Evaluating Firm-Level Expected-Return Proxies

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

Working Paper 15-022
Evaluating Firm-Level Expected-Return Proxies

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*

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

June 3, 2017

Abstract

We argue, from an extensive literature review, that in the vast majority of research settings, biases in alternative expected-return proxies (ERPs) are irrelevant. Therefore, in most settings, the choice between alternative ERPs should be based on an evaluation of their relative measurement-error variances. We develop a parsimonious evaluation framework that empirically estimates a given ERP’s cross-sectional and time-series measurement-error variances. We then apply this framework to five classes of firm-level ERPs nominated by recent studies, including factor-based ERPs from finance and implied costs of capital (ICC) estimates from accounting. Our analyses show ICCs are particularly useful in tracking time-series variations in expected returns. We also find broad support for a “fitted” or “characteristic-based” approach to ERP estimation.

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

*We thank David Burgstahler (editor), Bhaskaran Swaminathan of LSV Asset Management, and anonymous referees as well as seminar participants at Stanford University, Northwestern University, University of Chicago, University of California-Irvine, Louisiana State University, National University of Singapore, University of Pennsylvania (Wharton), University of Toronto, University of Michigan (Ross), Harvard University, Yale University, Columbia University, Indiana University, University of Washington (Seattle), University of Waterloo, and UCLA (Anderson), for helpful comments and suggestions on an earlier draft. We thank Kyle Thomas for excellent research assistance. An earlier version of this paper was titled “Evaluating Implied Cost of Capital Estimates.” Our Online Appendix is available via the following URL: Link to Online Appendix
I Introduction

Expected rates of return play a central role in many managerial and investment decisions that affect the allocation of scarce resources. Correspondingly, a substantial literature has arisen on the estimation of expected rates of return for individual equities and the economic drivers of expected-return variation. But while the importance of firm-level estimates is widely understood, there is little consensus on how such estimates should be made. As a result, the estimation techniques chosen by researchers vary widely across disciplines and studies, often without justification or discussion of alternative approaches.

Discordance 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. Absent an objective evaluative framework, it is extremely difficult for researchers to compare the relative merits of different firm-level ERPs and to adjudicate between them.

In this study, we develop an objective evaluative framework for ERPs. Our central contribution is a two-dimensional framework that can be used to empirically assess the relative quality of firm-level ERPs. Using a firm’s true, but unobservable, expected return as the normative benchmark, we define a given ERP’s deviation from this benchmark as its measurement error. Although measurement errors are themselves unobservable, we show it is possible to derive characteristics of the distribution of errors for each ERP, such that researchers can compare the relative performance of alternative proxies.

1 Typical empirical innovations include new forecasting techniques for earnings or the inclusion of additional asset-pricing factors. For example, Gebhardt et al. (2001) use a residual-income model and analysts’ earnings forecasts to estimate firms’ implied cost of equity capital. Subsequent researchers have modified this model by introducing the use of alternative growth forecasts (e.g., Easton and Monahan, 2005) and corrections for bias in analysts’ forecasts (e.g., Easton and Sommers, 2007), and/or by replacing analysts’ forecasts with mechanical earnings forecasts (e.g., Hou et al., 2012).
Two distributional properties of measurement errors are fundamental to ERP performance. The mean measurement error (also known as an ERP’s bias) is the average error over all observations (e.g., in a given cross section, or over time for a given firm). Conversely, the measurement error variance (MEV) is the spread of errors across observations. Our approach focuses on the latter, and is motivated by two key insights.

First, in the vast majority of research settings involving ERPs, the magnitude of bias in the estimate is either unimportant or completely irrelevant. Instead, researchers’ empirical tests typically rely on the relative magnitudes of ERPs—e.g., how differences in policies or firm characteristics are associated with differences in ERPs. For these applications, the ability of an ERP to “track” the variation in expected returns—thus minimizing measurement-error variance—is paramount for the precise estimation of effects of interest.\(^2\)

Our second key insight is that, although true expected returns are unobservable, it is a straightforward exercise to compute an empirical estimate of an ERP’s measurement-error variance. This estimation procedure is an important methodological contribution because, when combined with the first insight, it provides researchers with a simple way to objectively compare and evaluate the performance of a wide spectrum of firm-level ERPs.

Motivated by these two insights, we develop a two-dimensional framework that evaluates ERPs on the basis of their relative time-series and cross-sectional measurement-error variances (MEVs). This approach formalizes the intuition that well-performing ERPs should closely track expected returns both in the cross section (that is, cross-sectional variation in ERPs should reflect cross-sectional variation in firms’ expected returns) and over time (that is, time-series variation in a firm’s ERP should reflect variation in its expected returns over time). We show, both analytically and empirically, that these two dimensions of ERP performance are not redundant. Therefore, the optimal measure produced by this framework will depend, in part, on the relative importance that a researcher assigns to each dimension.

For example, researchers often study the variation in expected returns through a research

\(^2\)In our Online Appendix, we use simulation analysis to show that while measurement error variance affects researchers’ inferences by reducing the precision of her empirical tests, bias is irrelevant.
design that utilizes firm-fixed effects or difference-in-difference estimators. These studies identify the effect of interest through the time-series variation in the dependent variable, and call for a ERPs that exhibit the lowest time-series MEVs, or, whose time-series variation most closely tracks the underlying construct. Similarly, studies that identify the effects of interest through cross-sectional variation (e.g., using cross-sectional fixed effects or Fama-MacBeth regressions) calls for a proxy that exhibits the lowest cross-sectional MEV.

In each instance, the bias of the ERP (either in time series or in the cross section) is irrelevant, and is in fact generally neutralized by the research design (e.g., by the firm or cross-sectional fixed effect). This is because in most ERP-based research, the primary focus is on examining factors that influence variations in expected-returns over time or across firms. In these research settings, the absolute levels of the expected returns are typically of little or no relevance.

Table 1 illustrates this point. In this table, we summarize 81 studies culled from top Accounting and Finance journals since 1997 that feature different implementations of ICCs as ERPs. In an overwhelming majority of these studies (i.e., 83% = 67/81 of all papers), the empirical research design relies on the relative magnitudes of ERPs. For these studies, the precision of the treatment effect of interest calls for an ERP whose relative magnitudes (in the cross section or over time) closely matches those of true expected returns, i.e., an ERP exhibiting low MEV. For these studies, an ERP’s bias is in fact irrelevant.

Note that a MEV-based framework is not appropriate for all research settings. In particular, in studies where the absolute magnitudes of ERPs are of primary interest, biases in the ERP are inherently important and our framework would not be appropriate. However, in Table 1 we can identify only four such studies involving ICCs over the last 20 years. We discuss these papers, as well as the usefulness and limitations of the MEV criterion in other settings, in Section III of the paper.

Prior studies on ERP performance focus almost exclusively on cross-sectional tests, with
mixed results.\textsuperscript{3} We advance this literature by: (a) providing a parsimonious methodology for estimating ERPs’ MEVs, (b) introducing a time-series dimension to ERP performance evaluation, and (c) presenting new evidence on the relative performance of five major families of ERPs. Our empirical results highlight the particularly strong time-series performance of the implied-cost-of-capital (ICC) estimates featured in accounting studies in recent years.

We also compare and contrast our framework with other commonly-used alternative evaluation criteria. In particular, we make the case that selection of ERPs on the basis of minimum MEV is theoretically more appropriate than selection on the basis the minimum mean squared errors (MSE) criterion. This is because MSE punishes an ERP for both its bias and MEV, while for most research purposes only the latter matters.\textsuperscript{4}

A MEV approach is also more appropriate than the “slope coefficient” test that is common in the ERP literature in accounting and finance, wherein cross-sectional realized returns are regressed on alternative ERPs. In Section III, we show that while a perfect ERP would have an estimated slope coefficient of 1, it is decidedly not the case that an estimated slope of 1 implies an ideal ERP. In fact, it is easy to generate multiple noisy ERPs that all exhibit a slope of 1. This is because the estimated slope coefficient does not speak to the magnitude of the MEV of each ERP, and thus the slope coefficient and the “distance from 1” criterion do not generally help to differentiate the quality of ERPs in terms of MEVs.

To illustrate how researchers can implement our framework, we assess the relative performance of five families of ERPs (see Appendices I and II for a description of each family). These five ERP families are based either on traditional equilibrium asset-pricing theory or on a variation of ICCs. Collectively, they encompass all of the prototype classes of ERPs nominated by the academic literature in both finance and accounting over the past 50 years.

Three of the ERPs we test originate in traditional equilibrium asset-pricing theory (the

\textsuperscript{3}Most prior studies use regression-based tests check whether the slope coefficient from a cross-sectional regression of ex-post returns on an ex-ante expected-returns proxy yields a coefficient of one. See Section II for a review of these results and Section III for a discussion of their theoretical underpinnings.

\textsuperscript{4}Recall the MSE is the average of the squared measurement errors, and thus is the sum of the estimator’s error variance and its squared bias.
left-hand branch of the ERP tree depicted in Appendix I), in which non-diversifiable risk is priced, and where a firm’s ERP is a linear function of its sensitivity to each risk factor (the $\beta$’s) and the risk premium associated with the factor (the $\gamma$’s). We test a single-factor version based on the Capital Asset Pricing Model (CAPM) and a similar multi-factor version based on four empirically-inspired factors (FFF). We also test a characteristic-based expected-return estimate (CER), discussed in Lewellen (2015). Each firm’s CER estimate is a fitted value from the regression—that is, it is a linear function of the firm’s current characteristics. In calculating CER, the relative rankings in each firm characteristic is analogous to a firm’s factor exposure (e.g., the $\beta$’s) in calculating factor-based proxies.

In addition, we test two prototype ERPs from the ICC literature (the right-hand branch of the ERP tree in Appendix I). The implied cost of capital is the internal rate of return that equates a firm’s market value to the present value of its expected future cash-flows. We explore a composite ICC measure that averages across five implementations commonly used in prior research. Finally, we develop a new ERP prototype by computing a “fitted” version of the ICC measure that we refer to as FICC. This proxy is new to the literature, but it seems to us to reflect a natural progression in the evolution of ERPs. Like CER, FICC is based on an instrumental variable approach, whereby each firm’s ICC estimate is a “fitted” value from a regression that is linear in the firm’s current characteristics.

Our results show that characteristic-based ERPs (FICC, ICC, and CER) outperform traditional factor-based proxies (CAPM, FFF) in both cross-sectional and in time-series tests. Among the three characteristic-based proxies, CER, the proxy nominated by Lewellen (2015), performs best in cross-sectional tests; and FICC, the fitted implied-cost-of-capital proxy, performs best in the time-series tests. In fact, among all the ERPs we consider, only FICC and CER consistently perform better than a trivial ERP—i.e., a fixed constant for all firms in the cross section or for a given firm across time.

These five implied cost of capital estimates are based on the following studies: (1) Gebhardt et al. (2001), (2) Claus and Thomas (2001), (3) Ohlson and Juettner-Nauroth (2005), (4) Gordon and Gordon (1997), and (5) Easton and Monahan (2005).
CER and FICC are alike in that they are “fitted” ERPs. In each case, the ERP is computed by expressing either ICC or returns as a linear function of current-year firm characteristics, which can be thought of as mechanisms for filtering the noise in realized returns and ICCs, respectively. To the extent these firm attributes capture some systematic aspect of true expected returns, these fitted ERPs should perform better than either the factor-based ERPs or the ICC. Indeed, our findings lend support to the broader use of characteristic-based fitting methodologies as a means to reduce noise.

To further understand how methods for handling noise affect the performance of ERPs, we also consider the effect of winsorization. As discussed in Section III, by pulling in the tails of the distribution and thus reducing variance, the level of winsorization naturally affects the variance of ERP measurement errors. However, because winsorization also alters the relation between ERP and returns, the effect of winsorization on MEV may not be monotonic. We examine the impact of different winsorization levels on each of our candidate ERPs.

Our results show that a moderate amount of winsorization generally improves ERP performance. For example, the MEV drops for most ERPs with just 1% winsorization (i.e., when the ERP is winsorized at the 1st and 99th percentiles of the pooled distribution). The improvement is greater for the noisier ERPs, such as ICC, CAPM, and FFF. In the case of ICC, with 7% or more winsorization, its time-series performance beats a trivial estimator. As expected, the improvement is less pronounced for “fitted” ERPs. In fact, in the case of FICC, its mean time-series SVAR actually deteriorates with increased winsorization.

We also estimate the economic magnitude of the performance gain (in terms of percentage MEV reduction) attained by one ERP over another.\textsuperscript{6} For example, our analysis shows that CER (the best performing ERP in the cross section), attained a 68%, 15%, 82%, and 92% improvement (in terms of percent MEV reduction) over ICC, FICC, CAPM, and FFF, respectively. Similarly, we find FICC (the best performing ERP in time-series tests), attained

\textsuperscript{6}A key input needed for this estimation is the monthly variance of true (but unobservable) expected returns. In Section IV, we explain how we arrive at such an estimate, for both time series and cross sectional analyses. The results we report here are based on conservative assumptions as to the variance of true expected returns.
a 64%, 48%, 94% and 97% improvement over ICC, CER, CAPM and FFF, respectively. When compared to the trivial estimator (i.e., a constant), CER represents an improvement of 12% in the cross section, and FICC represents an improvement of 51% in time series. Given the importance of ERP tracking error in most research settings, we view these improvements as being quite economically significant.

