NBER WORKING PAPER SERIES

CORPORATE VALUATION:
AN EMPIRICAL COMPARISON OF DISCOUNTING METHODS

Nicolas Hommel
Augustin Landier
David Thesmar

Working Paper 30898
http://www.nber.org/papers/w30898

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2023

We especially thank Hui Chen, Niels Gormsen and Jérôme Taillard for their thoughts, as well as numerous seminar and conference participants. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w30898

Corporate Valuation: An Empirical Comparison of Discounting Methods
Nicolas Hommel, Augustin Landier, and David Thesmar
NBER Working Paper No. 30898
January 2023
JEL No. D24,G31

ABSTRACT

The key purpose of corporate finance is to provide methods to compute the value of projects. The baseline textbook recommendation is to use the Present Value (PV) formula of expected cash flows, with a discount rate based on the CAPM. In this paper, we ask what is, empirically, the best discounting method. To do this, we study listed firms, whose actual prices and expected cash flows can be observed. We compare different discounting approaches on their ability to predict actual market prices. We find that discounting based on expected returns (such as variants on the CAPM or multi-factor model), performs very poorly. Discounting with an Implied Cost of Capital (ICC), imputed from comparable firms, obtains much better results. In terms of pricing methods, significant, but small, improvements can be obtained by allowing, in a simple and actionable way, for a more flexible term structure of expected returns. We benchmark all of our results with flexible, purely statistical models of prices based on Random Forest algorithms. These models do barely better than NPV-based methods. Finally, we show that under standard assumptions about the production function, the value loss from using the CAPM can be sizable, of the order of 10%.

Nicolas Hommel
Department of Economics
Princeton University
Princeton, NJ 08544
nhommel@princeton.edu

Augustin Landier
HEC Paris
1 rue de la libération
78351 Jouy en Josas Cedex
France
landier@hec.fr

David Thesmar
MIT Sloan School of Management
100 Main Street, E62-632
Cambridge, MA 02142
and NBER
thesmar@mit.edu

A data appendix is available at http://www.nber.org/data-appendix/w30898

The CAPM is the cozy bedtime story that tells students and practitioners that the world is in good order and that they have learned something which will allow them to understand it. But the real world isn't like that. (Welch, 2021)

1 Introduction

Project valuation is central to capital budgeting decisions. To properly evaluate a project, the classic textbook recommendation (see for instance Berk and DeMarzo (2006)) consists in using a simplified version of the present value formula:

$$PV_t = \sum_{s \geq 1} \frac{E_t \pi_{t+s}}{(1+r)^s} \tag{1}$$

where $E_t \pi_{t+s}$ is a forecast of cash-flows and r is the project's expected return estimated using the CAPM. Consistent with this recommendation, this approach is now used pervasively by practitioners. Graham (2022) documents that, in 2021, 85% of large corporations use the CAPM to compute their cost of capital. Jagannathan et al. (2016), and more recently, Gormsen and Huber (2022b), show that firms' reported discount rates are well explained by realized betas on the market, with a slope close to standard values of the equity premium. Hence, firms do indeed use the CAPM in order to compute their discount rates.

However, this approach raises both conceptual and empirical concerns that are widely discussed in the literature and in classrooms around the world. On the empirical side, it is well-known, since the 1970s, that the CAPM does not fit the data well: the average realized return of high beta securities does not differ much from low beta securities (see for instance Frazzini and Pedersen (2014), Hong and Sraer (2016) and the references therein). On the conceptual side, if expected returns are time-varying and random, one needs to introduce horizon-dependent discount rates, which will generically differ from expected returns (see for instance Hughes et al. (2009) for a clear exposition). Thus, models of expected returns may provide noisy proxies of discount rates when project risk premia are random and time-varying. Thus, there is a gap between the last several decades of asset pricing research and the classroom recommendation of how to evaluate projects. This gap may have large real effects since corporate discounte rates affect investment decisions (Gormsen and Huber (2022a)).

In this paper, we explore different measures of project-level discount rates and ask how well they predict the cross-section of project *values*. Our testing ground is made of publicly listed firms in the U.S., for which we can measure both expected cash-flows and actual prices. We explore several discount rate models discussed in the classroom: the CAPM, a richer factor model, and even a characteristic-based models of expected returns. We also investigate a multiple valuation approach, an imputed Implied Cost of Capital (ICC) approach, as well as a similar approach which allows for time-varying expected returns. Using simple cross-validation techniques (training the discount rate model on a training sample, testing it on a separate validation sample), we find that a simple "imputed ICC" method delivers a very good out of sample fit. By contrast, valuations based on models of expected returns have a much poorer fit. We also compare all these predictions to an a-structural, purely statistical, prediction based on random-forest algorithms: We find that the fit of our best NPV-based approaches is almost as good as this "best possible forecast." Finally, we show that, under a simple model of firm behavior, the pricing error made by using the CAPM instead of our preferred method, can destroy a quantitatively large amount of value, of the order of 10%.

The details of our approach are the following. To set the stage, we first evaluate the cost of using the wrong discount rate in an off-the-shelf dynamic corporate finance model Hennessy and Whited (2005). When discussing the role of discount rates, papers typically take examples of individual projects (e.g. Jacobs and Shivdasani (2012)). They highlight that the value of long duration projects is sensitive to the cost of capital, which may lead to big capital budgeting mistakes. But in order to evaluate the *average* cost of taking the wrong discount rate, one needs to make assumptions about the set of projects available to firms. We do this by using a dynamic corporate finance model, which we calibrate to standard parameter values. We then compute the percent loss of value of using a cost of equity that differs from the true one. With standard parameters, we find that a 3 ppt error in the cost of capital leads to a 3.5 pct loss of value. In comparative static exercises, we find that the effect is much larger for firms that are (1) *less* constrained and (2) face *more* elastic product demand.

We then go to the data and proceed to compute the pricing errors of different valuation approaches. We take the perspective of a "project pricer" who wishes to apply the present value formula (1) using cash-flow forecasts to estimate the value of a firm. This person observes firm characteristics (we try several sets) but, of course, not the price itself, which is the variable we aim to predict. We use a merged sample between CRSP, Compustat, and IBES, which starts in 2000 and covers all firms for which at least three horizons of forecasts are available in IBES. The advantage of these data is that we observe firm characteristics and cash-flow forecasts, so that we can estimate the PV of these cash-flows, and then compare them to observed prices. We ask which PV formula best fits prices.

To measure expected cash-flows, we use subjective analyst forecasts, which are good forecasts of future earnings. But we also run robustness checks with unbiased statistical EPS forecasts. We then split our sample into two subsamples of equal sizes: a training sample and a validation sample. The training sample is used to estimate statistical models of discount rates, while the validation sample is used to compare the performance of these models in predicting the cross-section of firm *values*.

We test several simple discount rate models. For each one of these models, we pay particular attention to avoiding look-ahead bias – we only use data that are available to the fictitious "project pricer" at the time of valuation. The first group consists of three models that are based on the empirical asset pricing literature: variants on the CAPM, 4-factor model, and characteristics-based expected returns. We try several CAPM-based discount rates. Besides the simple CAPM using betas estimated on rolling windows, we also try incorporating the term structure of safe rates of return, and also implement the beta shrinkage suggested by Levi and Welch (2017). Our multi-factor model is the classic 4-factor model of Fama and French (1993) and Carhart (1997). Our characteristics-based model of expected returns is estimated by regressing, on past data, future returns on a slew of firm-level predictors documented in the existing empirical asset-pricing literature (Keloharju et al., 2020).

We then explore three alternative models, that are *not* based on expected returns. The first one is a multiple of earnings forecasts, projected on firm characteristics. The statistical projection is estimated on the training sample and using past information only. This projected multiple is then estimated for any firm in our validation sample and multiplied by the firm's one-year EPS forecast. The second NPV-based model uses an imputed "Implied Cost of Capital (ICC)". To do this, we first follow the large accounting literature initiated by Gebhardt et al. (2001), and compute the ICC as the internal rate of return as the discount rate that equals the market price to the present value of expected cash-flows. We then project, still in the training sample, the resulting ICC on firm characteristics. We finally use the fitted relation to estimate imputed ICC on the validation sample, which we inject into the present value formula. The third and last approach is a variant of the imputed ICC that flexibly allows for time-varying expected returns, at the cost of a Campbell-Shiller log linearization (Campbell, 2017). This last approach is roughly equivalent to discounting all future cash-flows at 5%, and adding a multiplicative term that minimizes in-sample MSE. This term contains the possible effect of time-varying expected returns.

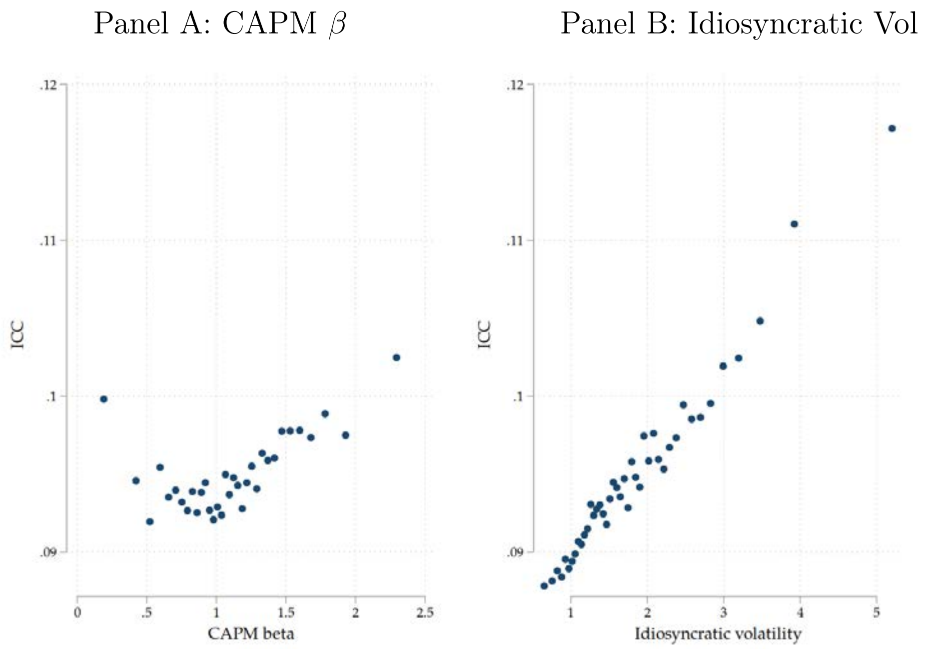
Finally, we report, for the sake of comparison, the performance of a fully flexible, purely statistical model of price prediction. This model uses the same observables as the NPV-based models, as well as cash-flow forecasts, but it does *not* assume that the PV formula holds. It is estimated using random-forest algorithms, and represents the best prediction one can obtain, if one is willing to give up the structure imposed by present value formula.

