity capital*, *equity cost of capital*, *risk premium*). Citation searching uses Google Scholar to find papers in leading journals that cite the following methodological papers in the implied-cost-of-capital literature: Botosan (1997), Claus and Thomas (2001), Easton (2004), Gebhardt et al. (2001), Gode and Mohanram (2003), Ohlson and Juettner-Nauroth (2005). The second column tabulates the number of papers that identify the parameters of interest using time-series or cross-sectional variation in the data and a breakdown of the research designs; the third column tabulates the corresponding percentages. A full listing of the papers can be found in the online appendix.

Category	Count	Percent
Total Papers	**98**	**100%**
A. Treatment-Effect Studies	75	77% of total
A1. Pure Time-Series Data	1	1% of A
A2. Pure Cross-Sectional Data	2	3% of A
A3. Panel Data	72	96% of A
A3[a]. Time FE/Firm FE/Changes	43	60% of A3
A3[b]. Fama-Macbeth	25	35% of A3
A3[c]. Diff-in-Diff	5	7% of A3
B. Equity-Risk-Premium Studies	5	5% of Total
C. Methodological or Forecasting Studies	18	18% of Total

Table 2. Monthly Expected-Return Proxies by Year

The upper panel of this table reports the annual mean and standard deviation of monthly expected-return proxies. The lower panel reports the pooled summary statistics on the annual mean and standard deviations. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2020). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return. A full description of each proxy appears in Section III and Appendix II.

Year	Obs	Mean				Std Deviation			
		ICC	CER	FICC	FER	ICC	CER	FICC	FER
1977	13,384	1.61%	0.84%	1.64%	1.23%	0.75%	0.33%	0.60%	4.32%
1978	27,274	1.58%	0.80%	1.60%	2.51%	0.73%	0.32%	0.60%	5.97%
1979	27,734	1.62%	0.94%	1.57%	3.04%	0.70%	0.36%	0.60%	5.83%
1980	28,905	1.48%	1.35%	1.52%	2.97%	0.75%	0.37%	0.61%	6.32%
1981	29,455	0.77%	1.38%	1.43%	0.81%	0.60%	0.42%	0.61%	5.65%
1982	29,761	0.95%	1.18%	1.43%	2.02%	0.67%	0.45%	0.60%	5.83%
1983	31,061	1.31%	1.84%	1.21%	2.53%	0.61%	0.35%	0.60%	5.55%
1984	31,453	1.25%	1.78%	1.29%	-0.18%	0.69%	0.36%	0.58%	5.24%
1985	34,154	1.36%	1.83%	1.26%	2.05%	0.65%	0.31%	0.59%	5.09%
1986	34,440	1.26%	1.78%	1.18%	0.89%	0.67%	0.31%	0.59%	5.71%
1987	34,312	1.21%	1.69%	1.18%	1.24%	0.65%	0.35%	0.58%	6.46%
1988	35,444	1.21%	1.47%	1.24%	1.98%	0.63%	0.41%	0.57%	5.04%
1989	37,005	1.17%	1.52%	1.17%	1.54%	0.67%	0.40%	0.55%	4.62%
1990	36,834	1.25%	1.15%	1.22%	-0.87%	0.74%	0.45%	0.56%	6.53%
1991	36,682	1.13%	1.18%	1.18%	2.82%	0.75%	0.48%	0.58%	5.73%
1992	36,710	1.01%	1.32%	1.16%	1.52%	0.68%	0.46%	0.55%	5.26%
1993	37,911	0.88%	1.01%	1.09%	1.76%	0.59%	0.41%	0.52%	4.64%
1994	43,210	0.92%	1.09%	1.09%	0.25%	0.59%	0.39%	0.47%	4.44%
1995	47,296	0.94%	1.10%	1.06%	2.43%	0.61%	0.37%	0.46%	4.50%
1996	49,041	0.98%	1.13%	0.98%	1.85%	0.61%	0.38%	0.44%	5.55%
1997	50,670	0.95%	1.16%	0.95%	2.13%	0.60%	0.40%	0.42%	6.10%
1998	51,576	0.92%	1.24%	0.94%	1.03%	0.59%	0.37%	0.43%	6.58%
1999	49,690	1.00%	1.21%	0.98%	1.72%	0.62%	0.40%	0.44%	6.55%
2000	48,006	1.00%	1.55%	0.94%	0.26%	0.62%	0.45%	0.45%	7.62%
2001	45,715	0.98%	1.45%	0.97%	1.48%	0.64%	0.46%	0.44%	7.38%
2002	44,088	1.00%	1.32%	0.95%	-0.51%	0.66%	0.44%	0.41%	7.03%
2003	42,461	0.90%	1.25%	0.92%	3.54%	0.63%	0.42%	0.39%	5.16%
2004	41,851	0.82%	1.42%	0.79%	1.84%	0.58%	0.39%	0.36%	4.94%
2005	41,199	0.81%	1.31%	0.80%	0.96%	0.56%	0.35%	0.36%	4.85%
2006	40,854	0.90%	1.26%	0.78%	1.86%	0.58%	0.33%	0.35%	4.58%
2007	40,262	0.90%	1.16%	0.79%	0.25%	0.58%	0.33%	0.35%	4.59%
2008	39,509	1.02%	0.94%	0.93%	-1.55%	0.66%	0.32%	0.40%	6.67%
2009	36,868	1.11%	0.78%	1.04%	3.02%	0.76%	0.34%	0.43%	7.45%
2010	36,533	1.01%	0.73%	0.88%	2.41%	0.69%	0.33%	0.41%	6.83%
2011	35,878	1.02%	0.75%	0.88%	0.00%	0.71%	0.29%	0.42%	5.88%
2012	35,500	1.07%	0.79%	0.94%	1.80%	0.72%	0.26%	0.44%	4.88%
2013	34,743	1.02%	0.82%	0.88%	2.99%	0.69%	0.21%	0.44%	4.33%
2014	34,635	0.98%	0.69%	0.85%	0.82%	0.67%	0.14%	0.44%	5.08%
2015	34,463	1.03%	0.58%	0.91%	-0.18%	0.72%	0.12%	0.48%	5.65%
2016	33,504	1.07%	0.46%	0.96%	1.94%	0.72%	0.08%	0.50%	5.84%
2017	32,261	0.96%	0.51%	0.89%	1.52%	0.68%	0.10%	0.51%	4.24%
2018	26,935	0.89%	0.75%	0.87%	0.70%	0.63%	0.18%	0.52%	5.18%
Mean	37,687	1.09%	1.22%	1.10%	1.49%	0.66%	0.37%	0.49%	5.65%
Median	36,772	1.01%	1.20%	1.05%	1.74%	0.66%	0.37%	0.46%	5.60%
Min	13,384	0.77%	0.69%	0.78%	-1.55%	0.56%	0.14%	0.35%	4.32%
Max	51,576	1.62%	1.84%	1.64%	3.54%	0.76%	0.48%	0.61%	7.62%
Std	7,713	0.22%	0.32%	0.24%	1.18%	0.06%	0.07%	0.09%	0.94%
Mean/Std	4.89	4.95	3.74	4.59	1.26	11.66	5.19	5.52	6.04

Table 3. Correlation between Expected-Return Proxies

This table reports the average monthly Pearson (Spearman) correlations above (below) the diagonal among the four expected-return proxies. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2020). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; the table is based on a common sample of observations with non-missing values in each of the four expected-return proxies. A full description of each proxy appears in Section III and Appendix II. We report t-statistics based on Newey-West-corrected standard errors in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1% respectively.

	ICC	CER	FICC	FER
ICC		0.424*** (17.34)	0.630*** (43.19)	-0.069*** (-9.59)
CER	0.457*** (21.98)		0.632*** (16.77)	-0.047*** (-6.15)
FICC	0.627*** (36.12)	0.647*** (17.90)		-0.037*** (-3.24)
FER	-0.074*** (-9.54)	-0.052*** (-6.74)	-0.041*** (-3.59)	

Table 4. Cross-Sectional Scaled Measurement-Error Variances

This table presents descriptive statistics for the cross-sectional scaled measurement-error variances (multiplied by 100) of four expected-return proxies, computed using Eq. (12) as follows:

$$SVar^{CS} = Var_t(\widehat{er}_{i,t}) - 2Cov_t(r_{i,t+1}, \widehat{er}_{i,t}),$$

where $Var_t(\widehat{er}_{i,t})$ is a given ERP's cross-sectional variance at time t and $Cov_t(r_{i,t+1}, \widehat{er}_{i,t})$ is the cross-sectional covariance between firms' ERPs at time t and their realized returns at $t+1$. Panel A reports distributional statistics of $SVar^{CS}$ for each expected-return proxy, using data from a sample of 497 calendar months during our 1977-2018 sample period. The rightmost column of this panel reports t-Statistics testing the null hypothesis that the mean $SVar^{CS}$ for a given expected-return proxy is 0. Panel B reports pair-wise comparisons of average $SVar^{CS}$ within the sample. Table values are negative (positive) when the expected-return proxy displayed in the leftmost column has a larger (smaller) scaled measurement-error variance than the expected-return proxy displayed in the topmost row. t-statistics testing the differences in means are reported in parentheses. t-statistics are estimated using Newey-West-corrected standard errors. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2020). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; the table is based on a common sample of observations with non-missing values in each of the four expected-return proxies. A full description of each proxy appears in Section III and Appendix II.