In summary, our empirical results support the broader adoption of a characteristic-based “fitted” approach to ERP estimation. These approaches are designed for, and appear successful in, removing noise. Another key lesson that emerges from these results is the superior performance of ICCs in time-series tests. This is an important but previously ignored dimension of ERP performance.

Consistent with Easton and Monahan (2005), we find that ICC-based ERPs generally do not perform well in the cross section. Yet at the same time, our results show that ICC-based ERPs actually perform quite well in time-series tests. For example, FICC—which handles the noise in ICC by projecting the measure onto a set of firm characteristics—can yield an over 50% reduction in time series MEV relative to a trivial estimator. Given the importance of time-series MEV as a source of noise in research designs that rely on within-firm variations, our findings suggest ICC-based ERPs can represent a significant improvement over the trivial estimator. Overall, these results offer a much more sanguine assessment of the ICC approach than prior literature (e.g., Easton and Monahan, 2005).

The rest of the paper is organized as follows. Section II discusses related literature. Section III presents the theoretical underpinnings of our performance metrics. Section IV provides details on our sample construction and empirical results. Section V discusses the Online Appendix and Section VI concludes.

\footnote{In the Online Appendix, we show that four of the five individual ICCs underlying our composite measure underperform a trivial estimator (i.e., a constant) in the cross section. However, we also find that three of the five individual ICCs underlying our composite measure actually outperform, in time series, all non-ICC-based measures that we consider.}
II Related Literature

A large and growing literature examines the impact of regulation, managerial decisions, and market design on firm-level expected returns.\(^8\) A central issue in these studies is the reliability of their measures of expected returns, which reduces the power of researchers’ tests and thus influences the precision of the estimated effects. Our evaluation framework provides a means to compare ERPs on the basis of their MEVs, which is critical to the precision of researchers’ tests. Our hope is that this framework will bring some clarity to the continued growth, and use, of potential ERPs.

The value of assessing ERPs within a two-dimensional framework is intuitive. In many decision contexts, such as investment and capital budgeting, we would like ERPs to reflect cross-sectional differences in true expected returns. In many other research contexts, however, it is also crucial for inter-temporal variations in a firm’s ERPs to reflect variations in the firm’s true expected returns—for example, when researchers use a difference-in-differences research design to study the impact of a regulatory change on a firm’s expected returns. In these settings the time-series dimension is more relevant, but existing performance tests do not assess the quality of ERPs along this dimension.

Unlike prior studies that focus on cross-sectional differences in ERP performance (e.g., Easton and Monahan, 2005), our time-series tests allow researchers to identify the most suitable ERP for tracking a firm’s expected-return variation over time in a particular context. Thus a key contribution of our paper is demonstrating how researchers can implement this critical second dimension of ERP performance evaluation.\(^9\)

\(^8\)For example, firm-level expected returns have been used variously to study the effect of disclosure levels (Botosan, 1997; Botosan and Plumlee, 2005), information precision (Botosan et al., 2004), legal institutions and security laws (Hail and Leuz, 2006; Daouk et al., 2006), cross-listings (Hail and Leuz, 2009), corporate governance (Ashbaugh et al., 2004), accrual quality (Francis et al., 2004; Core et al., 2008), taxes and leverage (Dhaliwal et al., 2005), internal control deficiencies (Ashbaugh-Skaife et al., 2009), voluntary disclosure (Francis et al., 2008), and accounting restatements (Hribar and Jenkins, 2004).

\(^9\)Two prior studies have examined the time-series properties of ICCs, but our work is quite distinct from theirs. First, Easton and Sommers (2007) examines the properties of aggregate risk premiums implied by ICCs and the role of analyst biases. Second, Pastor et al. (2008) assess the time-series relations between aggregate risk premiums and market volatility. Both studies focus on aggregate expected returns at the market level, and neither addresses the relative performance of alternative ERPs. In contrast, we focus on
The paper most closely related to ours is Easton and Monahan (2005), hereafter referred to as EM, which also proposes a methodology for comparing the cross-sectional performance of alternative ICCs. Although our cross-sectional MEV metric is conceptually similar to the EM approach, our paper is distinct from EM’s analysis in several ways. First, we argue and demonstrate that better-performing ERPs should track true expected returns not only in the cross section but also over time, thus allowing researchers to more comprehensively assess the relative performance of ERPs. Second, 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 calls for a proxy’s measurement errors to be uncorrelated with true expected returns. This requirement disqualifies a large class of ERPs from consideration.\(^{10}\) Third, our approach circumvents the requirement of the EM framework to estimate multiple firm-specific and cross-sectional parameters (e.g., cash-flow news), and is thus much simpler to implement empirically. In short, our study builds on the EM analysis by (a) extending the measurement-error approach to the time-series dimension, (b) simplifying the estimation of ERPs’ MEVs, and (c) broadening the applicability of this approach to different classes of ERPs.

In a related study, Botosan et al. (2011) proposes evaluating ERPs based on their associations with risk proxies. A central difference between our approach and that of Botosan et al. (2011) has to do with which empirical constructs are assumed to be valid risk proxies. Botosan et al. (2011) assumes that expected returns, as an economic construct, must exhibit certain associations with assumed risk proxies. Our approach relies on expected returns as a statistical construct, and our method relies on the properties of conditional expectations without appealing to other risk proxies.

---

\(^{10}\)As a simple example, any ERP that is a multiple of true expected returns would violate the assumption necessary to use the EM approach because the ERP’s error would be correlated with the true expected-return. Moreover, Wang (2015) provides evidence that ICCs contain measurement errors that are nonrandom and are correlated with expected returns.
III Theoretical Underpinnings

This Section begins with a simple decomposition of returns and a discussion of the role of ERP bias. We then derive a two-dimensional measurement-error-variance-based framework for evaluating ERPs.

Return Decomposition

We begin with a simple decomposition of realized returns:

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

(1)

where $r_{i,t+1}$ is firm $i$’s realized return in period $t + 1$, and $\delta_{i,t+1}$ is the firm’s unanticipated news or forecast error.\(^{11}\) In this framework, we define $er_{i,t}$ as the firm’s true but unobserved expectation of future returns conditional on publicly available information at time $t$, capturing all ex-ante predictability (on the basis of the information set) in returns. By the property of conditional expectations, $er_{i,t}$ is an optimal or efficient forecast.\(^{12}\)

It follows from this definition and the property of conditional expectations that a firm’s expected returns ($er_{i,t}$) cannot be systematically correlated with its forecast errors ($\delta_{i,t+1}$), in time series or in the cross section. 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, thereby violating the definition of an optimal forecast. Put differently, news, by definition, cannot be predictable.

Having thus defined our normative benchmark, we abstract away from the market-efficiency debate. If one subscribes to market efficiency, $er_{i,t}$ should only be a function of risk factors and of expected risk premiums associated with these factors. Conversely, in a behavioral framework, $er_{i,t}$ can also be a function of non-risk-related behavioral factors.

\(^{11}\)In Campbell (1991), Campbell and Shiller (1988a,b), and Vuolteenaho (2002), analyze the relative importance of cash flow and expected return component of news, which requires a log-linearization which is unnecessary for our purposes. The decomposition of Equation (1) follows directly from the “Decomposition Property” of conditional expectations (e.g., Angrist and Pischke, 2008).

\(^{12}\)This is the “Prediction Property” of conditional expectations (e.g., Angrist and Pischke, 2008).
Next, we introduce the idea of ERPs \( \hat{er}_{i,t+1} \), defined as the unobserved expected return \( er_{i,t+1} \) measured with error \( \omega_{i,t+1} \):

\[
\hat{er}_{i,t+1} = er_{i,t+1} + \omega_{i,t+1}.
\]

(2)

In concept, \( \hat{er}_{i,t+1} \) need not be an ICC estimate as defined in the accounting literature—it can be any ex-ante expected-return measure, including a firm’s Beta, its book-to-market ratio, or its market capitalization at the beginning of the period. The key is that, whatever the “true” expected-return may be, we do not observe it. What we can observe are empirical proxies that contain measurement error. Our goal here is to accurately evaluate these proxies capture the variation in (or the tracking of) \( er_{i,t} \).

Differences between alternative ERPs are reflected in the properties (time-series and cross-sectional) of their \( \omega \) terms. Comparisons between different ERPs are, therefore, comparisons of the distributional properties of the \( \omega \)'s they generate, over time and across firms. Statements we make about the desirability of one ERP over another are, in essence, expressions of preference with regard to the properties of the alternative measurement errors (i.e., the \( \omega \) terms) that each is expected to generate. In other words, when we choose one ERP over another, we are specifying the loss function (in terms of measurement error) that we find least distasteful or problematic. The selection of ERPs thus becomes a choice between the attractiveness of alternative loss functions, expressed over \( \omega \) space.

**Minimizing Measurement-Error Variance**

What properties of \( \omega \), or equivalently which alternative loss function, should researchers care most about? The answer depends on the role ERPs play in a given research setting. As we show below, in most research settings, minimizing the squared measurement error (i.e., MSE) loss function may not be the most conceptually appropriate; instead, researchers should choose to minimize measurement-error variance (MEV) in choosing ERPs.

MSE is the standard criterion in statistics for assessing the quality of estimators. It
incorporates two aspects of an ERP’s measurement error properties: bias and variance.

\[ \mathbb{E} [(e_{rt} - \hat{e}_{rt})^2] = \mathbb{E} [(\omega_t)^2] = \text{Var}(\omega_t) + \mathbb{E}(\omega_t)^2 \]  

Thus, two candidate ERPs with similar MSE can differ widely in terms of their MEVs \([\text{Var}(\omega)]\) and amount of bias \([\mathbb{E}(\omega_t)^2]\). Thus, an important issue is whether and when the bias term matters in the selection of ERPs.

In most studies involving firm-level ERPs, the focus of the research is on examining expected-return variations over time or across firms. Typically, researchers are interested in studying how expected returns are impacted by a particular regulatory policy or firm-level attributes. In these setting, a well-performing ERP should closely “track” variations in true expected returns. In fact, Cochrane (2011) refers to understanding the drivers of expected-return variation as the “central organizing question in current asset-pricing research.” In capturing such variations, an ERP’s MEV is primary, while its bias is often irrelevant.

To illustrate this point, Table 1 summarizes 81 studies culled from top Accounting and Finance journals since 1997 that examine variations in firm-level expected returns using ICC estimates. These studies feature many different implementations of ICCs as proxies of expected returns. In the overwhelming majority of these studies (i.e., 83% = 67/81 of all papers), the empirical research design relies on the relative magnitudes of ERPs—e.g., how differences in policies or firm characteristics are associated with differences in ERPs.

Collectively listed as group ‘A’ in Table 1, these studies rely on empirical tests that exploit either the cross-sectional variation in the data (e.g., by using cross-sectional data or by including cross-sectional fixed effects in panel data), or the time-series variation in the data (e.g., by using time-series data or by including time-fixed effects in panel data). In a few cases (5 studies), the researchers utilized a difference-in-difference design (e.g., by focusing on an intervening event or policy change that affects only a subset of firms).

For all these studies, the precision of the empirical test depends on the variance of an
ERP’s measurement error and not its bias. Indeed, an ERP’s bias is irrelevant because it has been neutralized by the research design (e.g., due to time or firm fixed effects). Therefore, for these applications the ability of an ERP to “track” the variation in expected returns—or minimizing MEV—is paramount.

To be sure, there are a few cases where ERP bias cannot be safely ignored. For example, in four studies (Claus and Thomas, 2001; Fama and French, 2002; Ashton and Wang, 2013; Fitzgerald et al., 2013), collectively listed as group ‘B’ in Table 1, the authors use an ICC-based approach to evaluate the magnitude of the overall market risk premium. The research objective in these studies is to measure the actual level of expected returns at the aggregate (market) level, so the absolute magnitude of the ERP is of primary interest. In these studies, research conclusions and inferences will of course be highly sensitive to the amount of bias in each ICC estimate, and a focus on minimizing MEV would not be appropriate.

Aside from the 71 studies noted above, Table 1 lists 10 other “mezzanine” studies, collectively listed as group ‘C’ in Table 1, where the primary focus is on the methodology for estimating ICCs. Studies in this category often characterize the absolute magnitude of ERPs in addition to conducting cross-sectional tests, e.g., through Fama-MacBeth regressions, to assess how ERPs are related to firm characteristics. In these cross-sectional tests, bias is again irrelevant. For such studies, therefore, we are agnostic as to whether MSE or MEV is a better evaluation criterion. This is because none of these studies address the role that ERPs might play in a given research setting. Until such a setting is specified, the choice of the loss function is quite arbitrary. Even if we grant that in these cases, MSE is an appropriate loss function, the main takeaway from Table 1 is that, given the current literature on ERPs, situations in which the MSE criterion is preferable to the MEV criterion are the exception rather than the rule.