We first find that, on the validation sample, models based on expected returns perform very poorly: this result holds for our various versions of the CAPM, as well as the four-factors model or characteristics-based models of expected returns. Second, we find that the imputed ICC-based model, and to some extent multiple valuation, perform much better. Third, a method based on the Campbell-Shiller approximation, which allows for a more flexible term-structure of expected returns, marginally improves over the imputed ICC method. This method is essentially equivalent to a 5% discounting of cash-flows with a multiplicative adjustment (designed to capture deviation from 5% and time-varying effects), which is fitted on comparable firms. The fact that this adjustment is multiplicative comes from the linearization, i.e. that log variations in expected returns are small. Fourth and last, we benchmark all these results with a random forest-based prediction of prices combining observables with forecasted EPS in a fully flexible way. Our last two NPV-based methods (imputed ICC and the Campbell-Shiller approximation) do almost as well as the statistical forecast, suggesting that, in the end, the structure imposed by the present value formula is not in itself detrimental to predicting prices.

We then challenge our findings with a series of robustness checks. In particular, one concern with our approach is that EPS forecasts may "influence" prices, so that using EPS forecasts to predict prices would overestimate the ability of the present value formula to explain the cross-section of prices. Another (almost opposite) worry is that EPS forecasts are only noisy and biased forecasts of future cash-flows: They are far from the true expected cash-flows that one needs in the NPV formula. To address these issues, we construct statistical forecasts that minimize the prediction error of EPS itself: Our baseline results are essentially unchanged. Another robustness check that we run is that we construct a more complex term structure of EPS forecasts. Again, our results do not change much.

Our paper is primarily a contribution to the discussion on capital budgeting, which tries to reconcile the textbook recommendation of CAPM-discounting of cash-flows, with the empirical failure of the CAPM. Dessaint et al. (2021) show that acquisitions of low beta targets tends to be more value-destroying, indicating that the use of the CAPM may destroy value through target overvaluation. On the opposite end, Stein (1996) and Gollier (2020) take a normative approach. The empirical failure of the CAPM is irrelevant for the decision-maker, as what really matters for her is to maximize her objective function. Under proper assumptions about this objective function, using the CAPM leads to the objective-maximizing rule, whether the CAPM holds or not in the data. Our approach is empirical, in that we seek to come up with the discounting rule that best predicts prices out of sample – without trying to get to the bottom of its economic foundation. Our approaches that are *not* based on expected returns implicitly assume that similar projects have similar discount factors for similarly risky cash-flows, which is, in a sense, consistent with the law of one price. But we do not impose more structure on the data – while traditional asset pricing models do.

Figure 1: ICC vs Risk



Note. This Figure shows binned scatter plots of ICC against two different measures of risk. ICC is the discount rate that equates prices to DCF of forecasted EPS (see paper for detailed procedure). Panel A shows the dependance on the CAPM β computed using the beta suite of WRDS. Panel B shows the dependence on the volatility of the residual from the Fama-French 4 factor model.

Our preferred discounting rule – the imputed ICC approach – does not rely on the CAPM, but it accounts for risk. In Figure 1, we show the binned scatter plots of the ICC of firms in our sample against two measures of risk: CAPM β and idiosyncratic volatility (see main text for construction details). Quite clearly, empirical discount rates implicit in valuations do not line up with the CAPM – or if anything with a very flat slope. They line up much better with idiosyncratic volatility, consistent with survey evidence (Jagannathan et al. (2016), Décaire (2021), Gormsen and Huber (2022b)). Overall then, our results suggest some adjustment for risk is useful, just not the one predicted by the CAPM. They thus echo recommendations from financial textbooks *before* the CAPM was invented. Dewing (1934) proposes four classes of risk (from low to very high), with rates ranging from 10 to 25%. He suggests that investors use these rates in order to "capitalize current earnings" (he does not advocate for the use of DCF, which was not well-know at the time). After the war, Dean (1951) introduces the DCF formula. He too speculates that the discount rate is the investor's "own estimate of his uncertainty", but does not give more detail.

Finally, given its focus on predicting project values, rather than returns, our paper is related to a variety of contributions in the literature. Cochrane (2011) observes that returns are noisy and intrinsically hard to predict, while prices are much more predictable. Décaire (2021) uses oil project-level data to back out their hurdle rates. He finds that such rates strongly correlate with idiosyncratic volatility, indicating that companies do not only use CAPM inputs to construct their discount rates (as Jagannathan et al. (2016) and Gormsen and Huber (2022b)). Cohen et al. (2009) is the most related paper to ours, since it uses firm-level data. Like them, our focus is to predict the cross-section of price levels (as opposed to returns), but our approach differs in two important ways. First, we take the perspective of an investor that uses cash-flow forecasts to evaluate a project. We do not use ex-post realized cash-flows but ex-ante forecasts. We then measure how close the resulting valuation is from effective prices *out of our estimation sample*. Another difference with Cohen et al. (2009) is that we test various discounting alternatives, while they focus on the cash flow-based CAPM. One of our approaches, like Brennan (1997) and Ang and Liu (2004), allows for time-varying expected returns, but we do so without restricting ourselves to a CAPM set-up and we use expectations from analyst forecasts rather than econometric forecasts based on past dynamics. In that sense, our paper is related to research by De la O and Myers (2021) who link aggregate price movements with aggregate data on cash-flow expectations. Two key differences are that (1) we test different ways of computing the PV and (2) we work with the cross-section of stocks. Another important key feature of our analysis – inspired by typical textbook recommendations – is our use of the full term structure of cash-flows forecasts in valuation. This differentiates our approach from Gupta and VanNieuwerburgh (2021), who impose a specific structure on the joint dynamics of the cash-flow process and discount factors. Another difference with this last paper is that our aim is to assess the *predictive* power of various valuation techniques. Most recently, De La O et al. (2022) connect prices with earnings and returns forecasts from sell-side analysts. They find that most of the cross-section of prices is explained by earnings forecasts. This finding echoes our result that a singe ICC already prices the cross-section of firm values reasonably well – although varying the ICC by industry and size improves the prediction.

Section 2 evaluates the cost of using the wrong cost of capital in a standard model of dynamic corporate finance. Section 3 presents the data we use. Section 4 describes in details the various discounting models that we are testing. Section 5 compares how well these models predict the cross-section of prices out-of-sample. Section 6 provides various robustness tests. Section 7 concludes.

2 Value Loss from Using Wrong Cost of Capital

A company which does not use the right cost of capital destroys value, by under- or over-investing. How much value does a given valuation error entail? Most existing papers focus on a single project (e.g. Jacobs and Shivdasani (2012)). In this case, valuation errors depend on project duration. When the firm uses the wrong cost of capital, it will destroy most value by over- or underinvesting in long-term projects. But such analysis does not help assess how much value is destroyed by the use of the wrong cost of capital.

Do evaluate such loss at the firm level, one needs to define the set of projects available to the firm. In this Section, we start with a simple, off-the-shelf, dynamic corporate finance model. This model has (1) a production function (which defines available projects) and (2) investment and financing frictions (which make capital structure decisions non-trivial). We calibrate this model with parameters from the literature. In this model, shareholders impose a cost of equity $r^\star$. We then measure the present value loss induced by choosing a different cost of equity $r \neq r^\star$.

2.1 Set-Up

Our model is a simplified version of Hennessy and Whited (2005). All the elements presented here are standard.

We omit firm and time indices to lighten notations. Log TFP $\tilde{z}$ follows an AR1 process of persistence ρ and innovation volatility σ. In the current period, the firm inherit from previous period capital stock k and one-period net debt d. Once productivity is revealed, it chooses labor ℓ, new capital k' and debt d'. Production technology is Cobb–Douglas and there is a one period time to build for capital. Finally, the firm is a monopoly facing demand with constant elasticity $\varphi > 1$. Current revenues are thus proportional to $\left(e^{\tilde{z}} k^\alpha \ell^{1-\alpha}\right)^{1-\frac{1}{\varphi}}$. Every period, labor adjusts without friction and wages are constant, so that operating profits $\pi(z,k) \propto e^z k^\theta$, with $\theta \equiv \frac{\alpha(\varphi-1)}{1+\alpha(\varphi-1)} < 1$ and $z \propto \tilde{z}$.

Investment and debt dynamics are also standard. Investment i is $k' - (1-\delta)k$. Capital adjustments costs are given by $\frac{\gamma}{2}\frac{(k'-(1-\delta)k))^2}{k}$. Debt is risk-free. The risk-free rate is r_f. Net debt can be negative in which case the firm holds cash. Firms pay corporate taxes at rate τ. Interest payments $r_f d$ and depreciation δk are tax deductible.

Finally, the firm faces financial frictions. There is an upper bound to debt given by the collateral constraint $d' \leq Mk'$, and an upper bound on cash given by $d' \geq -mk'$. Equity issuance is costly: Issuing \$1 worth of equity costs $\$1 + \lambda$. This is modelled as a multiplier $1 + \lambda$ to cash-flows when they are negative.

Overall, current period cash-flows are thus given by:

$$\Pi(z,k,d,k',d') = \Psi\left((1-\tau)\left(\pi(z,k) - \delta k - r_f d\right) + k - k' - \frac{\gamma}{2}\frac{i^2}{k} + d' - d\right)$$

where $\Psi(x) = \left(1 + \lambda \times \mathbf{1}_{\{x<0\}}\right) \times x$.

The cost of equity is $r^\star$. Investment and capital structure policies are the solution of the Bellman equation:

$$V(z,k,d)=\sup_{-mk'\leq d'\leq Mk'}\left\{\Pi\left(z,k,d,k',d'\right)+\frac{1}{1+r^{\star}}\mathbb{E}\left[V(z',k',d')\mid z\right]\right\},$$

We note $\kappa(z,k,d;r^{\star})$ and $\Delta(z,k,d;r^{\star})$ the optimal capital and debt choices. The dependence on $r^{\star}$ is highlighted as it will become important below. In the standard setting, firm value is given by $V(\cdot)$. But if the company does not choose $r^{\star}$ as its cost of equity, the result will differ, as we now discuss.