Panel A: Cross-Sectional Measurement-Error Variance (N=497)						
	Mean	P25	Median	P75	STD	t-stat
ICC	0.0003	-0.0070	0.0016	0.0112	0.0187	0.3652
CER	-0.0017	-0.0049	-0.0006	0.0028	0.0112	-3.3785
FICC	-0.0004	-0.0074	-0.0004	0.0086	0.0160	-0.5593
FER	0.1541	0.0626	0.1200	0.2032	0.2726	12.6005

Panel B: t-Statistics of Differences in Variances				
	ICC	CER	FICC	FER
ICC		-0.002	-0.001	0.154
		-(3.13)	-(1.69)	(11.83)
CER	0.002		0.001	0.156
	(3.13)		(1.95)	(12.13)
FICC	0.001	-0.001		0.154
	(1.69)	-(1.95)		(11.97)
FER	-0.154	-0.156	-0.154	
	-(11.83)	-(12.13)	-(11.97)	

Table 5. Time-Series Scaled Measurement-Error Variances

This table presents descriptive statistics for the time-series scaled measurement-error variances (multiplied by 100) of four expected-return proxies, computed using Eq. (11) as follows:

$$SVar^{TS} = Var_i(\widehat{er}_{i,t}) - 2Cov_i(r_{i,t+1}, \widehat{er}_{i,t}),$$

where $Var_i(\widehat{er}_{i,t})$ is a given ERP's time-series variance for firm i and $Cov_i(r_{i,t+1}, \widehat{er}_{i,t})$ is the time-series covariance between firm i's ERP at t and its realized returns in at $t+1$. Panel A reports distributional statistics of $SVar^{TS}$ for each expected-return proxy, using data from a sample of 11,620 unique firms with at least 20 observations in the 1977-2018 sample period. The rightmost column reports t-statistics testing the null hypothesis that the mean $SVar^{TS}$ for a given expected-return proxy is 0. Panel B reports pairwise comparisons of average $SVar^{TS}$ within the sample. Table values are negative (positive) when the expected-return proxy displayed in the leftmost column has a larger (smaller) scaled measurement-error variance than the expected-return proxy displayed in the topmost row. t-statistics testing the differences in means are reported in parentheses. t-statistics are estimated using Newey-West-corrected standard errors. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2020). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; the table is based on a common sample of observations with non-missing values in each of the four expected-return proxies. A full description of each proxy appears in Section III and Appendix II.

Panel A: Time-Series Measurement-Error Variance (N=11,620)						
	Mean	P25	Median	P75	STD	t-stat
ICC	-0.0089	-0.0147	-0.0038	0.0016	0.0335	-28.7271
CER	-0.0002	-0.0055	0.0000	0.0052	0.0188	-1.3847
FICC	-0.0073	-0.0105	-0.0042	-0.0009	0.0161	-48.4993
FER	0.2792	0.0971	0.2248	0.4042	0.4226	71.1859

Panel B: t-Statistics of Differences in Variances				
	ICC	CER	FICC	FER
ICC		0.009	0.002	0.288
		(24.92)	(6.43)	(70.27)
CER	-0.009		-0.007	0.279
	-(24.92)		-(30.75)	(68.91)
FICC	-0.002	0.007		0.286
	-(6.43)	(30.75)		(70.61)
FER	-0.288	-0.279	-0.286	
	-(70.27)	-(68.91)	-(70.61)	

Table 6. Subsample Variation in Scaled Measurement-Error Variances

This table presents descriptive statistics for the cross-sectional and time-series scaled measurement-error variances (multiplied by 100) of four expected-return proxies, for six sub samples. Panels A, B, C, and D focus on the subsamples of large, all-but-tiny, small, and tiny firms. A firm is large if its market capitalization at the beginning of the calendar year is larger than the NYSE median; otherwise it is considered small. A firm is tiny if its market capitalization at the beginning of the calendar year is below the 20th percentile of NYSE; otherwise it is part of the all-but-tiny subsample. Panels E and F focus on, respectively, the first half (1977-1997) and second half (1998-2018) of our sample. t-statistics are estimated using Newey-West-corrected standard errors, and test the null hypothesis that the mean SVAR is zero. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2020). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; the table is based on a common sample of observations with non-missing values in each of the four expected-return proxies. A full description of each proxy appears in Section III and Appendix II.

	Cross-Sectional SVARS				Time-Series SVAR			
	Mean	Median	STD	t-Stat	Mean	Median	STD	t-Stat
Panel A: Large Firms (T=497, N=6,785)								
ICC	0.0009	0.0019	0.0120	1.7481	-0.0028	-0.0019	0.0202	-11.4452
CER	-0.0009	0.0001	0.0091	-2.0894	0.0008	0.0005	0.0144	4.3682
FICC	0.0007	0.0015	0.0105	1.5786	-0.0045	-0.0031	0.0115	-32.2040
FER	0.1216	0.0880	0.2580	10.5101	0.3028	0.2395	0.3574	69.8112
Panel B: All-but-tiny Firms (T=497, N=10,020)								
ICC	0.0014	0.0029	0.0147	2.1418	-0.0055	-0.0029	0.0275	-19.9712
CER	-0.0012	-0.0001	0.0098	-2.6734	0.0005	0.0002	0.0162	2.9460
FICC	0.0010	0.0020	0.0134	1.7059	-0.0059	-0.0038	0.0139	-42.6132
FER	0.1357	0.1059	0.2629	11.5116	0.2809	0.2285	0.3757	74.8499
Panel C: Small Firms (T=497, N=7,008)								
ICC	-0.0002	0.0008	0.0201	-0.2595	-0.0120	-0.0062	0.0426	-23.6082
CER	-0.0018	-0.0015	0.0098	-4.1898	-0.0004	-0.0002	0.0227	-1.2985
FICC	-0.0017	-0.0009	0.0103	-3.7350	-0.0075	-0.0043	0.0193	-32.4092
FER	0.1966	0.1638	0.2953	14.8467	0.2669	0.2111	0.5045	44.2990
Panel D: Tiny Firms (T=497, N=3,182)								
ICC	-0.0014	-0.0006	0.0199	-1.5230	-0.0150	-0.0087	0.0471	-17.9110
CER	-0.0015	-0.0013	0.0127	-2.5601	-0.0015	-0.0010	0.0300	-2.8036
FICC	-0.0026	-0.0011	0.0101	-5.8224	-0.0082	-0.0041	0.0211	-21.9087
FER	0.2416	0.1937	0.3472	15.5156	0.2622	0.1930	0.5945	24.8727
Panel E: 1977-1997 (T=246, N=7,248)								
ICC	0.0000	0.0002	0.0163	0.0294	-0.0079	-0.0031	0.0287	-23.3984
CER	-0.0016	-0.0016	0.0066	-3.8979	0.0004	0.0003	0.0185	2.0423
FICC	-0.0001	-0.0009	0.0170	-0.1128	-0.0066	-0.0038	0.0129	-43.1083
FER	0.1465	0.1155	0.1421	16.1681	0.2514	0.1973	0.3636	58.8588
Panel F: 1998-2018 (T=251, N=7,384)								
ICC	0.0006	0.0029	0.0208	0.4385	-0.0097	-0.0039	0.0358	-23.2272
CER	-0.0018	0.0004	0.0144	-1.9435	-0.0014	-0.0003	0.0182	-6.5420
FICC	-0.0007	-0.0003	0.0150	-0.7139	-0.0078	-0.0042	0.0190	-35.3695
FER	0.1615	0.1228	0.3571	7.1645	0.2985	0.2438	0.4710	54.4602

Table 7. Scaled Measurement-Error Variances of Factor-Based ERPs

This table presents descriptive statistics for the cross-sectional and time-series scaled measurement-error variances (multiplied by 100) of specific variants of factor-based expected-return proxies. t-Statistics are estimated using Newey-West-corrected standard errors, and test the null hypothesis that the mean SVAR is zero. Each factor-based ERP is constructed as the inner product of estimated factor sensitivities (βs) and expected factor premiums (γs), based on a particular factor model. CAPM is based on the capital-asset pricing model; FF3 is based on the Fama and French (1993) three-factor model; FF4 augments FF3 with a momentum factor; FF5 is based on the Fama and French (2015) five-factor model; and FF6 augments FF5 with a momentum factor (Fama and French, 2018); $q4$ is the Hou et al. (2015) q-factor model; and $q5$ augments $q4$ with an expected growth factor (Hou et al., 2019). FER is a composite measure of factor-based expected returns that takes the equal-weighted average of $q5$ and FF6. Panels A-F report on factor-based expected-return proxies that vary in the estimation of factor sensitivities or expected future factor premiums. Panel A, B, and C use rolling-60-month, full-sample, and industry average of rolling-60-month firm-level βs respectively; each panel uses the prior-month's factor premiums as estimates of $gamma$s. Panels D and E use as estimates of γs rolling-60-month average factor premiums and cumulative-average factor premiums respectively; both use rolling-60-month βs. Panel F uses as estimates of γs factor premiums measured contemporaneously with realized returns (i.e., realized factor premiums) combined with rolling-60-month βs. We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; each panel is based on a common sample of observations with non-missing values in each of the expected-return proxies. A full description of each proxy appears in Section III and Appendix II.

	Cross-Sectional SVARS (T=497)				Time-Series SVARS (N=11,611)			
	Mean	Median	STD	t-Stat	Mean	Median	STD	t-Stat
Panel A: Rolling βs and Prior-Month Factor Premiums								
CAPM	0.0517	0.0421	0.0805	14.3052	0.1204	0.1015	0.2785	46.5811
FF3	0.1172	0.0911	0.1818	14.3705	0.2125	0.1772	0.3604	63.5473
FF4	0.1455	0.1101	0.2138	15.1692	0.2457	0.2052	0.3884	68.1558
FF5	0.1693	0.1385	0.2306	16.3649	0.2770	0.2307	0.4027	74.1135
FF6	0.1948	0.1526	0.2513	17.2828	0.3076	0.2541	0.4279	77.4449
$q4$	0.1506	0.1227	0.2252	14.9085	0.2554	0.2096	0.3963	69.4395
$q5$	0.1799	0.1483	0.2360	16.9876	0.2945	0.2380	0.4244	74.7596
FER	0.1541	0.1200	0.2726	12.6005	0.2792	0.2248	0.4226	71.1859
Panel B: Full-Sample βs and Prior-Month Factor Premiums								
CAPM	0.0353	0.0337	0.0762	10.3264	0.1062	0.0898	0.2730	41.9224
FF3	0.0607	0.0505	0.1595	8.4770	0.1482	0.1239	0.3430	46.5621
FF4	0.0739	0.0589	0.1933	8.5204	0.1643	0.1371	0.3478	50.9088
FF5	0.0776	0.0661	0.2202	7.8556	0.1726	0.1447	0.3462	53.7103
FF6	0.0886	0.0739	0.2440	8.0995	0.1857	0.1549	0.3666	54.5717
$q4$	0.0814	0.0665	0.2121	8.5583	0.1752	0.1438	0.3597	52.4776
$q5$	0.0971	0.0802	0.2207	9.8024	0.1976	0.1593	0.3822	55.7072
FER	0.0767	0.0612	0.2261	7.5670	0.1696	0.1403	0.3553	51.4487
Panel C: Industry-Average Rolling βs and Prior-Month Factor Premiums								
CAPM	0.0025	-0.0001	0.0397	1.3829	0.0648	0.0749	0.2323	30.0408
FF3	0.0058	0.0019	0.0668	1.9178	0.0704	0.0827	0.2611	29.0546
FF4	0.0062	0.0026	0.0721	1.9058	0.0687	0.0812	0.2605	28.4017
FF5	0.0058	0.0031	0.0819	1.5617	0.0695	0.0831	0.2639	28.3716
FF6	0.0062	0.0037	0.0857	1.6110	0.0682	0.0823	0.2641	27.8118
$q4$	0.0049	0.0021	0.0747	1.4480	0.0671	0.0815	0.2545	28.4171
$q5$	0.0057	0.0026	0.0764	1.6464	0.0681	0.0821	0.2553	28.7267
FER	0.0044	0.0013	0.0798	1.2185	0.0670	0.0804	0.2595	27.8209

