A Market-Based Cost of Capital*

Niels Joachim Gormsen Kilian Huber Theis Ingerslev Jensen

July 2026

Please click here for the latest version

Abstract

A firm's cost of capital is a central object in economics. It is often defined based on the expected long-run returns on the firm's assets in financial markets, but these expected returns are unobserved and there exists no method that delivers unbiased estimates out of sample. We provide a method that uses machine learning to predict future cash flows of firms and then derives long-run expected returns from observed market prices and the present value identity. We show that this method works better than predicting returns directly if the model determining cash flows is more stable than that determining expected returns, a condition that holds empirically. The resulting estimates are unbiased out of sample and have greater predictive power than existing alternatives. We use them to construct a firm-level "market-based cost of capital" (MCC). Consistent with the standard prediction that firms maximize their market value by using expected returns as their cost of capital, we find that firms whose perceived cost of capital is closer to the MCC have higher market values.

*We are grateful for helpful comments from Paul Fontanier, Stefano Giglio, Bryan Kelly, Eben Lazarus, Lasse Pedersen, Alp Simsek, and David Thesmar as well as many seminar and conference participants. We acknowledge support from the Danish National Research Foundation (grant no. DNRF167 and DNRF 193). Gormsen is at Copenhagen Business School and The University of Chicago; www.voices.uchicago.edu/gormsen. Huber is at The University of Chicago; www.kilianhuber.org. Jensen is at Yale University; www.tijensen.com.

1 Introduction

The cost of capital is often defined as the expected long-run return on a firm's debt and equity in financial markets. This cost of capital plays a key role in economic analyses. It is used by researchers to discount long-run cash flows and discipline asset pricing, macro, and corporate finance models. It is also used by firms in valuation and investment decisions and by investors in fundamental analysis and portfolio allocation decisions.

But expected returns are unobserved, so there is no readily available estimate of the cost of capital. Common methods for calculating the expected returns underlying the cost of capital include the Capital Asset Pricing Model (CAPM), the implied cost of capital (ICC) based on analyst forecasts, and factor models aimed at predicting short-run returns. These methods deliver biased estimates of long-run expected returns, and hence of the cost of capital. We set out to provide unbiased out-of-sample estimates of long-run expected stock returns that can be used to calculate the cost of capital.

Our method estimates long-run expected returns using expected cash flows. The method starts from the present value identity, which says that prices are equal to the expected value of future cash flows, discounted by the expected returns. To solve for the expected returns, we first use machine learning to forecast firms' expected cash flows at different horizons. We then use the observed equity market price to solve for the expected equity returns. We show conceptually that this indirect approach outperforms the more traditional method of predicting expected returns directly, as long as the model determining cash flows is more stable than the model determining expected returns. This condition holds in the data.

The resulting estimate of long-run expected returns is unbiased out-of-sample, in the sense that it predicts firm-level returns in financial markets without predictable errors. This finding contrasts with existing methods based on the CAPM, multi-factor models, the implied cost of capital (ICC), and direct machine learning of returns, all of which typically deliver biased estimates of long-run expected returns out-of-sample. Our estimates are also powerful in that they explain a relatively large fraction of realized returns.

We use the new estimates of expected returns to calculate a firm-level cost of capital, which we call the "market-based cost of capital" (MCC). In theory, firms can maximize their current market value by using the MCC as their cost of capital. Indeed, we find that firm equity market value is greater when a firm's perceived cost of capital (which it actually uses in its internal decisions) is closer to its MCC. Intuitively, firms with "too low" a perceived cost of capital undertake value-destroying investments, which lowers their market value; firms

with "too high" a cost of capital pass on investments that would have increased the value of the firm. In addition, the equity market responds more positively to a firm announcing new investments when the firm's perceived cost of capital is "too high" relative to the MCC, consistent with this firm investing too little. The findings support the classical intuition that the cost of capital is a crucial input into firms' decisions and that firms using an unbiased cost of capital, like the MCC, achieve higher market values.

We begin the paper by describing our approach to estimating long-run expected equity returns. Rather than predicting returns directly, as is commonly done in the literature, we predict firms' cash flows and back out expected returns indirectly from observed prices using the present value identity. We show that this indirect approach is advantageous because it has a smaller out-of-sample mean squared error than estimating returns directly when the dynamics of expected cash flows are more stable than the dynamics of expected returns. The intuition is that the indirect method only needs to predict cash flows and does not need to account for volatile expected return dynamics, which are instead inferred through the market price. Empirically, we find that models for cash flows are indeed substantially more stable than the models for expected returns.

We implement the indirect approach using a machine learning method. Specifically, we use a gradient-boosted decision tree ensemble to forecast cash flows at different future horizons. The method builds forecasts sequentially, with each new tree designed to improve on the errors made by the existing ensemble. Each cash flow forecast relies only on information available in the given year, so the forecasts are made out-of-sample. In addition, we avoid overfitting by checking the results of each iteration against an external validation dataset. As soon as the forecast error in the validation data increases, we stop the prediction procedure, ensuring that we do not overfit to random noise in the training data.

The new cash flow forecasts are available for roughly 16,000 firms listed in 93 countries. We forecast individual horizons up to ten years ahead and thereafter apply a terminal growth rate assumption based on country-level forecasts. The forecasts begin in 1984. We verify that, on average, the cash flow forecasts are strong predictors of future realizations. The forecasts explain around 40% of future variation at two-year horizons and around 15% at ten years. In comparison, existing cash flow forecasts relying on financial analysts explain around 10% at two years and close to 0% at ten years.

Using these cash flows, we can invert the present-value identity to get a firm-level estimate of long-run expected equity returns, which we call the market-based cost of equity (MCE). The MCE is an unbiased estimate of long-run expected returns out of sample. When we

regress realized ten-year equity returns on the MCE, the slope coefficient is 0.93 in the U.S. sample, 0.96 globally, and 1.01 outside the U.S., none of them statistically distinguishable from 1. The intercept is small and statistically not different from zero, so the MCE is an unbiased estimate of expected returns. The same regression produces a slope coefficient near zero for the CAPM and well below one for analyst-based ICC.

We also compare the MCE to a direct machine-learning forecast of returns that uses the same predictors and the same algorithm. The direct forecast is biased and dispersed, and its predictions vary considerably across model training vintages, consistent with the theoretical result that volatile asset pricing dynamics make the direct approach less stable. The MCE leads to a higher out-of-sample R^2 and smaller prediction errors than the alternatives. In fact, most of the other alternatives produce negative out-of-sample R^2, when compared with a naive estimator based on past returns, whereas the MCE method produces reliable out-of-sample predictability.

One use of the MCE is to calculate a firm's overall cost of capital. A firm's cost of capital can be interpreted as the discount rate used by equity markets to value the firm's expected cash flows. The cost of capital influences the allocation of capital in essentially all models in economics and finance. According to standard models and business school teaching, it is also the rate at which firms should evaluate new investments if they want to maximize their market value. Surveys indicate that the vast majority of large firms calculate an internal estimate of their firm-level cost of capital and use it as the starting point for their investment decisions. To calculate the firm-level cost of capital, we combine our newly estimated MCE with standard estimates of leverage, taxes, and the cost of debt, following the textbook formula for the weighted average cost of capital (WACC). We term the resulting estimate the market-based cost of capital (MCC).

We construct a firm-level "cost of capital deviation" by subtracting the MCC from the perceived cost of capital that is actually used internally by firms. Data on firm perceptions are collected by Gormsen and Huber (2025) who find that firms' long-run investment decisions are indeed associated with the observed perceived cost of capital.

A simple model generates two predictions about the relation between firm market value and the cost of capital deviation. First, there should be an inverted U relation, so that firms with a smaller deviation achieve higher market values for a given level of book equity. Intuitively, a firm whose perceived cost of capital is too high relative to the MCC is, on average, more conservative than the equity market and passes on investments the equity market would have rewarded. In contrast, a firm whose perceived cost of capital is too low

is more lenient and undertakes investments the equity market does not reward. In the data, we find exactly this relation. The coefficients suggest that a 1 percentage point decrease in the deviation raises equity market value by 7 to 9 log points.

Second, the model predicts that the equity market reacts more positively to new investments when the cost of capital deviation is larger. A positive deviation implies that the perceived cost of capital is "too high" relative to the MCC, so the firm invests too little relative to the market's view. We measure when firms announced large new investments using a database of firm-level news. The sample contains unambiguous announcements of concrete projects undertaken by the firm itself. In the data, the equity market returns in 5-day windows around the announcements are significantly more positive for firms with a larger cost of capital deviation, consistent with the model.

Our paper contributes to a growing literature on measuring and understanding expectations. This literature shows that agents do not have fully rational expectations about economic and financial variables (Bianchi et al. 2022; Bordalo et al. 2020, 2024; Coibion and Gorodnichenko 2015; Greenwood and Shleifer 2014), emphasizing the importance in measuring and modeling expectations directly. In this context, it is useful to have a rational benchmark for the expectations. Our measure can serve as a rational benchmark for expected stock returns, and our analysis of the perceived cost of capital illustrates how such a benchmark can be helpful for understanding the economic consequences of agents' biased perceptions (see also Bianchi et al. 2025).

Our paper complements recent work by Hommel et al. (2025) who evaluate how well different estimates of the cost of capital predict firm equity market values. Hommel et al. embrace the challenge of estimating a cost of capital, or discount rate, without using the price as an input. This captures an important feature of capital budgeting decisions, namely, that many projects do not have traded prices attached to them. Our approach differs in that we use the price as a key input into the market-based cost of capital, which is restrictive but helpful for improving the predictive performance of our measure. The two approaches are complementary, and they both show that the CAPM performs poorly, whereas methods based on machine learning are relatively more accurate.

Our paper also speaks to a recent literature on long-run expected returns. This literature disagrees on the extent to which long-run expected returns are predictable. Keloharju et al. (2021) argue that predictability in returns for the average characteristic in the literature is concentrated in the short horizon, and that the average characteristic does not predict returns after a few years. We do find that the MCE predicts returns more strongly in the

first few years than in the more distant years, but the MCE continues to predict returns even at the long horizon. This finding reflects that the MCE captures long-run expected returns better than the factors proposed in the existing literature (which are often designed to predict short-run returns). Recent studies by Cho and Polk (2024) and Van Binsbergen et al. (2023a) also find evidence of predictability in long-run returns. More generally, our paper connects to the literature that aims to explain returns using machine learning techniques (see Kelly and Xiu (2023) for a review and Bianchi et al. (2025) for a recent example).

Previous surveys reveal qualitative evidence about the eclectic methods used by firms to set their perceived cost of capital (e.g., Graham and Harvey 2001; Jacobs and Shivdasani 2012; Mukhlynina and Nyborg 2016; Jagannathan et al. 2016; Graham 2022). Gormsen and Huber (2025) find that the average firm does not use an unbiased perceived cost of capital, which can lead to capital misallocation, but they do not provide an unbiased cost of capital and do not estimate the relation between firm market values and cost of capital deviations.

2 Two Approaches for Estimating Long-Run Returns

We start by studying the advantage of calculating expected returns by combining prices and expected cash flows. We focus on a setting with a single firm and a 1-period forecasting horizon.

The traditional approach for estimating expected return revolves around forecasting regressions for returns. Assume expected returns depend on K observed state variables f,

$$\mathbb{E}_t(r_{t+1}) = \alpha^{\text{ER}} + \beta_t^{\text{ER}\prime} f_t, \tag{1}$$

where β_t^{ER} is a vector of coefficients on the K state variables. State variables are i.i.d. over time with unit variance and zero mean, and are mutually uncorrelated. To predict expected returns using this "ER" methodology, one regresses realized returns on the state variables, and calculates predicted returns for time $t+1$ as

$$\widehat{\mathbb{E}}_t^{\text{ER}}(r_{t+1}) = \bar{\alpha}^{\text{ER}} + \bar{\beta}^{\text{ER}\prime} f_t, \tag{2}$$

where $\bar{\alpha}^{\text{ER}}$ and $\bar{\beta}^{\text{ER}}$ are the population coefficients of the regression (i.e., values that OLS estimates converge towards as the training window grows). This ER approach has been widely used in return prediction.

An alternative to this approach, which is relevant when return horizons are long, is to

estimate expected cash flows and back out expected returns from prices. For simplicity, consider equity claims to cash flows that are paid out in one period (where one period may be long),

$$p_t = \frac{\mathbb{E}_t(CF_{t+1})}{1 + \mathbb{E}_t(r_{t+1})},$$

where $\mathbb{E}_t(CF_{t+1})$ is the time-t expected value of the free cash flow at time $t+1$. Assuming cash flows follow the model

$$\mathbb{E}_t(CF_{t+1}) = \alpha^{\text{CF}} + \beta_t^{\text{CF}\prime} f_t, \tag{3}$$

we can calculate expected returns as

$$1 + \mathbb{E}_t(r_{t+1}) = \frac{\mathbb{E}_t(CF_{t+1})}{p_t} = \frac{\alpha^{\text{CF}} + \beta_t^{\text{CF}\prime} f_t}{p_t}. \tag{4}$$

We call this second approach the "CF approach." To predict expected returns using this CF methodology, one regresses realized cash flows on the state variables, and calculates predicted returns for time $t+1$ as

$$\widehat{\mathbb{E}}_t^{\text{CF}}(r_{t+1}) = \frac{\bar{\alpha}^{\text{CF}} + \bar{\beta}^{\text{CF}\prime} f_t}{p_t} - 1, \tag{5}$$

where $\bar{\alpha}^{\text{CF}}$ and $\bar{\beta}^{\text{CF}}$ are again the population coefficients of the regression.

We assume that the asset pricing dynamics vary over time, while the cash flow dynamics are constant:

$$\beta_t^{\text{ER}} = \mu + \rho(\beta_{t-1}^{\text{ER}} - \mu) + \eta_t, \quad \eta_t \overset{iid}{\sim} (0, \sigma_\eta^2 I_K), \qquad \beta_t^{\text{CF}} = \beta^{\text{CF}}, \tag{6}$$

where μ is a K-dimensional constant vector, $\sigma_\eta^2 > 0$ is the innovation variance, I_K is the K-dimensional identity matrix, and $|\rho| < 1$, such that the processes are stationary. We assume that the loadings and factor realizations are independent.

The assumed dynamics capture an economic environment in which the fundamental cash flows of firms are relatively stable over time while asset pricing dynamics are volatile (as in, e.g., Campbell and Cochrane 1999). The volatile asset pricing dynamics can, for instance, arise from time variation in fundamental risk, behavioral demand, or frictions. Mechanically, the cash flow model can remain stable while the expected return model varies because the price absorbs the expected-return variation.

The volatile asset pricing dynamics result in forecast errors when using the ER method, but because the CF method only predicts the stable cash flows and absorbs the volatile asset pricing dynamics from the price, its estimates are unbiased. We summarize this finding in the proposition below, in which we measure accuracy by the mean squared error of the predicted return (all proofs are in Appendix A):

$$\text{MSE}^m \equiv \mathbb{E}\left[\left(\widehat{\mathbb{E}}_t^m(r_{t+1}) - r_{t+1}\right)^2\right], \qquad m \in \{\text{ER}, \text{CF}\}. \tag{7}$$

Proposition 1. *The MSE gain of the CF method over the ER method in the full population is*

$$\text{MSE}^{ER} - \text{MSE}^{CF} = K \text{ var}\left(\beta_t^{ER}\right). \tag{8}$$

The gain is positive and increases with the number of state variables and the variance of the return loadings.

Proposition 1 shows that the CF method produces smaller forecast errors when asset pricing dynamics are volatile and cash flow dynamics are stable. If we generalize the analysis to allow for volatility in cash flow dynamics, the MSE gain exists whenever the loadings in the ER regression are more volatile than those in the CF regression, a condition we verify empirically. Moreover, the gain is larger when there are more state variables, as each additional variable adds another fluctuating loading that the ER method approximates with a fixed coefficient.

Cash flows are substantially more predictable than returns, but this feature does not, in itself, lead to an advantage for the CF method. Both the ER and CF methods aim to predict realized returns, and in return space, both methods have the same R^2 when the correct parameter estimates are used. The advantage of the CF method does not come from cash flows being more predictable than returns, but from the model underlying the cash flow prediction being more stable.

Proposition 1 evaluates an ER method that does not attempt to account for the variation in return loadings. In practice, one could run the return regression over a shorter window, which would bring the estimated coefficient closer to the current one. Doing so would reduce the bias from stale coefficients but raise estimation variance, so there is a bias-variance trade-off. The CF approach does not face this trade-off. Another option is to proxy the variation directly. If one could perfectly model–or proxy for–the asset pricing dynamics, one could solve the ER approach. In our model, the price is such a proxy: interacting the (inverse of the) price with the state variables makes the ER coefficients constant and the ER approach

work as well as the CF approach. Whether one can obtain such an ER model is an empirical question that we turn to in the upcoming sections.

3 A Cash Flow-Based Estimate of Long-Run Expected Stock Returns

In this section, we lay out our cash-flow based estimator for long-run expected stock returns. Since the main application of these expected returns will be to calculate the cost of equity for firms, we will refer to these expected returns as the cost of equity. We refer to our new measure as the market-based cost of equity (MCE).

The starting point for our new method is the basic accounting expression for a firm's equity market value p_t:

$$p_t = \sum_{h=1}^{\infty} \frac{\mathbb{E}_t[\text{fcfe}_{t+h}]}{(1 + r_t^{\text{equity}})^h}, \tag{9}$$

where $\mathbb{E}_t[\text{fcfe}_{t+h}]$ is a forecast for the free cash flow accruing to equity holders in period $t + h$. Firm market value p_t is directly observed, but we do not easily observe cash flow forecasts that capture substantial variation. Our method focuses on the estimation of cash flow forecasts, which can then be used to derive the cost of equity.

3.1 Forecasting Cash Flows

We define an explicit forecasting horizon of H periods and a terminal value period to represent the infinite sum of cash flows in equation (9). We impose the terminal value condition that all firms' cash flows grow at the same constant rate g_t:

$$p_t = \sum_{h=1}^{H} \frac{\mathbb{E}_t[\text{fcfe}_{t+h}]}{(1 + r_t^{\text{equity}})^h} + \frac{\mathbb{E}_t[\text{fcfe}_{t+H}](1 + g_t)}{r_t^{\text{equity}} - g_t} \frac{1}{(1 + r_t^{\text{equity}})^H}. \tag{10}$$

In our estimation, we set $H = 10$. The terminal value condition can be motivated as an economic prior that competition eliminates above-average growth in the long run. Consistent with this view, the terminal growth rates used in valuation practice do not vary much across firms (Decaire and Guenzel 2023).

We forecast cash flows in every period t for all horizons $h \leq 10$. We employ a machine-learning technique based on decision tree learning. Decision tree learning essentially splits

firms into groups based on potential predictors of cash flows, and then estimates which groups are associated with higher future cash flows. However, single decision trees exhibit high variance and poor performance. We therefore build on the LightGBM model of Ke et al. (2017), which enables efficient and fast construction of an ensemble of decision trees built in a sequential way, in which each new decision tree is trained to predict the errors on the existing ensemble.[1] All forecasts are out-of-sample, meaning that the cash flow forecasts are based on information that was available at time t. We elaborate on our implementation details in Appendix B.

With the cash flow forecasts in hand, we can solve for r_t^{equity} in equation (10). This is our estimate of the "market-based cost of equity" (MCE). This MCE equates the discounted sum of expected cash flows with the observed market value. It is therefore the expected return on the firm's equity in financial markets.

Our method is inspired by seminal work calculating an "implied cost of capital" (e.g., Gebhardt et al. 2001; Claus and Thomas 2001; Easton 2004; Ohlson and Juettner-Nauroth 2005), which relies on a net present value formulation and analyst forecasts to derive the cost of equity. Since analyst forecasts are biased, the implied cost of capital is also biased and – as documented in Section 4 – produces negative out-of-sample R-squared.

3.2 Data

We construct a database for around 16,000 firms listed in 93 countries. We start the samples in 1950 for U.S. stocks and 1985 for non-U.S. stocks to ensure that accounting data are available. Data on market values and stock returns originally come from CRSP for U.S.-listed firms and from Compustat for non-U.S.-listed firms. The accounting data are from Compustat. We rely on the database provided by Jensen et al. (2023) through WRDS and GitHub. Finally, we add analyst forecasts from I/B/E/S, and we always use the median forecasts from the consensus file.

We exclude financial firms (i.e., groups 45–48 in the Fama and French 1997 classification) because free cash flows of financial firms are conceptually different from free cash flows of non-financial firms.[2] We also require firms to have a non-missing and positive accounting

[1] The model is a gradient-boosted decision tree ensemble (GBDT). Such GBDTs perform well on tabular data, outperforming deep learning methods across a wide variety of tabular benchmarks (Shwartz-Ziv and Armon 2022).

[2] We aim to estimate firms' unlevered free cash flows by separating cash flows due to operations from cash flows due to financing. However, this separation is difficult for financial firms because their business model is all about financing.

value of assets and shares outstanding.

3.3 Estimating the cost of equity (MCE)

Our method requires cash flow forecasts to determine the cost of equity. We forecast real cash flows in U.S. dollars. The forecasts are made in every month t for up to ten years in the future. The terminal growth rate for cash flows is the country-specific forward real GDP growth rate between years 4 and 5 from the IMF World Economic Outlook.[3]

We employ two measures of cash flows to equity holders: free cash flows to equity and residual income. Details on the data construction are in Appendix B.1.

When using the free cash flow to equity, we directly rely on (9). The free cash flow to equity is:

$$\text{fcfe}_t = e_t + (\text{d\&a}_t - \text{capx}_t) - \Delta\text{nwc}_t + \Delta\text{debt}_t, \tag{11}$$

where e_t is net income, $\text{d\&a}_t$ is depreciation and amortization, capx_t is capital expenditures, Δnwc_t is the change in net operating working capital, and Δdebt_t is net debt issuance. We forecast free cash flows to equity in $t+1,\ldots,t+10$, and solve for r_t^{equity} numerically.

When using the residual income measure (Edwards and Bell 1961; Ohlson 1995), we rely on a slightly modified net present value formulation:

$$p_t = b_t + \sum_{h=1}^{10} \frac{\mathbb{E}_t[\text{ri}_{t+h}]}{(1+r_t^{\text{equity}})^h} + \frac{\mathbb{E}_t[\text{ri}_{t+10}](1+g_t)}{r_t^{\text{equity}} - g_t} \frac{1}{(1+r_t^{\text{equity}})^{10}}, \tag{12}$$

where b_t is the firm's book equity and ri_t is residual income:

$$\text{ri}_t = e_t - r_t^{\text{equity}} b_{t-1}. \tag{13}$$

To estimate the cost of equity from this model, we forecast net income in $t+1,\ldots,t+10$, and infer the future book equity as:

$$b_t = b_{t-1} + e_t - d_t, \tag{14}$$

where d_t is the firm's dividends. We estimate dividends by multiplying our earnings forecast by the firm's most recent payout ratio (d_t/e_t).[4] The method for inferring the future book

[3] We use a forward rate rather than a cumulative rate because the terminal growth rate should reflect steady-state growth, uncontaminated by near-term cyclical effects.