2.2 Measuring the Value Loss

Assume the firm uses $r\neq r^{\star}$ as its estimate of the cost of equity. Then, the firm uses r to update, every period, its capital and debt. However, the present value of such a firm's equity uses the true discount rate $r^{\star}$. Thus, the value of this firm is given by:

$$\begin{aligned}W(r,r^{\star},z,k,d)\equiv\mathbb{E}&\left[\sum_{t=0}^{\infty}\frac{\Pi\left(z_t,k_t,d_t,k_{t+1},d_{t+1}\right)}{(1+r^{\star})^t}\,\middle|\,z_0=z,k_0=k,d_0=d\right]\\ k_{t+1}&=\kappa(z_t,k_t,d_t;r)\\ d_{t+1}&=\Delta(z_t,k_t,d_t;r)\end{aligned}$$

where, given our notations, $W(r^{\star},r^{\star},z,k,d)=V(z,k,d)$. In what follows, we focus on the average value of firms whose initial characteristics (z,k,d) are drawn from the stationary distribution μ, that is:

$$W(r,r^{\star})\equiv\mathbb{E}_{\mu}\left[W(r,r^{\star},z,k,d)\right]=\mathbb{E}_{\mu}\left[\sum_{t=0}^{\infty}\frac{\Pi\left(z_t,k_t,d_t,k_{t+1},d_{t+1}\right)}{(1+r^{\star})^t}\right].$$

where we choose as stationary distribution μ the distribution of states (z,k,d) for firms using $r^{\star}$, the benchmark rate. As a result, our value loss will take the steady state distribution as given.

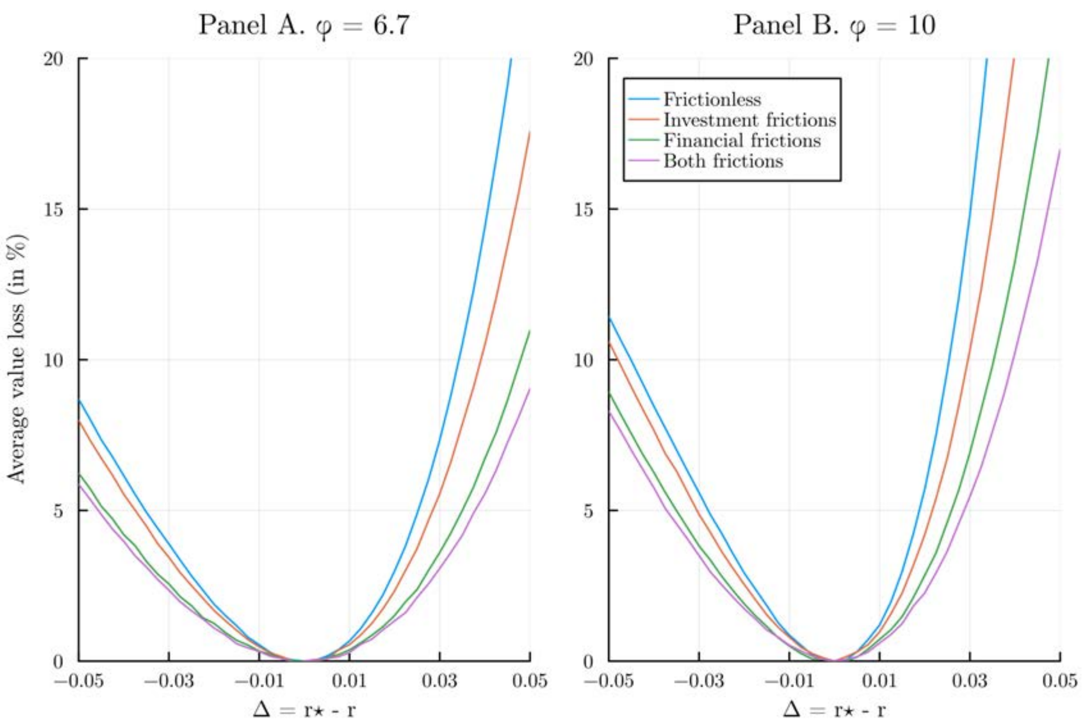
Given these notations, the value loss of using r instead of $r^{\star}$ is $\frac{W(r^{\star},r^{\star})-W(r,r^{\star})}{W(r^{\star},r^{\star})}>0$.

2.3 Results

Our detailed numerical procedure and parameters are reported in Appendix A. In Figure 2, we show the average value loss for each value of $\Delta\equiv r^{\star}-r$ (across values of r and $r^{\star}$, to simplify presentation).

Figure 2 shows that cost of capital errors can be quite costly. For instance, with standard demand elasticity (6.7, see e.g. Midrigan and Xu (2014)), a 3 ppt underestimation ($r=r^{\star}-.03$) leads to a 3.5% loss. As discussed above, the insight here is that a

Figure 2: Value loss of using the wrong discount rate



Note. This figure shows the value loss from using the wrong discount rate r when the true discount rate is $r^{\star}$. We report value losses under four scenarios: (1) In blue, a frictionless benchmark without debt ($\gamma = 0, \lambda = 0, M = 0$), (2) In red, investment frictions only ($\gamma = 0.1$, $\lambda = 0$, $M = 0$), (3) In green, financial frictions only ($\gamma = 0.0$, $\lambda = 0.1$, $M = 0.4$, $m = 0.25$), (4) In purple, both frictions ($\gamma = 0.1$, $\lambda = 0.1$, $M = 0.4$, $m = 0.25$).

larger price elasticity makes investment intrinsically more sensitive to the discount rate. Note that a 3 ppt difference is typical in our cross-section of firms. When comparing CAPM-predicted to our preferred "imputed IRR" measure, the standard deviation of the difference is approximately 3 ppt.

A second message from Figure 2 is that frictions dampen the value loss quite significantly. In our calibration, frictions divide the frictionless loss by two. The intuition is that a very constrained firms would never choose its capital stock, so that the value it loses from not taking the right discount rate is zero. Last, we experimenting with demand elasticity: 6.7 is more standard, while 10 is larger than average. We find that the value loss is more pronounced for larger values of φ. This is because decreasing returns to scale makes investment policy less sensitive to the cost of capital.

3 Data description

3.1 Baseline data

We combine data from Compustat, the Center for Research in Security Prices (CRSP) and IBES. Details on our methodology are provided in Appendix D.

— From the IBES Unadjusted Summary Files, we retrieve contemporaneous earnings per share (EPS) along with forecasts up to three fiscal years ahead. We denote these analyst forecasts by $F_t EPS_{it+h}$, where $h \in \{1, 2, 3\}$ is the horizon, i is the firm and t is the fiscal year. We use the IBES Unadjusted Summary files in order to avoid retroactive adjustments in a firm's number of shares which would not be reflected in the price per share of CRSP.

— We take data on stock prices, returns, and shares outstanding from the CRSP Monthly Stock Returns database. We start with ordinary common shares (share codes 10 and 11) that trade on the New York Stock Exchange, American Stock Exchange, and Nasdaq (exchange codes 1, 2 and 3). We further restrict our attention to stocks with nonmissing price and shares outstanding whose price is above one dollar per share and below five thousand dollars per share.

— Accounting data are from CRSP/Compustat Merged Fundamentals Annual. We first merge the CRSP and Compustat datasets together through the `permno` identifier, and then with IBES using the IBES CRSP Linking Table provided by WRDS.

— For monthly factor returns, we use the Fama-French series from WRDS for: Equity premium, SMB, HML and UMD. To calculate risk premia, we compute average monthly returns on five-year rolling windows.

— For factor loadings, we use the Beta Suite by WRDS. Betas are estimated from daily returns at the stock level on a 252 days rolling window for both the market model and the Fama–French–Carhart model.[1] We then compute monthly betas by averaging daily betas. Whenever we use it, idiosyncratic volatility is the volatility of the residual from the Fama–French–Carhart model.

— Finally, long-term GDP growth forecasts are taken from the Survey of Professional Forecasters maintained by the Philadelphia Federal Reserve. Every time we form forecasts, we use the most recent long-term GDP growth forecast available, defined as the sum of the median long-run inflation forecast (CPI10) and the median long-run real gross domestic product forecast (RGDP10). We show the recent evolution of expected growth in Appendix Figure A.1. For the riskless rate, we use the 5-year Treasury constant maturity rate (DGS5) from the Federal Reserve Economic Data (FRED) at the monthly frequency.

We restrict ourselves to firm-fiscal year observations for which EPS forecasts up to three fiscal years ahead are available and positive. Our panel starts in fiscal year 2000. Finally, we winsorize observations at the 1 percent and 99 percent thresholds to minimize the influence of outliers.[2]

3.2 Timing and look-ahead bias

In order to compute the present value of expected cash-flows at a given point in time, we need pay special attention to the timing of data availability to avoid look-ahead bias. Indeed, such look-ahead bias would lead us to overestimate the capacity of the PV formula to predict prices, and may differentially affect various discount rate adjustments.

We thus focus on forecasts that were released 3 months after the end of fiscal year t, that is, 3 months into fiscal year $t + 1$. This ensures that the accounting data for year t is public, and has therefore been incorporated into earnings forecasts. We denote by P_{it} the price of firm i at that time, bearing in mind that it uses information about fiscal year t, but is measured three months into fiscal year $t + 1$.

Our perspective is that of an investor who wants to perform corporate valuation on a project using cash-flow forecasts and possibly additional project-specific information. This fictitious investor will use the PV formula in order to estimate P_{it} using cash-flow forecasts up to three fiscal years ahead (F_tEPS_{it+1}, F_tEPS_{it+2} and F_tEPS_{it+3}) and a perpetual growth assumption after that, based on long-run GDP growth forecasts. We describe this formula in greater detail in Sections 4.4 and 5. We explore different variants of this formula as robustness checks.

[1] Of the 252 days, we require that at least 126 days of data be available to estimate betas.

[2] The results of this paper still hold if we winsorize observations that deviate from the median by more than five interquartile ranges, or if we remove observations instead of winsorising them.

Predictive models are trained in-sample on a rolling-window basis using accounting and market data available up to three months after the end of fiscal year t (hence data known at the time of formation of P_{it}). Both cash-flows forecasts and GDP growth forecast are the latest available at that time.

Returns are computed on a fiscal year basis. More specifically, we calculate returns for (fiscal) year $t+1$ between the last day of the 3rd month after the end of fiscal year t and the last day of the third month after the end of fiscal $t+1$. As a result, we can safely assume that accounts for fiscal t are known when predicting returns in $t+1$ – this point only matters for empirical models of expected returns.