Table 7. [Continued]

	Cross-Sectional SVARS (T=497)				Time-Series SVARS (N=11,611)			
	Mean	Median	STD	t-Stat	Mean	Median	STD	t-Stat
Panel D: Rolling βs and Rolling-Average Factor Premiums								
CAPM	0.0161	0.0161	0.0248	14.4818	0.0290	0.0202	0.0821	37.9027
FF3	0.0178	0.0176	0.0248	16.0361	0.0312	0.0211	0.0847	39.6804
FF4	0.0186	0.0183	0.0256	16.1915	0.0314	0.0214	0.0884	38.2327
FF5	0.0200	0.0193	0.0283	15.7413	0.0338	0.0229	0.0887	40.9974
FF6	0.0208	0.0204	0.0290	15.9608	0.0337	0.0233	0.0911	39.8200
$q4$	0.0187	0.0177	0.0229	18.2128	0.0313	0.0207	0.0881	38.2209
$q5$	0.0205	0.0202	0.0238	19.2221	0.0330	0.0219	0.0913	38.8268
FER	0.0190	0.0186	0.0255	16.5994	0.0319	0.0213	0.0880	38.9307
Panel E: Rolling βs and Cumulative-Average Factor Premiums								
CAPM	0.0146	0.0149	0.0191	16.9743	0.0220	0.0139	0.0774	30.5507
FF3	0.0157	0.0158	0.0185	18.9282	0.0236	0.0146	0.0790	31.9944
FF4	0.0161	0.0167	0.0184	19.5056	0.0238	0.0148	0.0811	31.5042
FF5	0.0173	0.0174	0.0215	17.8919	0.0249	0.0155	0.0806	33.2090
FF6	0.0177	0.0179	0.0212	18.6154	0.0252	0.0155	0.0827	32.6604
$q4$	0.0166	0.0161	0.0164	22.4898	0.0238	0.0148	0.0829	30.8712
$q5$	0.0183	0.0178	0.0182	22.3833	0.0256	0.0155	0.0853	32.2164
FER	0.0166	0.0166	0.0191	19.4399	0.0243	0.0151	0.0817	31.9481
Panel F: Rolling βs and Realized Factor Premiums								
CAPM	0.200	0.203	0.040	111.751	0.218	0.181	0.298	78.960
FF3	0.319	0.320	0.044	160.526	0.364	0.297	0.407	96.214
FF4	0.351	0.351	0.046	171.269	0.406	0.331	0.435	100.525
FF5	0.394	0.394	0.047	185.754	0.457	0.373	0.484	101.810
FF6	0.430	0.429	0.049	194.553	0.503	0.412	0.501	108.272
$q4$	0.351	0.352	0.046	170.436	0.406	0.327	0.471	92.930
$q5$	0.390	0.388	0.047	184.851	0.454	0.369	0.483	101.397
FER	0.366	0.366	0.045	179.654	0.427	0.346	0.467	98.470

Table 8. Scaled Measurement-Error Variances of Implied-Cost-of-Capital ERPs

This table presents descriptive statistics for the cross-sectional and time-series scaled measurement-error variances (multiplied by 100) of specific variants of implied-cost-of-capital proxies. t-Statistics are estimated using Newey-West-corrected standard errors, and test the null hypothesis that the mean SVAR is zero. Each implied-cost-of-capital proxy is the internal rate of returns equating a firm's stock price to the firm's discounted future earnings, based on a particular earnings-forecast type and a specific formation of the discounted-dividend model. GLS and CAT use variants of the residual-income valuation model, following Gebhardt et al. (2001) and Claus and Thomas (2001) respectively. AGR and PEG use variants of the abnormal-earnings-growth valuation model, following Ohlson and Juettner-Nauroth (2005) and Easton (2004). ICC is a composite measure that takes the equal-weighted average GLS, CAT, AGR, and PEG. Panel A reports on implied-cost-of-capital estimates that use earnings forecasts based on the mechanical cross-sectional forecast model of Hou et al. (2012). Panel B reports on implied-cost-of-capital estimates that use analysts' median earnings forecasts from IBES. Panel C reports on the mechanical-forecast-based composite ICC and the analyst-forecast-based composite ICC using a common sample of observations with non-missing values in both. We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; each panel is based on a common sample of observations with non-missing values in each of the expected-return proxies. A full description of each proxy appears in Section III and Appendix II.

	Cross-Sectional SVAR				Time-Series SVAR			
	Mean	Median	STD	t-Stat	Mean	Median	STD	t-Stat
Panel A: Mechanical-Forecast Based (T=497, N=11,584)								
GLS	-0.0011	-0.0004	0.0123	-1.9829	-0.0063	-0.0035	0.0206	-33.0169
CAT	0.0018	0.0033	0.0179	2.2198	-0.0068	-0.0029	0.0418	-17.5955
AGR	0.0120	0.0122	0.0243	10.9655	0.0070	-0.0005	0.0982	7.6445
PEG	0.0034	0.0046	0.0241	3.1281	-0.0073	-0.0035	0.0356	-21.8767
ICC	0.0007	0.0017	0.0188	0.7779	-0.0080	-0.0038	0.0310	-27.8120
Panel B: Analyst-Forecast Based (T=492, N=8,265)								
GLS	-0.0009	-0.0004	0.0075	-2.7407	-0.0032	-0.0024	0.0120	-24.0364
CAT	-0.0008	-0.0004	0.0069	-2.6988	-0.0027	-0.0020	0.0161	-15.3734
AGR	-0.0001	0.0007	0.0107	-0.3025	-0.0033	-0.0023	0.0355	-8.5395
PEG	0.0004	0.0009	0.0140	0.6732	-0.0029	-0.0019	0.0277	-9.6196
ICC	-0.0006	0.0000	0.0099	-1.3138	-0.0033	-0.0022	0.0185	-16.2249
Panel C: Common Sample (T=492, N=6,360)								
ICC (Mechanical)	0.0007	0.0017	0.0188	0.7779	-0.0031	-0.0020	0.0136	-18.3031
ICC (Analyst)	-0.0006	0.0000	0.0099	-1.3138	-0.0027	-0.0019	0.0104	-20.4945

Table 9. Scaled Measurement-Error Variances of Characteristic-Based ERPs

This table presents descriptive statistics for the cross-sectional and time-series scaled measurement-error variances (multiplied by 100) of specific variants of characteristic-based expected-return proxies. t-statistics are estimated using Newey-West-corrected standard errors, and test the null hypothesis that the mean SVAR is zero. We implement an ERP from Lewellen (2015), denoted JLR, based on three firm characteristics: the log of market capitalization (size), the log of book-to-market (bm), and stock returns dating from twelve months to two months prior to the forecast date (momentum). The other ERP we implement, denoted LPV and from Lyle and Wang (2015) and Chattopadhyay et al. (2020), is also based on three firm characteristics: bm, the log of gross return on equity (roe), and the mean of prior-month daily squared returns (vol). Each of these variants is constructed as a linear combination of firm characteristics, where the weights are estimated by historical cross-sectional regressions of realized returns on those characteristics. CER is a composite that takes the equal-weighted average of JLR and LPV. We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and on publicly available information at the conclusion of each calendar month, where the expected return corresponds to the following month's return; each panel is based on a common sample of observations with non-missing values in each of the expected-return proxies. A full description of each proxy appears in Section III and Appendix II.