[4] Our method for inferring the future book equity can result in negative book equity, in which case the

equity is inspired by Gebhardt et al. (2001). We then solve for r_t^{equity} numerically.

Our MCE measure is the average of the two cost of equity estimates (if only one is available, we use that estimate directly). Because the cash flow forecasts are in real terms, the implied discount rate is a real cost of equity. We convert it to nominal by adding the expected five-year forward inflation rate from the IMF World Economic Outlook. Further, we remove estimates where the implied real cost of equity exceeds 10%, because such high values are implausible and indicate that our model assumptions are invalid for the specific firm.[5] Further details are available in Appendix C.[6]

The explanatory variables in the cash flow prediction models are 153 characteristics from Jensen et al. (2023); 79 characteristics we derive from their data set related to current cash flow levels and changes; 12 characteristics based on analyst forecasts from IBES related to the expected cash flows for the firm over the next three years and the analyst long-term growth rate forecast, as well as the corresponding median within the firm's industry and country, respectively; 45 industry dummies based on the classification from Fama and French (1997), expanded to also include the computer software group, but without firms from the four financial groups; and 93 country dummies. In total, the models are fit on 244 continuous and 138 dummy characteristics. Further details about the input features are in Appendix B.2.[7]

All predictions are out-of-sample (OOS). The model is trained on an expanding training sample, with the last 10 years used as the validation sample. We select hyperparameters by evaluating 20 candidate sets, chosen to maximize entropy over a predefined search space spanning five specific hyperparameters, and select the ones with the lowest mean squared error in the validation sample. Further details on the model and hyperparameters are in Appendix B.3.

When the hyperparameters have been found, we re-estimate the model on the combined training and validation data. The first model is trained on a data set ending on December

residual income from (13) is not well defined. When projected book equity becomes negative, we set it to zero, which implies zero residual income from that point forward.

[5]For example, our valuation models assume an infinite firm lifetime, which is a poor assumption for firms that the market expects to go out of business soon.

[6]The Appendix contains additional information about the MCE. Specifically, Figure A11 shows the coverage, and Figure A12 show how the cost of equity vary over time.

[7]Some of the explanatory variables—especially those that require analyst data—are missing for a large fraction of our sample. Removing observations with missing explanatory variables would thus drastically reduce our sample and tilt it towards relatively large stocks with analyst coverage. Fortunately, LightGBM handles missing values natively by learning how missing values map to the dependent variable. As a result, we keep observations with missing values and delegate their handling to the learning algorithm.

31st, 1983, and we re-estimate the model parameters and hyperparameters every year. All forecasts are made OOS on a monthly basis, with the first forecast made on January 31st, 1984, and the last on December 31st, 2023. Further details on the sample split are in Appendix B.4.

4 Predicting Returns and Cash Flows

4.1 Predicting Returns

We start the analysis by studying the extent to which the MCE predicts realized stock returns. Since the MCE is meant to predict returns over a long horizon, we estimate the association between the MCE and future ten-year realized equity returns.

$$r^{\text{realized},i}_{t\to t+120} = \alpha_t + b \times \left[\left(1+\text{MCE}^i_t\right)^{10} - 1\right] + \varepsilon^i_{t+h}, \tag{15}$$

where $r^i_{t\to t+120}$ is the buy-and-hold realized return from t to $t+120$ months (ten years), α_t is a time fixed effect, and MCE^i_t is the market-based cost of equity converted to a ten-year horizon.

Implementation Implementing the long-horizon regression in (15) raises two challenges. The first is the unit of returns: the MCE is annualized and must be matched to the horizon of the dependent variable. Equation (15) does this by compounding the MCE to the return horizon. The alternative is to annualize the realized return on the left-hand side, but doing so biases the estimate. Ten-year realized returns are strongly right-skewed, and annualizing is a concave transformation that downweights extreme observations that are important for the mean (i.e., $E_t\left[R^{1/h}_{t\to t+h}\right] < \left(E_t[R_{t\to t+h}]\right)^{1/h}$ by Jensen's inequality).

The second challenge pertains to delisting. Over a ten-year horizon, around 50% of firms in CRSP delist (see Appendix E). The most common source of delisting is that the firm is bought or merged with another company, but another source is that firms fail. Calculating realized returns only for firms that survive for 10 years would lead to a survivorship bias in our analysis. To address this issue, we include all delisted firms in the sample, even if they delist during a ten-year period. We follow multiple strategies to impute returns, as discussed in Appendix E, and all lead to largely similar results. Our main approach is to impute returns after delisting using the median return for firms in the same country and month. The imputation approach has no impact on a firm that goes bankrupt and has a delisting

return of -100%, but ensures that firms that, say, get acquired continue to accumulate returns.

Main Regressions Table 1, Panel A, reports regressions of ten-year realized returns on the MCE, controlling for month fixed effects. The point estimates in the U.S. and the global sample are close to 1 and statistically indistinguishable from 1. The estimates are also significantly different from 0. Panel B reports the average error, that is, the ten-year realized minus predicted return. In the U.S. sample, the average error is a significant 25% (approximately 2% annualized). This result suggests that realized returns in the out-of-sample period exceeded investors' ex ante expectations in the U.S., consistent with Binsbergen et al. (2025). In the global and non-U.S. samples, however, the average errors are closer to zero and statistically insignificant, indicating no systematic errors overall.

For our main regression, the global regression in column (ii), the slope coefficient is 0.96, and the average forecast error is statistically insignificant. The MCE is thus an unbiased out-of-sample estimate of 10-year expected returns. The unbiased relation between MCE and realized returns is illustrated in Figure 1.

We calculate standard errors based on Driscoll-Kraay with a 120-month bandwidth. We find similar results using a Fama-MacBeth methodology with Newey-West standard errors (Table A8).

Table 1, Panel B, also shows the out-of-sample R^2 relative to three benchmark models, defined as:

$$R^2_{\text{bench}} = 1 - \frac{\sum(r_i - \hat{r}_i)^2}{\sum(r_i - \hat{r}_{i,\text{bench}})^2}, \tag{16}$$

where r_i is the realized return, $\hat{r}_i$ is the out-of-sample MCE prediction, and $\hat{r}_{i,\text{bench}}$ is the out-of-sample benchmark prediction. The more standard in-sample R^2 uses the in-sample mean instead of the benchmark prediction, but the in-sample mean is not available in real time and is therefore an inappropriate benchmark for our out-of-sample predictions. Instead, we consider three alternative cost of equity estimates that could have been used in real time. MKT is the ten-year risk-free rate at the estimation date plus a rolling estimate of the historical ten-year local market excess return based on data through this date. CAPM is the 10-year risk-free rate plus the same rolling estimate of the local market risk premium, multiplied by the stock's market beta, using the last 60 months of data. OWN extrapolates the firm's own past return, by compounding the firm's trailing geometric-mean monthly return (using all data through the estimation date, and requiring at least 60 months of history)

	10-year realized return		
	U.S.	Global	Global ex-U.S.
	(i)	(ii)	(iii)
Panel A: Regression estimates			
Cost of equity	0.93***	0.96***	1.01***
	(0.08)	(0.18)	(0.23)
Observations	1,106,464	2,610,489	1,504,025
Panel B: Predictive accuracy ($\alpha = 0, \beta = 1$)			
Avg. error	0.25	0.14	0.06
	(0.12)	(0.10)	(0.15)
OOS R^2_{CAPM} (%)	2.1	3.8	5.3
OOS R^2_{MKT} (%)	0.4	2.1	4.1
OOS R^2_{OWN} (%)	9.0	12.1	14.7

The table reports regressions of 10-year realized buy-and-hold returns on the current 10-year cost of equity, estimated separately for U.S., global, and global ex-U.S. samples. All regressions include month fixed effects. The average pricing error is the mean difference between realized returns and the predicted cost of equity. OOS R^2_{CAPM}, R^2_{MKT}, and R^2_{OWN} report out-of-sample predictive R^2, computed as $1 - \sum(r_i - \hat{r}_i)^2 / \sum(r_i - \hat{r}_{i,\text{bench}})^2$, where the benchmark prediction is the CAPM-implied cost of equity, the country-level expanding-window average market return, or an extrapolation of the firm's own past return. All benchmarks use only information available at the time of prediction, and each R^2 is computed on the subsample for which its benchmark is available. We report Driscoll-Kraay standard errors using a bandwidth of 120 months in parentheses. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

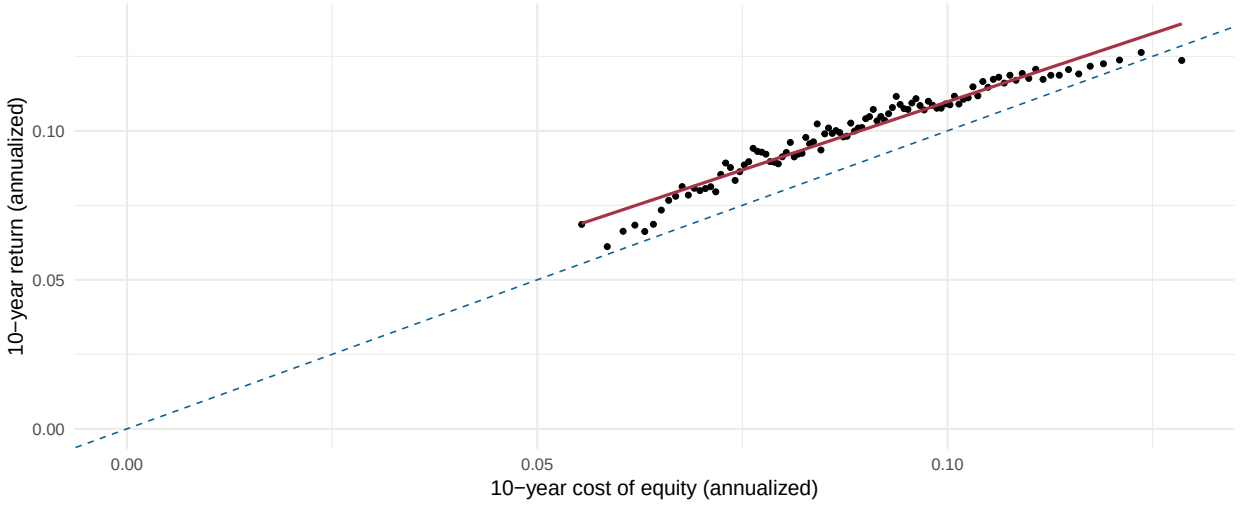
Table 1: Return predictability

into a ten-year return, shrinking it halfway toward the MKT estimate, and winsorizing it at the 5th and 95th percentiles each month. The out-of-sample R^2s are all positive, suggesting that an investor would achieve higher accuracy using the MCE than the historical market return, the CAPM, or a naive extrapolation of the firm's own past performance.

Figure 2 breaks the predictability of 10-year realized returns into the individual forward returns in each of the 120 months. The slope coefficients are above 1 in the early part of the term structure and decline in the horizon. This pattern is consistent with mean reversion in expected returns. Imagine a firm has an MCE of 10%, which is above the cross-sectional mean of around 8%. The MCE of 10% implies that, over the life of the firm, the expected returns are elevated by 2 percentage points above the mean. In the next few years, the

Figure 1
Average Realized Returns and the MCE

This figure shows the relationship between future 10-year realized returns and the ex ante MCE. Each month, stocks are sorted into 100 portfolios by their estimated cost of equity. For each portfolio, we compute the average MCE and the average realized buy-and-hold return over the next 10 years, and then average across months. Both axes are annualized. The solid red line shows the best linear fit, and the dashed blue line shows the 45-degree line. The sample includes all global stocks, 1983–2023.

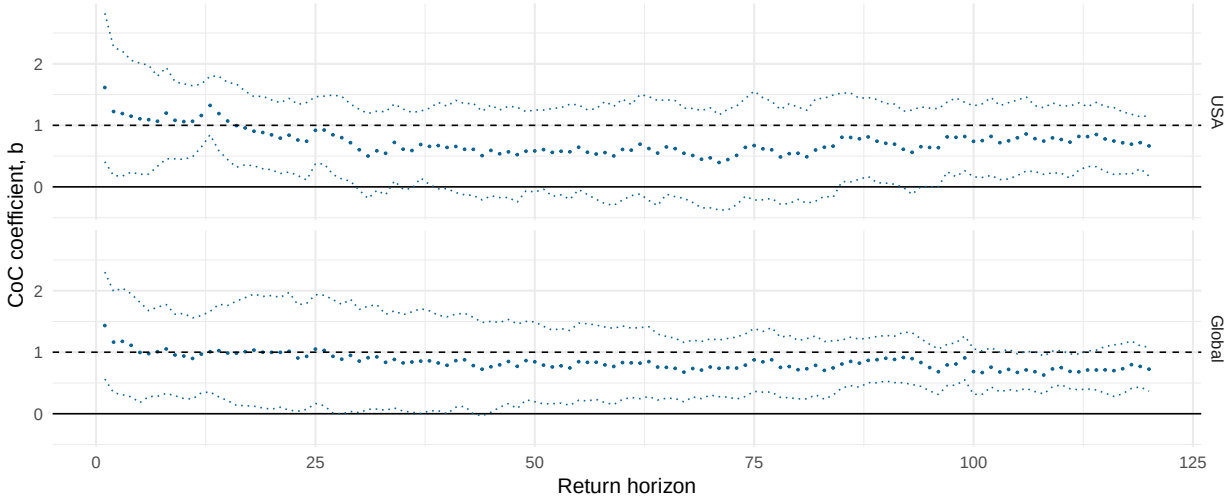


returns are elevated by more than 2 percentage points, and in the later years the returns are elevated by less than 2 percentage points, giving rise to the downward-sloping term structure for loadings of forward returns on the MCE.

The findings in Figure 2 run counter to the results in Keloharju et al. (2021). Keloharju et al. argue that established characteristics do not predict forward rates beyond year five and that, by extension, there is no variation in long-run expected returns. In Figure A8 we replicate their exact methodology, which is based on the CAPM alphas for Fama-French sorted on lagged characteristics. Using this approach, we replicate the decay that is also present in Figure 2, but we find that the alphas stabilize above zero. While Keloharju et al. (2021) find that forward alphas decay from 12% in year 0 to 2% from year 6, and 0% in year 10, the MCE has forward alphas that decay from 6% in year 0 to 3% in year 6 and 2% in year 10. The alphas remain statistically significant at the 10-percent level until year 7. The exact magnitudes do not compare, as we rely on the more conservative Fama and French methodology, whereas Keloharju et al. (2021) use decile sorts and a slightly longer sample. Overall, the results highlight the MCE's ability to predict long-run returns.

Figure 2
Forward return predictability

This figure reports the slope coefficient b from regressions of future excess stock returns on the MCE, $r_{t+h}^{\text{realized},i} = \alpha_t + b \times \left[(1 + \text{MCE}_t^i)^{1/12} - 1\right] + \varepsilon_{t+h}^i$, for return horizons $h = 1, 2, \ldots, 120$ months. All regressions include month fixed effects. The top row restricts the sample to U.S. firms; the bottom row uses the global sample. Dotted lines indicate 95% confidence intervals based on Driscoll-Kraay standard errors with a 120-month bandwidth. A coefficient of one (dashed line) implies that a one-percentage-point higher MCE maps one-for-one into higher realized returns.



4.2 MCE Versus Alternative Approaches

MLR Approach The most direct alternative to the MCE is to forecast returns instead of cash flows, which we call machine-learned returns (MLR). MLR uses the same prediction technology as the MCE—the same LightGBM model and the same predictors—but replaces the cash flow targets with the firm's realized ten-year return (details in Appendix D). The two methods thus differ only in what they forecast: the MCE forecasts cash flows and recovers the discount rate by inverting the present-value identity at the observed price, whereas the MLR forecasts the return directly. This is the empirical counterpart of the distinction in Section 2, where the MCE plays the role of the CF approach and the MLR that of the ER approach. Because the MCE lets the price absorb the dynamics in expected returns and only has to forecast the more stable cash flows, Proposition 1 implies that it should predict returns more accurately out of sample whenever expected-return dynamics are less stable than cash-flow dynamics.

The MLR methodology predicts returns, but unlike the MCE, it is a biased predictor of returns. As shown in Table 2 Panel A, the slope coefficient on realized 10-year returns on

	10-year realized return					
	LightGBM methods		Implied Cost of Capital methods			
	MCE	MLR	CT	GLS	PEG	OJ
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Panel A: Regression estimates						
Cost of equity	0.96***	0.50***	0.11	0.38**	0.10	0.07
	(0.18)	(0.16)	(0.13)	(0.17)	(0.10)	(0.10)
Observations	2,610,489	4,872,824	1,003,308	1,877,706	1,000,734	985,775
Panel B: Predictive accuracy ($\alpha = 0, \beta = 1$)						
Avg. error	0.14	-0.19	0.05	-0.00	-0.05	-0.56
	(0.10)	(0.28)	(0.08)	(0.08)	(0.10)	(0.08)
OOS R^2_{MKT} (%)	2.1	-0.5	-0.1	1.1	-0.5	-2.6
OOS R^2_{CAPM} (%)	3.8	0.9	1.4	2.9	1.3	-1.3
OOS R^2_{OWN} (%)	12.1	6.6	14.0	14.1	16.3	12.9
Panel C: MSE relative to MCE						
ΔMSE	—	0.94	0.25	0.12	0.34	0.88
$p^{\text{bs}}_{\text{ind}}$		0.000	0.000	0.004	0.000	0.000
$p^{\text{bs}}_{\text{time}}$		0.004	0.093	0.109	0.048	0.010

The table reports regressions of 10-year realized buy-and-hold returns on the current 10-year cost of equity for various estimation methods. MCE is the market-based cost of equity from this paper. MLR predicts 10-year cumulative excess returns directly using LightGBM, and then adds the current risk-free rate. CT, GLS, PEG, and OJ are analyst-based ICC estimates following Claus and Thomas (2001), Gebhardt et al. (2001), Easton (2004), and Ohlson and Juettner-Nauroth (2005), respectively. All regressions include month fixed effects. The average pricing error is the mean difference between realized returns and the predicted cost of equity. OOS R^2_{CAPM}, R^2_{MKT}, and R^2_{OWN} report out-of-sample predictive R^2, computed as $1 - \sum(r_i - \hat{r}_i)^2 / \sum(r_i - \hat{r}_{i,\text{bench}})^2$, where the benchmark prediction is the CAPM-implied cost of equity, the country-level expanding-window average market return, or an extrapolation of the firm's own past return (defined in the text). All benchmarks use only information available at the time of prediction. Panel C reports $\Delta\text{MSE} = \text{MSE}_{\text{method}} - \text{MSE}_{\text{MCE}}$, computed on the id-month pairs where both MCE and the comparison method have non-missing predictions. Positive values indicate that MCE has lower MSE. $p^{\text{bs}}_{\text{ind}}$ and $p^{\text{bs}}_{\text{time}}$ are one-sided bootstrap p-values for the null that $\Delta\text{MSE} \leq 0$, obtained by resampling Fama-French 49 industries and 10 equal-sized time blocks, respectively, with replacement (10,000 iterations). We report Driscoll-Kraay standard errors using a 120-month window in parentheses. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Return benchmarks – unbalanced – fixed effects

the MLR is 0.5. It is statistically different from both 0 and 1, implying that it is a biased predictor, but that it does predict returns. Figure 3 plots the average realized returns to portfolios sorted on the MLR, along with the ex ante MLR of the portfolios. The fitted line through the portfolios is well below the 45-degree line, which is the benchmark for unbiased expectations.

The MLR also has lower predictive power for returns than the MCE method. As shown in Panel B of Table 2, the MLR has lower out-of-sample R^2 relative to all three of the benchmarks that we consider. Moreover, it has negative out-of-sample R^2 relative to the market-benchmark, illustrating the difficulty of conventional return prediction exercises in beating the historical average. MCE, in contrast, has positive out-of-sample R^2 relative to the market benchmark. Finally, Panel C shows that the MSE for the MLR is significantly larger than the MSE for the MCE approach.[8]

Stability of MCE relative to MLR Why is the cash flow-based MCE method a better predictor of returns than the return-based MLR method? Our model in Section 2 suggests that MCE is superior because the model is more stable.

We investigate the stability of the model and confirm that the MCE method indeed is more stable than the MLR. For each method, we fix the set of stocks and generate predictions from every training vintage—models trained on data ending in 1984, 1985, ..., 2023—and compute the correlation between the predictions of each pair of vintages. For instance, in 2020, we use the models trained in the past to make predictions in 2020 based on 2020 characteristics of firms and we compare that to the predictions made by the most up-to-date model. A method built on a stable model produces similar predictions for all versions of the model, so its pairwise correlations are close to one. An unstable model, on the other hand, produces different predictions from different models and therefore has lower correlations.

Figure 4 reports the distributions. The MCE predictions are far more correlated across model vintages than the MLR predictions. The lowest set of correlations for the MCE predictions is above 0.5 and the average is around 0.8. In contrast, the MLR approach has correlations as low as 0.1 and the average correlation is around 0.6.

The results in Figure 4 imply that one can use any MCE model vintage and get fairly

[8]Appendix Appendix F shows that the MCE's advantage over the MLR is not merely an artifact of the MLR slope being far from one. Even when each forecast is rescaled by its in-sample regression slope (removing any disadvantage from a poorly scaled forecast) the MCE delivers the better fit. Moreover, the optimal in-sample combination of the two forecasts places almost four times as much weight on the MCE as on the MLR.

Figure 3
Return Forecast Benchmarking

This figure shows the relationship between future 10-year realized returns and the current 10-year cost of equity for different estimation methods. MCE is the market-based cost of equity from this paper. MLR predicts 10-year cumulative excess returns directly using LightGBM, and then adds the current risk-free rate. CT, GLS, PEG, and OJ are analyst-based ICC estimates following Claus and Thomas (2001), Gebhardt et al. (2001), Easton (2004), Ohlson and Juettner-Nauroth (2005), respectively. Each month, stocks are sorted into 100 portfolios by their estimated cost of equity. For each portfolio, we compute the average cost of equity and the average realized buy-and-hold return over the next 10 years, and then average across months. Both axes are annualized. The solid red line shows the best linear fit, and the dashed blue line shows the 45-degree line. The sample includes all global stocks, 1983–2023.