3.3 Validation vs training subsample

Our dataset is an annual panel that runs from 2000 to 2019. We split it in two halves (one for training and one for validation) as follows: every year, half of the newly entering firms are allocated to the training sample, and the other half is allocated to the validation sample. Once allocated to either subsample, firms stay there for as long as they remain in the Compustat universe. This sampling technique ensures that firms do not move across samples, so that a firm used to train the model has no chance of ending up in the validation sample (this adjustment turns out to change little to our results but is made to make sure the validation sample does not contain any firm that has been part of the training sample). We show the evolution of the sample size in Figure 3. Starting in 2006, we have over 1,000 firms per year in the overall sample. Both subsamples have the same size. Table A.3 reports summary statistics for the final sample.

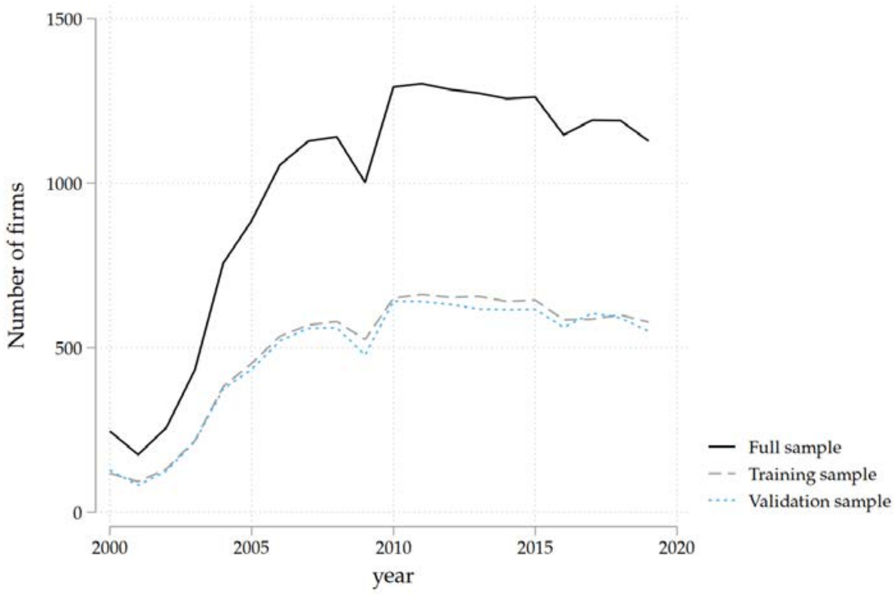
3.4 Cash-flow forecasts

To compute the present value of a stock, we follow the classic textbook approach and therefore need two ingredients: expected cash-flows and discount rates. We now describe expected cash-flows, and postpone the discussion of discount rates to the next section, since they are the core of our analysis. For expected cash-flows, we follow two approaches:

— **Analyst forecasts.** In our baseline analysis, we use the median consensus firm-level EPS forecast from the IBES Summary Files. Analyst forecasts may be biased and noisy but they also contain relevant information about future EPS that may be unobservable to the econometrician (e.g. DeSilva and Thesmar (2021)).

— **Statistical forecasts.** Since analyst forecasts may not be fully efficient (see for instance So (2013) and van Binsbergen et al. (2020)), we attempt to correct for noise and bias by forecasting EPS with an MSE-optimal combination of accounting variables and analysts' forecasts. We call statistical forecasts these corrected EPS forecasts.

Figure 3: Sample size: full, training and validation



Note. We show here the number of firms in our sample for every fiscal year in the period we consider. The construction of the two subsamples (training and validation) is explained in the text.

We estimate this predictive model on the training sample on a 5-year rolling window, up to fiscal year t. The variables we use for EPS prediction are described in column (3) of Table A.2. We use a lasso regression to estimate this unbiased forecast, and tune the penalty parameter through three-fold cross-validation. We run these regressions after preparing the data as described in Section 3.1 but before constructing the ICC, DR and multiple variables of Section 4 in order to maximize sample size.

In Table 1, we report information on the quality of EPS forecasts using the two approaches across horizons. The first line reports that well-known fact that analyst forecasts tend to be optimistic (the ex-post realization is on average less than forecasts). The optimism bias is small at short horizons, and stronger at longer horizons. On the second line, we see that from horizon 2 onward, compared to analysts' forecasts, statistical forecasts systematically display a smaller bias: this is by design, since these statistical forecast use subjective forecasts but seek to minimize the MSE. The specific history of macro-shocks during our period, and the lasso penalty, however, make statistical forecasts slightly biased. The third and fourth lines show that statistical forecasts also display a smaller variance: Statistical forecasts are less noisy than human ones. But of course, forecasting noise increases with horizon, as analyzed in DeSilva and Thesmar (2021).

We use EPS forecasts and transform them into flows to equity by multiplying them by a payout ratio that reflects both dividend distributions and share buybacks (the con-

Table 1: Performance of Analyst and Statistical EPS forecasts at three horizons

Horizon	$t+1$	$t+2$	$t+3$
Panel A: Mean forecast error			
Analyst Forecasts	-0.13	-0.35	-0.60
Statistical forecasts	0.02	0.10	0.20
Panel B: s.d. of forecast error			
Analyst Forecasts	0.96	1.44	1.77
Statistical forecasts	0.81	1.24	1.47
Panel C: R^2_{OOS}			
Analyst Forecasts	0.80	0.64	0.53
Statistical forecasts	0.86	0.75	0.70
Observations	4672	4672	4672

Note. We report out-of-sample statistics. Analyst forecasts correspond to consensus forecasts from IBES. For each forecasting rule F_t, we compute the forecast error as $EPS_{it+h} - F_t EPS_{it+h}$. Panels A and B report the means and s.d. of these errors, by rule and horizon. Panel C shows the OOS R^2 measured as:

$$R^2_{OOS} = 1 - \frac{\sum_{it} \left(EPS_{it+h} - F_t EPS_{it+h}\right)^2}{\sum_{it} \left(EPS_{it+h} - \widehat{EPS}_{it}\right)^2}$$

where $\widehat{EPS}_{it} = 0$, so the statistic R^2_{OOS} represents the MSE gain of the forecasting rule relative to using 0 as the forecast.

struction of this ratio, b_i, is described below in Section 4.4). We do not directly use dividend forecasts as firms have a larger and larger propensity to pay out their shareholders through share buybacks (Farre-Mensa et al., 2014). As a result, a very large share of expected dividend payouts are zero in our data. At three year horizon for instance, about 19% of the forecasted DPS are zero (see Figure A.2), and we do not have analyst forecasts of future buybacks.

In robustness we value equity based on discounting of two other types of cash-flows, which are also widely used: (1) First, we make statistical forecasts of *free cash flows to equity* (i.e. amounts from income that are available for distribution after investment decisions are made), and discount them as in the baseline. This is known as the *'flow to equity"* approach of valuation. Note that the flow to equity method and our baseline (discounted expected equity payouts) are theoretically equivalent as long as non-distributed cash-flows are reinvested at a return equal to the discount rate used. (2) Second, we use statistical forecasts of *total free cash flows to the firm*, which we use to compute a total firm value.

4 In-Sample Calibration of Valuation Models

We now describe the valuation models that we consider and how we calibrate their parameters in-sample. We take three different approaches. First, we follow the recipes found in most textbooks and use asset-pricing models to calculate discount rates as expected returns. Second, we propose alternative models *not* based on expected returns: a multiple approach, an internal rate of return approach, and a flexible loglinear model that accounts for non-flat term structure of expected returns. Third, we use machine learning techniques to compute (rescaled) prices based on observables: This last approach does not use any theoretical restriction. All models are calibrated using only information available at the time of valuation.

4.1 CAPM

Given the emphasis put on the CAPM in introductory finance classes, it is natural to use it as a benchmark. In this approach, we follow typical recommendations and assume a flat term structure of expected annual returns based on the formula:

$$\widehat{r}_{it+h+1} = r_{it+1}^{CAPM} = r_t^f + \beta_{it} \times 5\%$$

where h is the horizon, r_t^f is the yield on 5-year Treasury bonds (as described in the previous Section, the betas are retrieved from WRDS). We assume an equity risk premium of 5%. The data we use contain no look ahead bias.

This approach is the one typically recommended in most textbooks. One advantage is its simplicity, since in this case future expected returns are constant and the NPV formula simplifies into the classic formula of the intro (1):

$$PV_{it} = \sum_{h\geq 0} \frac{E_t \pi_{it+h+1}}{(1+r_{it+1}^{CAPM})^{h+1}}$$

We also explore alternative specifications that have been discussed to improved on the baseline approach. First, we allow for a term-structure in the risk-free rate and use the Fama and Bliss (1987) discount bonds from year one to five available from CRSP.

Second, we allow the equity risk premium to vary, rather than fixing it at 5% as we did in the baseline. To do this, we compute each firm discount rate as risk-free rate *plus* beta *times* an unknown equity risk premium. In the validation sample, we then look for the ERP that minimizes the in-sample pricing error. Figure A.4 shows this optimal ERP, which hovers between 4 and 10%, with an average of 6.6% – not far from the 5% we impose in our benchmark but now time-varying. This method ("flexible ERP CAPM") allows the CAPM to fit the *aggregate* pricing of equities at each point in time, so that its performance is purely evaluated on its ability to predict the cross-section.

Third, we implement the correction proposed in Levi and Welch (2017) which takes into account the fact that betas can be mismeasured and may vary overtime. The intuition consists of shrinking the betas towards the average beta of firms in the same size category:

$$\beta_{it}^{LW} = 0.65 \times \widehat{\beta}_{it} + 0.35 \times \text{Target}_{it}$$

where Target_{it} varies across market capitalization terciles: the smallest tercile has a target of 0.5, the middle 0.7, and the largest 0.9. This approach is shown by Levi and Welch (2017) to predict equity betas well in sample.

4.2 Fama-French-Carhart 4 factor model

Since the CAPM does not explain expected returns very well, richer factor models have been introduced to improve the in-sample fit of expected returns. This option is sometimes discussed in the classroom, when students raise the objection that the CAPM does not help predicting returns. We settle here for the 4-factor model of Fama and French (1993) and Carhart (1997) which includes the excess market return, SMB, HML and UMD. For firm i at date t, we thus calculate the following discount rate:

$$\begin{aligned}\widehat{r}_{it+h+1} = r_{it+1}^{FF} = r_t^f &+ \beta_{it}^{\text{Market}} \times (R_{m,t} - r_t^f) \\ &+ \beta_{it}^{\text{SMB}} \times \text{SMB}_{m,t} + \beta_{it}^{\text{HML}} \times \text{HML}_{m,t} + \beta_{it}^{\text{UMD}} \times \text{UMD}_{m,t}\end{aligned}$$

where the loadings and annual risk premia are computed using 5 year rolling windows as described in Section 3.