	Cross-Sectional (T=497)				Time-Series (N=11,611)			
	Mean	Median	STD	t-Stat	Mean	Median	STD	t-Stat
Panel A: Rolling Coefficients								
JLR	-0.0020	-0.0008	0.0139	-3.1617	0.0001	0.0003	0.0221	0.7293
LPV	-0.0008	0.0001	0.0105	-1.6904	-0.0013	-0.0005	0.0210	-6.7776
CER	-0.0018	-0.0009	0.0104	-3.8327	-0.0007	-0.0002	0.0189	-3.8570
Panel B: Cumulative Coefficients								
JLR	-0.0017	-0.0016	0.0155	-2.5083	0.0021	0.0005	0.0287	7.9610
LPV	-0.0006	-0.0003	0.0078	-1.5793	-0.0029	-0.0018	0.0122	-25.4581
CER	-0.0019	-0.0015	0.0078	-5.5197	-0.0010	-0.0010	0.0167	-6.3560

Table 10. Application: Re-Examining Prior Literature

This table reports the results of regressions of proxies of expected returns on multiple risk proxies studied in prior research. In each panel, the odd numbered columns replicate the original study's main cross-sectional-regression result using the ERP of choice—an ICC variant—from that study. The even numbered columns of each panel replace the ERP from the original study with the characteristic-based ERPs nominated by this study (CER), leaving all other aspects of the regression unchanged. Appendix II details the construction of CER. Panel A revisits the link between firms' distress risk and expected returns, and replicates the findings of Chava and Purnanandam (2010). We report coefficient estimates from pooled OLS regressions of ERPs on the main explanatory variables of interest, DRP_{hazard} (columns 1 and 2) and DRP_{EDF} (columns 3 and 4), are the annual percentile rankings of firms' default risk based on their hazard rate and the expected distance to default, respectively. Both measures range from 0 to 100, where higher values represent greater distress risk, and are obtained from the authors of Chava and Purnanandam (2010). $logta$ is the log of total assets, mtb is firms' market-to-book ratio, and $ret_{t-1,t}$ is the past month's stock return. Panel B revisits the link between idiosyncratic risk and expected returns, and replicates the main regression and replicates the findings of Fu (2009). We report coefficient estimates from pooled OLS regressions of ERPs on $IVOL_t$ (columns 1 and 2) and $IVOL_{t-1}$ (columns 3 and 4), where IVOL refers to the idiosyncratic volatility of the stock defined as the standard deviation of the daily residuals from a regression of returns on the Fama-French three factor model. $Ln(ME)$ is the log of firms' market capitalization; $Ln(BE/ME)$ is the log of book-to-market; $RET(-2, -7)$ is the compound gross return from 2 to 7 months prior to measurement of $IVOL_t$; $TURN$ is the average turnover and $CVTURN$ is the coefficient of variation of turnovers in the past 36 months. Panel C revisits the link between firms' information quality and expected returns, and replicates the findings of Francis et al. (2004) and Chen et al. (2015). We report coefficient estimates from pooled OLS regressions of ERPs on AQ (columns 1 and 2) and DQ (columns 3 and 4). AQ is a measure of accrual quality (Francis et al., 2004), computed as the cross-sectional decile rank in the time-series standard deviation of residuals from rolling 10-year firm-level regressions of total current accruals on cash flow from operations. AQ ranges from 1 to 10, where higher values indicate greater information risk or lower information quality. DQ is a measure of disclosure quality based on the level of disaggregation of financial data items in firms' annual reports (Chen et al., 2015). The measure ranges from 0 to 1, where higher values indicate lower information risk or better information quality, and is obtained from the authors of Chen et al. (2015). $Beta$ is estimated from rolling 60-month firm-level regressions, where excess returns are computed using the one-month treasury bill rate; $Size$ is defined as the log of a firm's market capitalization; and BM is the ratio of book value of equity to market value of equity. Standard errors in all regressions, reported in parentheses, are two-way cluster robust, clustering by firm and year. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1% respectively.

	Panel A: Distress Risk				
	GLS (Analyst) (1)	CER (2)		GLS (Analyst) (3)	CER (4)
DRP_{hazard}	0.0330*** (0.003)	0.0198*** (0.004)			
DRP_{EDF}				0.0345*** (0.003)	0.0217*** (0.005)
$logta$	-0.5691*** (0.068)	-0.5178*** (0.123)		-0.6069*** (0.065)	-0.5134*** (0.118)
mtb	-0.1907** (0.074)	-0.8775*** (0.101)		-0.1533 (0.104)	-0.8825*** (0.132)
$ret_{t-1,t}$	-4.2084*** (0.439)	-0.5028 (0.715)		-4.9964*** (0.488)	-1.0443 (0.743)
Time FE	Yes	Yes		Yes	Yes
Observations	38,269	38,269		34,741	34,741
R^2	0.1996	0.5446		0.2071	0.5421

Table 10. [Continued]

Panel B: Idiosyncratic Volatility

	GLS (Analyst) (1)	CER (2)	GLS (Analyst) (3)	CER (4)
$IVOL_t$	0.0314***	0.0101		
	(0.003)	(0.007)		
$IVOL_{t-1}$			0.0288***	0.0102
			(0.003)	(0.011)
$Ln(ME)$	-0.0107***	-0.0630***	-0.0118***	-0.0632***
	(0.003)	(0.007)	(0.003)	(0.006)
$Ln(BE/ME)$	0.1291***	0.2398***	0.1285***	0.2386***
	(0.006)	(0.020)	(0.006)	(0.020)
$Ret(-2,-7)$	-0.1552***	0.0459	-0.1603***	0.0441
	(0.012)	(0.039)	(0.013)	(0.039)
$Ln(TURN)$	0.0236***	-0.0005	0.0244***	-0.0005
	(0.003)	(0.005)	(0.003)	(0.004)
$Ln(CVTURN)$	-0.0436***	-0.0487***	-0.0419***	-0.0470***
	(0.006)	(0.010)	(0.006)	(0.009)
Time FE	Yes	Yes	Yes	Yes
Observations	688,384	688,384	664,216	664,216
R^2	0.3060	0.7913	0.2999	0.7912

Panel C: Information Quality

	PEG (Analyst) (1)	CER (2)	ICC (Mechanical) (3)	CER (4)
AQ	0.4940***	-0.0263		
	(0.065)	(0.021)		
DQ			-4.4211**	0.5481
			(1.666)	(0.630)
$Beta$	1.7673***	0.2324*	-0.5185**	0.1602
	(0.310)	(0.121)	(0.204)	(0.112)
$Size$	-0.4068***	-0.4113***	-2.6951***	-0.5059***
	(0.113)	(0.095)	(0.217)	(0.156)
BM	2.8374***	3.0632***	2.3121***	1.7983***
	(0.646)	(0.393)	(0.203)	(0.238)
Time FE	Yes	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Observations	2,377	2,377	33,588	33,588
R^2	0.2481	0.6603	0.3142	0.5426

Evaluating Firm-Level
Expected-Return Proxies

Online Appendix

Abstract

This Online Appendix contains supplementary results and arguments for the paper
"Evaluating Firm-Level Expected-Return Proxies." In this Online Appendix below, we
list and examine 84 studies culled from top finance and accounting journals between
1997 and 2016 that use implied cost of capital estimates (ICCs) as a primary depen-
dent variable of examination. We categorize them based on whether they are focused
on absolute magnitudes or relative magnitudes (treatment-effect studies) in expected
returns. We also establish the robustness of our results to alternative treatment of
extreme values.

JEL Classifications: G10, G11, G12, G14, M41

In the first part of this Online Appendix below, we categorize and list 98 studies culled from top finance and accounting journals between 1997 and 2016 that use implied cost of capital estimates (ICCs) as a primary dependent variable of examination. We identified articles published in the following five finance journals—*Journal of Corporate Finance*, *Journal of Financial and Quantitative analysis*, *Journal of Financial Economics*, *Journal of Finance*, and *Review of Financial Studies*—and five accounting journals—*Contemporary Accounting Research*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Review of Accounting Studies*, and *The Accounting Review*—by searching for keywords and citations.

Keyword searching involves searching full-text and abstracts in ABI-Proquest, Business Source Complete, and the historical archives of each journal for combinations of the terms "implied" and "ex ante" with variations on the term "cost of capital" (e.g., "cost of equity capital," "equity cost of capital," "risk premium"). Citation searching uses Google Scholar to find papers in the leading journals that cite the following methodological papers in the ICC literature: Botosan (1997), Claus and Thomas (2001), Easton (2004), Gebhardt et al. (2001), Gode and Mohanram (2003), Ohlson and Juettner-Nauroth (2005).

We first categorize these papers based on whether they are focused on absolute magnitudes or relative magnitudes (treatment-effect studies) in expected returns, or are focused on methodology or forecasting. Studies that are primarily interested in absolute magnitudes are interested in estimating the magnitude of the equity-risk premium (equity-risk-premium studies); studies that are primarily interested in relative magnitudes examine the impact of various factors—such as firm characteristics or policies—on firm-level or market-level expected returns. We then subcategorize the studies focused on relative magnitudes based on the primary empirical designs. We present below a full enumeration of these papers in their respective categories.

Total Papers: 98

A. Treatment-Effect Studies: 75 (77% of Total)
> *Papers*: *Ashbaugh-Skaife et al. (2009)*; *Attig et al. (2008)*; *Baginski and Rakow (2012)*; *Barth et al. (2008, 2013)*; *Ben-nasr et al. (2012)*; *Botosan (1997)*;

Botosan and Plumlee (2002, 2005); Botosan et al. (2011, 2004); Boubakri et al. (2014, 2012); Bratten et al. (2013); Callahan et al. (2012); Campbell et al. (2012); Cao et al. (2015); Chava and Purnanandam (2010); Chen et al. (2008, 2009, 2010, 2011b,a,c, 2013b, 2015); Core et al. (2006); Dai et al. (2013); Daske et al. (2008, 2013); Donangelo (2014); Dhaliwal et al. (2006, 2007, 2005, 2011, 2016); El Ghoul et al. (2011, 2013); Francis et al. (2004, 2008, 2005); Frank and Shen (2016); Beng Wee et al. (2016); Gordon and Wilford (2012); Hail and Leuz (2006, 2009); Hann et al. (2013); Hodder et al. (2006); Hou (2015); Hribar and Jenkins (2004); Hwang et al. (2013); Khurana and Raman (2004, 2006); Kim et al. (2012); Kim (2014); Krishnan et al. (2008); Lang et al. (2012); Lara et al. (2011); Lau et al. (2010); Lawrence et al. (2011); Lee et al. (2009); Li and Mohanram (2014); Siqi (2010); Liu and Natarajan (2012); McInnis (2010); Mohanram and Rajgopal (2009); Naiker et al. (2013); Ng and Rezaee (2015); Ogneva et al. (2007); Ortiz-Molina and Phillips (2014); Pástor et al. (2008); Rajan et al. (2007); Sengupta and Zhang (2015); Sikes and Verrecchia (2015); Zhou et al. (2016)

A1. Pure Time-Series Data: 1 (1% of A)

Papers: *Pástor et al. (2008)*

A2. Pure Cross-Sectional Data: 2 (3% of A)

Papers: *Baginski and Rakow (2012); Botosan (1997)*

A3. Panel Data: 72 (96% of A)