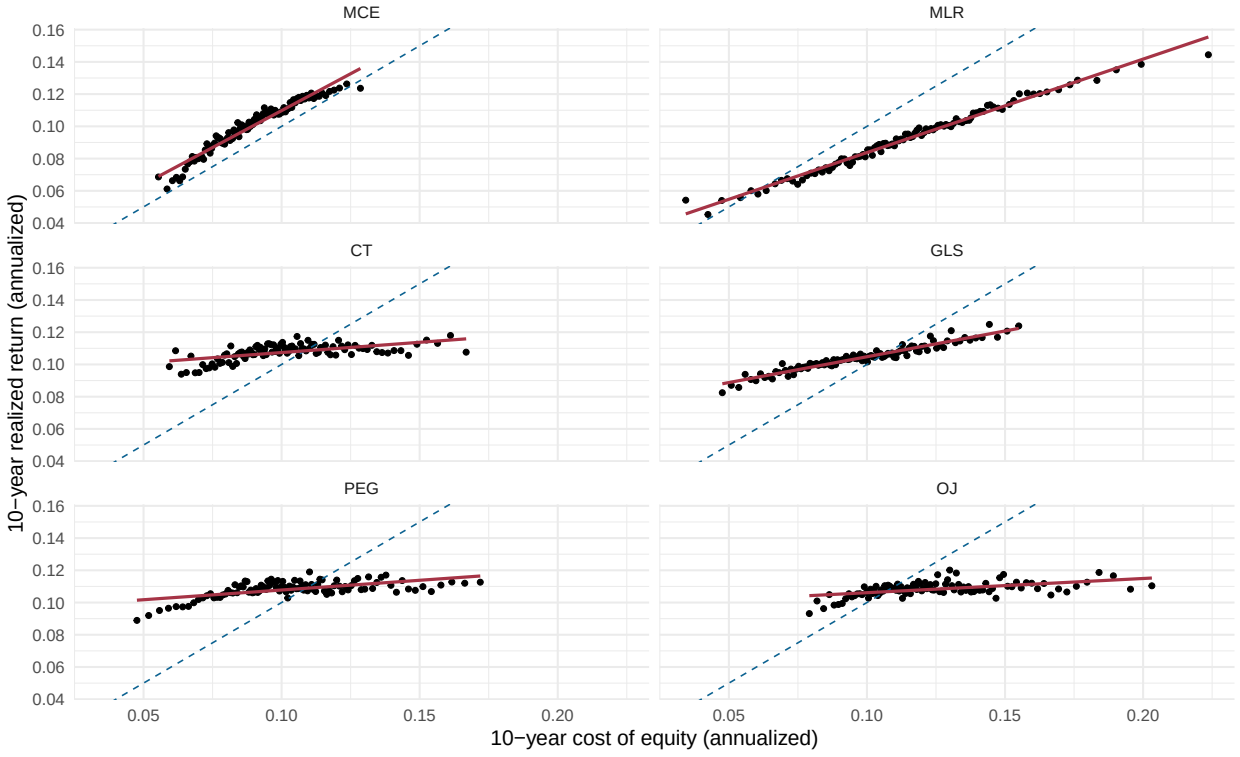
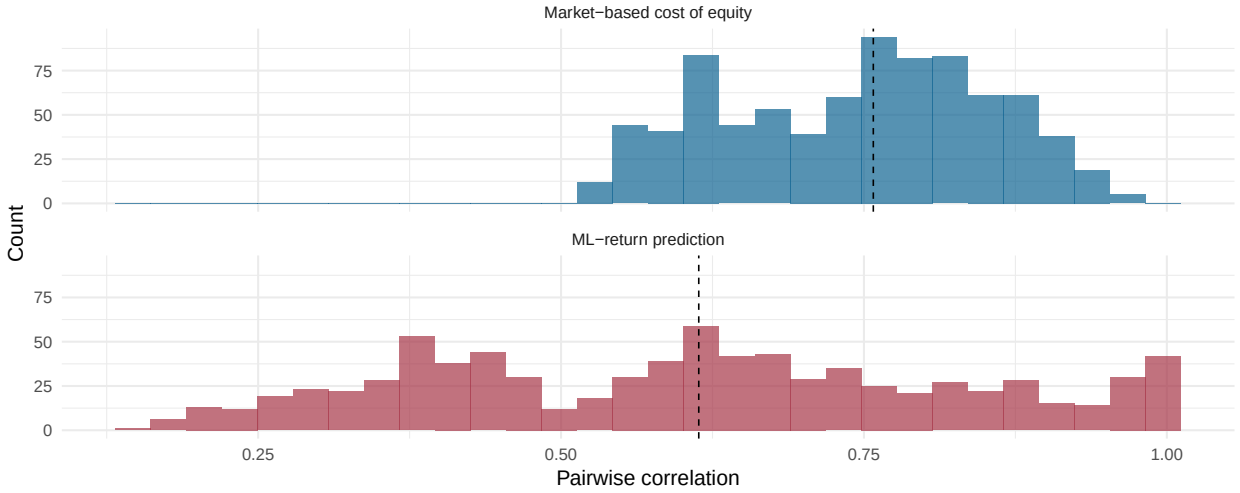


Figure 4
Model stability: MCE versus MLR

This figure compares the stability of MCE and MLR predictions across model training vintages. For each method, we generate predictions using every training vintage (models trained with data ending in 1984, 1985, . . . , 2023) applied to the same set of U.S. stocks. We then compute pairwise correlations between all vintage pairs, yielding one correlation per pair. The histograms show the distribution of these pairwise correlations, with the dashed line indicating the median. For MCE, the predictions are implied costs of equity derived from cash flow forecasts via discounted cash flow valuation. For MLR, the predictions are annualized 10-year return forecasts from LightGBM. Higher correlations indicate greater stability across training vintages.



similar predictions, whereas predictions can vary greatly across vintages for the MLR model. Consider the example of value firms. Value-like firms had high returns up until 1990, and thus a large risk premium in the 1990 model, but more modest returns since 1990, leading to smaller risk premia in the 2020 model. But the cash flow dynamics for value firms are more stable over the sample, and so the models do not change much between 1990 and 2020. The potential effects of changes in value premia are instead inferred directly from the price. The cash flow model can therefore better exploit the full dataset to obtain stable forecast of cash flows and, ultimately, expected returns.

At a mechanical level, the MLR method has biased estimates because it applies insufficient shrinkage to its out-of-sample forecasts. We examine this in Figure A10, which tracks the optimal ridge penalty λ across training vintages. The test-optimal λ is systematically higher than the validation-optimal λ, indicating that the degree of shrinkage chosen ex-ante is consistently too low relative to what would have been optimal ex-post.

Alternative Implied Cost of Capital Approaches Our MCE method is related to the implied cost of capital techniques used in the literature. Like the MCE, they back out the cost of equity from cash flow forecasts and the observed price, so they are also CF-based methods in the sense of Section 2. They differ from the MCE in the cash flow forecasts they use. The ICC methods rely on analyst forecasts and use biased estimates of expected cash flows (see upcoming Section 4). The MCE, on the other hand, uses state-of-the-art machine learning to obtain the best possible estimates of expected cash flows.

We consider four analyst-based ICC methods: Claus and Thomas (CT), Gebhardt, Lee, and Swaminathan (GLS), Easton (PEG), and Ohlson and Juettner-Nauroth (OJ). The exact descriptions of these methods can be found in Appendix D. Each method uses analyst earnings forecasts and extrapolates them beyond the analyst horizon with mechanical rules.

Table 2 shows that the ICC methods predict returns worse than the MCE. In Panel A, the slope on the ICC estimates is below one for all four methods. Panel B shows that their out-of-sample R^2 is lower than that of the MCE, and Panel C shows that their MSE is significantly larger. Figure 3 shows the same at the portfolio level. The MCE is substantially stronger than the ICC methods, even though both rest on the same valuation logic, because its cash flow forecasts are more accurate. We explore this prediction accuracy for cash flows in the upcoming section.

4.3 Cash flow predictability

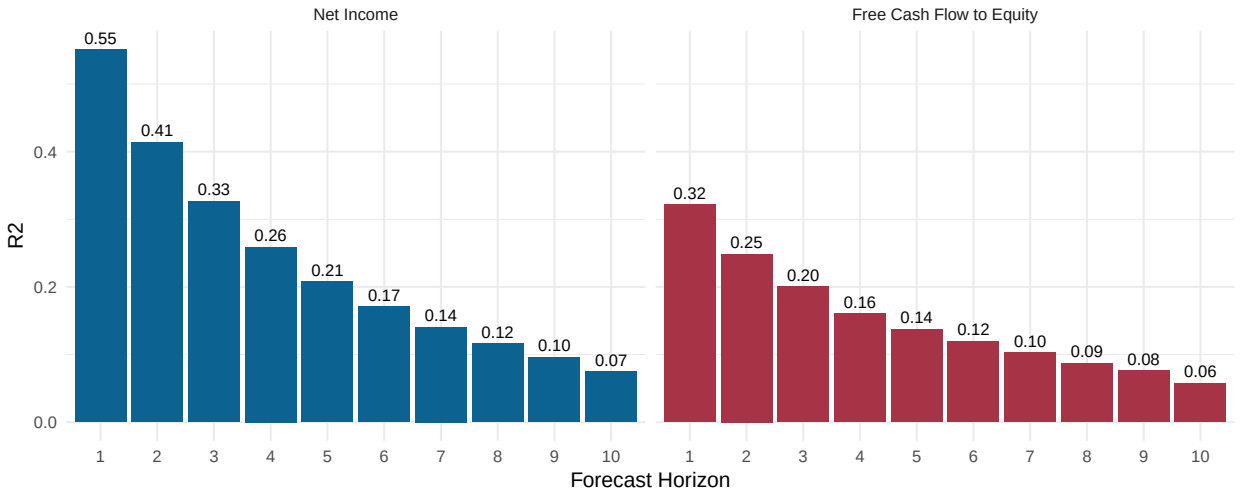
The ability of the MCE to predict returns rests on an ability to predict cash flows. In this subsection, we explore the ability to predict cash flows out of sample, and compare the ability to that of existing methods.

Figure 5 shows the out-of-sample R^2 of the cash flow prediction models for each of the two cash flow targets and ten forecast horizons. The R^2 is positive across all the outcome variables used in our prediction model. A positive R^2 is not guaranteed when conducting out-of-sample predictions. We predict short-horizon cash flows more accurately than long-horizon cash flows. Furthermore, net income is more predictable than free cash flows to equity because accounting rules smooth underlying cash flows. For example, a large investment strongly reduces free cash flows to equity in one year, but has weaker effects on net income over several years due to depreciation allowances. Figure A4 shows that the R^2 is relatively stable over time, with a slight upward trend.

We compare our cash flow forecasts to existing benchmarks from the literature. Few

Figure 5
MCE Cash Flows

This figure shows the out-of-sample R^2 of the cash flow forecasts used by the MCE model. We consider two different cash flow variables, as indicated by the panel header, and ten forecast horizons, as indicated by the x-axis. The results are based on global stocks. The sample is balanced, meaning that stocks must have non-missing realized and predicted cash flow for both variables and all ten horizons. Figure A3 shows the results from the unbalanced sample.



papers predict long-run cash flows, so we mainly consider earnings forecasts.[9] Our two main benchmarks come from implied cost of capital methods by Claus and Thomas (2001, CT) and Gebhardt et al. (2001, GLS). Both methods take analyst forecasts over the next two years as input, but differ in how they extrapolate cash flows beyond year two. CT assumes that cash flows grow at analysts' long-term growth rate forecasts, while GLS assumes that each firm's return on equity converges linearly to its corresponding industry average over a 12-year period. We also consider the random walk model, which assumes that all future cash flows are equal to the most recent realization. The random walk (RW) model of earnings is supported by early research in accounting (Ball and Watts 1972), and earnings surprises are often defined with respect to the RW model (see, e.g., Foster et al., 1984 and Martineau, 2022). For a detailed description of the benchmark models, see Appendix D.

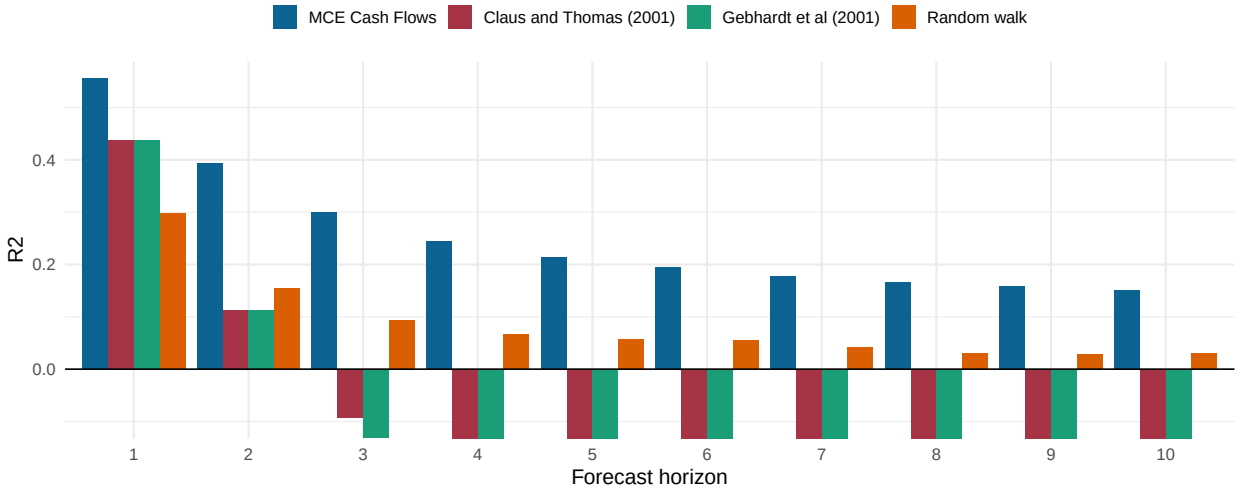
Figure 6 shows that the out-of-sample R^2 of the MCE cash flow forecasts exceeds that of all the benchmarks. At a one-year horizon, the difference relative to CT and GLS (which

[9] Analysts typically produce earnings forecasts one quarter or one year ahead. For papers trying to forecast short-run cash flows with machine learning, see Hillenbrand and McCarthy (2023), Van Binsbergen et al. (2023b), Zhang et al. (2024), and De Silva and Thesmar (2024). Among these, the longest horizon considered is four years ahead, by De Silva and Thesmar (2024).

at this horizon are effectively analyst forecasts) is relatively small.[10] However, by year three, both have a negative out-of-sample R^2. The RW model performs relatively poorly at short horizons but it beats the CT and GLS benchmarks at the three- to ten-year horizon. Overall, the results suggest that the MCE cash flow forecasts outperform existing methods in the literature—especially at the long horizons that are key for long-run expected returns. Figure A2 shows binned scatter plots of predictions relative to realizations for each of the models. The CT, GLS, and RW models have systematic biases for various parts of the distributions, whereas our method appears approximately unbiased.

Figure 6
Cash Flow Forecast Benchmarking

The figure shows the out-of-sample R^2 for net income forecasts at horizons 1 through 10 years for different methods. The MCE cash flows come from a LightGBM machine learning model trained on firm characteristics, explained in detail in Appendix B; the benchmarks are implied cash flow forecasts from the Claus and Thomas (2001) and Gebhardt et al. (2001) analyst-based ICC models as well as a random walk (last observed net income). All forecasts are scaled by total assets and winsorized at the 1st and 99th percentiles. The benchmark models are explained in detail in Appendix D. The y-axis starts at -0.1, even though some methods have lower R^2. The sample is balanced, requiring all methods to have non-missing predictions for each firm-month-horizon observation.



[10]The dependent variable in Figure 6 is for earnings realized over a constant yearly horizon, but the analyst forecasts used by CT and GLS are for the next fiscal year, which is typically closer than one year. Figure A9 shows that the ML model also beats analysts when we map their forecasts to a constant one-year horizon. It also shows that the analysts' out-of-sample R^2 improves when the forecast is demeaned to remove bias arising from well-known analyst conflicts of interest (Richardson et al. 2004). However, even after demeaning the forecasts and outcome, the ML model remains superior.

5 The Market-Based Cost of Capital

Our main application is to calculate the cost of equity and ultimately cost of capital for firms. In this section, we describe this object in detail.

5.1 The Cost of Capital and Expected Returns

Financial investors provide capital to the firm in the form of debt and equity. The market value of the debt provided to the firm is D and that of equity is E. The firm invests its capital in assets with an expected long-run rate of return r_t^{firm}. In addition, the firm deducts its payments to debt holders from its taxes, so it also generates tax savings of $\text{D} \times \tau \times r_t^{\text{debt}}$, where τ is the tax rate and r_t^{debt} the debt rate. In total, the expected return on capital invested in the firm is:

$$\text{ROR}_t = \frac{(\text{D+E}) \times r_t^{\text{firm}} + \text{D} \times \tau \times r_t^{\text{debt}}}{\text{D+E}} \tag{17}$$

$$= r_t^{\text{firm}} + \frac{\text{D}}{\text{D+E}} \tau r_t^{\text{debt}}. \tag{18}$$

Instead of investing in the firm, debt holders could purchase an alternative, external debt security that has the same characteristics as the firm's debt and generates an expected return $\widetilde{r_t^{\text{debt}}}$. Similarly, equity holders could invest in an external equity security that has the same characteristics as the firm's equity and generates an expected return of $\widetilde{r_t^{\text{equity}}}$. If all the firm's investors reallocated their capital to the external securities, their expected return would be:

$$\text{OCC}_t = \frac{\text{D} \times \widetilde{r_t^{\text{debt}}} + \text{E} \times \widetilde{r_t^{\text{equity}}}}{\text{D+E}}, \tag{19}$$

which is therefore the opportunity cost of the capital (OCC) invested in the firm.

Assuming the law of one price holds, investments with the same characteristics (e.g., risk and term) yield the same expected return. Hence, the expected return on the firm's debt equals the return on the external security, $r_t^{\text{debt}} = \widetilde{r_t^{\text{debt}}}$, which is also known as the firm's "cost of debt."[11] Similarly, the expected return on the firm's equity is $r_t^{\text{equity}} = \widetilde{r_t^{\text{equity}}}$, called

[11] This formulation follows the simplifying assumption, suggested in many textbooks and by practitioners, that the tax saving is proportional to the expected return on debt, whereas in reality the tax saving may depend on realized interest payments. Relaxing the assumption introduces additional terms, but leaves the basic intuition intact.

the firm's "cost of equity."

The law of one price also implies that the return on the capital invested in the firm (ROR) is identical to the return on investing the capital in external securities (OCC), as both investments have the same characteristics. By equalizing (18) and (19), we can solve for the expected return on the firm's assets under the law of one price:

$$r_t^{\text{firm}} = \frac{\text{D}}{\text{D+E}}(1-\tau)\, r_t^{\text{debt}} + \frac{\text{E}}{\text{D+E}}\, r_t^{\text{equity}} = \text{WACC}. \tag{20}$$

This formulation shows that the firm can use the expected returns on debt and equity to determine the expected return on its assets in the eyes of financial investors. This expected return r_t^{firm} is known as the firm-level after-tax cost of capital or the weighted average cost of capital (WACC). Intuitively, it can also be thought of as the funding cost of the firm's typical investment. The measurement in our paper estimates the WACC at the firm level. In the remainder of the paper, we use the terms "cost of capital" and WACC interchangeably.

We insert the MCE into the definition of the cost of capital (20). Combined with conventional measures of firm leverage, tax rates, and the cost of debt, we use the MCE to calculate an estimate of the firm-level cost of capital. We term it the "market-based cost of capital" (MCC) since it is consistent with observed market values. Estimating the MCC requires cash flow forecasts with strong predictive power. If we instead used weak forecasts, the resulting cost of capital would be severely biased.

5.2 Estimating the cost of debt

We observe traded bond yields for U.S. firms, so we can calculate the cost of debt as:

$$\mathbb{E}_t[r_{t+1}^{\text{debt}}] = \text{yield}_t - [\text{prob. of default}]_t \times (1 - [\text{recovery rate}]_t). \tag{21}$$

We compute the yield as the weighted average over all the firm's outstanding bonds using bond market value as weights. Data on bond yields are from Dick-Nielsen et al. (2023), based on dealer quotes over 1973-2001 and traded prices over 2002-2023. The probability of default is the average realized default rate of bonds with the same rating over the past 3 years, following Campello et al. (2008). The recovery rate is again rating-specific and from Altman and Kishore (1998). To adjust for differences in bond maturity, we subtract the

duration-matched risk-free rate and add the ten-year U.S. treasury yield:

$$r^{\text{debt}} = \mathbb{E}_t[r_{t+1}] - r^f_{\text{duration-matched}} + r^f_{10}. \tag{22}$$

This serves as our cost of debt estimate for firms with traded bonds.

We project the cost of debt for firms with traded bonds on their distance-to-default DtD_t, estimated as in Bharath and Shumway (2008):

$$r^{\text{debt}}_t = a_t + b_t \times \text{DtD}_t + \varepsilon. \tag{23}$$

We then use the estimates from this regression to predict the cost of debt for firms without traded bonds. Figure A13 shows how the cost of debt varies over time.

5.3 Estimating the firm-level cost of capital (MCC)

We calculate the MCC, our market-based estimate of the firm-level cost of capital, using the WACC formula (20). We directly observe the market value of equity and proxy for the market value of debt using the book value of debt.[12] We estimate the tax rate as an equal-weighted average of the firm's effective tax rate and the median effective tax rate of all firms in the same year and country. The firm's effective tax rate is Compustat item `txt` divided by `pi`, capped at zero and one, to control for outliers (e.g., due to carry-forward tax losses). Figure A14 shows how the firm-level cost of capital varies over time.

6 The Cost of Capital and Firm Value

Using a simple model, we formalize the classical intuition that firms achieve a higher market value when they use an unbiased cost of capital, like the MCC, in their investment decisions. We find an inverted U relationship in the data: firms have greater market value when their perceived cost of capital is closer to the MCC. In addition, firm market value responds more positively to the announcement of a new investment when the perceived cost of capital is relatively larger than the MCC. The findings suggest that equity markets value firms more highly when their investment decisions are more in line with the MCC.

[12] In some textbooks, the leverage measure is the firm's long-run target leverage (Ross et al. 2022). Under that view, one can view our market-based current leverage measure as a proxy for the target leverage.

6.1 Model of inverted U

We sketch a simple model to formalize how firms' perceptions about their cost of capital affect firms' capital allocation and equity market value.

We consider an infinite-horizon discrete-time model with periods $t = 0, 1, 2, \ldots$ and a cross-section of firms indexed by i. Each firm i owns capital stock $K_{i,t}$ which depreciates at rate $\gamma \in (0, 1)$ per period. The law of motion for capital is:

$$K_{i,t+1} = (1 - \gamma)K_{i,t} + I_{i,t}, \tag{24}$$

where $I_{i,t}$ is gross investment at date t.

In each period, capital generates stochastic output according to the production function:

$$Y_{i,t}(K_{i,t}) = (\alpha_i + \eta_{i,t})K_{i,t} - \frac{\beta_i}{2}K_{i,t}^2, \tag{25}$$

where $\alpha_i, \beta_i > 0$ are firm-specific productivity parameters and $\eta_{i,t}$ is an i.i.d. productivity shock with $E[\eta_{i,t}] = 0$. Capital in period 0 is 0, $K_{i,0} = 0$.