As in the benchmark CAPM application, we assume a flat term structure of expected returns. The improvement brought by this model is (1) the time varying market risk premium and (2) three additional factors, which are known to improve the in-sample fit of the cross-section of returns.

4.3 Characteristics-based expected returns

Here we follow Keloharju et al. (2020) and use characteristics to measure expected returns. Using a 10-year rolling window of data, we run the following regressions in the training sample to predict annual returns between $s+h$ and $s+h+1$ based on information available at date s

$$r_{is+h+1} = \beta_h X_{is} + \epsilon_{is,h} \tag{2}$$

where we vary the horizon h between 0 and 4. All the information necessary to run these regressions has to be available to the forecaster at date t – which means that we have less data to estimate long-run expected returns. The regressor X_{it} contains variables which are known to forecast returns at the annual frequency: dividend yield, net shares issue, BM, ROA, Amihud ratio, market leverage ratio, tangibility, growth, size (market capitalization), idiosyncratic. Detailed definitions are available in Appendix D. All variables are signed so that they predict larger returns as shown in the literature. These are an important subset of characteristics used by Keloharju et al. (2020). We do not use the characteristics directly; instead, we use their ranks normalized between $[-.5; +.5]$. This ensures all characteristics have the same mean and variance, and mitigates the influence of outliers.

We estimate Equation (2) using OLS and penalized regressions, and report the results in Table 2. In lines 2 to 4 of the Table, we let the data choose the optimal weights on each rank-normalized characteristic. The second line uses OLS. Lines 3 and 4 used penalized method (Lasso and Ridge). Penalization parameters are estimated through three-fold cross-validation. All estimation methods yield similar results; R^2 is very low at short horizons (1 and 2 years out), of the order of 1%. Three years out, the R^2 is almost zero, consistently with Keloharju et al. (2020).

We also try an alternative approach to predict returns. In this alternative approach (the "composite score"), we follow Keloharju et al. (2020) more directly and construct a composite score that is equal to the sum of these ranked characteristics – which are all signed to predict positive returns. Every year t, we then regress future returns up

Table 2: Predicting returns at various horizons

	R^2_{OOS}		
Horizon (years)	1	2	3
Composite score	0.003	0.006	0.008
OLS	0.018	0.000	–0.001
Lasso	0.018	0.002	–0.001
Ridge	0.016	0.001	–0.000

Note. We show here the out-of-sample predictive power of annual returns at three horizons. In the spirit of Gu et al. (2020), we construct R^2 measures as one minus MSE normalized by the MSE assuming returns would be equal to 9%

$$R^2_{OOS} = 1 - \frac{\sum_i (\widehat{r}_{it,h} - r_{it,h})^2}{\sum_i (r_{it,h} - .09)^2}$$

where $r_{it,h}$ is the realized return at horizon h. $\widehat{r}_{it,h}$ is the predicted return at the same horizon, predicted using a composite score, OLS, or penalized regressions. We shrink returns towards 9% for all methods. Penalization parameters are tuned using three-fold cross-validation.

to 5 year horizons on this score within the 10-year rolling window $[t-10; t]$, and use the estimated model to forecast returns from $t+1$ to $t+5$. Finally, we shrink returns towards 9%. This alternative method is a way to incorporate prior knowledge about the sign of each characteristics, and is a way to avoid overfitting on the relative weights of each characteristic. This method only estimates two scalars: the constant and slope of future returns w.r.t. the composite score. In order to maximize sample size, we use all the data available from the merged CRSP–Compustat sample (i.e., prior merging with IBES) from 1990 to 2019. We report the predictive power of this approach in the top line of Table 2. In the remainder of this paper, we use the composite score approach for its simplicity, though all methods yield similar results.

4.4 Imputed ICC

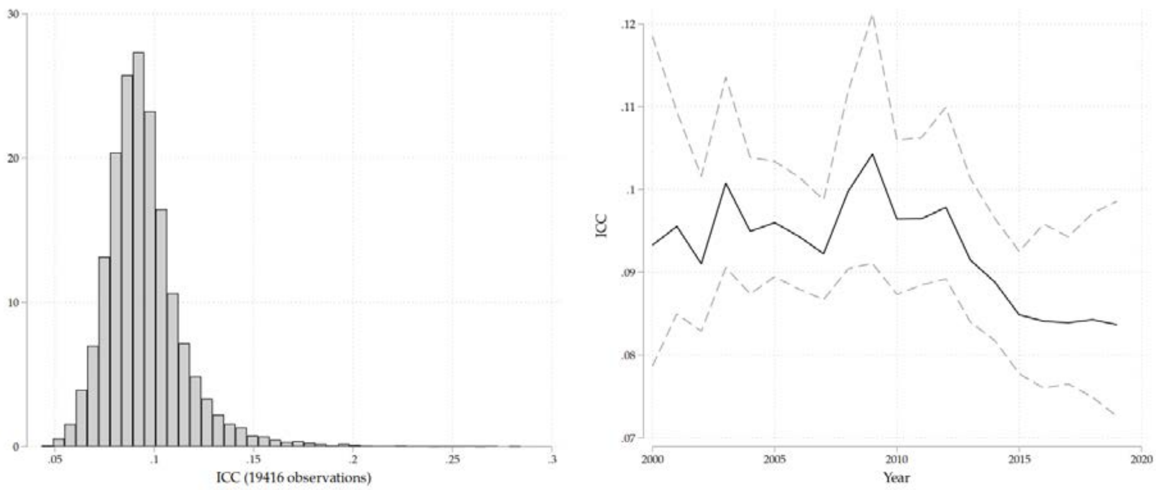
We implement the Imputed Implied Rate of Return (imputed ICC) methodology as follows. In the spirit of the accounting literature, beginning with Gebhardt et al. (2001), we define, for each firm i at date t, the present value function:

$$PV_{it}(r) = \frac{b_i F_t EPS_{it+1}}{1+r} + \frac{b_i F_t EPS_{it+2}}{(1+r)^2} + \frac{b_i F_t EPS_{it+3}}{(1+r)^3} \cdot \frac{1+r}{r-g_t}$$

where b_i is the payout ratio of the firm, $F_t EPS_{it+h}$ is the term structure of EPS forecasts and g_t is the long-run nominal GDP growth expectation from macro forecasters. The payout ratio b_i is the rolling industry average common payout ratio, defined as the sum of dividends (Compustat item `dvc`) and common stock repurchases (total buybacks prstkc minus preferred buybacks pstkrv), normalized by net income (when net income if positive, otherwise we report payout as missing). We winsorize payout ratios at 0 and 1 and then take the average at the industry level.

Similarly to Gebhardt et al. (2001), the ICC is defined as the positive solution to the equation $PV_{it}(r) = P_{it}$, where P_{it} is the stock price. Note that PV_{it} itself depends on the type of forecast used, so we get different ICCs for analysts' forecasts and statistical forecasts. We only keep values of the ICC that lie between 0 and 30%. The resulting distribution is well behaved and stable over time, with a median around 9% that slowly declines to 8% in the mid 2010s. We show the distribution of the ICC based on analysts' forecasts and its distribution in Figure 4.

FIGURE 4: Implied Cost of Capital



Note. The left panel shows the distribution of ICCs as defined in the main text. The right panel shows the time series evolution of the median ICC and two quartile breakpoints.

Once these ICCs have been obtained, we use three different lists of firm-level characteristics to predict there ICCs *in the training sample*. These lists can be viewed as information sets of varying size. We explore three lists:

1. **"Trivial"**: a simple constant – this is equivalent to assuming a single ICCs for all firms. It is an extreme model but already quite successful at predicting prices as we will see.

2. **"PE"**: log total assets and 2-digit GICS industry dummies. These variables are designed to simulate the information that an investor would have when making capital budgeting decision. This investor would not have access to past accounting or returns data, only project size and industry. This acronym stands for "Private Equity", as we look only at variables that would be available for a privately held company.

3. **"Corp. Acc."**: a large set of firm-level characteristics that includes: analysts' EPS forecasts from $t+1$ to $t+3$, realized EPS, ROA, idiosyncratic volatility, tangibility ratio, asset growth, book leverage, size, firm age, and cash ratio. We also include industry and year dummies. Details on the construction of each variable can be found in Appendix D. This set of variable is maximalist and can only be used to evaluate projects that are publicly listed. It is mostly here for reference.

We report the regression results in Table A.4, columns 2,5,8. We note that the in-sample R^2 is reasonably large: the "PE" variables (mostly, industry dummies) generate a R^2 of 0.24. Adding the additional firm-level characteristics increases the R^2 to 0.36.

Every year t, we use rolling windows of 5 years and each one of the three sets of firm-level characteristics to predict the cross-section of ICCs at the same date t in the training sample. We then use the estimated relation to predict ICCs in the validation sample. Finally, we inject the predicted ICC into the function $PV_{it}(\cdot)$ to compute predicted prices in the validation sample.

4.5 Earnings forecast multiple

Multiples can be viewed as a simplified version of the ICC approach describe above. Since they are widely used by practitioners, it is natural to test their predictive power, compared to the (slightly) more complex ICC.

As for the ICC, we use three sets of firm characteristics to minimize the MSE between observed log prices and the log multiple valuation. Let X_{it} be a set of characteristics, we are looking for parameters γ_t, every period, which are given by:

$$\widehat{\gamma}_t = \arg\min_{\gamma} \sum_i \left[\log P_{it} - \log\left(F_t EPS_{it} \cdot e^{X_{it}\gamma}\right)\right]^2$$

which is equivalent to a linear regression of the log price-to-earnings multiple on observables X_{it}. We report the results from these regressions in Table A.4, columns 1,4,7. It clearly appears that industry dummies and size already generate a sizable R^2 of .23; But the additional firm-level characteristics add another 20 ppt.

As for the ICC, we fit the model in the training sample and compute the quantity $\exp(X_{it}\widehat{\gamma}_t)$ in the validation sample. We then multiply it by EPS forecasts to obtain firm-level price estimates.

4.6 Allowing for flexible term structure through linearization

This approach is similar to the imputed ICC approach but allows for time-varying expected returns. We implement this through a Campbell and Shiller (1988) (hereafter CS) approximation. We start with the classic log-linearization of the present value formula:

$$p_{it} = \frac{k}{1-\rho} + \underbrace{\sum_{s\geq 0} \rho^s (1-\rho)\mathbb{E}_t d_{it+1+s}}_{D_{it}=\text{expected dividends}} - \underbrace{\sum_{s\geq 0} \rho^s \mathbb{E}_t \xi_{it+1+s}}_{DR_{it}=\text{discount rate}} \tag{3}$$

where p_{it} is the log price, d_{it} is the log dividend and ξ_{it} is the log gross return.