In summary, in most ERP studies, researchers are interested in how firm attributes, such as disclosure or corporate governance quality, affect firm-level expected returns. Meaningful inferences about magnitudes of these effects call for ERPs whose variations or differences
across firms or over time reflect those of true expected returns. In fact, excluding the primarily methodological papers, Table 1 indicates that minimizing MEV is a more appropriate evaluative criterion than minimizing MSE in 94% (67/71) of the research settings involving ICCs that we survey. Therefore, we argue that for most research purposes better ERPs should exhibit lower MEVs.

Measuring Time-Series and Cross-Sectional ERP Measurement-Error Variances

This Section develops a methodology for estimating an ERP’s cross-sectional and time-series MEVs. For brevity, we only summarize the key results here. Most of the technical details are relegated to the Technical Appendix.

**Time-Series Error Variance**

We show in this sub-section that it is possible to derive an empirically estimable and firm-specific measure, Scaled Time-Series Variance, denoted as \( SVar_i(\omega_{i,t}) \), that allows us to compare alternative ERPs in terms of their time-series MEV, even when the errors themselves are not observable. This analysis then provides the foundation for our comparison of ERPs.

The Technical Appendix provides a detailed derivation of \( Var_i(\omega_{i,t}) \) and \( SVar_i(\omega_{i,t}) \). Briefly, under the assumptions that (a) expected returns and measurement errors are jointly covariance-stationary,\(^{13}\) and (b) future news is unforecastable, we show in equation (T7) that the time-series MEV of a given ERP for firm \( i \) can be expressed as:

\[
Var_i(\omega_{i,t}) = Var_i(\hat{\epsilon}_r_{i,t}) - 2Cov_i(r_{i,t+1}, \hat{\epsilon}_r_{i,t}) + Var_i(\epsilon r_{i,t}),
\]

where \( Var_i(\hat{\epsilon}_r_{i,t}) \) is the time-series variance of a given ERP for firm \( i \), \( Var_i(\epsilon r_{i,t}) \) is the time-series variance of firm \( i \)'s expected returns, and \( Cov_i(r_{i,t+1}, \hat{\epsilon}_r_{i,t}) \) 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 that the (time-series) variance of a given

\(^{13}\)Note that this assumption is, in fact, less restrictive than the stationarity assumption needed to conduct the time-series regressions common in the literature.
ERP’s measurement error is increasing in the variance of the ERP \(Var_i(\hat{er}_{i,t})\). 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 that the variance of the error terms for a given ERP is decreasing in the covariance of the ERP and future returns \(Cov_i(r_{i,t+1}, \hat{er}_{i,t})\). This is also intuitive: to the extent 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 that the third term, \(Var_i(er_{i,t})\), 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 therefore does not play a role in relative performance comparisons. In other words, we need only the first two terms of (4) to determine which ERP exhibits lower time-series variance in measurement errors. Accordingly, we define the sum of the first two terms of (4) as the Scaled TS Variance:

\[
SVar_i(\omega_{i,t}) = Var_i(\hat{er}_{i,t}) - 2Cov_i(r_{i,t+1}, \hat{er}_{i,t}).
\]  

Note that \(SVar_i\) can be a negative number if the covariance between the ERP and future realized returns is sufficiently large and positive. In our empirical tests, we compute for each ERP and each firm the scaled error variance measure using (5), and then assess the time-series performance of ERPs based on the average of \(SVar_i\) across the \(N\) firms in our sample:

\[
AvgSVar_{TS}^{TS} = \frac{1}{N} \sum_i SVar_i(\omega_{i,t}).
\]  

For a given sample, ERPs that exhibit lower average time-series MEVs \(Var_i(\omega_{i,t})\) also exhibit lower average scaled TS variance \((AvgSVar_{TS}^{TS})\). All else equal, ERPs with lower time-series MEVs for a given sample are deemed to be of higher quality because time-series variation in the ERP is more likely to reflect the time-series variation in firms’ expected returns.
Note that AvgSVar\textsuperscript{TS} facilitates relative comparisons across ERPs. If we impose additional structure (that is, if we make stricter assumptions about the time-series behavior of expected returns), it is possible to obtain an empirically estimable absolute measure of the time-series MEV.\textsuperscript{14} We also note that this time-series framework can be utilized by researchers to select the best ERP for a specific firm on the basis of SVar\textsubscript{i}(\omega\textsubscript{i,t}).

\textit{Cross-Sectional Error Variance}

Although in certain research applications—e.g., those that identify the effects of interest based on the within-firm or time-series variation in the data—choosing the ERP with the lowest time-series SVAR is ideal, in other research contexts cross-sectional properties are more (or equally) important. For example, studies that identify causal parameters of interest based on across-firm or within-cross-section variation in the data, researchers should choose the ERP with lowest cross-sectional MEVs.

Employing similar logic, we show in Part B of the Technical Appendix that it is possible to derive an empirically estimable and proxy-specific measure, Average Scaled Cross-Sectional Variance (AvgSVar\textsuperscript{CS}), that allows us to compare the cross-sectional MEV of alternative ERPs. We show that the cross-sectional MEV of a given ERP for a given cross section \( t \) can be expressed as:

\[
\text{Var}_t(\omega\textsubscript{i,t}) = \text{Var}_t(\hat{\text{er}}\textsubscript{i,t}) - 2[\text{Var}_t(\text{er}\textsubscript{i,t}) + \text{Cov}_t(\text{er}\textsubscript{i,t}, \omega\textsubscript{i,t})] + \text{Var}_t(\text{er}\textsubscript{i,t}),
\]

(7)

where \( \text{Var}_t(\hat{\text{er}}\textsubscript{i,t}) \) is a given ERP’s cross-sectional variance at time \( t \), \( \text{Var}_t(\text{er}\textsubscript{i,t}) \) is the cross-sectional variance in firms’ expected returns at time \( t \), and \( \text{Cov}_t(r\textsubscript{i,t+1}, \hat{\text{er}}\textsubscript{i,t}) \) is the cross-sectional covariance between firms’ ERPs at time \( t \) and their realized returns in period \( t + 1 \).

Since the cross-sectional variance in firms’ expected returns—the last term—is invariant across different ERPs, relative comparisons of cross-sectional ERP MEV can be made by comparing the Scaled Cross-Sectional Variance:

\textsuperscript{14}This can be done, for example, by assuming that expected returns and ERP measurement errors follow AR(1) processes (e.g., Wang, 2015). The empirical tests in this paper, however, do not require such assumptions.
\[ S\text{Var}_t(\omega_{i,t}) = \text{Var}_t(\hat{e}r_{i,t}) - 2[\text{Var}_t(\text{er}_{i,t}) + \text{Cov}_t(\text{er}_{i,t}, \omega_{i,t})]. \] (8)

In particular, our empirical tests assess the cross-sectional performance of ERPs based on the average of \( S\text{Var}_t \) across the \( T \) cross-sections in our sample:

\[ \text{AvgSVar}^{CS} = \frac{1}{T} \sum_t \text{Var}_t(\hat{e}r_{i,t}) - 2[\text{Var}_t(\text{er}_{i,t}) + \text{Cov}_t(\text{er}_{i,t}, \omega_{i,t})]. \] (9)

Part B of the Technical Appendix shows that \( \text{AvgSVar}^{CS} \) can be estimated by

\[ \frac{1}{T} \sum_t \text{Var}_t(\hat{e}r_{i,t}) - 2\text{Cov}_t(r_{i,t+1}, \hat{e}r_{i,t}), \] (10)

following assumptions similar to those that characterize the time-series case above. Similar to the earlier intuition, Equation (10) 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.

**The Two-Dimensional Framework**

We emphasize that the two dimensions of our evaluation framework are not redundant. An ERP that performs well in time series may perform very 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 obscure the cross-sectional ordering of expected returns across firms. 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 proxies of 2 percent and 10 percent for stocks A and B, respectively. Suppose that a particular ERP model produces expected-returns proxies 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 misorders the stocks’ expected returns in the cross section and results in cross-sectional error variance.

Conversely, an ERP that performs well in the cross section may perform 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. Suppose again that the true expected returns of stocks A and B are always 10 percent and 2 percent, respectively. Now consider an ERP model that produces ERPs for A and B of 13 percent and 5 percent in certain years, and 10 percent and 2 percent in others. This ERP always correctly orders expected returns in the cross section and exhibits constant measurement errors for each firm in the cross section (i.e., zero MEV in each cross section), but produces time-series error variance. In this case, time-series variation in the proxies does not reflect variations in true expected returns, and instead reflects variations in measurement errors.

In sum, an ERP that is equal to true expected returns is clearly a “perfect” ERP in our framework. More broadly, as shown above, an ideal ERP in our framework may have non-zero measurement error, so long as these errors are (a) constant across time for a given firm and (b) constant in the cross section for all firms.

**Trivial Estimator as an Absolute Benchmark for ERP Performance**

As noted above, the variance of true expected returns, $Var(\epsilon_r)$, is unobservable and is constant across alternative ERPs, and thus we omit it from our evaluative framework. Therefore, we cannot directly estimate the absolute magnitude of the MEVs; however, because we can estimate them up to a constant (which is fixed across ERPs), comparing ERPs based on differences in SVARs identifies differences in MEVs. These have two important implications for our framework. First, ours is fundamentally a relative-performance framework. Second, our SVAR measure is bounded below not by 0, but by $-Var(\epsilon_r) < 0$, the SVAR we expect from the “perfect” ERP.

Although SVAR=0 is not an indication of optimal performance, it nevertheless remains a natural reference point in our framework. This is because a trivial estimator—one that specifies the ERP as a constant for all firms in the cross section and over time—produces
time-series and cross-sectional SVARs of 0. Thus, when examining a new ERP, a researcher can compare its SVAR to 0 to decide whether it constitutes an MEV-improvement over a trivial estimator.

**SVAR vs. Slope Coefficient Test**

Another technique commonly used to assess ERP performance is the so-called “slope coefficient” test. As implemented in the literature, this test typically involves cross-sectional regressions of future realized returns on each ERP proxy. An ERP with an estimated slope coefficient that is closer to 1 is generally deemed superior to one whose estimated slope is further from 1. Furthermore, although we have never seen its rationale stated, distance from 1 is presumably measured in absolute values (i.e., negative and positive distances from 1 are equally harmful to an ERP’s cause).

It is easy to see that a perfect ERP would have an estimated slope coefficient of 1. However, it is decidedly not the case that an estimated slope of 1 implies an ideal ERP. This is because the estimated slope coefficient does not speak to the magnitude of the MEV of each ERP. To build intuition for why the slope coefficient test differs from our MEV test, observe that the MEV can be related to the slope coefficient as follows:

\[ \text{Var}(\omega_t) = \text{Var}(er_t) + \text{Var}(\hat{er}_t) - 2\text{Cov}(r_{t+1}, \hat{er}_t). \]  

(11)

Because the coefficient from the slope coefficient test (\(\beta\)) is equal to \(\frac{\text{Cov}(r_{t+1}, \hat{er}_t)}{\text{Var}(\hat{er}_t)}\), we can rewrite \(\text{Var}(\omega_t)\) as:

\[ \text{Var}(\omega_t) = \text{Var}(er_t) + \text{Var}(\hat{er}_t)[1 - 2\beta]. \]  

(12)

Equation (12) illuminates the two main conceptual problems with the slope coefficient test. First, the estimated slope coefficient (\(\beta\)) does not generally reflect each ERP’s MEV

---

15Because future realized returns (\(r_{t+1}\)) are related to expected returns (\(er_t\)) through the decomposition property of conditional expectations, i.e., \(r_{t+1} = er_t + \text{news}_{t+1}\), a perfect ERP—with no measurement errors—implies a slope coefficient of 1.
[i.e., lower $\text{Var}(\omega_t)$]. Second, the “distance from 1” criterion is not monotonically related to each ERP’s MEV.

Consider the first point first. Equation (12) shows the relation between $\text{Var}(\omega_t)$ and $\beta$ depends on $\text{Var}(\hat{er}_t)$. Specifically, two ERPs with similar $\beta$ can have different MEVs if the ERPs have different variances. In fact, it is easy to generate multiple noisy ERPs that all exhibit a slope of 1\(^{16}\).

Ideally, we would also like ERPs to have a perfectly linear relationship with true expected returns: as it turns out, an ERP exhibits 0 MEV if, and only if, it produces a slope of 1 and has a perfect linear correlation (i.e., of 1) with expected returns,

$$\text{Corr}(\hat{er}_t, er_t) = \frac{\text{Cov}(er_t, \hat{er}_t)}{\text{Var}(\hat{er}_t)} \times \frac{\text{SD}(\hat{er}_t)}{\text{SD}(er_t)} = \beta \frac{\text{SD}(\hat{er}_t)}{\text{SD}(er_t)},$$

which is achieved only when the variance in the ERP is equal to that of expected returns. We can also see this in Equation (12): an ERP with a $\beta$ of 1 exhibits $\text{Var}(\omega_t) = \text{Var}(er_t) - \text{Var}(\hat{er}_t)$, which is minimized when the variances or the ERP and the underlying expected returns are the same.