Firm problem. Firm i maximizes the perceived present value of future cash flows using the perceived cost of capital $r_{i,t}^{\text{perc.}}$. The perceived cost of capital may differ from the firm-level cost of capital, r_i^{firm}, that financial investors use to discount firm i's cash flows. We model the deviation as:

$$r_{i,t}^{\text{perc.}} = r_i^{\text{firm}} + \delta_{i,t}. \tag{26}$$

For simplicity, we assume that r_i^{firm} is constant and that the cost of capital deviation, $\delta_{i,t}$, is orthogonal to r_i^{firm} and the fundamentals α_i and β_i.

Firm i chooses capital to maximize its perceived present value of cash flows:

$$\max_{\{K_{i,t}\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \frac{1}{(1 + r_i^{\text{perc.}})^t} \left[Y_{i,t}(K_{i,t}) - I_{i,t}\right]\right]. \tag{27}$$

Taking expectations over the productivity shocks and noting that firms choose a constant capital stock in the absence of adjustment costs, the firm's first-order condition implies that the optimal choice of capital, K_i^*, is given by:

$$K_i^* = \bar{K}_i - \frac{\delta_i}{\beta_i}, \tag{28}$$

where $\bar{K}_i \equiv (\alpha_i - r_i^{\text{firm}} - \gamma)/\beta_i$ is the capital the firm would choose if it used the correct discount rate ($r_i^{\text{perc.}} = r_i^{\text{firm}}$). In equilibrium, investment equals depreciation: $I_i^* = \gamma K_i^*$.

Market value. Investors price assets using a stochastic discount factor M_t. For firm i, the equity price satisfies:

$$P_i = \mathbb{E}\left[\sum_{t=0}^{\infty} M_t \cdot (Y_{i,t}(K_{i,t}) - I_{i,t})\right]. \tag{29}$$

We assume that each firm is too small to influence the pricing of risk in the economy, such that the stochastic discount factor is not influenced by the choice of capital of the individual firm. We assume the stochastic discount factor satisfies conditions under which the multi-period discount factor can be represented as $(1+r_i^{\text{firm}})^{-t}$ for a constant firm-specific discount rate r_i^{firm} that reflects both time preference and systematic risk. These assumptions yield the value function,

$$P_i(K) = \sum_{t=1}^{\infty} \frac{1}{(1+r_i^{\text{firm}})^t}[Y_i(K) - \gamma K] - K = \frac{Y_i(K) - \gamma K}{r_i^{\text{firm}}} - K. \tag{30}$$

Result 1 (An Inverted U Between Market Value and Cost of Capital Wedges)**.** *The market value of firm i is given by*

$$P_i(K_i^*) = P_i(\bar{K}_i) - \frac{\delta_i^2}{2\beta_i r_i^{firm}}, \tag{31}$$

where $P_i(\bar{K}_i)$ is the market value of the firm when the cost of capital deviation is zero, $\delta_i = r_i^{perc.} - r_i^{firm} = 0$.

Result 1 implies an inverted U relation: firm market value is maximized when firms use the same cost of capital as financial markets, r_i^{firm}, in their investment decisions. If firms use $r_i^{\text{perc.}}$ below or above r_i^{firm}, market value decreases. The inverted U is illustrated in Figure 7.

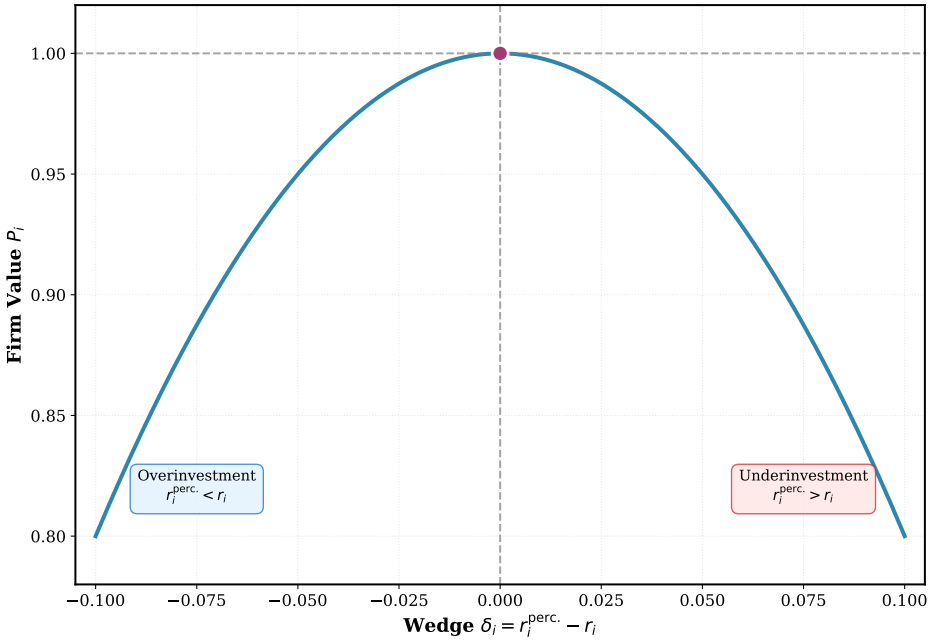
Our second result characterizes the relationship between investment and firm value in the cross-section of firms. From equation (28), we have $I_i^* = \gamma K_i^* = \gamma \bar{K}_i - \gamma \delta_i/\beta_i$. Since investment is proportional to capital, the correlation between investment and market value depends on the sign of the deviation.

Result 2 (Investment-value correlation)**.** *In the cross-section of firms, the correlation between investment and value depends on the sign of firms' cost of capital deviation*

Figure 7

Model: Inverted U between firm market value and deviations from MCC

This figure shows the relation between market value and cost of capital deviations in our model. The model is calibrated such that market value is \$1 when the cost of capital deviation is 0.



$$Corr(I_i, P_i) > 0 \text{ for firms for which } \delta_i > 0$$
$$Corr(I_i, P_i) < 0 \text{ for firms for which } \delta_i < 0$$

Result 2 shows that the comovement between market values and investment depends on the sign of the cost of capital deviation. Firms with a positive deviation underinvest, and higher investment is therefore met with a positive response from investors. Firms with a negative deviation overinvest, and higher investment is therefore met with a negative response from investors.

6.2 Empirical results on inverted U

We investigate the relation between firm equity market value and the cost of capital deviation in the data (result 1 in the model). To measure the deviation, we calculate the difference between a firm's perceived cost of capital and its MCC. The "perceived cost of capital" is the cost of capital that influences the firm's real investment decisions, measured by Gormsen and Huber (2025). The data are hand-collected from conference calls between firm managers,

financial investors, and analysts. They measure explicit statements by managers on the calls about their internal perceptions of their firm-level weighted average cost of capital. In the data, firms with a higher perceived cost of capital have higher returns on invested capital and lower investment rates, consistent with the view that the perceived cost of capital influences capital allocation in the long run. Figure A16 shows a histogram of the cost of capital deviation.

We study which firms use a given amount of book equity more effectively to generate a high market value. We therefore analyze whether firms with similar book equity are valued differently in equity markets depending on their cost of capital deviation. Figure 8 illustrates the average log market value for firms in different bins of the cost of capital deviation. The figure plots point estimates from a regression of log market value on the perceived cost of capital. The regression additionally controls for fixed effects for the firm's book equity decile interacted with country fixed effects, so the point estimates capture differences within firms of similar book equity size. In addition, we control for quarter-by-year-by-country fixed effects, so macroeconomic shocks do not bias the estimates.

Figure 8 reveals an inverted U relation. Firm market value is lower when the cost of capital perceived by the firm (perc. CoC) deviates more from the MCC. This finding suggests that the cost of capital used by firms is strongly associated with market value, consistent with the classical intuition about the opportunity cost of capital and modern models of firm behavior.

Table 3 suggests that the association between the cost of capital deviation and market value is statistically and economically significant. In columns 1 to 2, the regressor is the absolute value of the deviation. The point estimates are all significant at the 1 percent level and of similar magnitude in the U.S. and the full sample.

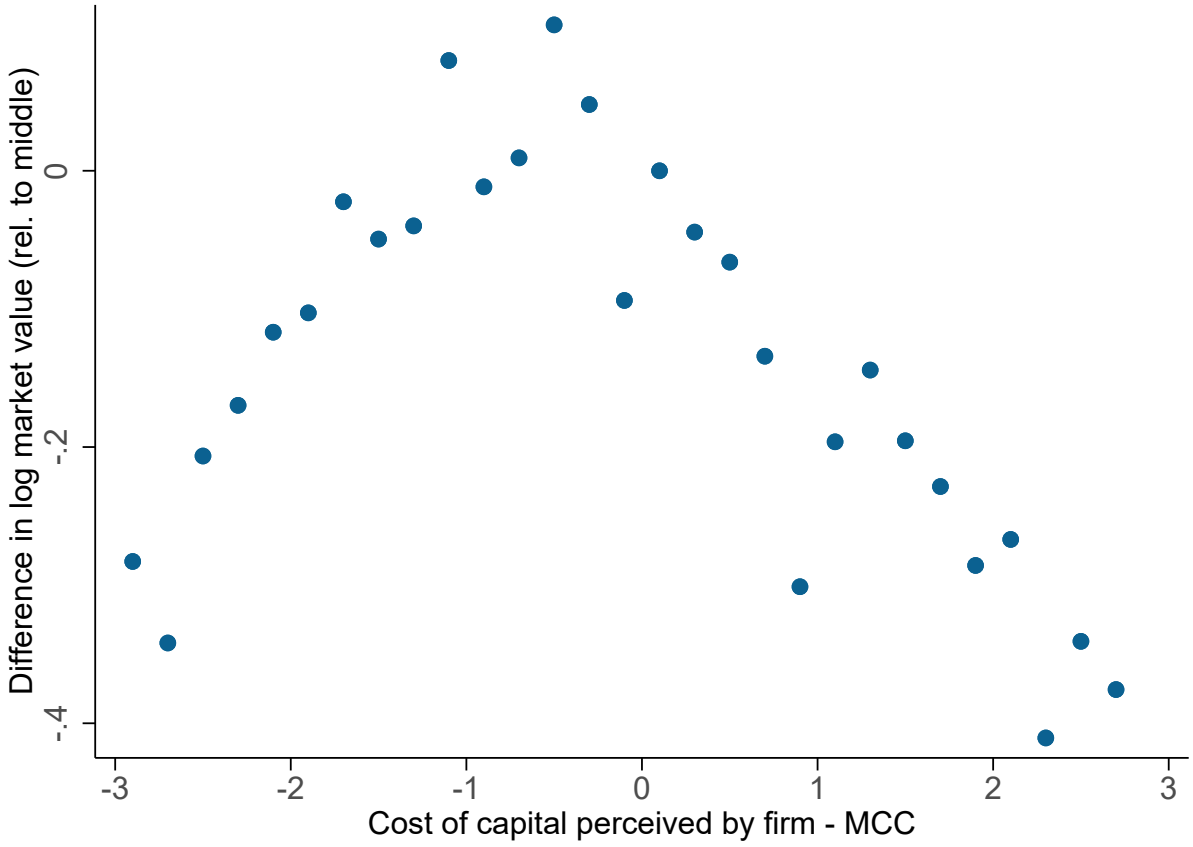
In columns 3 to 5, we use the simple subtraction as the regressor. A 1 percentage point increase in the deviation is associated with an 8.7 log point value increase in the range of relatively large negative deviations (below -0.3 percentage points) analyzed in column 3. In contrast, a 1 percentage point increase in the deviation is associated with a 7 log point value decrease in the range of relatively large positive deviations (above 0.3 percentage points) analyzed in column 4. The magnitude is thus similar on both sides. The point estimate is statistically insignificant when the deviations are close to zero in column 5, albeit statistically noisy. The pattern of results is consistent with the model and the inverted U shape in Figure 7.

Taken together, the empirical results suggest that firms using an unbiased perceived cost

Figure 8

Inverted U: firm market value and deviations from the MCC

The figure shows the relation between firm market value and the cost of capital deviation (i.e., the difference between perc. CoC and MCC). The perc. CoC is reported by the firm on conference calls and influences the firm's investment decisions. MCC is the market-based cost of capital from this paper. We construct bins for the cost of capital deviation in percentage points. The bins are 0.2 percentage points wide, except for the leftmost bin, which includes all observations below -2.8 (roughly the 1st percentile), and the rightmost bin, which pools all observations above 2.8. The figure plots point estimates from a regression of firm log market value on indicators for the bins. The point estimates give the difference in log market value relative to the middle bin (difference just above 0). The regression controls for fixed effects for book equity deciles-by-country and quarter-by-year-by-country. Book equity deciles are based on the firm's median decile in the distribution of book equity relative to other firms in the same quarter-year. Table 3 reports related regressions.



	(1) U.S.	(2) Global	(3) $\leq$ -0.3	(4) $\geq$ 0.3	(5) $>$ -0.3, $<$ 0.3
| Perc. CoC − MCC |	-0.081***	-0.096***			
	(0.019)	(0.018)			
Perc. CoC − MCC			0.087*	-0.070***	-0.173
			(0.049)	(0.025)	(0.210)
Observations	3,098	5,204	1,509	3,055	640
Within R^2	0.0185	0.0266	0.00924	0.0154	0.00240

The table reports regressions of firm log market value on the deviation between the cost of capital perceived by the firm and the MCC. Perc. CoC is the cost of capital reported by the firm on conference calls that influences the firm's investment decisions. MCC is the market-based cost of capital from this paper. The regressor in columns 1 to 2 is the absolute difference between Perc. CoC and MCC. The regressor in columns 3 to 5 is the simple subtraction. The sample includes only U.S. firms in column 1; all firms in column 2; only firms where the simple subtraction is at most -0.3 percentage points in column 3 (i.e., firms where the perc. CoC is low relative to the MCC); only firms where the simple subtraction is at least 0.3 in column 4; and firms where the simple subtraction is close to 0 in column 5. All specifications control for fixed effects for book equity deciles-by-country and quarter-by-year-by-country. Book equity deciles are based on the firm's decile in the distribution of book equity relative to other firms in the same quarter-year. Standard errors are clustered at the firm level. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Inverted U: firm market value and deviations from the MCC

of capital, like the MCC, are valued more positively in equity markets.

6.3 Empirical results on investment announcements

We analyze how firm equity market values respond to firms announcing new investment projects (result 2 in the model). If the firm's cost of capital deviation is above zero, the firm's perceived cost of capital is "too high" compared to the MCC and the firm invests relatively little compared to the cost of capital effectively used by the equity market. An announcement of a firm with a positive deviation should therefore lead to a relatively higher market value compared to an announcement of a firm with a negative deviation.

Investment announcements are available on the S&P Capital IQ Key Developments database, which reports all types of news about firms. We consider all firms for which we observe a perceived cost of capital and the MCC in the same quarter. We prompt Claude AI to extract all news items which are unambiguously new announcements of investment projects, not just minor updates on existing projects. We require the investment to be a concrete, specific project undertaken by the firm itself. The investment needs to be the

	(1) U.S.	(2) Global	(3) U.S.	(4) Global
(Perc. CoC − MCC) > 0	0.199**	0.199**		
	(0.100)	(0.098)		
Perc. CoC − MCC			0.032*	0.034*
			(0.019)	(0.019)
Observations	414	709	414	709
Within R^2	0.0121	0.0107	0.00472	0.00465

The table analyzes firms' equity market returns in windows around real investment announcements by the firm. Perc. CoC is the cost of capital reported by the firm on conference calls that influences the firm's investment decisions. MCC is the market-based cost of capital from this paper. The regressor in columns 1 to 2 is an indicator for a positive difference between Perc. CoC and MCC. The regressor in columns 3 to 4 is the simple subtraction. The outcome is the abnormal stock return of the firm (firm return minus market return) in a window of -3 to +1 days around an investment announcement reported on the Key Developments Dataset. The outcome is winsorized at the 1^{st} and 99^{th} percentiles. The sample includes only U.S. firms in columns 1 and 3 and all firms in columns 2 and 4. All specifications control for fixed effects for book equity deciles-by-country and quarter-by-year-by-country. Book equity deciles are based on the firm's decile in the distribution of book equity relative to other firms in the same quarter-year. Standard errors are clustered at the firm level. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Stock returns and investment announcements

central focus of the news, rather than described only in passing, which avoids classification errors and makes it more likely that the announcement reveals new information. Store openings, site breaking grounds, or product launches are not included, since the investment has already been made in the past in these cases. We also exclude M&A events because they are not typically evaluated by the equity market using the existing firms' MCC. Overall, these restrictions leave 709 firms in our estimation sample.

To analyze the equity market response, we construct abnormal stock returns as the firm's stock return minus the market return in a window of -3 days to 1 day around an investment announcement. We regress the abnormal stock return on an indicator for a positive cost of capital deviation. The point estimates in columns 1 and 2 of Table 4 suggest that the abnormal return for firms with a positive deviation is 0.2 percentage points higher, both in the U.S. and the global sample. The coefficients are significant at the 5 percent level.

We also find positive and significant coefficients when using the simple subtraction of

MCC from perceived cost of capital as the regressor in columns 3 and 4. The regressions include the baseline controls of book equity deciles and quarter-by-year fixed effects, interacted with country fixed effects, to ensure that size-specific differences across firms and macroeconomic shocks do not bias the results. In Table A9, we report similar results dropping the size controls in column 2 and using an unwinsorized abnormal stock return as the outcome in column 3.

The findings suggest that the equity market response to new investment announcements depends on the cost of capital deviation. The MCC seems to capture well the cost of capital effectively used by the equity market and therefore how positively the equity market responds to a new announcement, in line with the model. More generally, the announcement results further support the conclusion that the cost of capital used by firms is a key determinant of firm market values.

7 Conclusion

We estimate the firm-level cost of capital using market prices and new machine-learning forecasts of future cash flows. The key challenge is that the textbook cost of capital that firms should use in investment decisions is not directly observed. Existing empirical proxies, including the CAPM and analyst-based implied cost of capital measures, are biased. Our approach combines forecasts of future cash flows with a valuation equation that backs out the discount rate consistent with observed market values. The result is an unbiased market-based cost of capital (MCC).

Empirically, we show that the method performs well along two dimensions. First, the cash flow forecasts exhibit out-of-sample predictive power, especially relative to widely used benchmarks. Second, the market-based cost of equity strongly predicts future long-run realized returns. These findings suggest that the MCC captures variation in expected returns well.

We then address how the cost of capital used by firms matters for firm market values. Using hand-collected data on firms' perceived cost of capital, we find that market value is higher when firms' perceived cost of capital is closer to the MCC, consistent with the classical view that firms create value by aligning investment decisions with the opportunity cost of capital faced by investors.

References

Altman, E. I. and V. M. Kishore (1998): "Default and returns on high yield bonds: analysis through 1997," *NYU Salomon Center Report, January1995*.

Ball, R. and R. Watts (1972): "Some time series properties of accounting income," *The Journal of Finance*, 27, 663–681.

Bauer, M. D. and G. D. Rudebusch (2020): "Interest Rates under Falling Stars," *American Economic Review*, 110, 1316–54.

Bharath, S. T. and T. Shumway (2008): "Forecasting default with the Merton distance to default model," *The Review of Financial Studies*, 21, 1339–1369.

Bianchi, F., D. Q. Lee, S. C. Ludvigson, and S. Ma (2025): "The Prestakes of Stock Market Investing," Tech. rep., National Bureau of Economic Research.

Bianchi, F., S. C. Ludvigson, and S. Ma (2022): "Belief distortions and macroeconomic fluctuations," *American Economic Review*, 112, 2269–2315.

Binsbergen, J. v., S. Hua, J. Peeters, and J. Wachter (2025): "Is the United States a lucky survivor? A hierarchical Bayesian approach," *The Journal of Finance*, 80, 2355–2388.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110, 2748–82.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2024): "Belief overreaction and stock market puzzles," *Journal of Political Economy*, 132, 1450–1484.

Campbell, J. Y. and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

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

Cho, T. and C. Polk (2024): "Putting the price in asset pricing," *The Journal of Finance*, 79, 3943–3984.

Claus, J. and J. Thomas (2001): "Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets," *Journal of Finance*, 56, 1629–1666.

Coibion, O. and Y. Gorodnichenko (2015): "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 105, 2644–2678.

De Silva, T. and D. Thesmar (2024): "Noise in expectations: Evidence from analyst forecasts," *The Review of Financial Studies*, 37, 1494–1537.

Decaire, P. and M. Guenzel (2023): "What Drives Very Long-Run Cash Flow Growth Expectations?" *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Dick-Nielsen, J., P. Feldhütter, L. H. Pedersen, and C. Stolborg (2023): "Corporate bond factors: Replication failures and a new framework," *Available at SSRN 4586652*.

Easton, P. D. (2004): "PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital," *Accounting Review*, 79, 73–95.

Edwards, E. O. and P. W. Bell (1961): *The theory and measurement of business income*, Univ of California Press.

Eskildsen, M., M. Ibert, T. I. Jensen, and L. H. Pedersen (2024): "In search of the true greenium," *Available at SSRN*.

Fama, E. F. and K. R. French (1997): "Industry costs of equity," *Journal of financial economics*, 43, 153–193.

Foster, G., C. Olsen, and T. Shevlin (1984): "Earnings releases, anomalies, and the behavior of security returns," *Accounting Review*, 574–603.

Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan (2001): "Toward an Implied Cost of Capital," *Journal of Accounting Research*, 39, 135–176.

Gormsen, N. J. and K. Huber (2025): "Firms' Perceived Cost of Capital," .

Graham, J. R. (2022): "Presidential Address: Corporate Finance and Reality," *Journal of Finance*, 77, 1975–2049.

Graham, J. R. and C. R. Harvey (2001): "The Theory and Practice of Corporate Finance: Evidence from the Field," *Journal of Financial Economics*, 60, 187–243.

Greenwood, R. and A. Shleifer (2014): "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 27, 714–746.

Hillenbrand, S. and O. McCarthy (2023): "The Optimal Stock Valuation Ratio," *Available at SSRN 4288780*.

Hommel, N., A. Landier, and D. Thesmar (2025): "Corporate Valuation: An Empirical Comparison of Discounting Methods," .

Jacobs, M. T. and A. Shivdasani (2012): "Do You Know Your Cost of Capital?" *Harvard Business Review*, 90, 118–124.

Jagannathan, R., D. A. Matsa, I. Meier, and V. Tarhan (2016): "Why Do Firms Use High Discount Rates?" *Journal of Financial Economics*, 120, 445–463.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023): "Is there a replication crisis in finance?" *The Journal of Finance*, 78, 2465–2518.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu (2017): "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, 30.

Kelly, B. and D. Xiu (2023): "Financial machine learning," *Foundations and Trends in Finance*, 13, 205–363.

Keloharju, M., J. T. Linnainmaa, and P. Nyberg (2021): "Long-term discount rates do not vary across firms," *Journal of Financial Economics*, 141, 946–967.