We use this formula (3), observations of p_{it} and expected log dividends to back out DR_{it}, the sum of future returns, for firms in the training sample. The parameters k and ρ are classic constants that are functions of the average log dividend-price ratio in the sample. We use $\rho = 0.95$ (as in Campbell (2017)), which corresponds to a dividend-price ratio between 4% and 5%. To impute dividends from cash flows, we assume that expected log dividends $\mathbb{E}_t d_{it+1+h}$ of firm i in year t can be written as the industry-level payout ratio b_i (computed as in Section 4.4 using past data only) times expected earnings $F_t EPS_{t+1+h}$ so that $\mathbb{E}_t d_{it+1+h} = \log(b_i F_t EPS_{t+1+h})$.

Once we have these forecasts, we implement the usual two-step method. On the training sample, we first compute the contribution of discount rates as:

$$DR_{it} = p_{it} - \frac{k}{1-\rho} - (1-\rho)\left[\mathbb{E}_t d_{it+1} + \rho \mathbb{E}_t d_{it+2} + \frac{\rho^2}{1-\rho}\mathbb{E}_t d_{it+3} + \frac{\rho^3}{(1-\rho)^2}\ln(1+g_t)\right]$$

which we then project on firm-level characteristics available at date t in the training sample, using the three alternative information sets described above: “trivial”, “PE”, and “Corp. Acc.”. We report the results from these regressions in Table A.4. On the validation sample, we then use the predicted $\widehat{DR_{it}} = \widehat{\gamma} X_{it}$ and the above formula to estimate log prices:

$$\widehat{p}_{it} = \widehat{\gamma}_t X_{it} + \frac{k}{1-\rho} + (1-\rho)\left[\mathbb{E}_t d_{it+1} + \rho \mathbb{E}_t d_{it+2} + \frac{\rho^2}{1-\rho}\mathbb{E}_t d_{it+3} + \frac{\rho^3}{(1-\rho)^2}\ln(1+g_t)\right]$$

Overall, this approach flexibly captures the potential time-variation in future expected returns (it does not make any assumption about the structure of expected returns), but relies on an approximation (the log-linearization).

4.7 Machine learning approaches to valuation

Finally, we apply a-structural machine learning models to capital budgeting. Specifically, we train two forest-based models, one based on the random forest algorithm of Breiman (2001) and one on the local linear forest algorithm of Friedberg et al. (2021). These models non-parametrically estimate prices as a function of firm characteristics

$$p(x) = \mathbb{E}\left[p_{it} \mid X_{it} = x\right]$$

where p_{it} denotes the log price. Compared with previous models, the forest-based approach is more agnostic as it relaxes two key assumptions: (1) the structure of the present value formula; (2) the linearity of discount rates in firm characteristics. Recent examples of forests applications in finance include van Binsbergen et al. (2020) for earnings prediction and Gu et al. (2020) for returns prediction. Gu et al. (2020) show that, for returns, forest-based estimates compare favorably with penalized regression methods and are competitive with neural networks. These papers emphasize the importance of nonlinearities and interaction terms in finance as a source of additional predictive power over linear methods, which could prove important in our exercise.

At a high level, random forests are an ensemble method that combines predictions from many regression trees using bootstrap aggregation (bagging)[3]. Regression trees partition the data into "leaves" in order to deliver the best prediction in the sense of the mean squared error. Within each leaf, the outcome of interest is approximated by the average outcome. Because determining a globally optimal partition is computationally prohibitive, regression trees proceed recursively and use a local criterion: at each step, for each covariate, we look for the best two-fold partition of the data in the sense of the mean squared error. We then split the data in two halves using the most efficient partition. This procedure is repeated until the minimum leaf size reaches a pre-determined threshold. Regression trees are popular because they are easy to explain and interpret but they usually do not perform as well as other methods out of sample and are highly unstable due to their recursive-choice structure.

Random forests deliver significant improvements in prediction accuracy over regression trees by building many de-correlated trees and averaging out their predictions. Averaging out many tree-based estimates drives down the variance of the overall model and yields smoother estimates, improving the out-of-sample performance. More precisely, the idea is similar to that of the bootstrap: we draw B bootstrap samples and grow one tree per sample. The algorithm to grow a tree is as described above, except that at each step we use only m randomly drawn covariates, where m is a parameter of the model. At the end of the procedure, the random forest prediction at point x_0 is the average prediction over all trees

$$\widehat{p}(x_0) = \frac{1}{B}\sum_{b=1}^{B}\widehat{p}_b(x_0)$$

where $\widehat{p}_b(x_0)$ is the prediction made by the bth tree.

For an investor performing capital budgeting, machine learning models have two complementary interpretations. First, they can be viewed as a black box model that gives

[3]For a more detailed presentation, see the dedicated chapter in the book of Hastie et al. (2009). For theoretical results from the statistical literature, including asymptotics, see the review by Biau and Scornet (2016) and the recent article by Athey et al. (2019).

an idea of the best performance that can be achieved in the data. This performance can then be used as a benchmark for economically motivated models. Second, forest-based estimates can be interpreted as a data-driven method of multiples. Indeed, the estimator $\widehat{p}(x_0)$ can be rewritten as

$$\widehat{p}(x_0) = \sum_{i=1}^{N} \alpha_{it}(x_0) p_{it}$$

where the $\alpha_{it}(x_0)$ are firm-specific weights that sum to one (see Athey et al. (2019)). Therefore, the predicted price is a weighted average of the price of firms in the training sample. Under this interpretation, the random forests model is seen as a weight generator: given a set of observables x_0, it determines how each firm from the training sample should be weighted in the final price prediction.

The view of random forests as weight generators has motivated the introduction of generalized random forests, which we also implement as an alternative machine learning model. Specifically, we use the local linear forest algorithm of Friedberg et al. (2021) which uses the weights generated by a random forest to solve a local least squares problem. The advantage of this method is that it exploits the (assumed) smoothness of the true function $p(x)$ to discipline the RF algorithm, a bit like kernel estimator would. Intuitively, it uses the RF algorithm to identify weights $\alpha_{it}(x_0)$. Then, one runs a (penalized) linear regression between p_{it} and X_{it} using these weights. It essentially amounts to a linear approximation of the RF estimate among the set of comparable firms to x_0, the shape of this set being determined by the RF algorithm.

In our work, for random forests, we use the R package `ranger` by Wright and Ziegler (2017), which implements Breiman's algorithm along with measures of variable importance. We set $m = 14$, $B = 1,000$ and $\ell = 5$, where ℓ denotes the minimum leaf size. The values for B and ℓ are classical, and we choose the value for m on a grid using the out-of-bag mean squared errors from forests trained in-sample[4]. Local linear forests are implemented in the R package `grf`, which we use with near-default parameters ($m = 5$, $\ell = 5$, and $B = 1,000$). For both tree-based models, we draw half of the sample size in each bootstrap sample.

5 Baseline Results

[4] Practitioners have found that random forests are not very sensitive to parameter tuning, a point that was already made by Breiman (2001), although there is little guidance from theory, as discussed for instance in Section 2.4 of Biau and Scornet (2016).

5.1 Predicting prices

We are now ready to present our results. We calibrate our models in-sample and test them out-of-sample. We show results for the different types of models described in the previous Section: PV-based models based on expected returns, a multiple-based model, PV-based model based on imputed ICC, a model based on the Campbell–Shiller approximation, and fully nonparametric models based on random forests.

In PV-based models, the price of firm i in year t is computed according to the present-value formula

$$\widehat{P}_{it}^{m} = \frac{b_i F_t EPS_{it+1}}{1+\widehat{r}_{it+1}} + \frac{b_i F_t EPS_{it+2}}{(1+\widehat{r}_{it+1})(1+\widehat{r}_{it+2})} + \sum_{s\geq 3} \frac{b_i F_t EPS_{it+3}(1+g_t)^{s-3}}{\prod_{h=1}^{s}(1+\widehat{r}_{it+h})},$$

where m indexes a model. These models differ only in how they predict discount rates. In the CAPM, Fama–French, and imputed ICC models, discount rates are constant across horizons. In the characteristics-based model, discount rates are estimated up to a 5-year horizon, after which they are assumed to be constant. b_i is the payout ratio described earlier which we use to transform forecasted earnings into equity payout predictions (we later present, in robustness, results based on forecasts of firm free cash flows and forecasts of free cash flows to equity).

The next two models are similar in spirit but differ in their implementation. The multiple approach predicts the log price-to-earnings ratio, so that the predicted price can be written as $\widehat{P}_{it}^{\text{Multiple}} = F_t EPS_{it+1} \exp(X_{it}\widehat{\gamma}_t)$. Clearly, this model is not leveraging the full term-structure of cash-flow forecasts. The Campbell–Shiller model consists of predicting the contribution of discount rates at the firm-level, and then adding it to the contribution of expected cash-flows. It assumes that the linear approximation is valid, but makes not assumption about the dynamics of discount rates: it only assumes that these dynamics are ex-ante identical across firms with the identical observable characteristics.

Forest-based models are fully non-parametric models that directly estimate the log price of firms. Forest-based estimators can be viewed as weight generators. In the context of capital budgeting, this means that the estimated price of firm i at time t is a weighted average of comparable firms, where the weights are a function of observable characteristics X_{it}. Local linear forests are a refinement of this estimator that uses those weights to solve a local squares problem.

For all models except forest-based models, we drop observations whose predicted log price is below zero or above seven. For discounting models, we impose that the discount rate used in the Gordon growth formula be at least 1% greater than the long-term growth rate through winsorization.

5.2 Main results

For each model m, we compute the out-of-sample mean squared error

$$\text{MSE}^m = \frac{1}{N}\sum_{(i,t)}\left(\log P_{it} - \log \widehat{P}_{it}^m\right)^2.$$

We compute this measure *on the validation sample* with models estimated *on the training sample*. This measure can be naturally interpreted as the squared percentage error in predicting the price. Because out-of-sample mean squared error depends on how the observations were allocated across the training and testing sample, we repeat this exercise for 100 different train-test splits. We report average mean square errors across all splits, along with the corresponding standard deviations.