Our second concern is with the “distance from 1” criterion. By itself, “distance from 1” can provide little or no information about MEV. Because neither the slope coefficient nor the $\beta = 1$ criterion is informative about MEVs, it follows that the distance of $\beta$ from one is also uninformative of MEVs. In contrast, Equations (11) and (12) show that $S\text{Var}(\omega_t)$ incorporates the information about $\beta$ and interacts it with $\text{Var}(\hat{er}_t)$, so that the relation between $S\text{Var}(\omega_t)$ and $\text{Var}(\omega_t)$ is direct and monotonic. Thus, if the goal is to obtain the ERP with the lowest MEV, the $S\text{Var}(\omega_t)$ measure is preferred.

Finally, we note that although the regression-slope coefficient is not a good substitute for $S\text{Var}(\omega_t)$, the sign of the slope can still provide useful supplemental information. In theory, it is possible for ERPs with positive and negative slopes to produce the same MEVs. Thus,

\(^{16}\)For example, suppose a researcher estimates that an ERP’s slope coefficient as being 10, then simply scaling the ERP by a factor of 10 yields a slope coefficient of 1, although doing so actually increases MEV.
given two ERPs with the same $SVar(\omega_t)$, we would unambiguously prefer the one with a positive slope. As a result, researchers can use regression-slope coefficient to narrow the field of candidate ERPs to those that have a positive relation with returns and use our framework to decide on the appropriate ERP for a given research setting. Viewed through this lens, a positive relation between ERP and future returns is a necessary but not sufficient condition for assessing its MEV.

Taken together, this Section provides rationale for evaluating ERPs in a two-dimensional framework based on MEVs, and develops a parsimonious methodology under a set of minimalistic assumptions for implementing such a framework. Researchers can use equations (6) and (10) to gauge the relative performance of ERPs to determine the optimal choice for a given research context.

IV Empirical Implementation

A key strength of our two-dimensional evaluation framework is that it is easy to implement and is thus easily portable across research settings. To illustrate how researchers can empirically implement our framework, this Section supplements our theoretical analyses by evaluating the relative performance of five families of monthly ERPs. These five ERP groups are based either on traditional equilibrium asset-pricing theory or on some variation of the ICC approach featured in accounting studies in recent years. Collectively, they span all the prototype classes of ERPs nominated by the academic finance and accounting literature over the past 50 years.

Three of the ERPs we test originate in traditional equilibrium asset-pricing theory (the left-hand branch of the ERP tree in Appendix I), where non-diversifiable risk is priced. Specifically, we test a single-factor version based on the Capital Asset Pricing Model, denoted CAPM, and a similar multi-factor version based on Fama and French (1993), denoted FFF. We also test a characteristic-based ERP from Lewellen (2015), calibrated using realized returns and denoted CER, in which a firm’s factor exposure (e.g., $\beta$’s) reflect its relative
ranking in terms of each firm characteristic.

We also test two prototype ERPs from the ICC literature (the right-hand branch of the ERP tree in Appendix I). ICC is the internal rate of return that equates a firm’s market value to the present value of its expected future cash-flows. Finally, we develop a new ERP prototype by computing a “fitted” ICC measure based on an instrumental-variable approach, whereby each firm’s expected return proxy is a fitted estimate from the regression, i.e., a linear function of the firm’s current characteristics, denoted FICC. FICC can be thought of as a cleaned-up version of ICCs, where the noise in ICC estimates are filtered through a linear projection onto firm characteristics. See Appendix II for details on the construction of each of the ERPs used in this study.

Sample Selection

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 1977-2014. For each firm-month, we estimate five ERPs using data from the CRSP Monthly Stock file and, when applicable, firms’ most recent annual financial statements. To be included in our sample, each firm-month observation must include information on stock 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 five ERPs, detailed below. Our final sample consists of 1,644,434 firm-month observations. To remove the influence of extreme observations, each ERP is winsorized at the 5th and 95th percentiles of the pooled-sample distribution.

Descriptive Statistics

Table 2 reports the medians of five monthly ERPs for each year from 1977 through 2014. We compute ERPs for each firm-month in our main sample based on the stock price and on publicly available information as of the last trading day of each month. Our sample consists
of firm-months for which all five ERPs are non-missing. The number of firm-months varies by year, ranging from a low of 30,858 in 1978 to a high of 60,835 in 1998. The average number of firm-months per year is 43,275, indicating that ERPs are available for a broad cross section of stocks in any given year. The time-series means of the monthly median ERPs range from 1.02 percent (for ICC), to 1.86 percent (for FICC).

It is instructive to compare the two factor-based proxies (CAPM and FFF) with their non-factor-based counterparts (ICC, CER, and FICC). The monthly means for CAPM and FFF (1.66 percent and 1.57 percent) are similar to those of the non-factor-based proxies. However, the time-series standard deviation of the factor-based proxies is 3 to 5 times larger than the standard deviation of the non-factor-based proxies. Annual medians for CAPM range from -1.66 percent to 4.55 percent; in 5 out of 35 years the median of CAPM is negative, indicating that more than half of the monthly observations signal expected returns below zero. The high volatilities of CAPM and FFF likely reflect the instability of the market equity risk premium or their corresponding sensitivities estimated using historical realized returns.

Table 3 reports the average monthly Pearson (Spearman) correlations above (below) the diagonal among the five ERPs. We calculate correlations by month and then average them over the sample period. The table shows that the three non-factor-based proxies are highly correlated among themselves, as are the two factor-based proxies. However, we find no positive correlation across the two groups—none of the three non-factor-based proxies is positively correlated with the two factor-based proxies. In fact, the correlations between the non-factor-based and factor-based proxies are generally negative (consistent with earlier findings reported by Gebhardt et al., 2001).

**Comparison of Measurement-Error Variances**

Using this data on the five ERPs, we begin by assessing their cross-sectional properties. Specifically, we compute Equation (8) for each unique calendar-month-ERP pair, and report
the summary statistics for the cross-sectional SVAR from each ERP in Panel A of Table 4. For ease of reporting, we multiply SVAR estimates by 100. Table values in this panel represent distributional statistics of SVAR computed across the 455 months or cross sections in our 1977-2014 sample period. Panel B reports t-statistics based on Newey-West-adjusted standard errors corresponding to the pair-wise comparisons of cross-sectional SVARs within the sample of 455 months used in Panel A. The reported values in Panel B of Table 4 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 that the three non-factor-based proxies (ICC, CER, and FICC) generate significantly lower cross-sectional error variances than the two factor-based ERPs. Among the non-factor-based proxies, CER outperforms the remaining four proxies, indicating that CER is best suited to cross-sectionally rank firms in our sample in terms of their true expected returns.

Recall from Section III that a negative mean SVAR measure indicates the ERP has lower MEV than a trivial estimator (i.e., a fixed constant). Based on this useful reference point, Table 4 shows that only CER outperforms a trivial ERP in the cross section. Although ICC and FICC outperform the factor-based ERPs, neither outperform the trivial ERP in the cross section. Our findings are thus consistent with the findings and conclusions of Easton and Monahan (2005) that ICCs in general are unreliable in the cross section.\textsuperscript{17}

The second dimension of our evaluative framework compares ERPs on the basis of their time-series MEVs. Table 5 reports each ERP’s time-series scaled MEV by computing equation (5) for each unique firm-ERP pair. Panel A provides descriptive statistics for the times-series SVAR from each ERP. To construct this panel, we require each firm to have a minimum of 20 (not necessarily consecutive) months of data during our 1977-2014 sample period. Table values in this panel represent distributional statistics of SVAR (also multiplied

\textsuperscript{17}Our Online Appendix evaluates the individual ICCs that constitute our composite ICC measure. We show that one of the ICC measures, from Gebhardt et al. (2001), also achieves a cross-sectional SVAR below 0. However, the other four ICC measures produce cross-sectional SVARs above 0.
by 100) computed across 12,397 unique firms that met the minimum-data requirement. Panel B reports t-statistics corresponding to the pair-wise comparison of firm-specific measurement errors across the sample of 12,397 firms used in Panel A.

The results in Table 5 show that the three non-factor-based proxies (ICC, CER, and FICC) generate significantly lower time-series error variances than the other two, similar to the results from our cross-sectional tests. FICC in particular outperforms all other ERPs in time series, including the absolute benchmark of SVAR=0.

To understand differences in ERP performance, it is useful to consider the primary sources of noise for each ERP family. The “factor-based” ERPs (i.e., CAPM and FFF) rely on slope coefficients estimated from time-series return regressions using monthly returns and factor risk premiums estimated using historical premiums. To the extent that Beta estimates for firms or factor risk premiums at a given time are difficult to calibrate using the time-series of realized returns, (e.g., Fama and French, 1997; Lyle et al., 2013), the resultant ERPs will also be noisy. On the other hand, ICC is based on an internal-rate-of-return calculation that equates current price to a firm’s book value or earnings, forecasted earnings, and expected payouts. To the extent that the earnings forecasts and expected payouts are noisy, ICC will likewise contain noise.

CER and FICC are alike in that they are “fitted” ERPs. In each case, the ERP is computed by expressing either ICC or returns as a linear function of current-year firm characteristics, which can be thought of as mechanisms for filtering the noise in realized returns and ICCs, respectively. To the extent these attributes capture some systematic aspect of true expected returns, through their associations with realized returns or ICCs, these fitted ERPs should perform better than either the factor-based ERPs or the ICC. Indeed, our findings lend support for characteristic-based fitting methodologies that are designed for, and appear successful in, removing noise from their underlying measures.

To better understand how estimation noise affects ERP performance, we also consider the effects of winsorization. The level of winsorization is an important design choice in virtually
all empirical studies involving ERPs. In fact, one can view the choice of which ERP to use as a choice about which model, inputs, and winsorization point to use in computing the ERP for a given sample. Because winsorization pulls in the tails of the distribution and thus reduces variance, the level of winsorization naturally affects the variance of ERP measurement errors. However, because winsorization also alters the relation between ERP and returns, the effect of winsorization on MEV may not be monotonic. At one extreme, no winsorization will likely result in a noisy ERP due, for example, to noise in ERP inputs. At the other extreme, full winsorization—that is replacing all ERP values with the median of the pooled sample—will yield a trivial ERP (with SVAR=0).

In Table 6, we report the effect of winsorization on each of the five ERPs we consider. To construct this table, we winsorize each ERP at various levels based on the pooled-sample distribution, then calculate the resultant SVAR. The first column indicates the degree of winsorization. For example, 5% indicates that the ERP is winsorized at the 5th and 95th percentiles of the pooled-sample distribution. This is the default level of winsorization used in deriving our results in Tables 4 and 5.

Table 6 shows how the performance for each ERP varies as we change the winsorization cutoff point. Several interesting results emerge from this table. First, the value of winsorization is greater for the noisier estimates: ICC, CAPM, and FFF. Going from a winsorization of 5% to 10%, for example, has a much larger impact on ICC than it does on FICC or CER. This is not surprising as the “fitted” approach should already filter out much of the noise in CER and FICC. Second, even with heavy winsorization (e.g., 20% on each side), the factor-based ERPs (CAPM and FFF) never approach the performance achieved by the characteristic-based ERPs, suggesting that estimated factor sensitivities and factor risk premiums are extremely noisy.

The impact of winsorization on ICC is particularly interesting. Above 7% winsorization on each side, ICC begins to exhibit negative time-series SVARs, indicating it outperforms a trivial estimator in time series. In fact, it begins to outperform CER as well, making
it the second-best ERP (after FICC) in terms of time-series MEVs. On the other hand, winsorization appears to have a much smaller impact on the cross-sectional performance of ICC and FICC.

Figure 1 provides a succinct summary of our findings. The horizontal axis of the graph is time-series SVAR for each ERP; the vertical axis is cross-sectional SVAR. Each point on the graph represents a given ERP under either 5% or 10% winsorization. We plot the performance for the three best performing ERPs, namely FICC, CER, and ICC. For reference, the performance of a trivial estimator (i.e., either a cross-sectional or a time-series constant) is depicted by a dotted line. This figure shows that both CER and FICC have lower time-series MEV than a trivial estimator. Winsorization has a sharp effect on ICC but almost no effect on CER. In fact, ICC also outperforms the trivial estimator in terms of its time-series MEV at 10% winsorization. Among the plotted ERPs, only CER outperforms the trivial estimator on both dimensions at these levels of winsorization.