Martineau, C. (2022): "Rest in peace post-earnings announcement drift," *Critical Finance Review*, 11, 613–646.

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

Mukhlynina, L. and K. G. Nyborg (2016): "The Choice of Valuation Techniques in Practice: Education Versus Profession," .

Ohlson, J. A. (1995): "Earnings, book values, and dividends in equity valuation," *Contemporary accounting research*, 11, 661–687.

Ohlson, J. A. and B. E. Juettner-Nauroth (2005): "Expected EPS and EPS growth as determinantsof value," *Review of accounting studies*, 10, 349–365.

Richardson, S., S. H. Teoh, and P. D. Wysocki (2004): "The walk-down to beatable analyst forecasts: The role of equity issuance and insider trading incentives," *Contemporary accounting research*, 21, 885–924.

Ross, S., R. Westerfield, J. Jaffe, B. Jordan, and K. Shue (2022): *Corporate Finance*, McGraw-Hill Education, 13th ed.

Shwartz-Ziv, R. and A. Armon (2022): "Tabular data: Deep learning is not all you need," *Information Fusion*, 81, 84–90.

Van Binsbergen, J. H., M. Boons, C. C. Opp, and A. Tamoni (2023a): "Dynamic asset (mis) pricing: Build-up versus resolution anomalies," *Journal of Financial Economics*, 147, 406–431.

Van Binsbergen, J. H., X. Han, and A. Lopez-Lira (2023b): "Man versus machine learning: The term structure of earnings expectations and conditional biases," *The Review of financial studies*, 36, 2361–2396.

Zhang, Y., Y. Zhu, and J. T. Linnainmaa (2024): "Man versus Machine Learning Revisited," *Available at SSRN 4899584*.

Supplemental Appendix

Appendix A Proofs

Appendix A.1 Proof of Proposition 1

Setup. The proof uses the model and notation of Section 2. Additional notation that will be used in the proof is the cash flow error, $v_{t+1} \equiv CF_{t+1} - \mathbb{E}_t(CF_{t+1})$, which is mean-zero by construction and assumed i.i.d. with finite variance, and the return error, $u_{t+1} \equiv r_{t+1} - \mathbb{E}_t(r_{t+1})$. We also assume that the price is bounded away from zero, so that both mean-squared errors are finite.

Proof. We start by showing how the return error relates to the cash flow error, and that it disappears when evaluating the difference between the MSE of the ER and CF methods. Given the one-period nature of the problem, the realized return is:

$$r_{t+1} = \frac{CF_{t+1}}{p_t} - 1.$$

Subtracting the conditional expectation, the return error is the cash-flow error scaled by price:

$$\begin{aligned} u_{t+1} &= r_{t+1} - \mathbb{E}_t(r_{t+1}) \\ &= \left(\frac{CF_{t+1}}{p_t} - 1\right) - \left(\frac{\mathbb{E}_t(CF_{t+1})}{p_t} - 1\right) \\ &= \frac{v_{t+1}}{p_t}. \end{aligned} \tag{A1}$$

Each method's forecast error is therefore its error in the conditional mean minus the return error:

$$\widehat{\mathbb{E}}_t^m(r_{t+1}) - r_{t+1} = \underbrace{\left[\widehat{\mathbb{E}}_t^m(r_{t+1}) - \mathbb{E}_t(r_{t+1})\right]}_{\equiv e_m} - u_{t+1}.$$

The mean-squared error for a given method is therefore:

$$\mathrm{MSE}^m = \mathbb{E}\left[(e_m - u_{t+1})^2\right] = \mathbb{E}(e_m^2) - 2\,\mathbb{E}(e_m u_{t+1}) + \mathbb{E}(u_{t+1}^2).$$

The cross term, $2\,\mathbb{E}(e_m u_{t+1})$, is zero by the law of iterated expectations. The conditional-mean error e_m is known at time t: it is the difference between the method's prediction and the conditional mean $\mathbb{E}_t(r_{t+1})$, itself a function of time-t information. The return error, by contrast, has conditional mean zero by construction. Conditioning on time-t information first and taking out the known e_m:

$$\mathbb{E}(e_m u_{t+1}) = \mathbb{E}\left[\,\mathbb{E}_t(e_m u_{t+1})\right] = \mathbb{E}\left[e_m\ \mathbb{E}_t(u_{t+1})\right] = 0.$$

Furthermore, the last term, $\mathbb{E}(u_{t+1}^2)$, is identical for the two methods, so it cancels in the difference, leaving:

$$\mathrm{MSE}^{\mathrm{ER}} - \mathrm{MSE}^{\mathrm{CF}} = \mathbb{E}(e_{\mathrm{ER}}^2) - \mathbb{E}(e_{\mathrm{CF}}^2).$$

Next, we show that the CF method's error is zero. The population coefficients, $\bar{\alpha}^{\mathrm{CF}}$ and $\bar{\beta}^{\mathrm{CF}}$, minimize the expected squared residual:

$$Q(\alpha, \beta) = \mathbb{E}[(CF_{t+1} - \alpha - \beta' f_t)^2].$$

Differentiating with respect to α, setting the derivative to zero at the minimizers, and solving gives

$$\begin{aligned} 0 &= \mathbb{E}(CF_{t+1} - \bar{\alpha}^{\mathrm{CF}} - \bar{\beta}^{\mathrm{CF}\prime} f_t) \\ &= \mathbb{E}(CF_{t+1}) - \bar{\alpha}^{\mathrm{CF}} - \bar{\beta}^{\mathrm{CF}\prime}\,\mathbb{E}(f_t), \end{aligned}$$

which implies

$$\begin{aligned} \bar{\alpha}^{\mathrm{CF}} &= \mathbb{E}(CF_{t+1}) \\ &= \mathbb{E}(\alpha^{\mathrm{CF}} + \beta^{\mathrm{CF}\prime} f_t + v_{t+1}) \\ &= \alpha^{\mathrm{CF}} + \beta^{\mathrm{CF}\prime}\,\mathbb{E}(f_t) + \mathbb{E}(v_{t+1}) \\ &= \alpha^{\mathrm{CF}}, \end{aligned}$$

Proceeding in the same way for $\bar{\beta}^{\mathrm{CF}}$ gives:

$$\begin{aligned} 0 &= \mathbb{E}[f_t(CF_{t+1} - \bar{\alpha}^{\mathrm{CF}} - \bar{\beta}^{\mathrm{CF}\prime} f_t)] \\ &= \mathbb{E}(f_t CF_{t+1}) - \bar{\alpha}^{\mathrm{CF}}\,\mathbb{E}(f_t) - \mathbb{E}[f_t(\bar{\beta}^{\mathrm{CF}\prime} f_t)] \\ &= \mathbb{E}(f_t CF_{t+1}) - \mathbb{E}(f_t f_t')\bar{\beta}^{\mathrm{CF}} \\ &= \mathbb{E}(f_t CF_{t+1}) - I_K \bar{\beta}^{\mathrm{CF}}, \end{aligned}$$

which implies

$$\begin{aligned}\bar{\beta}^{\mathrm{CF}} &= \mathbb{E}(f_t CF_{t+1}) \\ &= \mathbb{E}[f_t(\alpha^{\mathrm{CF}} + \beta^{\mathrm{CF}\prime} f_t + v_{t+1})] \\ &= \mathbb{E}(f_t \alpha^{\mathrm{CF}}) + \mathbb{E}(f_t f_t')\beta^{\mathrm{CF}} + \mathbb{E}(f_t v_{t+1}) \\ &= I_K \beta^{\mathrm{CF}} \\ &= \beta^{\mathrm{CF}}.\end{aligned}$$

The CF method therefore recovers both true coefficients, and, as a result, its conditional-mean error is zero:

$$\begin{aligned}e_{\mathrm{CF}} &= \widehat{\mathbb{E}}_t^{\mathrm{CF}}(r_{t+1}) - \mathbb{E}_t(r_{t+1}) \\ &= \left(\frac{\bar{\alpha}^{\mathrm{CF}} + \bar{\beta}^{\mathrm{CF}\prime} f_t}{p_t} - 1\right) - \left(\frac{\alpha^{\mathrm{CF}} + \beta^{\mathrm{CF}\prime} f_t}{p_t} - 1\right) \\ &= 0. \end{aligned} \tag{A2}$$

The CF method recovers the true coefficients for cash flows because they are assumed to be time-invariant (so the population regression can recover them exactly), and accounts for the time variation in the return loadings because the current price enters its return forecast.

For the ER method, we similarly have $\bar{\alpha}^{\mathrm{ER}} = \mathbb{E}(r_{t+1})$, and it correctly recovers the expected return intercept:

$$\begin{aligned}\bar{\alpha}^{\mathrm{ER}} &= \mathbb{E}(r_{t+1}) \\ &= \alpha^{\mathrm{ER}} + \mathbb{E}(\beta_t^{\mathrm{ER}})' \mathbb{E}(f_t) + \mathbb{E}(u_{t+1}) \\ &= \alpha^{\mathrm{ER}}.\end{aligned}$$

We also have $\bar{\beta}^{\mathrm{ER}} = \mathbb{E}(f_t r_{t+1})$, but now the ER method does not recover the correct beta loading:

$$\begin{aligned}\bar{\beta}^{\mathrm{ER}} &= \mathbb{E}(f_t r_{t+1}) \\ &= \alpha^{\mathrm{ER}} \mathbb{E}(f_t) + \mathbb{E}(f_t f_t' \beta_t^{\mathrm{ER}}) + \mathbb{E}(f_t u_{t+1}) \\ &= \mathbb{E}(f_t f_t') \, \mathbb{E}(\beta_t^{\mathrm{ER}}) \\ &= I_K \mu \\ &= \mu,\end{aligned}$$

where the third line follows from the independence of loadings and predictors, and the last

term vanishes by iterated expectations, $\mathbb{E}(f_t u_{t+1}) = \mathbb{E}[f_t\ \mathbb{E}_t(u_{t+1})] = 0$. The ER method does not recover the true coefficient for return because the population regression model assumes that the loadings are constant over time, even though they are in fact time-varying.

As a result, the ER method's conditional-mean error is non-zero:

$$\begin{aligned} e_{\text{ER}} &= \widehat{\mathbb{E}}_t^{\text{ER}}(r_{t+1}) - \mathbb{E}_t(r_{t+1}) \\ &= \left(\bar{\alpha}^{\text{ER}} + \bar{\beta}^{\text{ER}\prime} f_t\right) - \left(\alpha^{\text{ER}} + \beta_t^{\text{ER}\prime} f_t\right) \\ &= \left(\alpha^{\text{ER}} + \mu' f_t\right) - \left(\alpha^{\text{ER}} + \beta^{\text{ER}\prime}{}_t f_t\right) \\ &= -(\beta_t^{\text{ER}} - \mu)' f_t \\ &= -g_t' f_t, \end{aligned}$$

where the last line defines the staleness gap g_t. The mean-squared error of the ER method is then:

$$\begin{aligned} \mathbb{E}(e_{\text{ER}}^2) &= \mathbb{E}[(g_t' f_t)^2] \\ &= \mathbb{E}[g_t' f_t f_t' g_t] \\ &= \mathbb{E}[g_t'\, \mathbb{E}(f_t f_t') g_t] \\ &= \mathbb{E}[g_t' I_K g_t] \\ &= \mathbb{E}\left(\sum_{k=1}^{K} g_{k,t}^2\right) \\ &= \sum_{k=1}^{K} \mathbb{E}(g_{k,t}^2), \end{aligned}$$

where the third line follows from the independence of loadings and predictors, and $g_{k,t}$ is the staleness gap for the kth loading. Each $g_{k,t}$ is mean-zero, so $\mathbb{E}(g_{k,t}^2)$ is the unconditional variance, which for a stationary AR(1) process is

$$\begin{aligned} \mathbb{E}(g_{k,t}^2) &= \text{var}(\beta_t^{\text{ER}}) \\ &= \frac{\sigma_\eta^2}{1-\rho^2}. \end{aligned}$$

Since each loading coordinate has the same variance, the mean-squared error is then just K

times this variance:

$$\begin{aligned}\mathbb{E}(e_{\text{ER}}^2) &= \sum_{k=1}^{K} \mathbb{E}(g_{k,t}^2) \\ &= K \text{ var}(\beta_t^{\text{ER}}). \end{aligned} \tag{A3}$$

Taking (A3) and subtracting (A2) recovers Equation (8) from Proposition 1. The MSE gain of the CF method is positive and increasing in K because $\text{var}(\beta_t^{\text{ER}}) = \frac{\sigma_\eta^2}{1-\rho^2}$ is positive, since $\sigma_\eta^2 > 0$ and $|\rho| < 1$.

□

Appendix B Machine-Learned Cash Flow Forecasts

This section documents the implementation details of the machine learning models used to forecast firm-level cash flows. We cover the output targets, the explanatory variables, the model specification, the hyperparameter tuning procedure, temporal sample splits, and the final refit strategy.

Appendix B.1 Output targets

We forecast two cash flow measures—net income (NI) and free cash flow to equity (FCFE)—at each of 10 annual horizons, yielding 20 target variables per firm-month. The following describes how each target is constructed.

- The valuation equations assume that the first cash flow occurs exactly one year from now, and that cash flows thereafter occur in one-year increments. However, fiscal year ends do not necessarily allow for this one-year constant horizon. We handle this by defining a constant-horizon cash flow over 1 year that interpolates between the adjacent accounting reports using inverse-distance weighting: for a target date falling between two fiscal year-ends, the weight on each observation is proportional to its proximity to the target date. When only one fiscal year observation is available, it is used directly. Equal weighting arises as the special case when the target date is equally far from both fiscal year-ends.

- Fiscal year ends differ across companies, which can create a mismatch when comparing annual earnings across companies. Our constant-horizon approach, described in the previous bullet point, handles this issue.

- The model draws on data from 93 countries, so to address currency mismatches, we convert all cash flows to U.S. dollars using the exchange rate at fiscal period-end. We obtain exchange rates from Compustat.

- The inflation rate varies a lot over our sample. Therefore, we adjust the dependent variable for the realized inflation between the forecast date and the cash flow realization, using the `CPIAUCNS` series from FRED. This means our model effectively forecasts real cash flows.

- When forecasting long-run cash flow, we face the problem that not all firms available at the forecast date have realized cash flows over all forecasting horizons. For example, the company may have been delisted or taken private. One approach would be to fit the model only to observations from firms with non-missing realized cash flows, but this would introduce a survivorship bias.[A1] Instead, we keep all firms that were alive on the forecast date, and impute missing cash flows by carrying forward the last non-missing cash flow with the firm's stock return, if available, and zero otherwise. By using the stock return, we also incorporate delisting returns, which is a flexible way of handling both good and bad delisting reasons.[A2]

- Equity issuances after the forecast date used, say, to acquire another company, could create a mechanical boost in future cash flow (the dependent variable) that would not necessarily accrue to the firm's current owners (because their ownership share would be diluted due to the equity issuance). Therefore, we divide the realized cash flows at all points by the number of shares outstanding, such that the dependent variable is the realized cash flow from owning one share at the forecast date.

- Working with a per-share variable, we need to adjust for stock splits that mechanically change the per-share value without any change in the fundamentals (e.g., a two-for-one split would cut the cash flow per share in half, but since owners get two shares for each share they owned, they are fundamentally unaffected). We make this adjustment using the `CFACSHR` variable from CRSP for U.S. stocks, and the `AJEXDI` variable from Compustat for non-U.S. stocks.

- Per share variables are somewhat arbitrary in the sense that the company can choose its number of shares outstanding, and this choice will lead to a difference in scale across

[A1] For example, if firms typically drop out of the sample for performance-related reasons, conditioning on firms that survive would introduce an upward bias in our cash flow forecasts.

[A2] For example, a firm that goes bankrupt will typically see its cash flows approach zero, whereas a firm that gets acquired will not.

companies. To handle this issue, we divide all realized cash flows by the most recent assets per share at the forecast date.

- Cash flow variables are computed from both annual and quarterly Compustat. When both are available for the same firm-date, the annual value is used.
- All target variables are winsorized at the 1st and 99th percentiles within each horizon-month combination.

Appendix B.2 Input features

Each prediction model uses 244 continuous and 138 categorical input features, grouped into four categories. Table A1 provides the full list of features by category.

JKP characteristics (148). We use 153 firm characteristics from Jensen et al. (2023), of which 5 are excluded as duplicates of our cash flow features (`gp_at`, `ocf_at`, `dbnetis_at`, `at_turnover`, `eqnetis_at`). These cover a broad range of anomaly categories, including momentum, value, profitability, investment, and trading frictions. Most variables are kept at their raw values, but three non-stationary features—`market_equity`, `dolvol_126d`, and `age`—are cross-sectionally rank-transformed each month. We refer to Jensen et al. (2023) for detailed definitions.

Cash flow features (84). We construct 21 additional accounting items from Compustat, each divided by total assets to normalize scale: sales, cost of goods sold, gross profit, SG&A (with and without R&D), R&D, EBITDA, depreciation & amortization, EBIT, net income, operating cash flow, free cash flow, free cash flow to equity, unlevered free cash flow, capital expenditures, change in net operating working capital, net debt issuance, dividends, unlevered net income, interest expense, and the marginal tax rate. For each of these 21 variables, we compute 1-, 2-, and 3-year changes, yielding 63 additional features that capture recent trends. The total is $21 + 63 = 84$ cash flow features.

IBES analyst forecasts (12). We use consensus EPS forecasts for fiscal years 1–3 and the long-term growth rate (LTG) from I/B/E/S, each scaled by assets per share. For each of these 4 variables, we also compute the industry median (Fama-French 49) and the country median, yielding 12 features in total. Industry and market medians require at least 10 non-missing observations.

Categorical features. Fama-French 49 industry classification (45 active categories, excluding the 4 financial industries) and country (93 countries) are passed as native LightGBM categoricals rather than one-hot encoded.

Missing values. LightGBM handles missing values natively by learning optimal split directions for missing inputs. We therefore retain all observations regardless of missing explanatory variables.

Table A1: Input Features to Cash Flow Prediction Models

Category	Variables
JKP (148)	age, aliq_at, aliq_mat, ami_126d, at_be, at_gr1, at_me, be_gr1a, be_me, beta_60m, beta_dimson_21d, betabab_1260d, betadown_252d, bev_mev, bidaskhl_21d, capex_abn, capx_gr1, capx_gr2, capx_gr3, cash_at, chcsho_12m, coa_gr1a, col_gr1a, cop_at, cop_atl1, corr_1260d, coskew_21d, cowc_gr1a, debt_gr3, debt_me, dgp_dsale, div12m_me, dolvol_126d, dolvol_var_126d, dsale_dinv, dsale_drec, dsale_dsga, earnings_variability, ebit_bev, ebit_sale, ebitda_mev, emp_gr1, eq_dur, eqnpo_12m, eqnpo_me, eqpo_me, f_score, fcf_me, fnl_gr1a, gp_atl1, inv_gr1, inv_gr1a, iskew_capm_21d, iskew_ff3_21d, iskew_hxz4_21d, ival_me, ivol_capm_21d, ivol_capm_252d, ivol_ff3_21d, ivol_hxz4_21d, kz_index, lnoa_gr1a, lti_gr1a, market_equity, mispricing_mgmt, mispricing_perf, ncoa_gr1a, ncol_gr1a, netdebt_me, netis_at, nfna_gr1a, ni_ar1, ni_be, ni_inc8q, ni_ivol, ni_me, niq_at, niq_at_chg1, niq_be, niq_be_chg1, niq_su, nncoa_gr1a, noa_at, noa_gr1a, o_score, oaccruals_at, oaccruals_ni, ocf_at_chg1, ocf_me, ocfq_saleq_std, op_at, op_atl1, ope_be, ope_bel1, opex_at, pi_nix, ppeinv_gr1a, prc, prc_highprc_252d, qmj, qmj_growth, qmj_prof, qmj_safety, rd5_at, rd_me, rd_sale, resff3_12_1, resff3_6_1, ret_12_1, ret_12_7, ret_1_0, ret_3_1, ret_60_12, ret_6_1, ret_9_1, rmax1_21d, rmax5_21d, rmax5_rvol_21d, rskew_21d, rvol_21d, sale_bev, sale_emp_gr1, sale_gr1, sale_gr3, sale_me, saleq_gr1, saleq_su, seas_11_15an, seas_11_15na, seas_16_20an, seas_16_20na, seas_1_1an, seas_1_1na, seas_2_5an, seas_2_5na, seas_6_10an, seas_6_10na, sti_gr1a, taccruals_at, taccruals_ni, tangibility, tax_gr1a, turnover_126d, turnover_var_126d, z_score, zero_trades_126d, zero_trades_21d, zero_trades_252d
CF levels (21)	sale_at, cogs_at, gp_at, xsga_at, xrd_at, sga_at, ebitda_at, dp_at, ebit_at, ni_at, ocf_at, fcf_at, fcfe_at, ufcf_at, capx_at, dNOWC_at, dvt_at, uni_at, xint_at, dbnetis_at, mtr
CF changes (63)	1-, 2-, and 3-year changes of each of the 21 cash flow level variables (e.g., sale_at_ch1, sale_at_ch2, sale_at_ch3)

Input Features to Cash Flow Prediction Models (continued)

Category	Variables
IBES (12)	`eps_fcst1`, `eps_fcst2`, `eps_fcst3`, `ltg`, `eps_fcst1_ind`, `eps_fcst2_ind`, `eps_fcst3_ind`, `ltg_ind`, `eps_fcst1_mkt`, `eps_fcst2_mkt`, `eps_fcst3_mkt`, `ltg_mkt`
Industry (45)	Fama-French 49 industry classification, excluding the 4 financial industries (banks, insurance, real estate, trading)
Country (93)	Country indicator

This table lists the input features used in the cash flow prediction models, grouped by category. The number in parentheses shows the number of variables. JKP variables are firm characteristics from Jensen et al. (2023). CF levels are accounting items scaled by total assets. CF changes are 1-, 2-, and 3-year changes of the CF level variables. IBES variables are I/B/E/S consensus analyst forecasts at the firm, industry, and market level. Industry and country are native LightGBM categorical features.

Appendix B.3 ML model

Model We use gradient-boosted decision trees implemented in LightGBM (Ke et al., 2017), with boosting type `gbdt`, a regression objective, and RMSE as the evaluation metric. We fit a separate model for each cash flow measure (free cash flow to equity and net income) at each of the ten forecast horizons, for a total of 20 models. Each model pools observations across all countries.

Hyperparameter search space We search over five hyperparameters using a 20-point max-entropy space-filling design. Table A2 summarizes the search space.

Two-stage tuning For each annual training window, we tune hyperparameters in two stages:

1. **Stage 1 (tree structure).** We evaluate all 20 candidate hyperparameter sets with a learning rate of $\eta_1 = 0.3$ and up to 1,000 boosting rounds. Training stops early if the validation RMSE does not improve for 25 consecutive rounds. The hyperparameter set with the lowest validation RMSE is selected.