The results are shown in Table 3. The first four asset-pricing driven approaches do not require projection on observables, but only model-based predictions of expected returns. The characteristics-based approach only uses the equal-weighted average of characteristic ranks, and fits a linear regression in rolling windows. To predict the discount rate rates using the other approaches, we use the three sets of variables previously described, except for the machine-learning approaches which require at least one feature. For forest-based models only, we include EPS forecasts in the "PE" information set, in order to make them competitive w.r.t. the PV-based approaches.

Overall, the results in Table 3 suggest that factor models do quite poorly at predicting the cross-section of prices in the validation sample. The mean squared error of factor models is above 0.5 for all models. The ERP CAPM is the best version of expected returns models, but even when endowed more flexibility to fit the aggregate time-series of valuations, the CAPM method is largely dominated by the competing approaches. One intuition for this is that the cross-section of ICCs as we estimate them is only weakly correlated with β's (as shown in Table A.5), so any discount rate based on β's alone has little chance of doing as well as our imputed ICC method. This also explains why the characteristics model performs better than baseline CAPM: Essentially, as returns are hard to predict, the predicted expected returns have little variation, which limits the noise in the discount rate compared to CAPM. However, the characteristics based method is dominated by the flexible ERP CAPM, which has a better fit on the time-series of aggregate valuations.

The fit of other models is reported using three different sets of conditioning variables. Across all models, having access to more conditioning variables (enlarging the information set) decreases, as expected, the out-of-sample MSE. Going from a trivial information set to a rich corporate accounts information set decreases the MSE by at least 25% for every model. Out of the three NPV-based models (multiple, ICC, and loglinear models), the multiple model performs the worst, which suggests that the term-structure of EPS

Table 3: Performance of various models in predicting log prices on the benchmark sample

	Performance			
Model	Trivial	PE	Corp.acc.	Factor
CAPM				0.919
				(0.015)
CAPM (including a term-structure)				0.967
				(0.016)
CAPM (flexible ERP)				0.584
				(0.011)
CAPM (LW)				1.093
				(0.015)
Fama-French-Carhart				1.555
				(0.016)
Characteristics-based predicted returns				0.742
				(0.032)
Multiple	0.316	0.271	0.209	
	(0.009)	(0.009)	(0.007)	
ICC	0.178	0.154	0.138	
	(0.005)	(0.005)	(0.005)	
Loglinear model	0.169	0.148	0.130	
	(0.005)	(0.005)	(0.005)	
Random forest		0.140	0.119	
		(0.005)	(0.004)	
Local linear forest		0.139	0.114	
		(0.004)	(0.004)	

Note. This table reports the average MSE across models and information sets. Given a train-test split, we restrict our sample to observations for which the log price prediction of every model is available. The MSE of model m is defined as

$$MSE^m = \frac{1}{N} \sum_{(i,t)} \left(\log P_{it} - \log \widehat{P}_{it}^m \right)^2$$

We repeat this procedure 100 times on different train-test splits and report the average MSE across all splits along with standard deviations. On average, there are 7,345 observations in this restriction of the validation sample.
Each line corresponds to a different model and each column corresponds to a different information set. Column (1) uses a constant only, column (2) uses industry dummies and log total assets ("PE" information set) and column (3) a large set of firm-level characteristics. In all models, expected cash-flows are given by analyst forecasts up to year 3 and then a perpetual growth assumption.

forecasts does contain valuable information. The ICC allows to account for the term structure of EPS forecasts but not that of discount rates. It performs well, even with a "trivial" (constant) or a small "PE" (size + industries) set of variables. The loglinear model does slightly better. This suggests that allowing for time-varying discount rates can lead to more accurate valuations. Put differently, it seems that the approximation error introduced by the CS approximation is more than offset by this additional flexibility.

The difference in MSEs between small characteristic size ("PE") and larger characteristics set ("Corp. Acc.") highlights the key role of the full term structure of EPS forecasts. "PE" only contains size and industry dummies, so the multiple only takes into account next year forecasts. "Corp. Acc." contains a host of firm-level characteristics, including the full term-structure of forecasts. This is why the multiple approach improves vastly (from an MSE of 0.316 to 0.209): It is allowed to used EPS forecasts two and three years out. The ICC does, however, not improve that much, so the distance between the two narrows. A multiple approach that can use "as much" information as the PV-based imputed ICC approach becomes quite competitive. The importance of the information in future forecasts will be highlighted on several occasions below.

Machine learning models are competitive with NPV-based models, with local linear forests outperforming all NPV-based models when the set of features is large enough. This suggests that there is some value to allowing for richer interactions between variable if investors have access to enough information about projects. As discussed previously, the performance of forest-based models also attests to the strength of the method of comparables, since RF can be viewed as a weight generator.

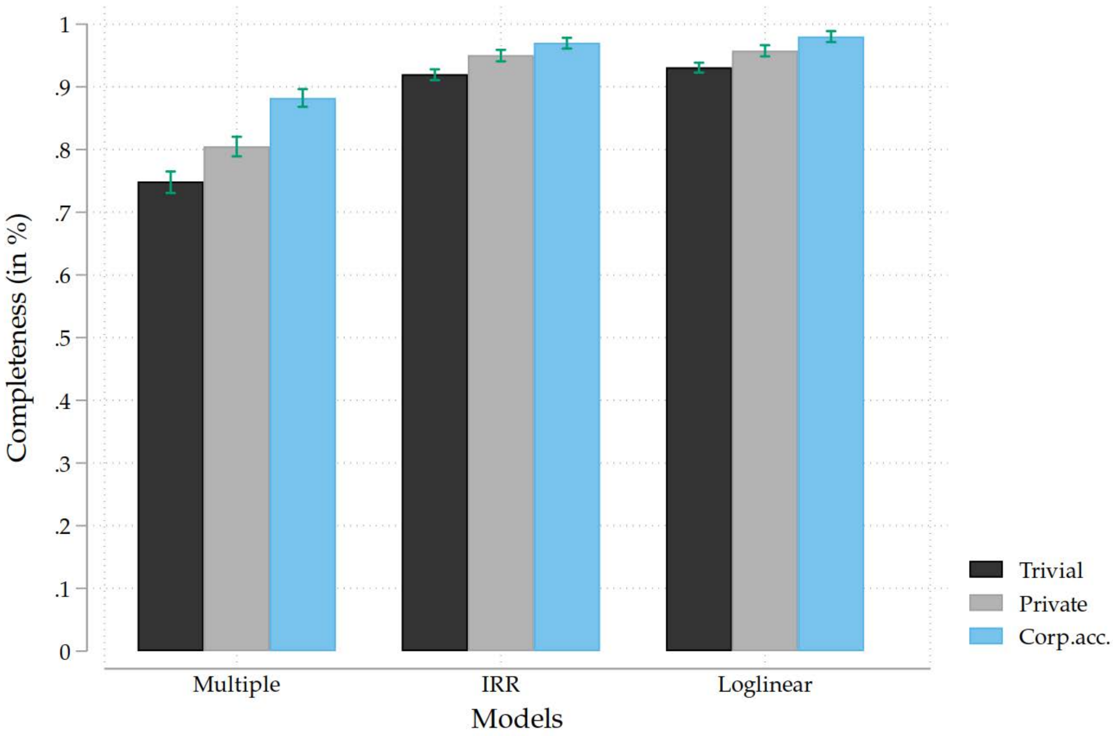
Another way to report the results of Table 3 is to view machine learning models as a black box that gives an estimate of the maximum predictability of prices in the data. This idea is formalized by Fudenberg et al. (2022), who introduce a metric of completeness defined by

$$\frac{\text{MSE}_{\text{benchmark}} - \text{MSE}_{\text{model}}}{\text{MSE}_{\text{benchmark}} - \text{MSE}_{\text{irreducible}}}$$

where MSE_m denotes the mean-squared error of model m. The completeness of a model is bounded between zero and one. It captures how much a model improves on a benchmark model, relative to how much improvement is achievable given observables. A completeness of 1 corresponds to a model being as precise in forecasting as black-box statistical forecasts. A completeness of zero corresponds to the benchmark model.

Figure 5 plots the completeness of models, with the CAPM used as the benchmark model and local linear forest estimated with a large information set used as a proxy for irreducible error. It is apparent that even with a trivial information set, all NPV-based model outperform the CAPM. Multiple-based methods become relatively better when more information is available, with a completeness above 80%. Finally, the imputed ICC and the flexible loglinear model perform best. They are almost 90% complete with a trivial information set (which is the best performance of the multiple) and reach near 100% completeness with maximal information. This last point means that the structure imposed by the present value formula does not "hurt predictability": a purely statistical model does barely better than the imputed ICC-based NPV, and not meaningfully better than the log-linear model.

Figure 5: Completeness of predictive models



Note. This figure reports the median of the completeness metric of Fudenberg et al. (2022) over 100 train-test splits. Confidence intervals at the 90% level are displayed. Completeness is computed as

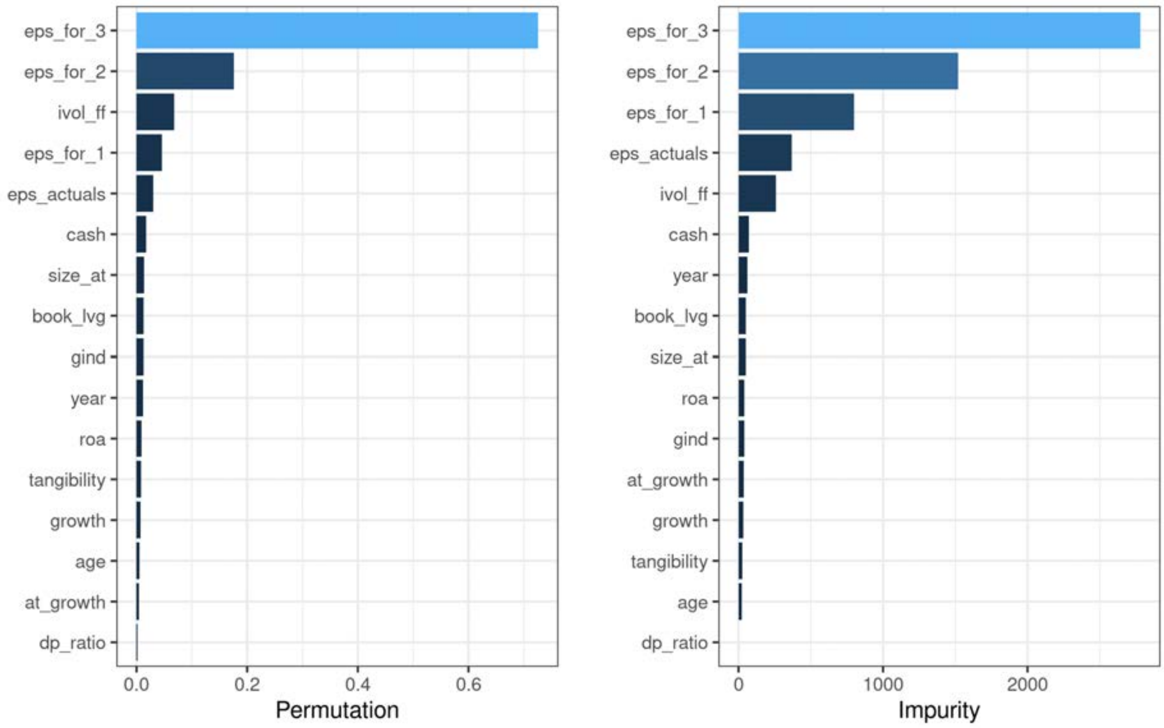
$$\frac{\text{MSE}_{\text{benchmark}} - \text{MSE}_{\text{model}}}{\text{MSE}_{\text{benchmark}} - \text{MSE}_{\text{irreducible}}}$$

where MSE_m denotes the mean-squared error of model m. The irreducible error is proxied by the error of the local linear forest model trained using corporate accounts information. The benchmark model is the CAPM.