Papers: *Ashbaugh-Skaife et al. (2009); Attig et al. (2008); Barth et al. (2008, 2013); Ben-nasr et al. (2012); Botosan and Plumlee (2002, 2005); Botosan et al. (2011, 2004); Boubakri et al. (2014, 2012); Bratten et al. (2013); Callahan et al. (2012); Campbell et al. (2012); Cao et al. (2015); Chava and Purnanandam (2010); Chen et al. (2008, 2009, 2010, 2011b,a,c, 2013b, 2015); Core et al. (2006); Dai et al. (2013); Daske et al. (2008, 2013); Dhaliwal et al. (2006, 2007, 2005, 2011, 2016); Donangelo (2014); El Ghoul et al. (2011, 2013); Francis et al. (2004, 2008, 2005); Frank and Shen (2016); Beng Wee et al. (2016); Gordon and Wilford (2012); Hail and Leuz (2006, 2009); Hann et al. (2013); Hodder et al. (2006); Hou (2015); Hribar and Jenkins (2004); Hwang et al. (2013); Khurana and Raman (2004, 2006); Kim et al. (2012); Kim (2014); Krishnan et al. (2008); Lang et al. (2012); Lara et al. (2011); Lau et al. (2010); Lawrence et al. (2011); Lee et al. (2009); Li and Mohanram (2014); Siqi (2010); Liu and Natarajan (2012); McInnis (2010); Mohanram and Rajgopal (2009); Naiker et al. (2013); Ng and Rezaee (2015); Ogneva et al. (2007); Ortiz-Molina and Phillips (2014); Rajan et al. (2007); Sengupta and Zhang (2015); Sikes and Verrecchia (2015); Zhou et al. (2016)*

A3(a). Time- or firm-fixed effects, or changes specifications: 43 (60% of A3)

Papers: *Attig et al. (2008); Barth et al. (2008); Ben-nasr et al. (2012); Boubakri et al. (2014, 2012); Bratten et al. (2013); Campbell et al. (2012); Cao et al. (2015); Chen et al. (2009, 2010, 2011b,a,c, 2013b, 2015); Daske et al. (2008, 2013); Dhaliwal et al. (2007, 2005, 2011, 2016); Donangelo (2014); El Ghoul et al. (2013, 2011); Francis et al. (2005); Frank and Shen (2016); Beng Wee et al. (2016); Gordon and Wilford (2012); Hail and Leuz (2006, 2009); Hann et al. (2013); Hribar and Jenkins (2004); Hwang et al. (2013); Khurana and Raman (2004); Kim (2014); Lang et al. (2012); Lawrence et al. (2011); Ortiz-Molina and Phillips (2014); Naiker et al. (2013); Ng and Rezaee (2015); Ogneva et al. (2007); Sengupta and Zhang (2015); Zhou et al. (2016)*

A3(b). Fama-MacBeth Regressions: 25 (35% of A3)

Papers: *Ashbaugh-Skaife et al. (2009); Barth et al. (2013); Botosan and Plumlee (2002, 2005); Botosan et al. (2011, 2004); Chava and Purnanandam (2010); Chen et al. (2008, 2013b); Core et al. (2006); Dhaliwal et al. (2006); Francis et al. (2004, 2008); Hodder et al. (2006); Hou (2015); Khurana and Raman (2006); Krishnan et al. (2008); Lara et al. (2011); Lau et al. (2010); Lee et al. (2009); Li and Mohanram (2014); Liu and Natarajan (2012); McInnis (2010); Mohanram and Rajgopal (2009); Rajan et al. (2007)*

A3(c). Difference-in-differences specifications: 5 (7% of A3)

Papers: *Callahan et al. (2012); Dai et al. (2013); Kim et al. (2012); Naiker et al. (2013); Siqi (2010)*

B. Equity-Risk-Premium Studies: 5 (5% of Total)

Papers: *Ashton and Wang (2013); Claus and Thomas (2001); Fama and French (2002); Ritter and Warr (2002); Fitzgerald et al. (2013)*

C. Methodological or Forecasting Studies: 18 (18% of Total)

Papers: *Baginski and Wahlen (2003); Easton (2004); Easton and Sommers (2007); Easton et al. (2002); Campbell and Thompson (2008); Campello et al. (2008); Chen et al. (2013a); Gebhardt et al. (2001); Gode and Mohanram (2003); Hou et al. (2012); Huang and Ritter (2009); Jones and Tuzel (2013); Larocque (2013); Li et al. (2013); Lyle and Wang (2015); Mohanram and Gode (2013); Nekrasov and Ogneva (2011); Palazzo (2012)*

In our main tests in the paper, we winsorize observations of our pooled sample at the 5th and 95th percentile. The additional tables presented in this Online Appendix explore the impact of how we handle extreme values, which we implemented in two steps. The first is that we trimmed extreme values of ERPs rather than winsorizing to mitigate the concern that the extreme 10 percent of the distribution has 2 values. The second is that we identified and treated extreme ERPs each month, rather than once for the pooled sample, to mitigate concerns about potential look ahead biases. These tests show our main inferences are qualitatively unchanged using this alternative approach.

References

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

Ashton, D. and Wang, P. 2013. Terminal valuations, growth rates and the implied cost of capital. *Review of Accounting Studies*, 18(1):261–290.

Attig, N., Guedhami, O., and Mishra, D. 2008. Multiple large shareholders, control contests, and implied cost of equity. *Journal of Corporate Finance*, 14(5):721–737.

Baginski, S. and Rakow, K. 2012. Management earnings forecast disclosure policy and the cost of equity capital. *Review of Accounting Studies*, 17(2):279–321.

Baginski, S. P. and Wahlen, J. M. 2003. Residual income risk, intrinsic values, and share prices. *The Accounting Review*, 78(1):327.

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

Barth, M. E., Konchitchki, Y., and Landsman, W. R. 2013. Cost of capital and earnings transparency. *Journal of Accounting & Economics*, 55(2/3):206–224.

Ben-nasr, H., Boubakri, N., and Cosset, J.-C. 2012. The political determinants of the cost of equity: Evidence from newly privatized firms. *Journal of Accounting Research*, 50(3):605–646.

Beng Wee, G., Jimmy, L., Chee Yeow, L., and Shevlin, T. 2016. The effect of corporate tax avoidance on the cost of equity. *The Accounting Review*, 91(6):1647–1670.

Botosan, C. 1997. Disclosure level and the cost of equity capital. *The Accounting Review*, 72(3):323–349.

Botosan, C., Plumlee, M., and Wen, H. 2011. The relation between expected returns, realized returns, and firm risk characteristics. *Contemporary Accounting Research*, 28(4):1085–1122.

Botosan, C., Plumlee, M., and Yuan, X. 2004. The role of information precision in determining the cost of equity capital. *Review of Accounting Studies*, 9:233–259.

Botosan, C. A. and Plumlee, M. A. 2002. A re-examination of disclosure level and the expected cost of equity capital. *Journal of Accounting Research*, 40(1):21–40.

Botosan, C. A. and Plumlee, M. A. 2005. Assessing alternative proxies for the expected risk premium. *The Accounting Review*, 80(1):21–53.

Boubakri, N., El Ghoul, S., and Saffar, W. 2014. Political rights and equity pricing. *Journal of Corporate Finance*, 27:326–344.

Boubakri, N., Guedhami, O., Mishra, D., and Saffar, W. 2012. Political connections and the cost of equity capital. *Journal of Corporate Finance*, 18(3):541–559.

Bratten, B., Choudhary, P., and Schipper, K. 2013. Evidence that market participants assess recognized and disclosed items similarly when reliability is not an issue. *The Accounting Review*, 88(4):1179–1210.

Callahan, C. M., Smith, R. E., and Spencer, A. W. 2012. An examination of the cost of capital implications of fin 46. *The Accounting Review*, 87(4):1105–1134.

Campbell, J. L., Dhaliwal, D. S., and Schwartz, W. C. 2012. Financing constraints and the cost of capital: Evidence from the funding of corporate pension plans. *Review of Financial Studies*, 25(3):868–912.

Campbell, J. Y. and Thompson, S. B. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Campello, M., Chen, L., and Zhang, L. 2008. Expected returns, yield spreads, and asset pricing tests. *The Review of Financial Studies*, 21(3):1297–1338.

Cao, Y., Myers, J., Myers, L., and Omer, T. 2015. Company reputation and the cost of equity capital. *Review of Accounting Studies*, 20(1):42–81.

Chattopadhyay, A., Lyle, M. R., and Wang, C. C. 2019. Expected stock returns worldwide: A log-linear present-value approach. *Working Paper*.

Chava, S. and Purnanandam, A. 2010. Is default risk negatively related to stock returns? *Review of Financial Studies*, 23(6):2523–2559.

Chen, H., Chen, J. Z., Lobo, G. J., and Wang, Y. 2011a. Effects of audit quality on earnings management and cost of equity capital: Evidence from China. *Contemporary Accounting Research*, 28(3):892–925.

Chen, H. J., Kacperczyk, M., and Ortiz-Molina, H. 2011b. Labor unions, operating flexibility, and the cost of equity. *Journal of Financial and Quantitative Analysis*, 46(1):25–58.

Chen, K., Chen, Z., and Wei, K. 2009. Legal protection of investors, corporate governance, and the cost of equity capital. *Journal of Corporate Finance*, 15(3):273–289.

Chen, K. C., Chen, Z., and Wei, K. J. 2011c. Agency costs of free cash flow and the effect of shareholder rights on the implied cost of equity capital. *Journal of Financial and Quantitative Analysis*, 46(1):171–207.

Chen, L., Da, Z., and Zhao, X. 2013a. What drives stock price movements? *The Review of Financial Studies*, 26(4):841–876.

Chen, L., Petkova, R., and Zhang, L. 2008. The expected value premium. *Journal of Financial Economics*, 87(2):269–280.

Chen, S., Miao, B., and Shevlin, T. 2015. A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports. *Journal of Accounting Research*, 53(5):1017–1054.