2. **Stage 2 (iteration count).** We take the winning hyperparameter set from Stage 1 and reduce the learning rate to $\eta_2 = 0.05$. The model is trained for up to 10,000 boosting rounds with the same early stopping rule (25 rounds). This stage determines the optimal number of boosting iterations at the lower learning rate.

Parameter	Range	Scale
`feature_fraction`	$[1/P,\ 1]$	Linear
`num_leaves`	$[1,\ 127]$	Linear
`min_data_in_leaf`	$[1,\ 500]$	Linear
`bagging_fraction`	$[0.2,\ 1.0]$	Linear
`lambda_l2`	$[10^{-4},\ 10^{5}]$	Log

This table reports the hyperparameter search space for LightGBM. Each model is tuned over a 20-point max-entropy space-filling design. P denotes the total number of features. `num_leaves` $= 1$ to 127 corresponds to tree depths of 1 to 7 via $2^d - 1$. Two additional parameters are held fixed: `lambda_l1` $= 0$ (no L1 regularization) and `max_depth` $= -1$ (no explicit depth limit, as tree complexity is controlled through `num_leaves`).

Table A2: Hyperparameter Search Space

The two-stage design balances computational cost with model quality: the coarse first stage efficiently identifies good tree structures, while the fine second stage precisely calibrates the ensemble size.

Appendix B.4 Train, validation, and test splits

We create a temporal train-validation-test sample split, where the train and validation data are used to find hyperparameters, and the test sample is used to make out-of-sample forecasts.

- **Test set.** A one-year window of observations for which we generate forecasts (e.g., January–December 1984 for the first split).
- **Validation set.** All observations whose target realization period ends before the test set begins, going back 10 years.
- **Training set.** All remaining observations whose target realization period ends before the validation set begins.

Once the hyperparameters have been selected, we refit the model on the combined training and validation data using the optimal tree parameters from the first stage hyperparameter tuning, the learning rate of 0.05, and the optimal number of rounds from the second stage hyperparameter tuning.

The model parameters are updated every year, using an expanding training window. Specifically, every year, we add one year to the training set, while the validation and test set

start and end dates are increased by one year. For the first split, all pre-test data ends on December 31, 1983, and splits are re-estimated annually through December 31, 2023. This yields 40 annual re-estimations per cash flow measure and horizon.

Appendix C The Market-Based Cost of Equity

This section documents the implementation choices involved in computing the market-based cost of equity (MCE) from the machine-learned cash flow forecasts, using the valuation frameworks in (9) and (12).

Terminal growth rate. The terminal growth rate for cash flows is the country-specific 5-year forward real GDP growth rate (that is, the expected real growth between years $t+4$ and $t+5$) from the IMF World Economic Outlook. We use a forward rate rather than a cumulative rate because the terminal value should reflect steady-state growth. For example, during the Global Financial Crisis, cumulative forecasts over any horizon were depressed by the immediate recession, whereas the forward rate between years 4 and 5 was much less affected, better reflecting long-run growth prospects. The IMF series begins in April 1990; for earlier dates, we use the April 1990 forecast for each country.

Forecast smoothing. Before computing the MCE, the 10-horizon raw per-share forecasts for each firm-month are smoothed by projecting them onto a quadratic basis $[1, h, h^2]$ via OLS, where $h = 1, \ldots, 10$ is the forecast horizon.

Half-year discounting. Cash flows are discounted at $(1+r)^{h-0.5}$ rather than $(1+r)^h$, reflecting the assumption that cash flows arrive at the midpoint of each period rather than at the end.

Root-finding. The cost of equity is found via vectorized bisection with a convergence tolerance of 10^{-8} and a maximum of 100 iterations. The search interval is $[\max(-0.5,\ g + 0.001),\ 0.5]$, where g is the terminal growth rate. This ensures that the discount rate strictly exceeds the terminal growth rate, as required for the terminal value to be meaningful.

Validity screening. Cost of equity estimates are set to missing when any of the following conditions hold: (a) the net present value increases with the discount rate; (b) no root exists because the NPV has the same sign at both bounds; or (c) the implied real cost of equity exceeds 10%.

Real to nominal conversion. Because the cash flow forecasts are in real terms and the terminal growth rate is a real GDP growth rate, the root-finding procedure yields a real cost of equity. We convert to a nominal cost of equity by adding the expected U.S. inflation rate. Expected inflation is the 5-year forward CPI inflation rate from the IMF World Economic Outlook, matching the source used for the terminal growth rate. Since all cash flows are denominated in U.S. dollars, we use the U.S. forecast for all firms. For dates before the IMF series begins (April 1990), we use the trend inflation rate from Bauer and Rudebusch (2020).

Combining NI and CF estimates. The combined cost of equity first demeans each model type (residual income and free cash flow to equity) within each country-type-month group to remove level differences across models, averages the demeaned values, and then adds back the overall country-month mean. When only one model produces a valid estimate for a given firm-month, that estimate is used directly.

Non-U.S. start date. Non-U.S. firms before January 1996 are excluded from the MCE sample. The non-U.S. accounting data starts in 1985, so this screen ensures that the cash flow models have at least ten years of non-U.S. data in their training data.

Appendix D Returns and Cash Flow Benchmarks

In this section we describe the methods that we use as benchmarks for our cash flow and returns forecasts.

Appendix D.1 Analyst-based ICC methods

We implement four analyst-based implied cost of capital methods following Eskildsen et al. (2024): Ohlson and Juettner-Nauroth (2005, OJ) and Easton (2004, PEG), which are based on the dividend discount model, and Gebhardt et al. (2001, GLS) and Claus and Thomas (2001, CT), which are based on the residual income model.

All four methods use analyst forecasts of earnings per share (EPS) from the I/B/E/S consensus file (median forecast). We denote the one- and two-year EPS forecasts as $\hat{\mathrm{E}}_t[e_{t+1}]$ and $\hat{\mathrm{E}}_t[e_{t+2}]$, the long-term growth forecast as LTG, the book equity per share as b_t, and the stock price as p_t. When the two-year EPS forecast is unavailable, we set $\hat{\mathrm{E}}_t[e_{t+2}] = \hat{\mathrm{E}}_t[e_{t+1}] \times 1.15$. We estimate the payout ratio as dividends divided by earnings from the last fiscal year. For firms with negative earnings, we follow Mohanram and Gode (2013) and set the payout ratio to 6% of total assets.

OJ. The Ohlson and Juettner-Nauroth (2005) method provides a closed-form solution for the implied cost of capital:

$$\text{ICC}^{OJ} = A + \sqrt{A^2 + \frac{\hat{\text{E}}_t[e_{t+1}]}{p_t} \times (\text{STG} - \lambda)}, \tag{A4}$$

where

$$A = \frac{1}{2}\left(\lambda + \frac{\hat{\text{E}}_t[d_{t+1}]}{p_t}\right) \quad \text{and} \quad \text{STG} = \max\left[\left(\frac{\hat{\text{E}}_t[e_{t+2}] - \hat{\text{E}}_t[e_{t+1}]}{\hat{\text{E}}_t[e_{t+1}]} \times \text{LTG}\right)^{\frac{1}{2}}, \text{LTG}\right]. \tag{A5}$$

Here, $\hat{\text{E}}_t[d_{t+1}]$ is next year's expected dividend, computed as the payout ratio times $\hat{\text{E}}_t[e_{t+1}]$. The parameter λ captures expected long-run economic growth and is set to the yield on a ten-year U.S. treasury bond minus 3%.[A3]

PEG. The Easton (2004) method simplifies the OJ formula by setting $\lambda = 0$ and ignoring dividends:

$$\text{ICC}^{PEG} = \sqrt{\frac{\hat{\text{E}}_t[e_{t+1}]}{p_t} \times \text{STG}}, \tag{A6}$$

with STG computed as in (A4).

Residual income framework. The GLS and CT methods share a common framework based on the residual income model:

$$p_t = b_t + \sum_{h=1}^{H}\left(\frac{\hat{\text{E}}_t\left[(\text{ROE}_{t+h} - r)b_{t+h-1}\right]}{(1+r)^h}\right) + \text{TV}, \tag{A7}$$

where $\text{ROE}_{t+h} = \hat{\text{E}}_t[e_{t+h}]/b_{t+h-1}$ and book value evolves via clean surplus accounting: $b_t = b_{t-1} + e_t - d_t$, with $d_t = e_t \times$ payout-ratio. The two methods differ in the forecast horizon H, the earnings forecasting procedure, and the terminal value specification.

GLS. The Gebhardt et al. (2001) method forecasts residual income over 12 years, and values the terminal residual income as a constant perpetuity. The implied cost of capital ICC^{GLS} solves:

$$p_t = b_t + \sum_{h=1}^{11}\left(\frac{\hat{\text{E}}_t\left[(\text{ROE}_{t+h} - \text{ICC}^{GLS})b_{t+h-1}\right]}{(1+\text{ICC}^{GLS})^h}\right) + \frac{\hat{\text{E}}_t\left[(\text{ROE}_{t+12} - \text{ICC}^{GLS})b_{t+11}\right]}{\text{ICC}^{GLS}(1+\text{ICC}^{GLS})^{11}}. \tag{A8}$$

[A3] Ohlson and Juettner-Nauroth (2005) define STG as the two-year growth in earnings, $(\hat{\text{E}}_t[e_{t+2}] - \hat{\text{E}}_t[e_{t+1}])/\hat{\text{E}}_t[e_{t+1}]$. To obtain a more stable estimate, we follow Mohanram and Gode (2013) and compute STG as the geometric mean of short- and long-term growth.

For years 1–2, EPS forecasts come directly from I/B/E/S. For years 3–12, ROE converges linearly from the analyst-implied ROE_{t+2} to the industry median ROE. We compute the industry median ROE using the 49 industries from Fama and French (1997), measured in U.S. dollars across all global firms with valid data.[A4]

CT. The Claus and Thomas (2001) method uses $H = 5$ explicit forecast periods and a growing perpetuity as terminal value. The implied cost of capital ICC^{CT} solves:

$$p_t = b_t + \sum_{h=1}^{5} \left(\frac{\hat{\text{E}}_t \left[(\text{ROE}_{t+h} - \text{ICC}^{CT}) b_{t+h-1} \right]}{(1 + \text{ICC}^{CT})^h} \right) + \frac{\hat{\text{E}}_t \left[(\text{ROE}_{t+5} - \text{ICC}^{CT}) b_{t+4} \right] (1+g)}{(\text{ICC}^{CT} - g)(1 + \text{ICC}^{CT})^5}, \quad \text{(A9)}$$

where g is the terminal growth rate, set equal to the yield on a ten-year U.S. treasury bond minus 3%. For years 1–2, EPS forecasts come from I/B/E/S. For years 3–5, EPS is grown from the second-year forecast using the LTG forecast from I/B/E/S.

Appendix D.2 The Market and the CAPM

The Market and CAPM models are used as benchmarks for our return forecasts. They make predictions over a ten-year horizon, matching the horizon in our main return benchmark analysis.

MKT. The market benchmark (which we use to compute out-of-sample R^2) is a country-level prediction based on the expanding-window average of realized equity risk premia. For each country, we compute rolling 10-year cumulative value-weighted market returns and subtract the 10-year risk-free rate compounded from the start of each window to obtain a 10-year equity risk premium (ERP):

$$\text{ERP}_{c,t} = \prod_{s=1}^{120} (1 + r^{mkt}_{c,t-120+s}) - (1 + r^{f,10}_{t-119})^{10}, \quad \text{(A10)}$$

where $r^{mkt}_{c,s}$ is the monthly value-weighted market return for country c and $r^{f,10}_{t-119}$ is the annualized 10-year risk-free rate at the start of the rolling window.

We then take an expanding-window average $\overline{\text{ERP}}_{c,t}$ of the realized 10-year ERP for each country, capped to the interval $[0, 3]$. Countries with insufficient history use the U.S. expanding-window average as a fill-in. The market benchmark prediction for 10-year cumulative returns is:

$$\hat{R}^{\text{MKT}}_{c,t} = (1 + r^{f,10}_t)^{10} - 1 + \overline{\text{ERP}}_{c,t}. \quad \text{(A11)}$$

[A4] Most ICC papers estimate industry ROE using U.S. firms only. Since our sample includes non-U.S. firms, we use a global industry median for consistency.

CAPM. The CAPM benchmark is a firm-level prediction that scales the country-level equity risk premium by each firm's market beta. We estimate β_i as the 60-month rolling CAPM beta from firm-level returns, capped to $[0, 3]$. The CAPM prediction for 10-year cumulative returns is:

$$\hat{R}_{i,t}^{\text{CAPM}} = (1 + r_t^{f,10})^{10} - 1 + \beta_i \times \overline{\text{ERP}}_{c,t}, \tag{A12}$$

where $\overline{\text{ERP}}_{c,t}$ is the same country-level expanding-window average used in the market benchmark.

Appendix D.3 Machine-learned returns (MLR)

A key return benchmark is a machine learning model that predicts returns directly, using a machine learning setup that closely follows our approach for forecasting cash flows (Appendix B), with the following differences:

1. **Dependent variable.** The target is the 10-year cumulative excess return, defined as the cumulative stock return from year t to $t + 10$ minus the compounded 10-year risk-free rate. We add the risk-free rate back to the model's predictions to obtain cumulative return forecasts.

2. **Single model.** We fit one model covering the full 10-year horizon, rather than 20 separate models (two cash flow measures $\times$ ten horizons). Each cash flow model targets a one-year window $[t+h-1,\ t+h]$; the return model targets the full $[t,\ t+10]$ window.

3. **Learning rates.** Using the same learning rate as for the cash flow forecasts gave poor results. We therefore lowered the learning rate to $\eta_1 = 0.01$ (vs. 0.3) for the first hyperparameter search stage and $\eta_2 = 0.001$ (vs. 0.05) for the second. The hyperparameter search space is otherwise identical.

4. **Validation buffer.** Since the return target spans ten years, we add an extra 10-year buffer between the validation and test sets. Without this buffer, validation observations near the test boundary would have targets that overlap the test period, contaminating the out-of-sample evaluation. The one-year cash flow targets do not require such a buffer.

5. **Data processing.** For non-U.S. firms, returns come from Compustat and some illiquid stocks have clear data errors. We therefore winsorize returns at the 0.1%/99.9% level before training.

Appendix E The Role of Delisting and Imputation

For our return predictability tests, such as the one in Table 1, we regress the stock's realized return over the next ten years on the MCE today. But many firms delist during the ten-year period, so we do not observe their realized returns over the full horizon. We handle this issue by imputing missing returns with the cross-sectional median. In this section, we assess the sensitivity of our results to alternative imputation methods. The result that the MCE predicts returns with a slope close to one is insensitive to this choice when the dependent variable is the cumulative return from t to $t+120$ months, but sensitive when the dependent variable is the single return realized in month $t+120$. The reason is that low-MCE firms delist far more often than high-MCE firms, and only about one-third of low-MCE firms survive the full ten years on average. This has a limited effect on the cumulative return estimates, where we can keep the observable return and only impute after the delisting, but a larger effect on the single-period return, because there is no observable return.

Appendix E.1 Methods for handling delisted firms

For a stock that delists at some time $t+d$ with $d \in \{1, 2, \ldots, 119\}$, our goal is to create a complete return series between t and $t+120$. To impute its return beyond the delisting point, we create an imputed return $r_t^{\text{imp(j)}}$ for imputation method j. We consider four imputation methods: setting missing returns to zero ("zero"), the risk-free rate ("risk-free"), the median return within a country and month ("median"), or the average return within a country and month ("mean").

To compute a delisted stock's cumulative return over a ten-year period, we take the product of its observed and imputed return:

$$r_{t\to t+120}^{i(j)} = \left(\prod_{h=1}^{d}\left[1+r_{t+h}^{i}\right]\right)\left(\prod_{h=d+1}^{120}\left[1+r_{t+h}^{\text{imp(j)}}\right]\right) - 1, \tag{A13}$$

where $r_{t\to t+120}^{i(j)}$ is the ten-year realized total return of stock i for imputation method j, and r_t^i is the stock's observed return, including its delisting return.

We also consider two methods that do not require any imputation. The first uses only the observable return as the dependent variable, and uses an MCE that matches the survival horizon as the explanatory variable ("match"). Finally, we also report the selection-biased case where we discard firms that delist between t and $t+120$ ("No imputation").

Appendix E.2 Ten-year return predictability robustness

Table A3 shows the slope parameter estimate for ten-year cumulative realized returns regressed on the MCE:

$$r^{i(j)}_{t\to t+h(j)} = \alpha_t + b \times \left[\left(1+\text{MCE}^i_t\right)^{k(j)} - 1\right] + \varepsilon^{i(j)}_{t+h(j)}, \tag{A14}$$

where $h(j)$ is the horizon, which is 120 for all methods except match, where it is d, and the MCE exponent is $k(j) = h(j)/12$.

The slope is large, positive, significant, and close to one across most methods in both the U.S. and globally. The outliers are the non-imputation methods. Removing delisted firms using the non-imputation methods yields a noticeably lower slope, particularly in the U.S., suggesting that the resulting selection bias correlates with the MCE—an assertion that we will provide evidence for in the next section. By contrast, the match method delivers a noticeably higher parameter estimate.

	U.S.			Global		
	Slope	S.E.	N	Slope	S.E.	N
Zero	1.00***	0.11	1,106,464	1.00***	0.17	2,610,489
Risk-free	0.99***	0.10	1,106,464	0.99***	0.16	2,610,489
Median (baseline)	0.93***	0.08	1,106,464	0.96***	0.18	2,610,489
Mean	0.83***	0.06	1,106,464	0.90***	0.17	2,610,489
Match	1.41***	0.10	1,106,464	1.27***	0.12	2,610,489
No imputation	0.44*	0.23	584,079	0.85***	0.24	1,744,048

The table reports regressions of ten-year cumulative buy-and-hold returns on the current ten-year cost of equity, estimated separately for the U.S. and global samples, under alternative rules for the return of a firm after it delists. "Zero" sets the missing returns to zero; "Risk-free" to the one-month risk-free rate; "Median" (the baseline used in the main text) to the median return within a country and month; "Mean" to the average return within a country and month. "Match" keeps every firm but measures both the dependent variable and the cost of equity over the firm's observed survival horizon. "No imputation" discards firms that delist between t and $t+120$. All regressions include month fixed effects. We report Driscoll-Kraay standard errors using a bandwidth of 120 months. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table A3: cumulative return predictability robustness

Table A4 shows the slope parameter estimate for the return in month $t+120$ regressed on the MCE:

$$r^{i(j)}_{t+120} = \alpha_t + b \times \left[\left(1+\text{MCE}^i_t\right)^{1/12} - 1\right] + \varepsilon^{i(j)}_{t+120}, \tag{A15}$$

where $r^{i(j)}_{t+120}$ is the observable return, if available, and otherwise the imputed return. We include the same methods as before, except for match, which is designed specifically for the

cumulative return regression.

The single-period results are less robust than the cumulative ones, especially in the U.S. Excluding the firms that delist drives the parameter estimate negative, while imputing with the mean lowers the estimate to an insignificant 0.13—a substantial difference relative to the baseline estimate of 0.67. The effect is smaller globally, where the estimate goes from 0.72 under the baseline to 0.42 and 0.44 with mean and no imputation, respectively. By contrast, imputing with zero or the risk-free rate increases the estimate both in the U.S. and globally.

	U.S.			Global		
	Slope	S.E.	N	Slope	S.E.	N
Zero	0.98***	0.35	1,106,464	0.91***	0.27	2,610,489
Risk-free	0.84**	0.35	1,106,464	0.84***	0.28	2,610,489
Median (baseline)	0.67***	0.23	1,106,464	0.72***	0.25	2,610,489
Mean	0.13	0.26	1,106,464	0.42	0.26	2,610,489
No imputation	-0.13	0.50	584,079	0.44	0.34	1,744,048

The table reports regressions of the single realized return in month $t + 120$ on the current one-month cost of equity, estimated separately for the U.S. and global samples, under alternative rules for the return of a firm after it delists. "Zero" sets the missing returns to zero; "Risk-free" to the one-month risk-free rate; "Median" (the baseline used in the main text) to the median return within a country and month; "Mean" to the average return within a country and month. "No imputation" discards firms that delist between t and $t + 120$. All regressions include month fixed effects. We report standard errors clustered by month. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table A4: single-period return predictability robustness

Figure A1 shows how the estimates vary across horizons from $t+1$ to $t+120$. The main finding is that most methods agree on horizons of two years or less, but diverge thereafter. Again, the cumulative return slopes are more similar than the single-period ones, and so are the global relative to the U.S. results.

Appendix E.3 MCE correlates with delisting

The difference between the no imputation and imputation results suggests that the MCE correlates with a firm's delisting probability. To show this explicitly and analyze the reason, we leverage the fact that CRSP provides detailed delisting reasons. CRSP covers only U.S. stocks, so this analysis is limited to the U.S. sample. Concretely, we sort stocks into ten MCE deciles each month, and then record whether the firm delists in the next 120 months and, if so, its delisting reason.

Figure A1
Return-predictability slope by horizon across delisting methods

Each panel plots the slope b from the return-predictability regression at every horizon $h = 1, \ldots, 120$ months, for the U.S. and global samples (columns) and for the cumulative buy-and-hold return $r^i_{t\to t+h}$ and the single return r^i_{t+h} (rows). The cumulative-return slope is estimated on the h-year cost of equity, and the single-return slope on the one-month cost of equity. Each line is one rule for handling delisted firms, as in Tables A3 and A4: "zero", "risk-free", "median", "mean", and "no imputation" fill or drop the post-delisting returns, while "match" (cumulative only) regresses the observed return through delisting on the cost of equity over the matched survival horizon. All regressions include month fixed effects; the dashed line marks a slope of one.