We can also estimate the economic magnitude of the performance gain (in terms of percent MEV reduction) attained by one ERP over another. An important input variable needed for this estimation is $\text{Var}(er)$, the variance of the true (but unobservable) expected returns. To obtain these estimates, we observe that the variance of our best performing ERPs—CER and FICC—represent 0.1%-0.3% of the cross-sectional or time-series variance in monthly realized returns on average. Specifically, whereas the median cross-sectional variance in monthly returns is 0.0180, the median cross-sectional variance of CER and FICC are 0.0000325 (= 0.2% of 0.0180) and 0.0000583 (= 0.3% of 0.0180). Similarly, whereas the median time-series variance in monthly returns is 0.0216, the median cross-sectional variance of CER and FICC are .0000191 (= 0.08% of 0.0216) and .0000157 (= 0.07% of 0.0216). Assuming conservatively that 1% of realized variance is represented by the variance in expected returns, and applying this ratio to the median variance in monthly realized returns, we can derive a rough estimate of 0.000180 for cross sectional $\text{Var}(er)$, and 0.000216 for
time series $Var(\omega)$. With these estimates of $Var(\omega)$, we can then compute the absolute magnitude of the mean MEV, $Var(\omega)$, for each ERP: $Var(\omega) = Var(\omega) + SVar(erp)$. Finally, we use these estimates of MEV to compute the proportional improvement from one ERP to another.

Following this procedure, we find that CER’s performance in the cross section represents a 68.1%, 15.4%, 82.4%, and 92.4% improvement over ICC, FICC, CAPM, and FFF (in terms of reduction in cross sectional MEV). Similarly, FICC’s superior performance in time series represents a 64.0%, 48.0%, 94.2%, and 97.0% improvement over ICC, CER, CAPM, and FFF (in terms of the reduction in time series MEV). Furthermore, each of these two best-performing ERPs significantly improves over a trivial estimator, whose MEV is simply the variance of expected returns. Specifically, we find that the cross sectional performance of CER represents an improvement of 11.6% over that of the trivial estimator, while the time series performance of FICC represents an improvement of 51.2% over that of the trivial estimator.

Taken together, our findings provide a more sanguine assessment of ICC-based ERPs than prior literature. Our results are consistent with the conclusions of Easton and Monahan (2005), in that most ICC-based ERPs do not appear to be reliably better in the cross section than a trivial estimator. However, our findings further demonstrate the usefulness of ICCs in time series, a dimension of performance previously overlooked by the literature. Specifically, we show that, when the noise in ICCs is appropriately handled—either through a “fitted” approach or through winsorization—they become some of the best-performing and most useful ERP prototypes that we consider. Our conjecture is that the earnings and payout forecasts that are inputs into ICCs contain measurement errors that distort the cross-sectional ordering of firms in terms of their expected returns. These measurement errors are relatively stable over time, leading to better performance for ICC-based ERPs in time series.

Finally, we hasten to point out that the conclusions drawn from our empirical analyses are

---

18These estimates are conservative in the sense that larger $Var(\omega)$ estimates decreases the magnitude of the proportional improvement in MEVs.
based on the sample used in the analysis (e.g., Ecker et al., 2013) and on the types of ERPs tested. The performance results presented here are intended to illustrate how researchers can implement our two-dimensional evaluation framework for ERPs. We have not conducted an exhaustive test of all ERPs, and we do not intend these empirical findings to limit future explorations into novel ways to develop new ERPs.

V Online Appendix

To supplement the evidence presented in this paper, we also tabulate the following results in our Online Appendix for brevity:

1. **Influence of MEV on Empirical Tests**: To illustrate the influence of MEVs on researchers’ inferences, we simulate an underlying structure of expected returns. Our simulation suggests that while MEV affects researchers’ inferences by reducing the precision of her empirical tests, bias is irrelevant. We also show that, by conflating bias with MEV, MSE can lead to improper selection of ERPs.

2. **Observation Weights**: The analysis above compares the performance of alternative ERPs within an equal-weighted framework that treats each cross-sectional and time-series observation equally, a choice made for parsimony. However, the appropriate weights that a researcher uses when evaluating ERPs may depend on a specific research setting or agenda. To gauge the influence of alternative weights, we re-ran our main cross-sectional and time-series tests using two alternative weighted approaches based on: (1) the variance of realized returns and (2) the number of observations, where both weights correspond to a given cross section (month) and time series (firm). The results show our inferences about the relative performance of certain ERPs are unchanged when applying these alternative weights, suggesting that our relative comparisons are fairly stable across firms and time.

3. **Individual ICC Estimates**: We separately evaluate, within our two-dimensional framework, each of the five individual ICCs that constitute our composite ICC measure. The efficient frontier for this expanded set is defined by three of the individual ICC proxies and FICC, which display the lowest scaled time-series MEVs, and CER, which displays the lowest scaled cross-sectional MEVs. Interestingly, we find that three of the five ICCs exhibit a negative time-series SVAR, representing some of the best performing ERPs in time series among all the candidates we consider.
VI Conclusion

Estimates of expected returns play a central role in many managerial and investment decisions that affect the allocation of scarce resources in society. In fact, Cochrane (2011) refers to understanding the drivers of the variation in expected returns as the “central organizing question in current asset-pricing research.” This study addresses a key problem in the literature that relies on estimates of expected returns: how to assess the relative performance of expected-returns proxies (ERPs) when returns are noisy.

Our paper demonstrates the importance of evaluating the time-series performance of ERPs, whereas most prior studies focus on cross-sectional performance evaluation. Evaluating an ERP’s time-series performance is crucial in numerous contexts, such as when researchers use a difference-in-differences research design to study the impact of a regulatory change on a firm’s expected returns. We derive a two-dimensional evaluation framework that allows for the assessment of both time-series and cross-sectional measurement-error variances of ERPs.

Using a firm’s true but unobservable expected return as the normative benchmark, we define an ERP’s deviation from this benchmark as its measurement error. Although measurement errors are unobservable, we show it is possible to derive characteristics of the distribution of errors for each ERP such that researchers can compare the relative performance of alternative proxies. By establishing a rigorous evaluative framework, our findings help researchers select the appropriate ERP for a given context, and thus establish a minimum bar for what should be demanded from new entrants in the vast and still growing pool of firm-level expected-return proxies.
References


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 \( \text{AvgSVar}^{TS} \). Our 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 \((e_{r_{i,t+1}})\), 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., \(\delta_{i,t+1} = r_{i,t+1} - e_{r_{i,t}}\)) 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 (T1):

\[
\hat{e}r_{i,t+1} = e_{r_{i,t+1}} + \omega_{i,t+1}.
\]

(T1)

Taking the time-series variance on both sides of (T1) and re-organizing terms, a firm i’s time-series variance in ERP measurement errors can be written as

\[
Var_i(\omega_{i,t}) = Var_i(\hat{e}r_{i,t}) + Var_i(e_{r_{i,t}}) - 2Cov_i(e_{r_{i,t}}, \hat{e}r_{i,t}),
\]

(T2)

which can be re-expressed as

\[
Var_i(\omega_{i,t}) = Var_i(\hat{e}r_{i,t}) - 2[Var_i(e_{r_{i,t}}) + Cov_i(e_{r_{i,t}}, \omega_{i,t})] + Var_i(e_{r_{i,t}}).
\]

(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 needs only to compare the first two terms of (T3), which we refer to collectively as the Scaled Time-Series Variance of an ERP’s measurement errors of firm i’s expected returns \([SVar_i(\omega_{i,t})])\).

\[
SVar_i(\omega_{i,t}) = Var_i(\hat{e}r_{i,t}) - 2[Var_i(e_{r_{i,t}}) + Cov_i(e_{r_{i,t}}, \omega_{i,t})].
\]

(T4)

\[\text{A stochastic vector process } \{y_t\}_{t \geq 1} \text{ is covariance-stationary if (a) } E[y_t] = \mu \text{ for all } t, \text{ and (b) } E(y_t - \mu)(y_{t-j} - \mu) = \sum_j \text{ for all } t \text{ and any } j. \text{ That is, the mean and autocovariances do not depend on the date } t.\]

\[\text{Note that } Var_i(\omega_{i,t}) = SVar_i(\omega_{i,t}) + Var_i(e_{r_{i,t}}).\]
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 \(\text{Var}(\text{er}_{i,t})\) and the time-series covariance between the firm’s expected returns and the ERP measurement errors \(\text{Cov}(\text{er}_{i,t}, \omega_{i,t})\). 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 \(\text{Cov}(r_{i,t+1}, \hat{\text{er}}_{i,t}) = \text{Var}(\text{er}_{i,t}) + \text{Cov}(\text{er}_{i,t}, \omega_{i,t})\). To obtain this result, note that realized returns is the sum of the expected returns and news:

\[
r_{i,t+1} = \text{er}_{i,t} + \delta_{i,t+1}.
\] (T5)

We define \(\text{er}_{i,t}\) to be 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), \(\text{er}_{i,t}\) 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, thereby violating the efficiency or “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:

\[
\text{Cov}(r_{i,t+1}, \hat{\text{er}}_{i,t}) = \text{Cov}(\text{er}_{i,t} + \delta_{i,t+1}, \text{er}_{i,t} + \omega_{i,t})
= \text{Var}(\text{er}_{i,t}) + \text{Cov}(\text{er}_{i,t}, \omega_{i,t}),
\] (T6)

where the first equality follows from the return decomposition of (T5) and the definition of ERP (T1), and the last equality follows from assumption \(A2\), which implies that \(\text{Cov}(\delta_{i,t+1}, \text{er}_{i,t}) = \text{Cov}(\omega_{i,t+1}, \delta_{i,t+1}) = 0\).

**Step 3. Estimating \(\text{AvgSVar}^{TS}\)**

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

\[
\text{Var}(\omega_{i,t}) = \text{Var}(\hat{\text{er}}_{i,t}) - 2\text{Cov}(r_{i,t+1}, \hat{\text{er}}_{i,t}) + \text{Var}(\text{er}_{i,t}),
\] (T7)

so that

\[
\text{SVar}(\omega_{i,t}) = \text{Var}(\hat{\text{er}}_{i,t}) - 2\text{Cov}(r_{i,t+1}, \hat{\text{er}}_{i,t}).
\] (T8)

The first term of \(\text{SVar}(\omega_{i,t})\) shows that, all else equal, an ERP’s MEV is increasing in the variance of the ERP. The second term of \(\text{SVar}(\omega_{i,t})\) 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 (T8) expresses \(\text{SVar}(\omega_{i,t})\) in terms of two empirically observable variables \(\{\hat{\text{er}}_{i,t}, r_{i,t+1}\}\). These variables can be computed empirically, with consistency achieved under standard regularity conditions.\(^{21}\) Our empirical tests compute, for each ERP and each firm,

\(^{21}\)The following regularity conditions are sufficient to ensure that sample time-series variances and
the relative error-variance measure using (T8), and assess the time-series performance of ERPs based on the average of $SVar_i$ across the $N$ firms in our sample:

$$AvgSVar^{TS} = \frac{1}{N} \sum_i SVar_i(\omega_{i,t}).$$  \hspace{2cm} (T9)

Notice also that $SVar_i(\omega_{i,t}) \geq -Var_i(\epsilon_{r,i,t})$, because $Var_i(\omega_{i,t}) = SVar_i(\omega_{i,t}) + Var_i(\epsilon_{r,i,t})$ and $Var_i(\omega_{i,t}) \geq 0$. Therefore, $-Var_i(\epsilon_{r,i,t})$ is the minimum bound for our empirically estimable Scaled Time-Series Variance measure. In other words, if we have an ICC that measures expected returns perfectly, then $SVar_i(\omega_{i,t}) = -Var_i(\epsilon_{r,i,t})$.

### Part B. Estimating Cross-Sectional ICC Measurement-Error Variance

Here 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 $SVar_i(\omega_{i,t})$

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

$$Var_i(\omega_{i,t}) = Var_i(\hat{\sigma}_{r,i,t}) + Var_i(\epsilon_{r,i,t}) - 2Cov_i(\epsilon_{r,i,t}, \hat{\sigma}_{r,i,t}),$$  \hspace{2cm} (C1)

which can be re-expressed as

$$Var_i(\omega_{i,t}) = Var_i(\hat{\sigma}_{r,i,t}) - 2[Var_i(\epsilon_{r,i,t}) + Cov_i(\epsilon_{r,i,t}, \omega_{i,t})] + Var_i(\epsilon_{r,i,t}).$$  \hspace{2cm} (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 needs only to compare the first two terms.
of (C2), which we refer to collectively as the Scaled CS Variance of an ERP’s measurement errors \(SVar_t(\omega_{i,t})\).

\[
SVar_t(\omega_{i,t}) = Var_t(\hat{e}r_{i,t}) - 2[Var_t(er_{i,t}) + Cov_t(\hat{e}r_{i,t}, \omega_{i,t})].
\]

(C3)

Notice that 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 \(Cov_t(\hat{e}r_{i,t}, \omega_{i,t})\).