Chen, Z., Dhaliwal, D. S., and Xie, H. 2010. Regulation fair disclosure and the cost of equity capital. *Review of Accounting Studies*, 15(1):106–144.

Chen, Z., Huang, Y., and Wei, K. J. 2013b. Executive pay disparity and the cost of equity capital. *Journal of Financial and Quantitative Analysis*, 48(3):849–885.

Claus, J. and Thomas, J. 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts from domestic and international stock markets. *Journal of Finance*, 56(5):1629–1666.

Core, J. E., Guay, W. R., and Rusticus, T. O. 2006. Does weak governance cause weak stock returns? an examination of firm operating performance and investors' expectations. *The Journal of Finance*, 61(2):655–687.

Dai, Z., Shackelford, D. A., Zhang, H. H., and Chen, C. 2013. Does financial constraint affect the relation between shareholder taxes and the cost of equity capital? *The Accounting Review*, 88(5):1603–1627.

Daske, H., Hail, L., Leuz, C., and Verdi, R. 2008. Mandatory IFRS reporting around the world: Early evidence on the economic consequences. *Journal of Accounting Research*, 46(5):1085–1142.

Daske, H., Hail, L., Leuz, C., and Verdi, R. 2013. Adopting a label: Heterogeneity in the economic consequences around IAS/IFRS adoptions. *Journal of Accounting Research*, 51(3):495–547.

Dhaliwal, D., Heitzman, S., and Zhen Li, O. 2006. Taxes, leverage, and the cost of equity capital. *Journal of Accounting Research*, 44(4):691–723.

Dhaliwal, D., Judd, J. S., Serfling, M., and Shaikh, S. 2016. Customer concentration risk and the cost of equity capital. *Journal of Accounting and Economics*, 61(1):23–48.

Dhaliwal, D., Krull, L., Li, O., and Moser, W. 2005. Dividend taxes and implied cost of equity capital. *Journal of Accounting Research*, 43(5):675–708.

Dhaliwal, D., Krull, L., and Li, O. Z. 2007. Did the 2003 Tax Act reduce the cost of equity capital? *Journal of Accounting Economics*, 43(1):121–150.

Dhaliwal, D., Li, O. Z., Tsang, A., and Yang, Y. G. 2011. Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review*, 86(1):59–100.

Donangelo, A. 2014. Labor mobility: Implications for asset pricing. *The Journal of Finance*, 69(3):1321–1346.

Easton, P. and Sommers, G. 2007. Effect of analysts' optimism on estimates of the expected rate of return implied by earnings forecasts. *Journal of Accounting Research*, 45(5):983–1015.

Easton, P., Taylor, G., Shroff, P., and Sougiannis, T. 2002. Using forecasts of earnings to simultaneously estimate growth and the rate of return on equity investment. *Journal of Accounting Research*, 40(3):657–676.

Easton, P. D. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The accounting review*, 79(1):73–95.

El Ghoul, S., Guedhami, O., Ni, Y., Pittman, J., and Saadi, S. 2013. Does information asymmetry matter to equity pricing? Evidence from firms' geographic location. *Contemporary Accounting Research*, 30(1):140–181.

El Ghoul, S., Guedhami, O., and Pittman, J. 2011. The role of IRS monitoring in equity pricing in public firms. *Contemporary Accounting Research*, 28(2):643–674.

Fama, E. F. and French, K. R. 2002. The equity premium. *Journal of Finance*, 57(2):637–659.

Fama, E. F. and French, K. R. 2015. A five-factor asset pricing model. *Journal of Financial Economics*, 116:1–22.

Fama, E. F. and French, K. R. 2018. Choosing factors. *Journal of Financial Economics*, 128(2):234–252.

Fitzgerald, T., Gray, S., Hall, J., and Jeyaraj, R. 2013. Unconstrained estimates of the equity risk premium. *Review of Accounting Studies*, 18(2):560–639.

Francis, J., Khurana, I., and Pereira, R. 2005. Disclosure incentives and effects on cost of capital around the world. *The Accounting Review*, 80:1125–1162.

Francis, J., LaFond, R., Olsson, P. M., and Schipper, K. 2004. Costs of equity and earnings attributes. *The Accounting Review*, 79(4):967–1010.

Francis, J., Nanda, D., and Olsson, P. 2008. Voluntary disclosure, earnings quality, and cost of capital. *Journal of Accounting Research*, 46.1:53–99.

Frank, M. Z. and Shen, T. 2016. Investment and the weighted average cost of capital. *Journal of Financial Economics*, 119(2):300–315.

Gebhardt, W., Lee, C., and Swaminathan, B. 2001. Toward an implied cost of capital. *Journal of Accounting Research*, 39(1):135–176.

Gode, D. and Mohanram, P. 2003. Inferring the cost of capital using the Ohlson-Juettner model. *Review of Accounting Studies*, 8(4):399–431.

Gordon, L. A. and Wilford, A. L. 2012. An analysis of multiple consecutive years of material weaknesses in internal control. *The Accounting Review*, 87(6):2027–2060.

Hail, L. and Leuz, C. 2006. International differences in the cost of equity capital: Do legal institutions and securities regulation matter? *Journal of Accounting Research*, 44(3):485–531.

Hail, L. and Leuz, C. 2009. Cost of capital effects and changes in growth expectations around U.S. cross-listings. *Journal of Financial Economics*, 93(3):428–454.

Hann, R. N., Ogneva, M., and Ozbas, O. 2013. Corporate diversification and the cost of capital. *The Journal of Finance*, 68(5):1961–1999.

Hodder, L. D., Hopkins, P. E., and Wahlen, J. M. 2006. Risk-relevance of fair-value income measures for commercial banks. *The Accounting Review*, 81(2):337–375.

Hou, K., Mo, H., Xue, C., and Zhang, L. 2019. q^5. *Working Paper*.

Hou, K., Van Dijk, M., and Zhang, Y. 2012. The implied cost of capital: A new approach. *Journal of Accounting and Economics*, 53:504–526.

Hou, K., Xue, C., and Zhang, L. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705.

Hou, Y. 2015. The role of diversification in the pricing of accruals quality. *Review of Accounting Studies*, 20(3):1059–1092.

Hribar, P. and Jenkins, N. 2004. The effect of accounting restatements on earnings revisions and the estimated cost of capital. *Review of Accounting Studies*, 9(2):337–356.

Huang, R. and Ritter, J. R. 2009. Testing theories of capital structure and estimating the speed of adjustment. *Journal of Financial and Quantitative analysis*, 44(2):237–271.

Hwang, L.-S., Lee, W.-J., Lim, S.-Y., and Park, K.-H. 2013. Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. *Journal of Accounting Economics*, 55(2/3):148–167.

Jones, C. S. and Tuzel, S. 2013. Inventory investment and the cost of capital. *Journal of Financial Economics*, 107(3):557–579.

Khurana, I. K. and Raman, K. K. 2004. Litigation risk and the financial reporting credibility of Big 4 versus non-Big 4 audits: Evidence from Anglo-American countries. *The Accounting Review*, 79(2):473–495.

Khurana, I. K. and Raman, K. K. 2006. Do investors care about the auditor's economic dependence on the client? *Contemporary Accounting Research*, 23(4):977–1016.

Kim, A. 2014. The value of firms' voluntary commitment to improve transparency: The case of special segments on euronext. *Journal of Corporate Finance*, 25:342–359.

Kim, Y., Li, H., and Li, S. 2012. Does eliminating the form 20-F reconciliation from IFRS to U.S. GAAP have capital market consequences? *Journal of Accounting and Economics*, 53(1):249–270.

Krishnan, G., Srinidhi, B., and Su, L. 2008. Inventory policy, accruals quality and information risk. *Review of Accounting Studies*, 13(2/3):369–410.

Lang, M., Lins, K. V., and Maffett, M. 2012. Transparency, liquidity, and valuation: International evidence on when transparency matters most. *Journal of Accounting Research*, 50(3):729–774.

Lara, J. G., Osma, B. G., and Penalva, F. 2011. Conditional conservatism and cost of capital. *Review of Accounting Studies*, 16(2):247–271.

Larocque, S. 2013. Analysts' earnings forecast errors and cost of equity capital estimates. *Review of Accounting Studies*, 18(1):135–166.

Lau, S. T., Ng, L., and Zhang, B. 2010. The world price of home bias. *Journal of Financial Economics*, 97(2):191–217.

Lawrence, A., Minutti-Meza, M., and Ping, Z. 2011. Can Big 4 versus non-Big 4 differences in audit-quality proxies be attributed to client characteristics? *The Accounting Review*, 86(1):259–286.

Lee, C., Ng, D., and Swaminathan, B. 2009. Testing international asset pricing models using implied costs of capital. *Journal of Financial and Quantitative Analysis*, 44(2):307–335.

Lewellen, J. 2015. The cross section of expected stock returns. *Critical Finance Review*, (4):1–44.

Li, K. and Mohanram, P. 2014. Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies*, 19(3):1152–1185.

Li, Y., Ng, D. T., and Swaminathan, B. 2013. Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics*, 110(2):419–436.

Liu, X. G. and Natarajan, R. 2012. The effect of financial analysts' strategic behavior on analysts' forecast dispersion. *The Accounting Review*, 87(6):2123–2149.

Lyle, M. R. and Wang, C. C. 2015. The cross section of expected holding period returns and their dynamics: A present value approach. *Journal of Financial Economics*, 116(3):505–525.

McInnis, J. 2010. Earnings smoothness, average returns, and implied cost of equity capital. *The Accounting Review*, 85(1):315–341.

Mohanram, P. and Gode, D. 2013. Removing predictable analyst forecast errors to improve implied cost of equity estimates. *Review of Accounting Studies*, 18(2):443–478.

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

Naiker, V., Navissi, F., and Truong, C. 2013. Options trading and the cost of equity capital. *The Accounting Review*, 88(1):261–295.

Nekrasov, A. and Ogneva, M. 2011. Using earnings forecasts to simultaneously estimate firm-specific cost of equity and long-term growth. *Review of Accounting Studies*, 16(3):414–457.