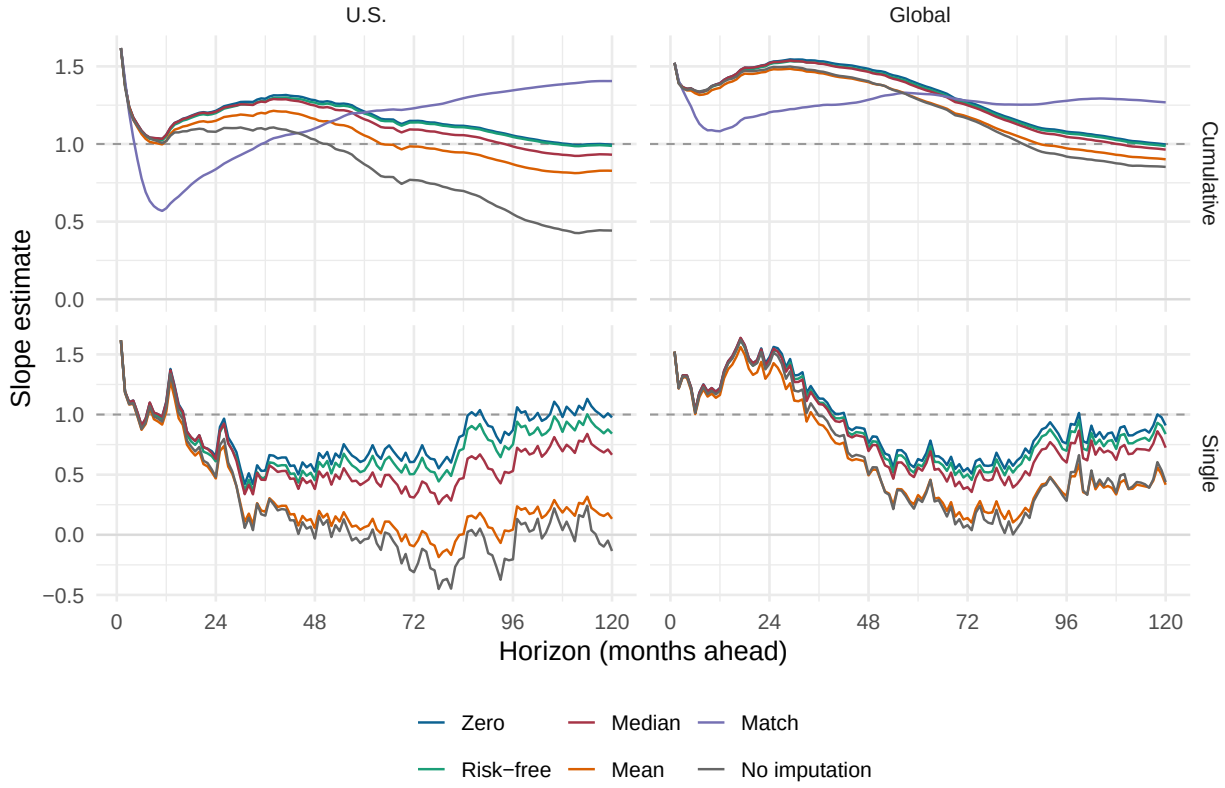


Table A5 shows the fraction of U.S. firms that delist within each MCE group, along with the reason why. Delisting is strongly correlated with the MCE: the low MCE group has a delisting rate of 62.7%, while the rate for the high MCE is 29.1%. The fraction of "bad" delistings is also higher in the low MCE group. Specifically, the fraction of firms in the worst category—dropped—is 40.7% in the low group and 28.4% in the high one. Similarly, the fraction of firms in the best category—merger—is 57.8% for the low group and 67.1% in the high one.[A5]

[A5] The dropped category is the worst group in the sense that its average delisting returns of −14% and its

			composition of those delistings (%)			
MCE decile	N	Delist 10y %	Merger	Exchange	Liquidation	Dropped
1 (low)	113,499	62.7	57.8	1.1	0.4	40.7
2	113,298	59.6	65.4	1.4	0.4	32.8
3	113,280	57.9	68.5	1.7	0.3	29.5
4	113,308	55.5	71.0	2.2	0.4	26.5
5	113,343	51.7	72.3	2.7	0.3	24.6
6	113,222	46.5	73.7	2.7	0.4	23.2
7	113,284	41.9	73.9	3.2	0.4	22.5
8	113,314	36.4	74.1	3.2	0.5	22.2
9	113,264	30.7	72.3	3.7	0.5	23.4
10 (high)	113,467	29.1	67.1	3.8	0.7	28.4

Each month, U.S. firms are sorted into ten cost-of-equity (MCE) deciles. For each firm-month, we record whether the firm delists over the next ten years, and "Delist 10y" is the resulting share. The next four columns show the reason, by putting delistings into four groups measured using CRSP: "Merger," the firm is acquired by or combined with another entity; "Exchange," it moves to a different exchange; "Liquidation," it liquidates and returns capital to shareholders; and "Dropped," it is removed for reasons like bankruptcy, low price, or failure to meet listing requirements. "Dropped" is the worst category as measured by its average delisting return and return in the year prior to delisting, while "merger" is the best on both measures. Exchanges and liquidations are comparatively rare and carry near-zero average delisting returns. The sample is U.S. firms.

Table A5: delisting and its reasons by cost-of-equity decile

Appendix F Comparing MCE and MLR

This section examines how the MCE and the machine-learned return forecast (MLR) relate to each other as predictors of 10-year realized returns, and how they perform in combination. All results in this section use the global sample underlying Table 1, restricted to firm-months where both the MCE and the MLR are available.

Table A6 reports panel regressions of 10-year realized buy-and-hold returns on the 10-year MCE, the MLR, and both jointly, with month fixed effects and Driscoll-Kraay standard errors. Panel B reports the mean squared error and the out-of-sample R^2 relative to the market benchmark implied by each column's fitted values. Table A7 reports the fit of various combinations of the two estimates, $w_{\text{MCE}}\text{MCE} + (1 - w_{\text{MCE}})\text{MLR}$, as the weight w_{MCE} on the MCE varies from zero to one in increments of 0.1.

average returns in the year before delisting of -44% are the lowest among the four groups. By contrast, the merger group is the best, with comparable numbers of 1.4% and 46.4%. These two groups also account for the vast majority of delistings, while exchange changes and liquidations are much rarer.

	10-year realized return		
	(i)	(ii)	(iii)
Panel A: Regression estimates			
MCE	0.96***		0.85***
	(0.18)		(0.20)
MLR		0.31**	0.23**
		(0.13)	(0.11)
Observations	2,610,475	2,610,475	2,610,475
Panel B: In-sample fit			
MSE	24.14	24.20	24.11
Within R^2 (%)	0.52	0.28	0.67
R^2_{MKT} (%)	3.27	3.05	3.42
Panel C: Out-of-sample fit ($\alpha = 0, \beta = 1$)			
OOS MSE	24.44	25.37	—
OOS R^2_{MKT} (%)	2.10	-1.65	—

The table reports panel regressions of 10-year realized buy-and-hold returns on the current 10-year cost of equity from the MCE and MLR methods. All regressions include month fixed effects and are estimated on the same sample. In Panel B, MSE is the mean squared error of each column's fitted values (including the estimated month fixed effects), Within R^2 is the within R^2 of each regression, and R^2_{MKT} is computed as $1-\sum(r_i-\hat{r}_i)^2/\sum(r_i-\hat{r}_{i,\text{MKT}})^2$, where $\hat{r}_i$ is the column's fitted value and the benchmark prediction $\hat{r}_{i,\text{MKT}}$ is the country-level expanding-window average market return. Because the fitted values embed the slope and month fixed effects estimated on the test sample, all Panel B statistics measure in-sample fit. Panel C reports the same statistics for the raw forecast without estimated parameters, i.e., imposing $\alpha = 0$ and $\beta = 1$; columns (i) and (ii) correspond to the $w_{\text{MCE}} = 1$ and $w_{\text{MCE}} = 0$ rows of Table A7, and the joint column has no single raw counterpart. We report Driscoll-Kraay standard errors using a 120-month window in parentheses. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Return predictability – MCE and MLR

Appendix G Additional Figures and Tables

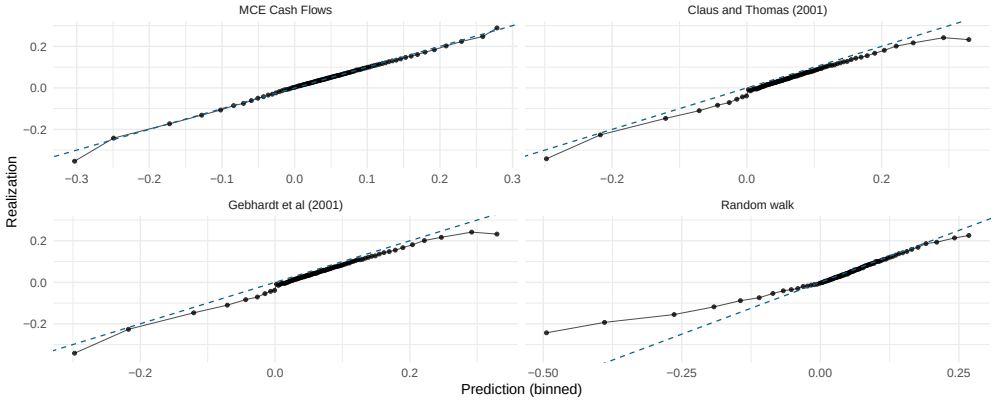
w_{MCE}	$1 - w_{\text{MCE}}$	IS within R^2 (%)	OOS MSE	OOS R^2_{MKT} (%)
0.0	1.0	0.28	25.37	-1.65
0.1	0.9	0.31	25.15	-0.74
0.2	0.8	0.35	24.95	0.05
0.3	0.7	0.40	24.78	0.72
0.4	0.6	0.46	24.64	1.27
0.5	0.5	0.52	24.53	1.71
0.6	0.4	0.59	24.46	2.02
0.7	0.3	0.65	24.41	2.22
0.8	0.2	0.67	24.39	2.30
0.9	0.1	0.62	24.40	2.26
1.0	0.0	0.52	24.44	2.10

The table reports the fit of various combinations of the MCE and the MLR. Columns are labeled IS or OOS depending on whether the statistic relies on parameters estimated on the evaluation sample. IS within R^2 is the in-sample within R^2 from a panel regression of realized returns on the combination with month fixed effects, where the slope and fixed effects are estimated on the evaluation sample. OOS MSE is the mean squared error of the raw combination for 10-year realized buy-and-hold returns; it uses no estimated parameters. OOS R^2_{MKT} is computed from the raw combination as $1 - \sum(r_i - \hat{r}_i(w))^2 / \sum(r_i - \hat{r}_{i,\text{MKT}})^2$, where the benchmark prediction is the country-level expanding-window average market return. The sample is the same across all rows and is identical to the sample in Table A6.

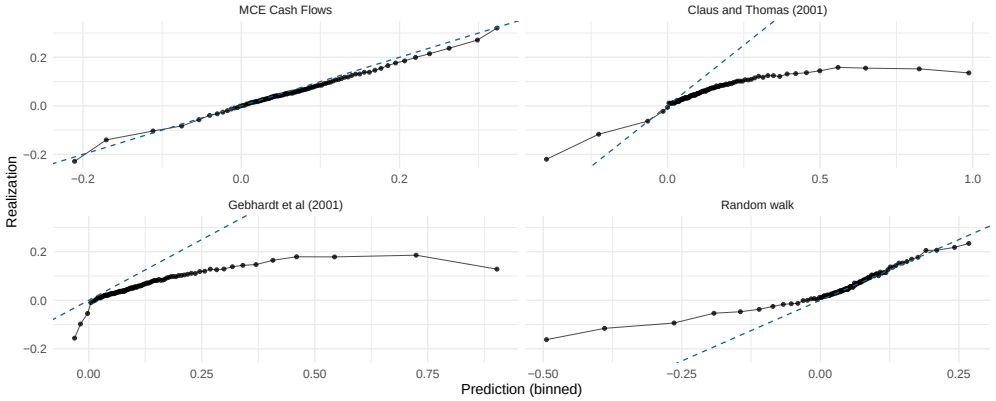
Table A7: Combinations of MCE and MLR

Figure A2
Cash flow predictability

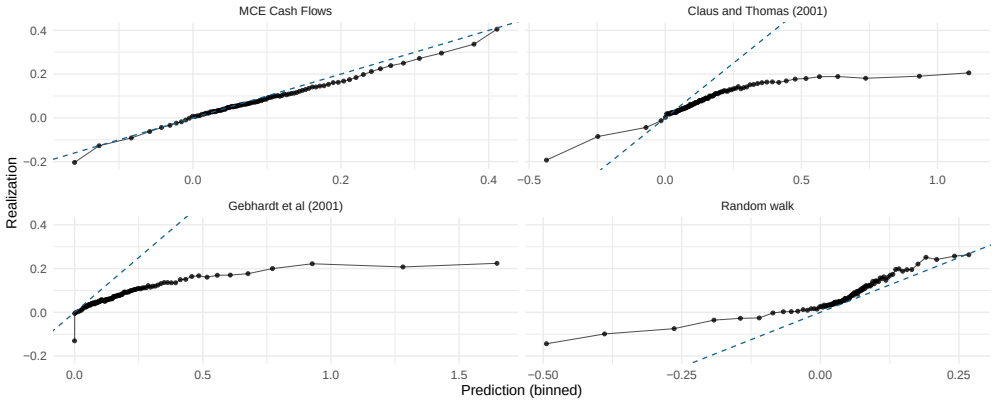
This figure shows binned scatter plots of realized earnings divided by assets at t on cash flow predictions at t, for forecast horizons $h = 1, 5, 10$ years.



Panel A. $h = 1$



Panel B. $h = 5$



Panel C. $h = 10$

Figure A3
Cash flow forecasts (unbalanced)

This figure shows the out-of-sample R^2 of the cash flow prediction models. We consider three different cash flow variables, as indicated by the panel header, and ten forecast horizons, as indicated by the x-axis. The sample is unbalanced, meaning that we use all data available and that the samples differ across variables and horizons. Figure 5 shows the results from the balanced sample.

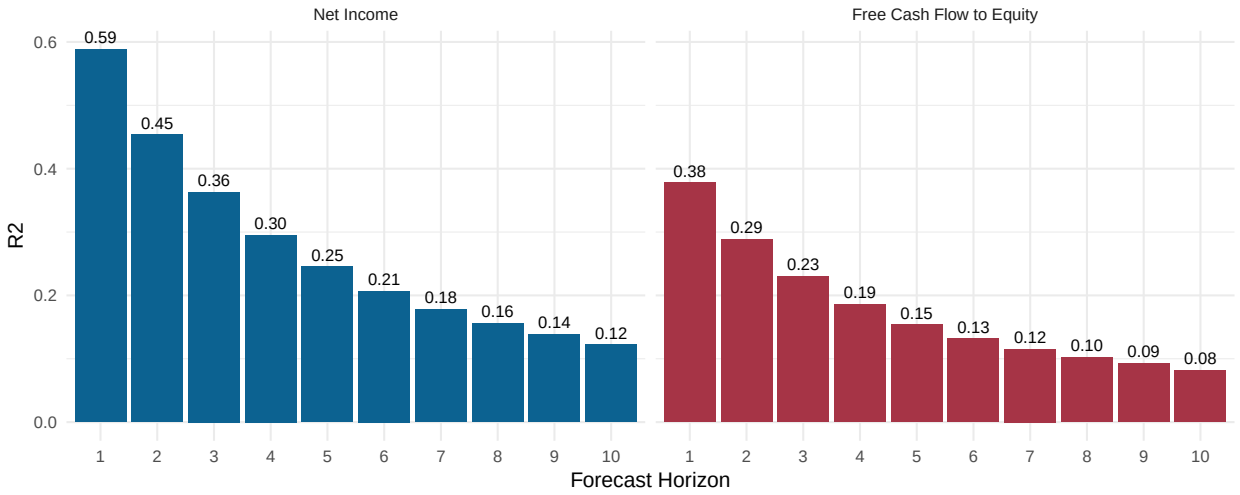


Figure A4
Cash flow forecasts over time (balanced)

This figure shows the out-of-sample R^2 of the cash flow prediction models every month. The panel header shows the cash flow variable, and the x-axis is the cash flow realization date. The sample is balanced. Figure A5 shows the results from the unbalanced sample.

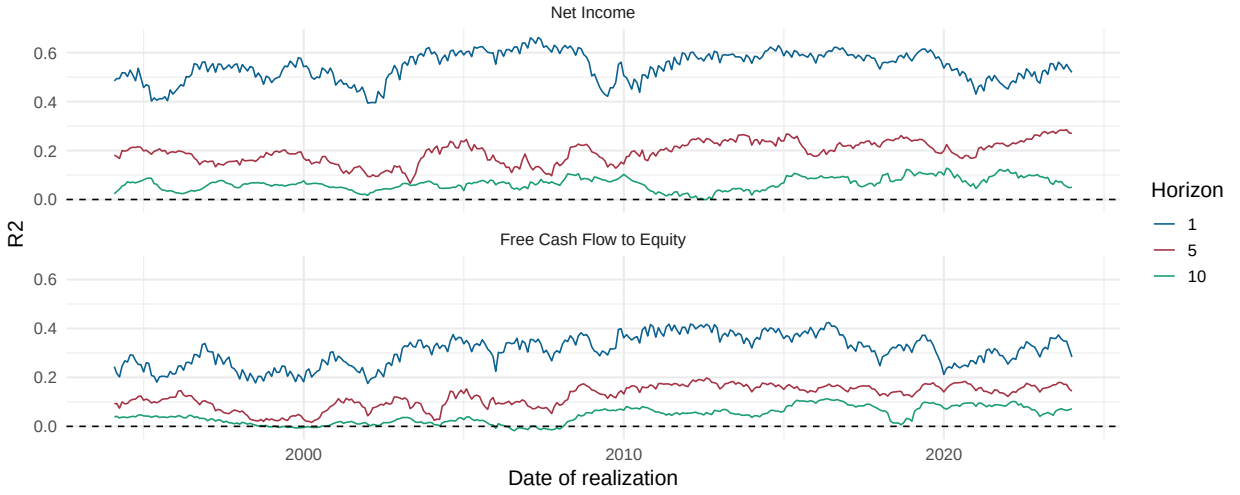


Figure A5
Cash flow forecasts over time (unbalanced)

This figure shows the out-of-sample R^2 of the cash flow prediction models every month. The panel header shows the cash flow variable, and the x-axis is the cash flow realization date. The sample is unbalanced. Figure A4 shows the results from the balanced sample.

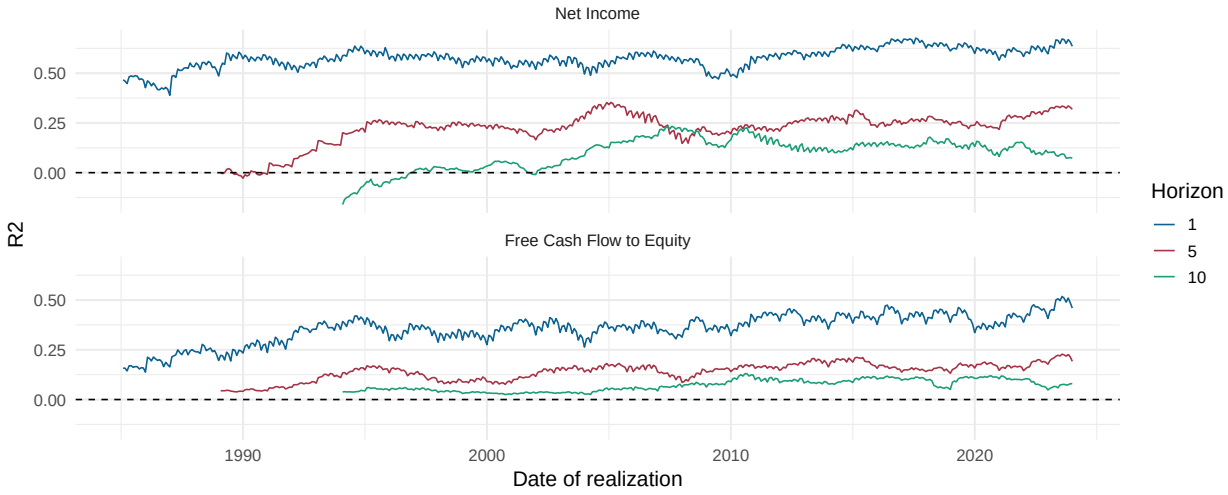


Figure A6
Long-run returns and the MCE (U.S. sample)

This figure shows the relationship between future 10-year realized returns and the current 10-year cost of equity. Specifically, each month we sort stocks into 100 portfolios by their cost of equity, and then compute the average 10-year cost of equity and the average realized buy-and-hold return over the next 10 years. We then average these two numbers for each portfolio over the full sample, 1983–2023. For ease of interpretation, the cost of equity and realized returns have been annualized. The solid red line shows the best linear fit, and the dotted blue line shows the 45-degree line. The sample is **U.S. stocks.**

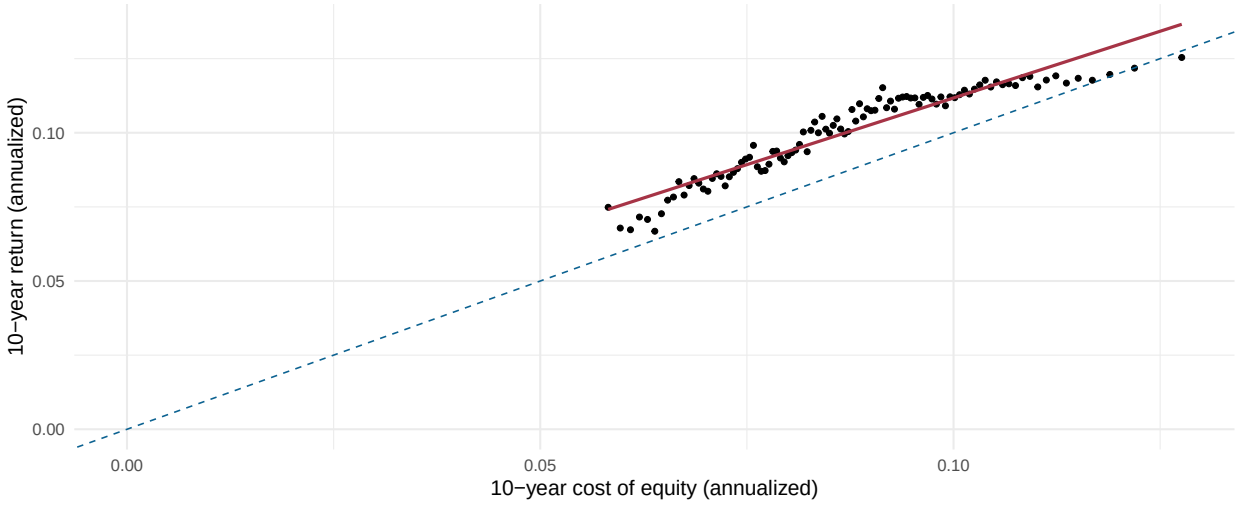


Figure A7
Long-run returns and the MCE (non-U.S. sample)

This figure shows the relationship between future 10-year realized returns and the current 10-year cost of equity. Specifically, each month we sort stocks into 100 portfolios by their cost of equity, and then compute the average 10-year cost of equity and the average realized buy-and-hold return over the next 10 years. We then average these two numbers for each portfolio over the full sample, 1983–2023. For ease of interpretation, the cost of equity and realized returns have been annualized. The solid red line shows the best linear fit, and the dotted blue line shows the 45-degree line. The sample consists of **non-U.S. stocks** listed in developed or emerging markets, as classified by MSCI.