**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:

\[
AvgSVar^{CS} = \frac{1}{T} \sum_{t} Var_t(\hat{e}r_{i,t}) - 2[Var_t(\hat{e}r_{i,t}) + Cov_t(\hat{e}r_{i,t}, \omega_{i,t})].
\]

(C4)

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

\[
\frac{1}{T} \sum_{t} Cov_t(r_{i,t+1}, \hat{e}r_{i,t}) = \frac{1}{T} \sum_{t} Cov_t(\hat{e}r_{i,t} + \delta_{i,t+1}, \hat{e}r_{i,t} + \omega_{i,t})
\]

\[
= \frac{1}{T} \sum_{t} [Var_t(\hat{e}r_{i,t}) + Cov_t(\hat{e}r_{i,t}, \omega_{i,t})],
\]

(C5)

where the first equality follows from the realized returns decomposition (T5) and the definition of expected-returns proxy (T1), and the last equality follows from the assumption (A2) that “news” cannot exhibit systematic forecastability.\(^{24}\)

\(\text{Note that } Var_t(\omega_{i,t}) = SVar_t(\omega_{i,t}) + Var_t(\hat{e}r_{i,t})\)

\(\text{Note that for any given cross section it might be possible for realized news and measurement errors to be correlated, but this cannot be true systematically (i.e., across many cross-sections) by the definition of news.}\)
Appendix I. Family Tree of Expected-Return Proxies

Firm-Level Expected-Return Proxy (ERP)

Empirical Asset Pricing Approach

- Equilibrium Pricing: only non-diversifiable risks are priced
- A firm’s ERP is a linear function of its sensitivity to each factor ($\beta$’s), and the price of the factor ($\gamma$’s)

Implied-Cost-of-Capital Approach

- The share price reflects the PV of the expected CF to shareholders
- Agnostic with respect to the source of risk. ERP is the IRR that equates expected future CF to current price

Basic Premise

- Factor Identification (which factors matter?)
- Risk Premium for each factor (the $\gamma$’s)
- Firm factor loadings (the $\beta$’s)

Estimation Challenges

- CF forecasting assumptions: Future Earnings/FCF/Dividends? Terminal value estimation?
- Assumes a constant discount rate (what about inter-temporal variation in a firm’s expected returns?)

Representative ERP Variables

- CAPM (Sharpe, 1964; Lintner, 1965)
- FFF (Fama and French, 1993; Fama and French, 1996)
- CER (Lewellen, 2015; Chattopadhyay et al., 2015)

- ICC (Gebhardt et al., 2001; Hou et al., 2012)
- FICC (Fitted ICC; a new proxy)
Appendix II. Summary of Expected-Return Proxies (ERP)

This table summarizes five expected-return proxies (ERPs). For each of the five proxies (listed in the first column), the table provides a short description (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 estimated (in the fourth column), and cites related prior studies (in the last column). A full description of each proxy appears in Section IV.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Factor risk-premium (Gamma) estimation</th>
<th>Factor-loading (Beta) estimation</th>
<th>Related prior studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPM</td>
<td>One-factor expected-return proxy based on the Capital Asset Pricing Model (CAPM)</td>
<td>Based on the realized risk premium over the preceding 12 months</td>
<td>Based on time-series regressions using realized returns over the preceding 60 months</td>
<td>Sharpe (1964); Lintner (1965)</td>
</tr>
<tr>
<td>FFF</td>
<td>A four-factor expected-return proxy based on realized returns and each firm’s estimated sensitivity to four-factors (MKT, SMB, HML, and UMD)</td>
<td>Based on each factor’s realized risk premium over the preceding 12 months</td>
<td>Based on time-series regressions using realized returns over the preceding 60 months</td>
<td>Fama and French (1993); Fama and French (1996)</td>
</tr>
<tr>
<td>CER</td>
<td>Characteristic-based expected-return proxy whereby a firm’s exposure to a factor is simply its scaled rank on that characteristic (Size, BTM, Momentum)</td>
<td>Based on ten-year average Fama-MacBeth coefficients for each characteristic</td>
<td>Based on current year firm characteristics</td>
<td>Lewellen (2015); Chattopadhyay et al. (2015)</td>
</tr>
<tr>
<td>ICC</td>
<td>The internal rate of return (IRR) that equates a firm’s forecasted cash-flows to its current market price</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Gebhardt et al. (2001); Claus and Thomas (2001); Ohlson and Juettner-Nauroth (2005); Gordon and Gordon (1997); Easton and Monahan (2005); Hou et al. (2012)</td>
</tr>
<tr>
<td>FICC</td>
<td>Characteristic-based expected-return proxy whereby each firm’s ICC is “fitted” to a set of firm characteristics</td>
<td>Based on ten-year average Fama-MacBeth coefficients for each characteristic</td>
<td>Based on current year firm characteristics</td>
<td>New proxy</td>
</tr>
</tbody>
</table>

Factor-Based Expected-Return Proxies

Our empirical tests include two estimates of expected returns to those derived from standard factor models: CAPM and a four-factor model based on Fama and French (1993) that adds a momentum factor (the UMD factor obtained from Ken French’s data library). At the end of each calendar month \( t \), we estimate the expected one-month-ahead returns as

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

for each factor model (with \( J = 1, 4 \) factors), where \( rf_{t+1} \) is the risk-free rate in period \( t+1 \), \( \hat{\beta}_j \) are the estimated factor sensitivities (estimated in time series for each firm using monthly stock and factors’ returns over the 60 months prior to the forecast date), and \( f_{j,t} \) are the corresponding factors in period \( t \). Expected monthly factor returns are estimated based on trailing average 60-month factor returns.

We denote the capital-asset pricing model and a four-factor Fama-French type model as CAPM and FFF respectively. We estimate CAPM for each firm at the end of each calendar month using historical factor sensitivities. Specifically, we first estimate each firm’s Beta to the market factor using the prior 60 months’ data (from \( t-1 \) to \( t-60 \)). CAPM is then obtained by multiplying the estimated Beta by the most recent 12 months’ compounded annualized market-risk premium (provided by Fama and French) and adding the risk-free rate. Similarly, FFF represents a four-factor based ERP computed using the Mkt-Rf, HML, SMB, and UMD factors and 60-month rolling Beta estimates.
Implied Cost of Capital (ICC)

An increasingly popular approach for estimating firm-level expected returns is to calculate a firm’s implied cost of capital, which reflects the discount rate that investors use to value a firm’s future cash flows, as implied by the current market price of owning a firm’s equity. This study was, in part, motivated by the observation that there are several popular approaches for estimating a firm’s ICC. Thus, rather than select a specific implementation, we examine the performance of ICC estimates as ERPs by taking an equal-weighted average of five common models and refer to this average as “ICC”. These five estimates are those based on (1) Gebhardt et al. (2001), (2) Claus and Thomas (2001), (3) Ohlson and Juettner-Nauroth (2005), (4) Gordon and Gordon (1997), and (5) Easton and Monahan (2005). As reported in Wang (2015), among the papers that study expected returns using ICCs and that were published in the top accounting and financial journals from 1997 to 2014, 46% use a equal-weighted composite ICC.

Specific details regarding the specific structure of each measure can be found in our Online Appendix, however, to provide an overview of the ICC approach, we present a specific example below based on the approach in Gebhardt et al. (2001). This approach is a practical implementation of the residual income valuation model that employs a specific forecast methodology, forecast period, and terminal value assumption. Specifically, the time-\(t\) ICC expected-returns proxy for firm \(i\) is the rate that solves

\[
P_{i,t} = B_{i,t} + \sum_{n=1}^{11} \frac{E_t[NI_{i,t+n}]}{E_t[B_{i,t+n-1}]} \left(1 + r_e\right)^n - r_e \frac{E_t[NI_{i,t+12}]}{E_t[B_{i,t+11}]} - r_e \frac{E_t[B_{i,t+11}]}{E_t[B_{i,t+11}]},
\]

where \(E_t[NI_{i,t+n}]\) is the \(n\)-year-ahead forecast of earnings estimated using the approach in Hou et al. (2012). We estimate the book value per share, \(B_{i,t+n}\), using the clean surplus relation, and apply the most recent fiscal year’s dividend-payout ratio \((k)\) to all future expected earnings to obtain forecasts of expected future dividends, i.e., \(E_t[D_{t+n+1}] = E_t[NI_{t+n+1}] \times k\). The resulting measure reflects the discount rate implied by market prices that investors use when valuing a firm’s future cash flow stream.

We implement each model using the earnings forecast approach outlined in Hou et al. (2012). We compute ICC as of the last trading day of each calendar month for all U.S. firms (excluding ADRs and those in the Miscellaneous category in the Fama-French 48-industry classification scheme), combining monthly prices and total-shares data from CRSP and annual financial-statements data from Compustat.

Characteristic-Based Expected-Return Proxies

Following Lewellen (2015), we calculate a characteristic-based ERP, which we denote as CER, by first estimating a firm’s factor loadings to three characteristics. This measure is based on an instrumental-variable approach, whereby each firm’s returns are regressed on a vector of firm characteristics in each cross section (i.e., using Fama-MacBeth regressions); then the historical average of the estimated slope coefficients is applied to a given forecast period’s observed firm characteristics to obtain a proxy of expected future returns.

We also compute a “fitted” expected-return proxy, which uses ICCs instead of returns as the dependent variable. We refer to this fitted version of ICC, which represents a “fitted” value using historically estimated Fama-MacBeth coefficients, as FICC. We apply rolling 10-year Fama-MacBeth coefficients in our implementations of FICC and CER.
Figure 1. Efficient Frontier for Best Performing ERPs

This figure summarizes our main empirical findings for the best performing ERPs. The horizontal axis of the graph is their time-series scaled mean error variance (SVAR); the vertical axis is their cross-sectional SVAR. Each point on the graph represents a given ERP under either 5% or 10% winsorization. The figure is labeled such that 5% (10%) indicates that a given ERP is winsorized at the 5th (10th) and 95th (90th) percentiles of its pooled distribution. We plot the performance for three ERPs, namely FICC, CER, and ICC. For reference, the performance of a trivial estimator (i.e., either a cross-sectional or a time-series constant) is depicted by a dotted line. Under the minimal mean error variance (MEV) performance criterion, the best performing ERPs form an efficient frontier to the lower left corner of the graph.
This table reports the total number of papers that examine the magnitudes and variations in ICCs. The papers tabulated were published since 1997 in the following accounting and finance journals: *The Accounting Review, Journal of Accounting and Economics, Journal of Accounting Research, Review of Accounting Studies, Contemporary Accounting Research, Accounting Horizons, Journal of Finance, Journal of Financial Economics, Review of Financial Studies,* and *Journal of Corporate Finance.* These articles are collated by searching for keywords and citations. Citation searching uses Google Scholar to find papers in the 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, data as well as varying fixed effects for regression models with ICC as the dependent variable.

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Percent</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Papers</td>
<td>81</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>A. Relative Magnitude</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1. Pure Time Series Data</td>
<td>1</td>
<td>1% of A</td>
<td>Pástor et al. (2008)</td>
</tr>
<tr>
<td>A2. Pure Cross Sectional Data</td>
<td>2</td>
<td>3% of A</td>
<td>Baginski and Rakow (2012); Botosan (1997)</td>
</tr>
<tr>
<td>A3. Panel Data</td>
<td>64</td>
<td>96% of A</td>
<td></td>
</tr>
<tr>
<td>A3[a]. Time FE/Firm FE/Changes</td>
<td>38</td>
<td>50% of A</td>
<td>Attig et al. (2008); Barth et al. (2008); Ben-nan et al. (2012); Botosan and Plummer (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); Chen et al. (2011, 2009, 2015); Zhihong et al. (2010); Dai et al. (2013); Daske et al. (2008, 2013); DHALIWAL et al. (2006, 2007, 2005, 2011); EL GHOL et al. (2013, 2011); 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); Hibar and Jenkins (2004); Hwang et al. (2013); Khurana and Raman (2004, 2006); Kim et al. (2012); Kim (2014); KRISHNAH et al. (2008, 2013); LAG et al. (2012); Garcia Lara et al. (2011); Lai et al. (2010); Lawrence et al. (2011); Li and Mohanram (2014); Sugio (2010); LIU and NATARAJAN (2012); McInnis (2010); Mohanram and Rajgopal (2009); NAIKER et al. (2013); OGNEVA et al. (2007); PÁSTOR et al. (2008); RAJAN et al. (2007); SENGUPTA and ZHANG (2015); SIKES and VERRECCHIA (2015); DHALIWAL et al. (2016); ZHOU et al. (2016)</td>
</tr>
<tr>
<td>A3[b]. Fama-Macbeth</td>
<td>21</td>
<td>33% of A</td>
<td></td>
</tr>
<tr>
<td>A3[c]. Diff-in-Diff</td>
<td>5</td>
<td>8% of A</td>
<td>Callahan et al. (2012); Dai et al. (2013); Kim et al. (2012); Sugio (2010); NAIKER et al. (2013)</td>
</tr>
<tr>
<td>B. Absolute Magnitudes</td>
<td>4</td>
<td>5% of Total</td>
<td>ASHTON and WANG (2013); CLAUS and THOMAS (2001); FAMA and FRENCH (2002); FITZGERALD et al. (2013)</td>
</tr>
<tr>
<td>C. Primarily Methodological</td>
<td>10</td>
<td>12% of Total</td>
<td>BAGINSKI and WALEHN (2003); EASTON (2004); EASTON and SOMMERS (2007); EASTON et al. (2002); GEBHARDT et al. (2001); GODE and MOHANRAM (2003); HOU et al. (2012); LAROCQUE (2013); NEKRASOV and ONGEVA (2011); MOHANRAM and GODE (2013)</td>
</tr>
</tbody>
</table>
Table 2. Monthly Expected-Return Proxies by Year

This table reports the median of monthly expected-return proxies. ICC is a composite measure of implied cost of capital by taking the equal-weighted average of five commonly proxies based on (1) Gebhardt et al. (2001), (2) Claus and Thomas (2001), (3) Ohlson and Juetter-Nauroth (2005), (4) Gordon and Gordon (1997), and (5) Easton and Monahan (2005); CER is a characteristic-based expected-return proxy derived from a historical regression of realized returns on a firm’s size, book-to-market, and return momentum where historically estimated ten-year average Fama-MacBeth coefficients are applied to current firm characteristics; FICC is analogously defined from a regression of ICC on a firm’s size, book-to-market, and return momentum; CAPM is the firm’s expected-return derived from the Capital Asset Pricing Model; and FFF is the four-factor model using the market, small-minus-big (SMB), high-minus-low (HML), and up-minus-down (UMD) factors. A full description of each proxy appears in Section IV. 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, where the expected-return corresponds to the following month. Proxies are treated as missing if they are either below zero or above 100 percent.