Ng, A. C. and Rezaee, Z. 2015. Business sustainability performance and cost of equity capital. *Journal of Corporate Finance*, 34:128–149.

Ogneva, M., Subramanyam, K. R., and Raghunandan, K. 2007. Internal control weakness and cost of equity: Evidence from SOX Section 404 disclosures. *The Accounting Review*, 82(5):1255–1297.

Ohlson, J. and Juettner-Nauroth, B. E. 2005. Expected EPS and EPS growth as determinants of value. *Review of Accounting Studies*, 10(2):349–365.

Ortiz-Molina, H. and Phillips, G. M. 2014. Real asset illiquidity and the cost of capital. *Journal of Financial and Quantitative Analysis*, 49(1):1–32.

Palazzo, B. 2012. Cash holdings, risk, and expected returns. *Journal of Financial Economics*, 104(1):162–185.

Pástor, L., Sinha, M., and Swaminathan, B. 2008. Estimating the intertemporal risk-return tradeoff using the implied cost of capital. *Journal of Finance*, 63(6):2859–2897.

Rajan, M., Reichelstein, S., and Soliman, M. 2007. Conservatism, growth, and return on investment. *Review of Accounting Studies*, 12(2/3):325–370.

Ritter, J. R. and Warr, R. S. 2002. The decline of inflation and the bull market of 1982–1999. *Journal of Financial and Quantitative Analysis*, 37(1):29–61.

Sengupta, P. and Zhang, S. 2015. Equity-based compensation of outside directors and corporate disclosure quality. *Contemporary Accounting Research*, 32(3):1073–1098.

Sikes, S. and Verrecchia, R. 2015. Dividend tax capitalization and liquidity. *Review of Accounting Studies*, 20(4):1334–1372.

Siqi, L. 2010. Does mandatory adoption of international financial reporting standards in the European Union reduce the cost of equity capital? *The Accounting Review*, 85(2):607–636.

Zhou, Q., Tan, K. J. K., Faff, R., and Zhu, Y. 2016. Deviation from target capital structure, cost of equity and speed of adjustment. *Journal of Corporate Finance*, 39:99–120.

Table O1. Monthly Expected-Return Proxies by Year

The upper panel of this table reports the annual mean and standard deviation of monthly expected-return proxies. The lower panel reports the pooled summary statistics on the annual mean and standard deviations. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2019). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return. In contrast to the results in the main paper, we truncate extreme values of ERPs for this table at the top and bottom 5 percent of the distribution each month, rather than once for the pooled sample, to mitigate concerns about potential look ahead biases. To construct the sample for these tests, we start with the sample from our main paper and then truncate each ERP individually (without requiring non-missing values of all 4 ERPs). A full description of each proxy appears in our main paper.

		Mean				Std Deviation			
Year	Obs	ICC	CER	FICC	FER	ICC	CER	FICC	FER
1977	13,381	1.66%	0.82%	1.88%	1.17%	0.78%	0.30%	0.71%	3.99%
1978	27,210	1.62%	0.77%	1.80%	1.80%	0.73%	0.30%	0.69%	7.91%
1979	27,766	1.65%	0.93%	1.75%	3.03%	0.69%	0.32%	0.69%	6.02%
1980	29,004	1.49%	1.35%	1.67%	2.60%	0.73%	0.31%	0.68%	7.61%
1981	29,389	0.58%	1.38%	1.52%	0.73%	0.65%	0.36%	0.63%	5.31%
1982	29,807	0.83%	1.18%	1.50%	2.36%	0.71%	0.40%	0.59%	6.00%
1983	31,013	1.28%	2.03%	1.22%	2.64%	0.52%	0.48%	0.55%	5.33%
1984	31,498	1.19%	1.86%	1.31%	-0.10%	0.69%	0.37%	0.53%	5.01%
1985	34,231	1.34%	1.90%	1.28%	2.24%	0.58%	0.31%	0.54%	4.87%
1986	34,397	1.23%	1.84%	1.19%	0.80%	0.59%	0.30%	0.54%	5.40%
1987	34,252	1.17%	1.73%	1.18%	-0.03%	0.56%	0.32%	0.52%	10.11%
1988	35,377	1.17%	1.49%	1.25%	2.09%	0.53%	0.33%	0.50%	4.53%
1989	36,951	1.13%	1.54%	1.18%	1.57%	0.58%	0.31%	0.48%	3.96%
1990	36,828	1.24%	1.14%	1.23%	-1.33%	0.73%	0.43%	0.49%	6.84%
1991	36,632	1.11%	1.17%	1.19%	3.12%	0.70%	0.45%	0.52%	5.70%
1992	36,809	0.96%	1.33%	1.16%	1.64%	0.57%	0.39%	0.49%	5.10%
1993	37,939	0.82%	0.99%	1.09%	1.70%	0.44%	0.36%	0.45%	3.81%
1994	43,199	0.86%	1.08%	1.09%	0.20%	0.45%	0.32%	0.40%	3.68%
1995	47,371	0.88%	1.10%	1.05%	2.41%	0.48%	0.30%	0.39%	3.73%
1996	49,031	0.93%	1.12%	0.98%	1.82%	0.48%	0.31%	0.37%	5.04%
1997	50,798	0.89%	1.15%	0.94%	2.30%	0.47%	0.34%	0.36%	5.85%
1998	51,520	0.86%	1.24%	0.93%	0.29%	0.45%	0.30%	0.36%	8.13%
1999	49,774	0.95%	1.21%	0.98%	2.06%	0.49%	0.33%	0.38%	7.03%
2000	47,991	0.94%	1.60%	0.93%	0.18%	0.50%	0.43%	0.40%	10.53%
2001	45,585	0.92%	1.47%	0.97%	1.44%	0.51%	0.41%	0.37%	9.74%
2002	44,048	0.94%	1.32%	0.95%	-0.83%	0.54%	0.37%	0.34%	8.39%
2003	42,482	0.84%	1.25%	0.91%	3.94%	0.50%	0.36%	0.32%	5.54%
2004	41,985	0.76%	1.44%	0.79%	1.77%	0.45%	0.33%	0.30%	4.57%
2005	41,139	0.76%	1.31%	0.80%	0.93%	0.44%	0.28%	0.30%	4.38%
2006	40,859	0.85%	1.26%	0.78%	1.87%	0.47%	0.26%	0.30%	4.12%
2007	40,154	0.85%	1.16%	0.79%	0.14%	0.47%	0.27%	0.30%	4.06%
2008	39,466	0.98%	0.93%	0.93%	-3.08%	0.56%	0.26%	0.34%	9.06%
2009	36,904	1.08%	0.74%	1.04%	3.60%	0.69%	0.30%	0.36%	9.94%
2010	36,654	0.96%	0.69%	0.88%	2.44%	0.57%	0.31%	0.34%	7.12%
2011	35,866	0.97%	0.72%	0.87%	0.00%	0.60%	0.25%	0.35%	6.55%
2012	35,486	1.02%	0.78%	0.93%	1.75%	0.63%	0.22%	0.36%	4.50%
2013	34,782	0.96%	0.82%	0.87%	2.97%	0.57%	0.17%	0.37%	3.74%
2014	34,673	0.92%	0.69%	0.83%	0.80%	0.55%	0.11%	0.37%	4.47%
2015	34,444	0.98%	0.58%	0.89%	-0.24%	0.62%	0.10%	0.41%	5.16%
2016	33,455	1.02%	0.45%	0.94%	2.01%	0.63%	0.10%	0.43%	5.88%
2017	32,237	0.90%	0.50%	0.86%	1.56%	0.55%	0.10%	0.45%	3.39%
2018	26,762	0.83%	0.75%	0.84%	0.55%	0.50%	0.15%	0.46%	5.05%
Mean	37,691	1.04%	1.22%	1.12%	1.40%	0.57%	0.32%	0.45%	5.99%
Median	36,819	0.96%	1.19%	1.05%	1.73%	0.56%	0.32%	0.39%	5.36%
Min	13,381	0.58%	0.69%	0.78%	-3.08%	0.44%	0.11%	0.30%	3.68%
Max	51,520	1.66%	2.03%	1.88%	3.94%	0.78%	0.48%	0.71%	10.53%
Std	7,715	0.25%	0.36%	0.29%	1.42%	0.10%	0.07%	0.12%	2.01%
Mean/Std	4.89	4.12	3.42	3.83	0.98	5.75	4.47	3.64	2.98

Table O2. Correlation between Expected-Return Proxies

This table reports the average monthly Pearson (Spearman) correlations above (below) the diagonal among the four expected-return proxies. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2019). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; the table is based on a common sample of observations with non-missing values in each of the four expected-return proxies. In contrast to the results in the main paper, we truncate extreme values of ERPs for this table at the top and bottom 5 percent of the distribution each month, rather than once for the pooled sample, to mitigate concerns about potential look ahead biases. To construct the sample for these tests, we start with the sample from our main paper and then truncate each ERP individually (and not require non-missing values of all 4 ERPs). A full description of each proxy appears in our main paper. We report t-statistics based on Newey-West-corrected standard errors in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1% respectively.