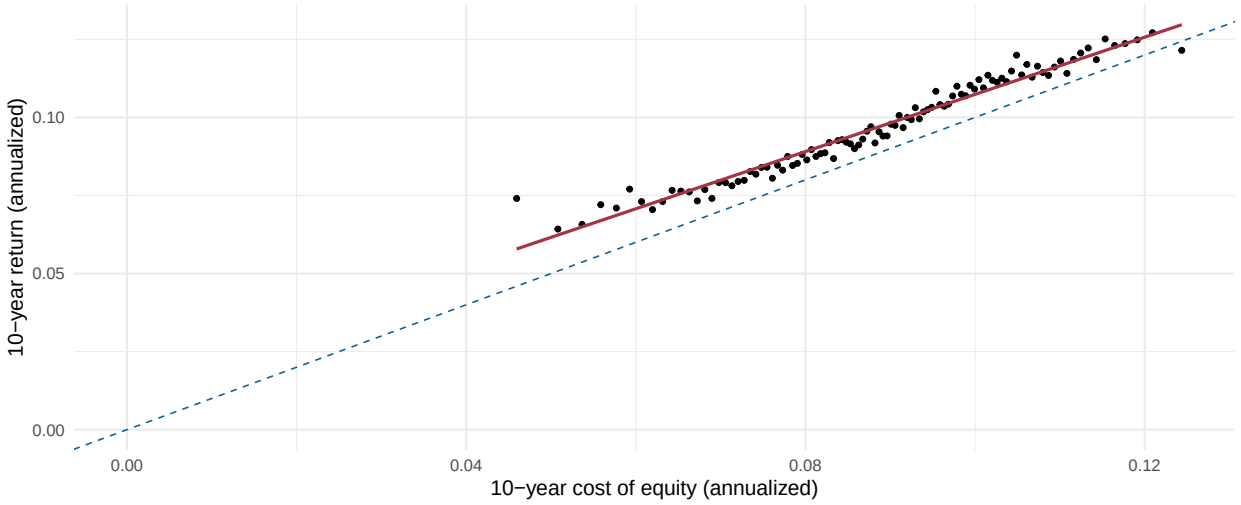


Figure A8
Forward return predictability: MCE versus Keloharju et al.

This figure compares the decay of forward CAPM alphas across return horizons for the market-based cost of equity (MCE) along with the results for Keloharju et al. (2021). Each month, stocks are sorted into long–short portfolios based on (lags of) the MCE signal using the Fama and French (1993) methodology. We estimate the CAPM alpha of forward returns at horizons $h = 1, \ldots, 10$ years (the "lag of the MCE characteristic underlying the portfolio sort"). Panel A plots the annualized alphas (% per year) and Panel B the corresponding t-statistics. The MCE is shown with blue squares (left axis) and the Keloharju et al. characteristics with grey circles (right axis); note the separate vertical scales. In Panel B, the dashed and dotted horizontal lines mark the 10% and 5% two-sided significance thresholds on the MCE (left) axis. The MCE alpha stabilizes above zero and remains significant at the 10% level through year 7, whereas the Keloharju et al. alpha decays toward zero. The t-statistics are not directly comparable across the two series because the MCE is available for a shorter sample period. The sample is 1984–2023 for the MCE factor.

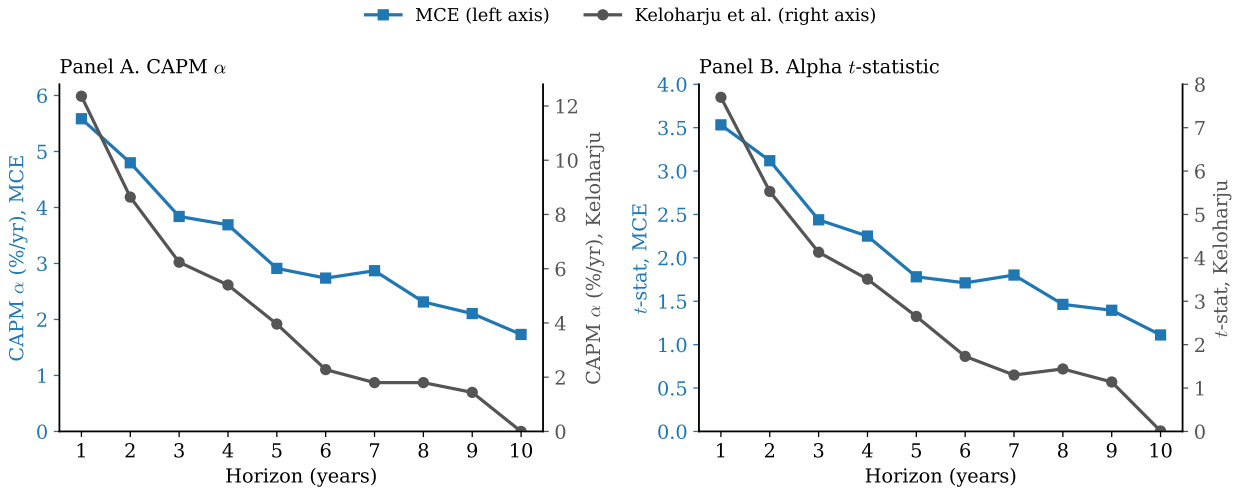


Figure A9
ML forecasts versus analysts

This figure compares the out-of-sample R^2 of one-year-ahead net income forecasts from the machine learning model, IBES analyst consensus forecasts, and a random walk (last observed net income). To ensure a comparable evaluation, the sample is restricted to analyst forecasts where the fiscal year end is exactly one year from the forecasting time, matching the constant-horizon property of the ML predictions. All forecasts and realizations are scaled by total assets and winsorized at the 1st and 99th percentiles. Firm-months where IBES and Compustat actual EPS disagree are excluded, and the sample requires non-missing predictions from all three methods. "Raw forecasts" computes R^2 as $1-\text{MSE}/\text{Var}(y)$. "De-meaned forecast" first subtracts the cross-sectional mean from both predictions and realizations before computing R^2, in an attempt to remove the well-known analyst level bias.

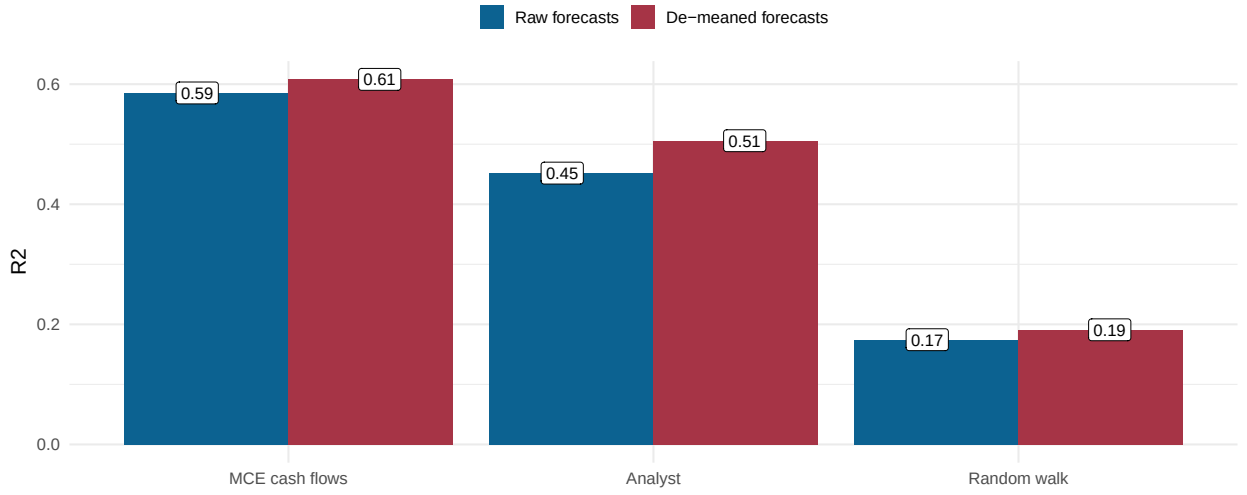


Figure A10
Optimal shrinkage analysis for return prediction

This figure shows the optimal ridge regression penalty parameter λ for a return prediction model that uses the same features as MLR but replaces LightGBM with ridge regression to isolate the role of shrinkage. For each training end date, we fit a ridge regression with λ selected from a grid of 1,000 values spaced evenly on a log scale from 10^0 to 10^6. The validation-optimal λ minimizes MSE on the validation set, while the test-optimal λ minimizes MSE on the out-of-sample test set. Features are cross-sectionally standardized (percentile-ranked, demeaned, and scaled to unit standard deviation) each month.

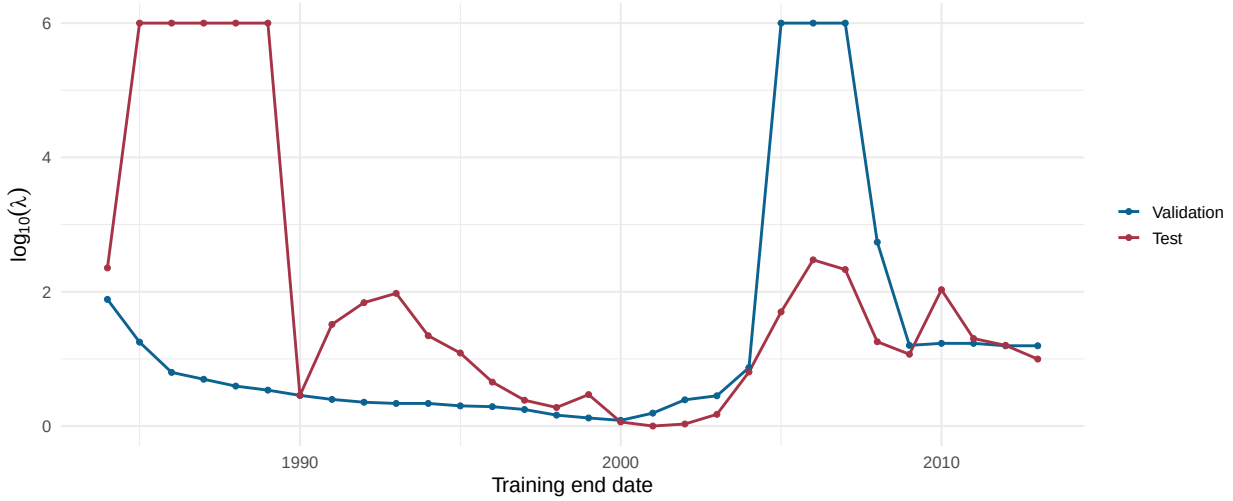


Figure A11
MCE Coverage

This figure shows the number of stocks with a non-missing MCE estimate, split by where the stock is listed. We only include firms that are from developed and emerging market countries, as classified by MSCI.

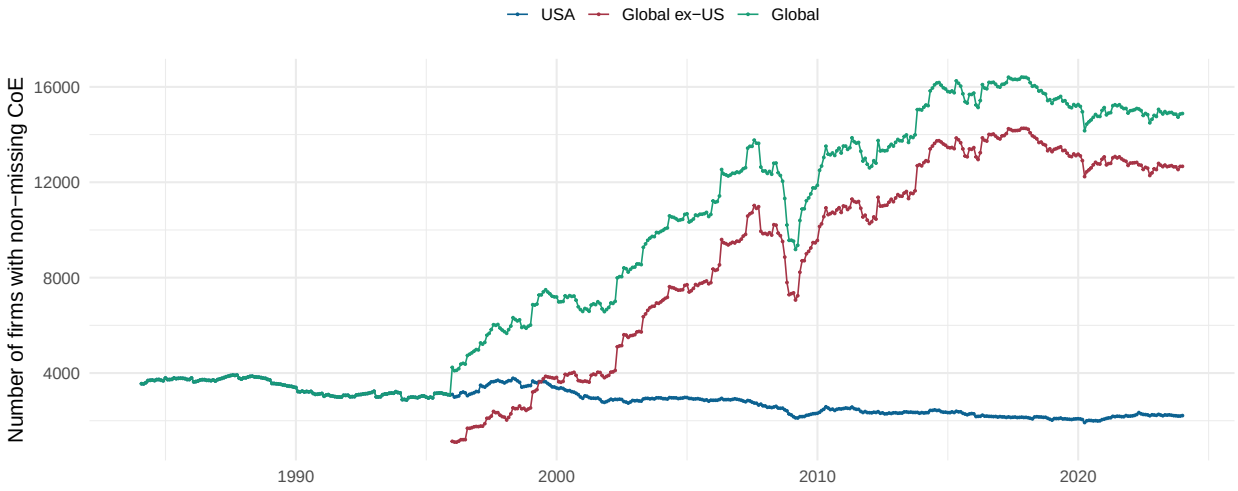


Figure A12
Time variation in the market's nominal MCE

This figure shows the value-weighted nominal market-based cost of equity (MCE).

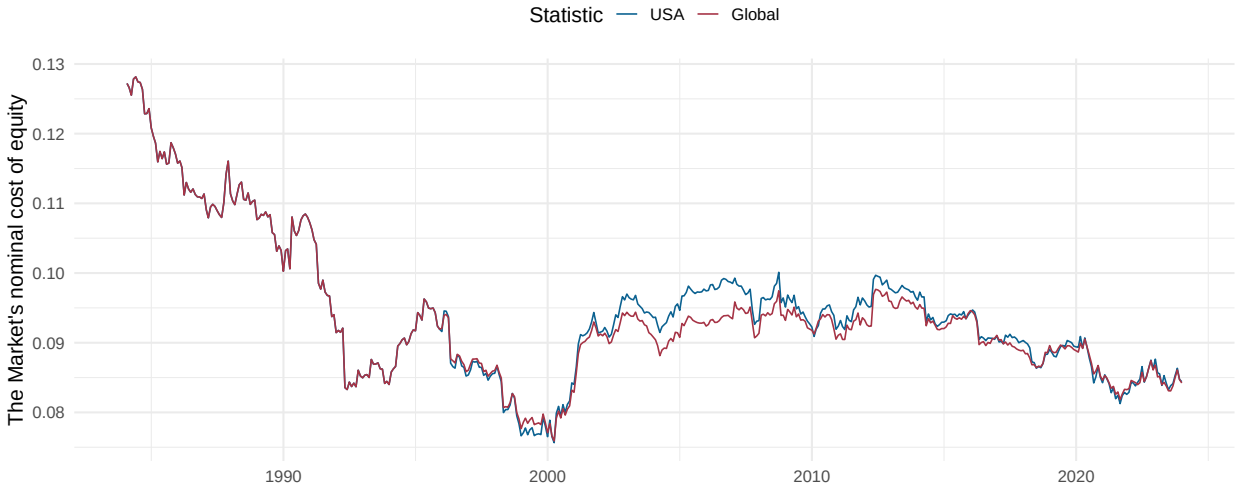


Figure A13
Time variation in the market's nominal MCD

This figure shows the value-weighted nominal market-based cost of debt (MCD).

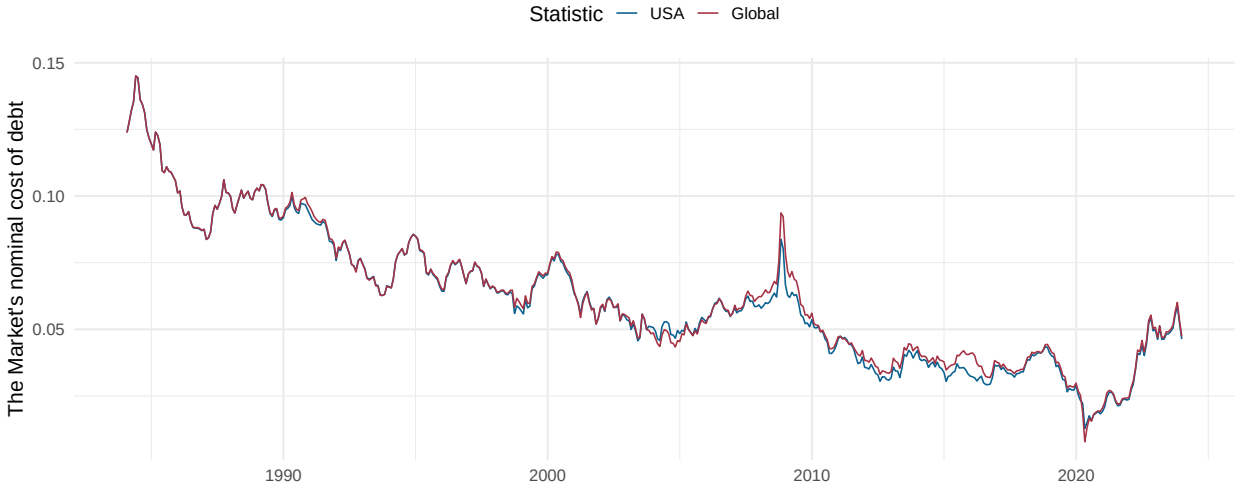


Figure A14
Time variation in the market's nominal MCC

This figure shows the value-weighted nominal market-based cost of capital (MCC).

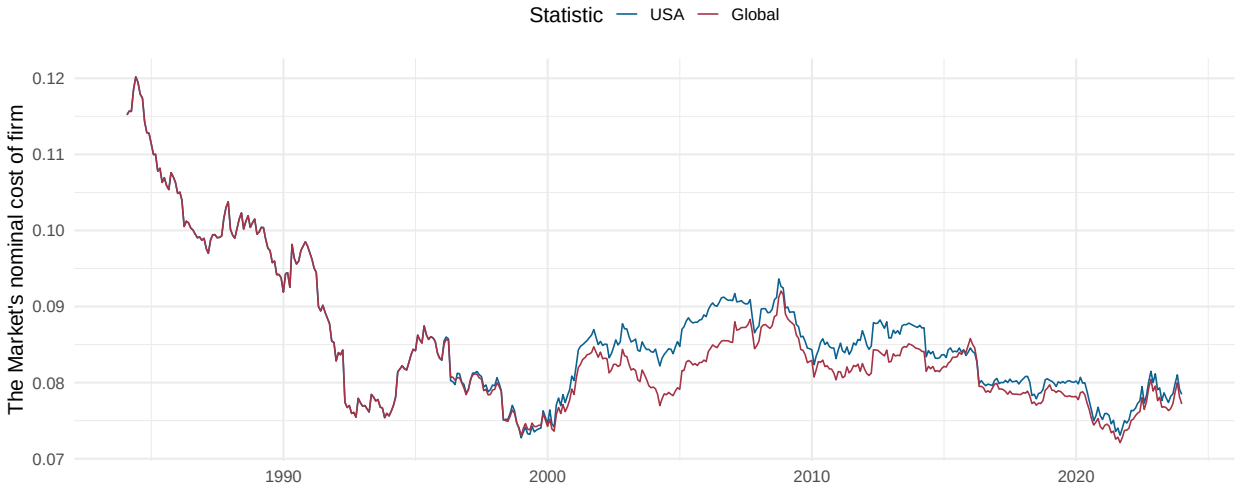
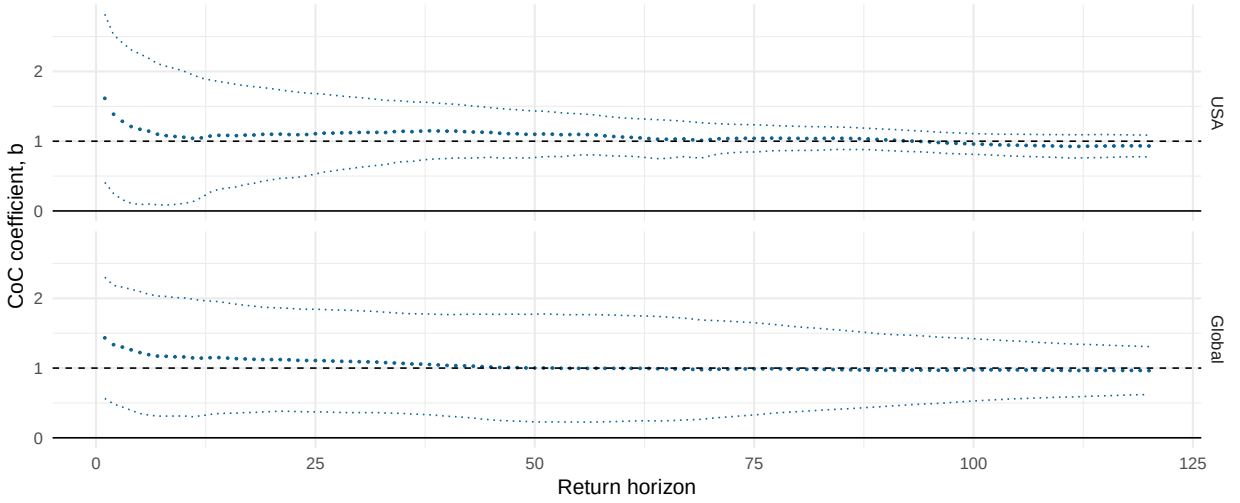


Figure A15
Buy-and-hold return predictability

This figure reports the slope coefficient b from regressions of future buy-and-hold stock returns on the MCE, $r^{\text{realized},i}_{t\to t+h} = \alpha_t + b \times \left[(1+\text{MCE}^i_t)^{h/12} - 1\right] + \varepsilon^i_{t+h}$, for return horizons $h = 1, 2, \ldots, 120$ months. All regressions include month fixed effects. The top row restricts the sample to U.S. firms; the bottom row uses the global sample. Dotted lines indicate 95% confidence intervals based on Driscoll-Kraay standard errors with a 120-month bandwidth. A coefficient of one (dashed line) implies that a one-percentage-point higher MCE maps one-for-one into higher realized returns.



	10-year realized return		
	U.S.	Global	Global ex-U.S.
	(i)	(ii)	(iii)
Cost of equity	0.93***	0.91***	0.94***
	(0.08)	(0.15)	(0.23)
Months	363	363	220

The table reports Fama-MacBeth regressions of 10-year realized buy-and-hold returns on the current 10-year cost of equity, estimated separately for the U.S., global, and global ex-U.S. samples. Each formation month, we run a cross-sectional OLS of the cumulative 10-year realized return on the compounded cost of equity $\left[(1+\text{MCE}_t^i)^{10}-1\right]$, and average the monthly slopes over the sample. Because 10-year returns overlap across consecutive months, t-statistics use Newey-West standard errors with 120 lags. The slope is statistically indistinguishable from one in all three samples, consistent with the panel estimates in Table 1. Standard errors are in parentheses. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Return predictability – Fama-MacBeth

	(1)	(2)	(3)
	Global	Global	U.S.
(Perc. CoC − MCC) > 0	0.199**	0.142*	0.170**
	(0.098)	(0.084)	(0.084)
Observations	709	709	709
Within R^2	0.0107	0.00680	0.00983

Column 1 replicates column 2 of Table 4. Column 2 controls for only quarter-by-year-by-country fixed effects, not book equity deciles-by-country. Column 3 uses the unwinsorized stock return as the outcome. Standard errors are clustered at the firm level. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Robustness: stock returns and investment announcements

Figure A16
Histogram of deviations from the MCC

The figure shows a histogram of the cost of capital deviation (i.e., the difference between perc. CoC and MCC in percentage points). The perc. CoC is reported by the firm on conference calls and influences the firm's investment decisions. MCC is the market-based cost of capital from this paper. The rightmost bin combines observations above 5 percentage points.