<table>
<thead>
<tr>
<th>Year</th>
<th>Obs</th>
<th>ICC</th>
<th>CER</th>
<th>FICC</th>
<th>CAPM</th>
<th>FFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>30,858</td>
<td>1.56%</td>
<td>0.86%</td>
<td>2.12%</td>
<td>-0.07%</td>
<td>1.11%</td>
</tr>
<tr>
<td>1978</td>
<td>30,839</td>
<td>1.48%</td>
<td>0.78%</td>
<td>1.92%</td>
<td>2.82%</td>
<td>2.81%</td>
</tr>
<tr>
<td>1979</td>
<td>31,630</td>
<td>1.51%</td>
<td>0.90%</td>
<td>1.86%</td>
<td>3.34%</td>
<td>3.50%</td>
</tr>
<tr>
<td>1980</td>
<td>33,220</td>
<td>1.53%</td>
<td>1.36%</td>
<td>1.78%</td>
<td>4.55%</td>
<td>3.51%</td>
</tr>
<tr>
<td>1981</td>
<td>33,376</td>
<td>1.04%</td>
<td>1.43%</td>
<td>1.61%</td>
<td>0.57%</td>
<td>1.19%</td>
</tr>
<tr>
<td>1982</td>
<td>34,387</td>
<td>1.21%</td>
<td>1.11%</td>
<td>1.73%</td>
<td>0.91%</td>
<td>1.06%</td>
</tr>
<tr>
<td>1983</td>
<td>36,245</td>
<td>1.31%</td>
<td>2.19%</td>
<td>1.36%</td>
<td>2.50%</td>
<td>3.02%</td>
</tr>
<tr>
<td>1984</td>
<td>36,268</td>
<td>1.33%</td>
<td>1.78%</td>
<td>1.59%</td>
<td>0.58%</td>
<td>-0.21%</td>
</tr>
<tr>
<td>1985</td>
<td>36,360</td>
<td>1.30%</td>
<td>1.86%</td>
<td>1.61%</td>
<td>2.13%</td>
<td>1.97%</td>
</tr>
<tr>
<td>1986</td>
<td>38,984</td>
<td>1.24%</td>
<td>1.81%</td>
<td>1.57%</td>
<td>1.95%</td>
<td>1.23%</td>
</tr>
<tr>
<td>1987</td>
<td>38,785</td>
<td>1.25%</td>
<td>1.67%</td>
<td>1.72%</td>
<td>2.16%</td>
<td>1.56%</td>
</tr>
<tr>
<td>1988</td>
<td>40,407</td>
<td>1.33%</td>
<td>1.41%</td>
<td>1.97%</td>
<td>1.06%</td>
<td>1.23%</td>
</tr>
<tr>
<td>1989</td>
<td>42,095</td>
<td>1.26%</td>
<td>1.53%</td>
<td>1.91%</td>
<td>1.71%</td>
<td>1.57%</td>
</tr>
<tr>
<td>1990</td>
<td>41,233</td>
<td>1.45%</td>
<td>1.15%</td>
<td>2.12%</td>
<td>-0.61%</td>
<td>-0.69%</td>
</tr>
<tr>
<td>1991</td>
<td>40,779</td>
<td>1.30%</td>
<td>1.25%</td>
<td>2.08%</td>
<td>2.33%</td>
<td>2.90%</td>
</tr>
<tr>
<td>1992</td>
<td>40,898</td>
<td>1.14%</td>
<td>1.39%</td>
<td>2.06%</td>
<td>0.99%</td>
<td>1.60%</td>
</tr>
<tr>
<td>1993</td>
<td>42,755</td>
<td>0.93%</td>
<td>1.05%</td>
<td>2.04%</td>
<td>1.23%</td>
<td>1.97%</td>
</tr>
<tr>
<td>1994</td>
<td>49,993</td>
<td>1.04%</td>
<td>1.09%</td>
<td>2.11%</td>
<td>1.01%</td>
<td>0.39%</td>
</tr>
<tr>
<td>1995</td>
<td>55,075</td>
<td>0.94%</td>
<td>1.10%</td>
<td>2.09%</td>
<td>2.91%</td>
<td>2.54%</td>
</tr>
<tr>
<td>1996</td>
<td>58,078</td>
<td>0.90%</td>
<td>1.11%</td>
<td>1.98%</td>
<td>2.05%</td>
<td>2.02%</td>
</tr>
<tr>
<td>1997</td>
<td>59,199</td>
<td>0.84%</td>
<td>1.13%</td>
<td>1.89%</td>
<td>2.62%</td>
<td>2.14%</td>
</tr>
<tr>
<td>1998</td>
<td>60,835</td>
<td>0.76%</td>
<td>1.16%</td>
<td>1.83%</td>
<td>2.58%</td>
<td>1.57%</td>
</tr>
<tr>
<td>1999</td>
<td>58,592</td>
<td>0.81%</td>
<td>1.15%</td>
<td>1.86%</td>
<td>1.51%</td>
<td>0.79%</td>
</tr>
<tr>
<td>2000</td>
<td>56,157</td>
<td>0.88%</td>
<td>1.57%</td>
<td>1.81%</td>
<td>-0.47%</td>
<td>0.01%</td>
</tr>
<tr>
<td>2001</td>
<td>52,808</td>
<td>0.87%</td>
<td>1.46%</td>
<td>1.82%</td>
<td>0.54%</td>
<td>1.68%</td>
</tr>
<tr>
<td>2002</td>
<td>50,231</td>
<td>0.90%</td>
<td>1.27%</td>
<td>1.81%</td>
<td>-0.16%</td>
<td>-0.44%</td>
</tr>
<tr>
<td>2003</td>
<td>47,027</td>
<td>0.83%</td>
<td>1.21%</td>
<td>1.88%</td>
<td>2.24%</td>
<td>3.41%</td>
</tr>
<tr>
<td>2004</td>
<td>46,239</td>
<td>0.58%</td>
<td>1.41%</td>
<td>1.73%</td>
<td>2.28%</td>
<td>2.14%</td>
</tr>
<tr>
<td>2005</td>
<td>45,886</td>
<td>0.56%</td>
<td>1.21%</td>
<td>1.80%</td>
<td>1.93%</td>
<td>1.16%</td>
</tr>
<tr>
<td>2006</td>
<td>45,091</td>
<td>0.64%</td>
<td>1.18%</td>
<td>1.80%</td>
<td>2.36%</td>
<td>2.00%</td>
</tr>
<tr>
<td>2007</td>
<td>45,661</td>
<td>0.65%</td>
<td>1.03%</td>
<td>1.79%</td>
<td>1.50%</td>
<td>0.59%</td>
</tr>
<tr>
<td>2008</td>
<td>44,591</td>
<td>0.82%</td>
<td>0.70%</td>
<td>2.17%</td>
<td>-1.66%</td>
<td>-1.20%</td>
</tr>
<tr>
<td>2009</td>
<td>42,036</td>
<td>0.97%</td>
<td>0.61%</td>
<td>2.43%</td>
<td>3.81%</td>
<td>3.62%</td>
</tr>
<tr>
<td>2010</td>
<td>41,326</td>
<td>0.82%</td>
<td>0.59%</td>
<td>1.98%</td>
<td>2.56%</td>
<td>2.71%</td>
</tr>
<tr>
<td>2011</td>
<td>40,860</td>
<td>0.73%</td>
<td>0.61%</td>
<td>1.93%</td>
<td>0.33%</td>
<td>-0.40%</td>
</tr>
<tr>
<td>2012</td>
<td>41,069</td>
<td>0.73%</td>
<td>0.71%</td>
<td>1.95%</td>
<td>2.03%</td>
<td>1.81%</td>
</tr>
<tr>
<td>2013</td>
<td>39,915</td>
<td>0.71%</td>
<td>0.72%</td>
<td>1.62%</td>
<td>3.32%</td>
<td>2.86%</td>
</tr>
<tr>
<td>2014</td>
<td>32,746</td>
<td>0.65%</td>
<td>0.61%</td>
<td>1.38%</td>
<td>1.55%</td>
<td>0.87%</td>
</tr>
</tbody>
</table>

Mean 43,275 1.02% 1.19% 1.86% 1.66% 1.57%  
Median 41,280 0.93% 1.16% 1.86% 1.94% 1.59%  
Min 30,839 0.56% 0.59% 1.36% -1.66% -1.20%  
Max 60,835 1.56% 2.19% 2.43% 4.55% 3.62%  
Std 8,374 0.30% 0.38% 0.22% 0.22% 1.24%  
Mean/Std 5.17 3.43 3.10 8.57 1.28 1.27
Table 3. Correlation between Expected-Return Proxies

This table reports the average monthly Pearson (Spearman) correlations above (below) the diagonal among the five expected-return proxies. ICC is a composite measure of implied cost of capital by taking the equal-weighted average of five commonly proxies based on (1) Gebhardt et al. (2001), (2) Claus and Thomas (2001), (3) Ohlson and Juettner–Nauroth (2005), (4) Gordon and Gordon (1997), and (5) Easton and Monahan (2005); CER is a characteristic-based–return proxy derived from a historical regression of realized returns on a firm’s size, book-to-market, and return momentum where historically estimated ten-year average Fama-MacBeth coefficients are applied to current firm characteristics; FICC is analogously defined from a regression of ICC on a firm’s size, book-to-market, and return momentum; CAPM is the firm’s expected-return derived from the Capital Asset Pricing Model; and FFF is the four-factor model using the market, small-minus-big (SMB), high-minus-low (HML), and up-minus-down (UMD) factors. A full description of each proxy appears in Section IV. *p*-values are shown in parentheses corresponding to one-tailed tests under the alternative hypothesis that expected-return proxies should be positively correlated.

<table>
<thead>
<tr>
<th></th>
<th>ICC</th>
<th>CER</th>
<th>FICC</th>
<th>CAPM</th>
<th>FFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.259</td>
<td>0.389</td>
<td>-0.118</td>
<td>-0.079</td>
<td></td>
</tr>
<tr>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>CER</td>
<td>0.458</td>
<td></td>
<td>0.658</td>
<td>-0.112</td>
<td>-0.066</td>
</tr>
<tr>
<td>(0.00)</td>
<td></td>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>FICC</td>
<td>0.639</td>
<td>0.679</td>
<td></td>
<td>-0.249</td>
<td>-0.144</td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td>-0.175</td>
<td>-0.121</td>
<td>-0.246</td>
<td></td>
<td>0.600</td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>FFF</td>
<td>-0.119</td>
<td>-0.074</td>
<td>-0.145</td>
<td>0.597</td>
<td></td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td>(0.00)</td>
</tr>
</tbody>
</table>
### Table 4. Cross-Sectional Scaled Measurement-Error Variances

This table presents descriptive statistics for the cross-sectional variances of scaled measurement errors (multiplied by 100) of five expected-return proxies. Scaled measurement-error variances are calculated for each unique firm-expected/return proxy pair using Equation (10) as follows:

$$\text{AvgSVar}^{CS} = \frac{1}{T} \sum_{t} \text{Var}_t(\hat{\epsilon}_{r,t}) - 2\text{Cov}_t(r_{1,t+1}, \hat{\epsilon}_{r,t}).$$

where $\text{Var}_t(\hat{\epsilon}_{r,t})$ is a given ERP’s cross-sectional variance at time $t$ and $\text{Cov}_t(r_{1,t+1}, \hat{\epsilon}_{r,t})$ is the cross-sectional covariance between firms’ ERPs at time $t$ and their realized returns in period $t+1$. Panel A reports summary statistics for the error variance from each model, using data from a sample of 455 calendar months during our 1977–2014 sample period. Table values in this panel represent descriptive statistics for the error variance from each expected-return proxy computed across these 455 months. ICC is a composite measure of implied cost of capital by taking the equal-weighted average of five commonly proxies based on (1) Gebhardt et al. (2001), (2) Claus and Thomas (2001), (3) Ohlson and Juettner-Nauroth (2005), (4) Gordon and Gordon (1997), and (5) Easton and Monahan (2005); CER is a characteristic-based expected-return proxy derived from a historical regression of realized returns on a firm’s size, book-to-market, and return momentum where historically estimated ten-year average Fama-MacBeth coefficients are applied to current firm characteristics; FICC is analogously defined from a regression of ICC on a firm’s size, book-to-market, and return momentum; CAPM is the firm’s expected-return derived from the Capital Asset Pricing Model; and FFF is the four-factor model using the market, small-minus-big (SMB), high-minus-low (HML), and up-minus-down (UMD) factors. A full description of each proxy in Section IV. Panel B reports $t$-statistics based on Newey-West-adjusted standard errors corresponding to the pair-wise comparisons of average cross-sectional scaled measurement-error variances within the sample of 455 months used in Panel A. 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. *, **, and *** indicate two-tailed significance at the 10%, 5%, and 1% levels respectively.