	ICC	CER	FICC	FER
ICC		0.373***	0.547***	-0.048***
		(20.50)	(32.79)	-(7.66)
CER	0.413***		0.585***	-0.041***
	(24.74)		(15.16)	-(6.40)
FICC	0.555***	0.599***		-0.031***
	(27.28)	(15.83)		-(3.28)
FER	-0.053***	-0.042***	-0.033***	
	-(7.91)	-(6.47)	-(3.54)	

Table O3. Cross-Sectional Scaled Measurement-Error Variances

This table presents descriptive statistics for the cross-sectional scaled measurement-error variances (multiplied by 100) of four expected-return proxies, computed as follows:

$$SVar^{CS} = Var_t(\widehat{er}_{i,t}) - 2Cov_t(r_{i,t+1}, \widehat{er}_{i,t}),$$

where $Var_t(\widehat{er}_{i,t})$ is a given ERP's cross-sectional variance at time t and $Cov_t(r_{i,t+1}, \widehat{er}_{i,t})$ is the cross-sectional covariance between firms' ERPs at time t and their realized returns at $t+1$. Panel A reports distributional statistics of $SVar^{CS}$ for each expected-return proxy, using data from a sample of 497 calendar months during our 1977-2018 sample period. The rightmost column of this panel reports t-Statistics testing the null hypothesis that the mean $SVar^{CS}$ for a given expected-return proxy is 0. Panel B reports pair-wise comparisons of average $SVar^{CS}$ within the sample. Table values are negative (positive) when the expected-return proxy displayed in the leftmost column has a larger (smaller) scaled measurement-error variance than the expected-return proxy displayed in the topmost row. t-statistics testing the differences in means are reported in parentheses. t-statistics are estimated using Newey-West-corrected standard errors. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2019). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; the table is based on a common sample of observations with non-missing values in each of the four expected-return proxies. In contrast to the results in the main paper, we truncate extreme values of ERPs for this table at the top and bottom 5 percent of the distribution each month, rather than once for the pooled sample, to mitigate concerns about potential look ahead biases. To construct the sample for these tests, we start with the sample from our main paper and then truncate each ERP individually (and not requiring non-missing values of all 4 ERPs). A full description of each proxy appears in our main paper.

Panel A: Cross-Sectional Measurement-Error Variance (N=497)

	Mean	P25	Median	P75	STD	t-stat
ICC	-0.0003	-0.0068	0.0008	0.0085	0.0164	-0.3830
CER	-0.0016	-0.0043	-0.0007	0.0023	0.0104	-3.3666
FICC	-0.0006	-0.0072	-0.0003	0.0071	0.0145	-0.9659
FER	0.1440	0.0264	0.0710	0.1493	0.4126	7.7823

Panel B: t-Statistics of Differences in Variances

	ICC	CER	FICC	FER
ICC		-0.001	0.000	0.144
		-(2.22)	-(0.94)	(6.26)
CER	0.001		0.001	0.146
	(2.22)		(1.54)	(6.36)
FICC	0.000	-0.001		0.145
	(0.94)	-(1.54)		(6.30)
FER	-0.144	-0.146	-0.145	
	-(6.26)	-(6.36)	-(6.30)	

Table O4. Time-Series Scaled Measurement-Error Variances

This table presents descriptive statistics for the time-series scaled measurement-error variances (multiplied by 100) of four expected-return proxies, computed as follows:

$$SVar^{TS} = Var_i(\widehat{er}_{i,t}) - 2Cov_i(r_{i,t+1}, \widehat{er}_{i,t}),$$

where $Var_i(\widehat{er}_{i,t})$ is a given ERP's time-series variance for firm i and $Cov_i(r_{i,t+1}, \widehat{er}_{i,t})$ is the time-series covariance between firm i's ERP at t and its realized returns in at $t+1$. Panel A reports distributional statistics of $SVar^{TS}$ for each expected-return proxy, using data from a sample of 11,620 unique firms with at least 20 observations in the 1977-2018 sample period. The rightmost column reports t-statistics testing the null hypothesis that the mean $SVar^{TS}$ for a given expected-return proxy is 0. Panel B reports pairwise comparisons of average $SVar^{TS}$ within the sample. Table values are negative (positive) when the expected-return proxy displayed in the leftmost column has a larger (smaller) scaled measurement-error variance than the expected-return proxy displayed in the topmost row. t-statistics testing the differences in means are reported in parentheses. t-statistics are estimated using Newey-West-corrected standard errors. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2019). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; the table is based on a common sample of observations with non-missing values in each of the four expected-return proxies. In contrast to the results in the main paper, we truncate extreme values of ERPs for this table at the top and bottom 5 percent of the distribution each month, rather than once for the pooled sample, to mitigate concerns about potential look ahead biases. To construct the sample for these tests, we start with the sample from our main paper and then truncate each ERP individually (and not requiring non-missing values of all 4 ERPs). A full description of each proxy appears in our main paper.

Panel A: Time-Series Measurement-Error Variance (N=11,611)						
	Mean	P25	Median	P75	STD	t-stat
ICC	-0.009	-0.014	-0.004	0.002	0.029	-31.386
CER	0.000	-0.006	0.000	0.005	0.018	-1.279
FICC	-0.007	-0.011	-0.005	-0.001	0.015	-53.586
FER	0.366	0.117	0.279	0.507	0.551	71.652

Panel B: t-Statistics of Differences in Variances				
	ICC	CER	FICC	FER
ICC	0.000	0.008	0.001	0.375
		(26.67)	(5.42)	(64.76)
CER	-0.008		-0.007	0.366
	-(26.67)		-(32.42)	(64.21)
FICC	-0.001	0.007		0.374
	-(5.42)	(32.42)		(64.69)
FER	-0.375	-0.366	-0.374	
	-(64.76)	-(64.21)	-(64.69)	

Table O5. Sub-Sample Variation in Scaled Measurement-Error Variances

This table presents descriptive statistics for the cross-sectional and time-series scaled measurement-error variances (multiplied by 100) of four expected-return proxies, for six sub samples. Panels A, B, C, and D focus on the sub-samples of large, all-but-tiny, small, and tiny firms. A firm is large if its market capitalization at the beginning of the calendar year is larger than the NYSE median; otherwise it is considered small. A firm is tiny if its market capitalization at the beginning of the calendar year is below the 20th percentile of NYSE; otherwise it is part of the all-but-tiny sub-sample. Panels E and F focus on, respectively, the first half (1977-1997) and second half (1998-2018) of our sample. t-statistics are estimated using Newey-West-corrected standard errors, and test the null hypothesis that the mean SVAR is zero. ICC is a composite measure of the implied cost of capital—based on mechanical forecasts of earnings following Hou et al. (2012)—that takes the equal-weighted average of four commonly used proxies: (1) GLS (Gebhardt et al., 2001), (2) CAT (Claus and Thomas, 2001), (3) PEG (Easton, 2004), and (4) AGR (Ohlson and Juettner-Nauroth, 2005). CER is a composite measure of two characteristic-based expected-return proxies—derived from historical cross-sectional regressions of realized returns on firm characteristics—that takes the equal-weighted average of two proxies: JLR (Lewellen, 2015) and LPV (Lyle and Wang, 2015; Chattopadhyay et al., 2019). FICC is a fitted ICC measure, derived from historical cross-sectional regressions of ICC on size, book-to-market, and return momentum. FER is a composite measure of factor-based expected returns—computed as the inner product of estimated factor sensitivities and expected factor premiums—that takes the equal-weighted average of two variants: the Hou et al. (2015) q-factor model augmented with an expected growth factor (Hou et al., 2019), and the Fama and French (2015) five-factor model augmented with a momentum factor (Fama and French, 2018). We compute a firm-specific expected-return estimate for each stock in our sample based on the stock price and publicly available information at the conclusion of each calendar month, so that the proxy is an expectation of the following month's return; the table is based on a common sample of observations with non-missing values in each of the four expected-return proxies. In contrast to the results in the main paper, we truncate extreme values of ERPs for this table at the top and bottom 5 percent of the distribution each month, rather than once for the pooled sample, to mitigate concerns about potential look ahead biases. To construct the sample for these tests, we start with the sample from our main paper and then truncate each ERP individually (and not requiring non-missing values of all 4 ERPs). A full description of each proxy appears in our main paper.

| | Cross-Sectional SVARS | | | | | Time-Series SVARS | | | |
	Mean	Median	STD	t-Stat		Mean	Median	STD	t-Stat
Panel A: Large Firms (T=497, N=6,772)									
ICC	0.001	0.002	0.011	1.947		-0.003	-0.002	0.017	-13.845
CER	-0.001	0.000	0.008	-2.286		0.001	0.000	0.013	5.288
FICC	0.001	0.001	0.009	1.408		-0.004	-0.003	0.011	-34.264
FER	0.132	0.064	0.409	7.181		0.405	0.298	0.517	64.496
Panel B: All-but-Tiny Firms (T=497, N=10,019)									
ICC	0.001	0.002	0.013	2.099		-0.005	-0.003	0.023	-22.383
CER	-0.001	0.000	0.009	-2.820		0.001	0.000	0.015	3.844
FICC	0.001	0.002	0.012	1.490		-0.006	-0.004	0.013	-46.084
FER	0.138	0.066	0.410	7.488		0.373	0.282	0.500	74.666
Panel C: Small Firms (T=497, N=7,008)									
ICC	-0.001	0.000	0.018	-1.084		-0.011	-0.006	0.038	-25.600
CER	-0.002	-0.001	0.009	-4.120		0.000	0.000	0.022	-1.515
FICC	-0.002	-0.001	0.009	-3.732		-0.008	-0.005	0.018	-35.623
FER	0.160	0.083	0.427	8.373		0.350	0.259	0.640	45.769
Panel D: Tiny Firms (T=497, N=3,182)									
ICC	-0.002	-0.001	0.018	-2.208		-0.015	-0.008	0.044	-18.664
CER	-0.001	-0.001	0.012	-2.502		-0.001	-0.001	0.027	-2.927
FICC	-0.002	-0.001	0.009	-5.635		-0.008	-0.005	0.020	-22.693
FER	0.174	0.100	0.452	8.608		0.321	0.231	0.680	26.726
Panel E: 1977-1997 (T=246, N=7,248)									
ICC	0.000	0.000	0.015	-0.347		-0.008	-0.003	0.028	-22.840
CER	-0.002	-0.002	0.006	-4.233		0.000	0.000	0.017	2.013
FICC	-0.001	-0.001	0.016	-0.534		-0.007	-0.004	0.014	-42.785
FER	0.107	0.067	0.161	10.398		0.274	0.222	0.349	66.800
Panel F: 1998-2018 (T=251, N=7,381)									
ICC	0.000	0.002	0.018	-0.209		-0.009	-0.004	0.029	-27.085
CER	-0.002	0.000	0.013	-1.809		-0.001	0.000	0.018	-5.955
FICC	-0.001	0.000	0.013	-0.878		-0.007	-0.004	0.015	-39.843
FER	0.180	0.079	0.556	5.134		0.446	0.313	0.724	52.891