#### Panel A: Cross-Sectional Measurement-Error Variance (N=455)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>P25</th>
<th>Median</th>
<th>P75</th>
<th>STD</th>
<th>$t$-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.032</td>
<td>0.010</td>
<td>0.030</td>
<td>0.051</td>
<td>0.040</td>
<td>17.015</td>
</tr>
<tr>
<td>CER</td>
<td>-0.002</td>
<td>-0.008</td>
<td>-0.001</td>
<td>0.005</td>
<td>0.014</td>
<td>-3.136</td>
</tr>
<tr>
<td>FICC</td>
<td>0.001</td>
<td>-0.010</td>
<td>0.001</td>
<td>0.013</td>
<td>0.022</td>
<td>0.771</td>
</tr>
<tr>
<td>CAPM</td>
<td>0.073</td>
<td>0.026</td>
<td>0.060</td>
<td>0.115</td>
<td>0.103</td>
<td>15.098</td>
</tr>
<tr>
<td>FFF</td>
<td>0.190</td>
<td>0.082</td>
<td>0.149</td>
<td>0.248</td>
<td>0.265</td>
<td>15.286</td>
</tr>
</tbody>
</table>

#### Panel B: $t$-Statistics of Differences in Variances

<table>
<thead>
<tr>
<th></th>
<th>ICC</th>
<th>CER</th>
<th>FICC</th>
<th>CAPM</th>
<th>FFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>-0.034</td>
<td>-0.031</td>
<td>0.041</td>
<td>0.158</td>
<td></td>
</tr>
<tr>
<td>CER</td>
<td>0.034</td>
<td>-0.003</td>
<td>0.075</td>
<td>0.192</td>
<td></td>
</tr>
<tr>
<td>FICC</td>
<td>0.031</td>
<td>-0.003</td>
<td>0.072</td>
<td>0.189</td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td>-0.041</td>
<td>-0.072</td>
<td>0.072</td>
<td>0.118</td>
<td></td>
</tr>
<tr>
<td>FFF</td>
<td>-0.158</td>
<td>-0.189</td>
<td>-0.118</td>
<td>0.925</td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Time-Series Scaled Measurement-Error Variances

This table presents descriptive statistics for the time-series variances of scaled measurement errors (multiplied by 100) of five expected-return proxies. Scaled measurement-error variances are calculated for each unique firm-expected-return proxy pair using Equation (6) as follows:

\[ \text{AvgSVar}_{TS} = \frac{1}{N} \sum_{t} SVar_i(\omega_{i,t}) \]

where \( SVar_i(\omega_{i,t}) \) is the scaled measurement-error variance of firm \( i \). Panel A reports summary statistics for the error variance from each model, using a sample of 12,397 unique firms with a minimum of 20 (not necessarily consecutive) months of data during our 1977-2014 sample period. Table values in this panel represent descriptive statistics for the error variance from each expected-return proxy computed across these 12,397 firms. ICC is a composite measure of implied cost of capital by taking the equal-weighted average of five commonly proxies based on (1) Gebhardt et al. (2001), (2) Claus and Thomas (2001), (3) Ohlson and Juettner-Nauroth (2005), (4) Gordon and Gordon (1997), and (5) Easton and Monahan (2005); CER is a characteristic-based expected-return proxy derived from a historical regression of realized returns on a firm’s size, book-to-market, and return momentum where historically estimated ten-year average Fama-MacBeth coefficients are applied to current firm characteristics; FICC is analogously defined from a regression of ICC on a firm’s size, book-to-market, and return momentum; CAPM is the firm’s expected-return derived from the Capital Asset Pricing Model; and FFF is the four-factor model using the market, small-minus-big (SMB), high-minus-low (HML), and up-minus-down (UMD) factors. A full description of each proxy appears in Section IV. Panel B reports \( t \)-statistics corresponding to the pair-wise comparisons of average firm-specific scaled measurement-error variances within the sample of 12,397 firms used in Panel A. 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. *, **, and *** indicate two-tailed significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Time-Series Measurement-Error Variance (N=12,397)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>P25</th>
<th>Median</th>
<th>P75</th>
<th>STD</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.008</td>
<td>-0.016</td>
<td>-0.002</td>
<td>0.024</td>
<td>0.103</td>
<td>8.337</td>
</tr>
<tr>
<td>CER</td>
<td>-0.001</td>
<td>-0.008</td>
<td>-0.001</td>
<td>0.005</td>
<td>0.022</td>
<td>-6.946</td>
</tr>
<tr>
<td>FICC</td>
<td>-0.011</td>
<td>-0.017</td>
<td>-0.006</td>
<td>-0.001</td>
<td>0.026</td>
<td>-47.707</td>
</tr>
<tr>
<td>CAPM</td>
<td>0.160</td>
<td>0.033</td>
<td>0.121</td>
<td>0.250</td>
<td>0.295</td>
<td>60.473</td>
</tr>
<tr>
<td>FFF</td>
<td>0.332</td>
<td>0.119</td>
<td>0.257</td>
<td>0.464</td>
<td>0.434</td>
<td>85.286</td>
</tr>
</tbody>
</table>

Panel B: \( t \)-Statistics of Differences in Variances

<table>
<thead>
<tr>
<th></th>
<th>ICC</th>
<th>CER</th>
<th>FICC</th>
<th>CAPM</th>
<th>FFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td></td>
<td>-0.009</td>
<td>-0.019</td>
<td>0.152</td>
<td>0.325</td>
</tr>
<tr>
<td>CER</td>
<td>0.009</td>
<td>-0.010</td>
<td>-0.010</td>
<td>0.161</td>
<td>0.334</td>
</tr>
<tr>
<td>FICC</td>
<td>(9.41)</td>
<td>(5.41)</td>
<td>(30.00)</td>
<td>(54.22)</td>
<td>(69.72)</td>
</tr>
<tr>
<td>CAPM</td>
<td>(20.22)</td>
<td>(30.00)</td>
<td>(56.78)</td>
<td>(70.40)</td>
<td></td>
</tr>
<tr>
<td>FFF</td>
<td>(47.88)</td>
<td>(54.22)</td>
<td>(56.78)</td>
<td>(46.74)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ICC</th>
<th>CER</th>
<th>FICC</th>
<th>CAPM</th>
<th>FFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td></td>
<td>-0.152</td>
<td>-0.161</td>
<td>-0.171</td>
<td>0.172</td>
</tr>
<tr>
<td>CER</td>
<td>(47.88)</td>
<td>(54.22)</td>
<td>(56.78)</td>
<td>(46.74)</td>
<td></td>
</tr>
<tr>
<td>FICC</td>
<td>(64.83)</td>
<td>(69.72)</td>
<td>(70.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td></td>
<td>-0.325</td>
<td>-0.334</td>
<td>-0.343</td>
<td>-0.172</td>
</tr>
<tr>
<td>FFF</td>
<td>(64.83)</td>
<td>(69.72)</td>
<td>(70.40)</td>
<td>(46.74)</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Impact of Winsorization

This table reports cross-sectional and time-series variances of scaled measurement errors of each ERP when varying the extent of winsorization. The figure is labeled such that 1% indicates that the data is winsorized in the cross section or time series at the 1st and 99th percentiles. Scaled measurement-error variances are calculated for each unique firm/expected-return proxy pair as follows:

$$AvgSVar^{CS} = \frac{1}{T} \sum_{t} Var_{t}(\hat{e}_{i,t}) - 2Cov_{t}(r_{i,t+1}, \hat{e}_{i,t})$$

where $Var_{t}(\hat{e}_{i,t})$ is a given ERP’s cross-sectional variance at time $t$ and $Cov_{t}(r_{i,t+1}, \hat{e}_{i,t})$ is the cross-sectional covariance between firms’ ERPs at time $t$ and their realized returns in period $t+1$. Time-series measurement-error variances are defined analogously when summarized over all firms in a given calendar month, spanning 455 months from 1977 through 2014 as follows:

$$AvgSVar^{TS} = \frac{1}{N} \sum_{i} SVar_{i}(\omega_{i,t})$$

where $SVar_{i}(\omega_{i,t})$ is the scaled measurement-error variance firm $i$. Firm-specific scaled measurement-error variances are calculated based on a sample of 12,397 unique firms that meet our data requirements. Cross-sectional tests correspond to summary statistics for the error variance from each model, using a sample of 12,397 unique firms with a minimum of 20 (not necessarily consecutive) months of data during our 1977-2014 sample period. ICC is a composite measure of implied cost of capital by taking the equal-weighted average of five commonly proxies based on (1) Gebhardt et al. (2001), (2) Claus and Thomas (2001), (3) Ohlson and Juettner-Nauroth (2005), (4) Gordon and Gordon (1997), and (5) Easton and Monahan (2005); CER is a characteristic-based expected-return proxy derived from a historical regression of realized returns on a firm’s size, book-to-market, and return momentum where historically estimated ten-year average Fama-MacBeth coefficients are applied to current firm characteristics; FICC is analogously defined from a regression of ICC on a firm’s size, book-to-market, and return momentum; CAPM is the firm’s expected-return derived from the Capital Asset Pricing Model; and FFF is the four-factor model using the market, small-minus-big (SMB), high-minus-low (HML), and up-minus-down (UMD) factors. A full description of each proxy appears in Section IV.

| Winsorization (%) | ICC | CER | FICC | CAPM | FFF | ICC | CER | FICC | CAPM | FFF |
|-------------------|-----|-----|------|------|-----|-----|-----|------|------|-----|-----|
| 0                 | 0.0943 | -0.0015 | 0.0018 | 0.1408 | 0.3781 | 0.0531 | 0.0001 | -0.0124 | 0.3338 | 0.7267 |
| 1                 | 0.0546 | -0.0019 | 0.0015 | 0.1117 | 0.2891 | 0.0248 | -0.0011 | -0.0121 | 0.2613 | 0.5404 |
| 2                 | 0.0497 | -0.0020 | 0.0013 | 0.0974 | 0.2534 | 0.0212 | -0.0012 | -0.0118 | 0.2252 | 0.4641 |
| 3                 | 0.0464 | -0.0020 | 0.0011 | 0.0872 | 0.2278 | 0.0187 | -0.0013 | -0.0116 | 0.1989 | 0.4102 |
| 4                 | 0.0418 | -0.0020 | 0.0010 | 0.0792 | 0.2073 | 0.0151 | -0.0013 | -0.0113 | 0.1778 | 0.3675 |
| 5                 | 0.0320 | -0.0021 | 0.0008 | 0.0726 | 0.1903 | 0.0077 | -0.0013 | -0.0111 | 0.1601 | 0.3324 |
| 6                 | 0.0213 | -0.0021 | 0.0007 | 0.0670 | 0.1754 | 0.0003 | -0.0014 | -0.0108 | 0.1449 | 0.3022 |
| 7                 | 0.0187 | -0.0021 | 0.0005 | 0.0622 | 0.1622 | -0.0013 | -0.0014 | -0.0106 | 0.1317 | 0.2758 |
| 8                 | 0.0163 | -0.0021 | 0.0004 | 0.0579 | 0.1504 | -0.0027 | -0.0014 | -0.0103 | 0.1199 | 0.2522 |
| 9                 | 0.0135 | -0.0021 | 0.0002 | 0.0540 | 0.1399 | -0.0042 | -0.0014 | -0.0101 | 0.1092 | 0.2313 |
| 10                | 0.0111 | -0.0021 | 0.0001 | 0.0505 | 0.1302 | -0.0052 | -0.0014 | -0.0099 | 0.0997 | 0.2123 |
| 15                | 0.0063 | -0.0021 | -0.0004 | 0.0368 | 0.0922 | -0.0061 | -0.0014 | -0.0087 | 0.0631 | 0.1396 |
| 20                | 0.0040 | -0.0020 | -0.0007 | 0.0269 | 0.0651 | -0.0057 | -0.0013 | -0.0076 | 0.0382 | 0.0900 |