5.3 Which observables are the most important?

We can gain additional insight on the contribution of each covariate by looking at measures of variable importance defined from random forests. We report two measures of variable importance (permutation and impurity) in Figure 6. Both measures of importance give very similar results.

Figure 6: Variable importance



Note. This figure reports two measures of variable importance associated with the random forests model applied to the prediction of log prices. Permutation importance is defined as the increase in mean squared error by randomly permuting observations for variable j. The prediction error is measured on out-of-bag data. Impurity importance corresponds to the cumulated decrease in impurity (defined as the sum of squared errors) attributed to splitting over variable j. This decrease is computed over each tree separately and then averaged over the entire forest. Impurity measures are corrected using the method proposed by Nembrini et al. (2018).

Figure 6 contains two insights. First, even though it is the most predictive variable, the one-year ahead EPS forecast is not enough to predict future prices: the full term structure of IBES forecasts is informative. Second, beyond all three EPS forecasts, the other key variable is a measure of risk: idiosyncratic volatility. This suggests that the term-structure cash-flows and risk are the two key components in the predictable part of prices, in line with the PV formula. Most other variables only have a negligible explanatory power.

6 Robustness Checks

This Section shows that our headline results are robust to numerous potential concerns.

6.1 Statistical forecasts

One concern with our analysis is that the expectations of investors, who are pricing the stocks in our sample, may differ from analyst forecasts. Another worry is that analyst forecasts directly affect prices, so that using these forecasts to explain prices is a bit tautological: a capital budgeter would not be in a position to affect the PV of a project by simply putting a number on it.

In this Section, we address these concerns by using statistical forecasts of cash-flows instead. Statistical forecasts are chosen to minimize the MSE of realized EPS. As described in Section 3, we estimate statistical forecasts of EPS based on IBES forecasts and corporate accounts characteristics, using LASSO. We then implement the different models exactly as described previously, but using statistical forecasts instead of IBES forecasts. We compute the MSE on the validation sample restricted to observations for which statistical forecasts are available up to three years ahead, and the log price prediction of every model is available.

We report the results in Table 4.[5] The broad ranking of models does not change, with factor models performing quite poorly and NPV-based models (multiple, ICC and loglinear) doing well. Within NPV-based models, the multiple is now comparable to the imputed ICC and the loglinear model. It outperforms the imputed ICC method, especially when a large set of firm-level characteristics ("Corp. Acc.") are available. This is not too surprising, as this large set of characteristics contains all three EPS forecasts. As a result, the multiple estimation is a linear version of the Random Forest approaches. With the smaller set of variables ("PE"), then ICC and multiple much closer to one another: in this

[5]To simplify presentation, we do not report the MSEs using analyst forecasts. This is because they are quantitatively and qualitatively similar. Details on our sample selection procedure are presented in Appendix D.

Table 4: Performance of various models in predicting log prices with statistical EPS forecasts

	Performance			
Model	Trivial	PE	Corp.acc.	Factor
CAPM				0.811
				(0.017)
CAPM (including a term-structure)				0.851
				(0.017)
CAPM (flexible ERP)				0.563
				(0.015)
CAPM (LW)				0.947
				(0.015)
Fama-French-Carhart				1.306
				(0.030)
Characteristics-based predicted returns				0.992
				(0.020)
Multiple	0.164	0.140	0.105	
	(0.003)	(0.005)	(0.004)	
ICC	0.205	0.176	0.147	
	(0.006)	(0.014)	(0.013)	
Loglinear model	0.197	0.169	0.127	
	(0.004)	(0.009)	(0.006)	
Random forest		0.100	0.089	
		(0.003)	(0.003)	
Local linear forest		0.099	0.086	
		(0.003)	(0.003)	

Note. This table reports the average MSE across models and information sets. Given a train-test split, we restrict our sample to observations for which statistical forecasts are available up to three years ahead. We further require that the log price prediction of every model be available. The MSE of model m is defined as

$$MSE^m = \frac{1}{N} \sum_{(i,t)} \left(\log P_{it} - \log \widehat{P}_{it}^m \right)^2$$

We repeat this procedure 100 times on different train-test splits and report the average MSE across all splits along with standard deviations. On average, there are 3,872 observations in this restriction of the validation sample.

Each line corresponds to a different model and each column corresponds to a different information set. Column (1) uses a constant only, column (2) uses industry dummies and log total assets ("PE" information set) and column (3) a large set of firm-level characteristics. In all models, expected cash-flows are given by analyst forecasts up to year 3 and then a perpetual growth assumption.

case, the imputed ICC method uses the full term structure of forecast, while the multiple approach does not, as the multiple is calibrated using size and industry dummies only.

In Table 4 as in Table 3, the loglinear model is still, by a small margin, the model that performs best. We report completeness measures in Figure A.3 of the Appendix. The bottom line is that the imputed ICC methodology, which is easy to implement, outperforms standard factor models and is competitive with other NPV methods even with very limited information available. A simple multiple-based approach performs well when many feature are observed.

6.2 Free cash flows

As a baseline, we chose to discount forecasts of *payout to equity* computed as earnings forecasts multiplied by an industry-level payout ratio (b_i). We chose this baseline because analyst forecasts are focused on earnings, many dividend forecasts are zero and share buy-back forecasts are not available. However, we can also provide valuations based on discounting of other forms of expected free cash-flows that are commonly used in practice. We study equity valuations based on (1) discounting expected *free cash flows to the firm*, and (2) expected *free cash flows to equity*. To define these two free cash flow variables, we use in each case a simple OLS model estimated in rolling windows, which uses past observables (the lag of the predicted free cash flow, IBES earnings forecasts, and the "corporate accounting" information set). The two types of free cash flows are defined in a standard manner using Compustat (cf. Appendix D). Table A.8 documents the statistical properties of forecast errors based on these estimates. For (2) we use the same discount rate as in our baseline; For (1), we define firm value as the market value of equity plus the book value of total debt and use for the CAPM formula a discount rate defined by

$$e_{it} \times \widehat{r_{it}^{CAPM}} + (1 - e_{it}) \times \text{AA Bond yield} \times (1 - \text{Tax rate})$$

where e_{it} is the share of equity in total firm value, the AA Bond yield is the ICE BofA AA US Corporate Yield Index from FRED, and the tax rate is set to 35%. The other pricing methods are then similarly adapted to these expected free cash-flows and provide estimates of equity prices. As before, we compare out-of-sample the relative precision of these estimates. The results are reported in Table A.9. Two conclusions emerge. The first one is the confirmation that here again the ICC approach strongly dominates the CAPM and other expected returns based approaches. The second one is that using expected payouts (our baseline approach) leads to more precise estimates in all methods, which justifies using it as our baseline. This might be due to the fact that using a variable that is directly proportional to the estimates of analysts on the numerator of our discounting formulas optimally exploits the information contained in IBES.

6.3 Adjusting for leverage

In this Section, we explore whether leverage adjustment changes our analysis, and find that it does not. So far, we have left leverage adjustment out of the discussion. In practice, however, projects may have a different leverage than comparables. The CAPM method we use accounts for this as firm-level equity betas should scale up with leverage. But, in the ICC, multiple and loglinear approaches, the additional risk related to leverage should in principle be accounted for. Existing textbooks make the following three-step recommendation, which is valid irrespective of the model used to compute equity discount rate r_e (it relies on the Modigliani-Miller conservation of value principle):

1. **Unlever the cost of equity of comparables**: For all comparable firms, unlever the cost of equity r_e^* with the following formula:

$$r^* = \frac{E^*}{E^* + D^*} r_e^* + \frac{D^*}{E + D^*} r_d^*$$

   where r_d^* is the cost of debt, D^* and E^* the *market values* of equity and debt of comparables.

2. **Compute the average unlevered cost of capital across comparables**. In our setting, this would mean regressing r on firm observables in the training sample. In most applications, where r_e is assumed well described by the CAPM, this would mean computing the average unlevered beta of comparables.

3. **Relever the cost of capital with the project's own capital structure**. This means using the following formula:

$$r_e = r^* + \frac{D}{E}\left(r^* - r_d\right)$$

   where r_d is an assumed cost of debt for the project, and $\frac{D}{E}$ is the project's own gearing ratio.

Thus, if a project is more levered than comparables, its equity will have a higher discount rate. What would be a an equivalent approach in our setting? A problem with the above approach is the well-known circularity problem. Even if the capital budgeter knows D, she does not know the project's equity value E, since finding it out is the goal of valuation.

To explore whether leverage is empirically important in our non-expected returns based PV-based methods, we thus implement the following exercise. To the "PE" characteristics (size and industry dummies), we add the *book* leverage ratio, which is technically not the right ratio but has the advantage of not containing any information on the price. So we then implement our method: First, we regress ICC, multiple, or loglinear models on industry dummies, size and book leverage in the training sample; Second, we use this model to predict the ICC, multiple or discount rate adjustment in the validation sample. Finally, we use the prediction to compute predicted prices in the validation sample, and compute the MSE of each model.

We show the results in Table A.7: There, we report MSEs using the "PE" variables and MSEs with "PE+leverage", for each one of the three PV-based models that do not use models of expected returns. We find that it makes absolutely no difference. This reflects the fact that leverage has a small explanatory power on multiples, ICC, and the loglinear models in the regressions presented in Table A.4. From this analysis, we draw the conclusion that leverage adjustments do not empirically improve the prediction of the cross-section of values.

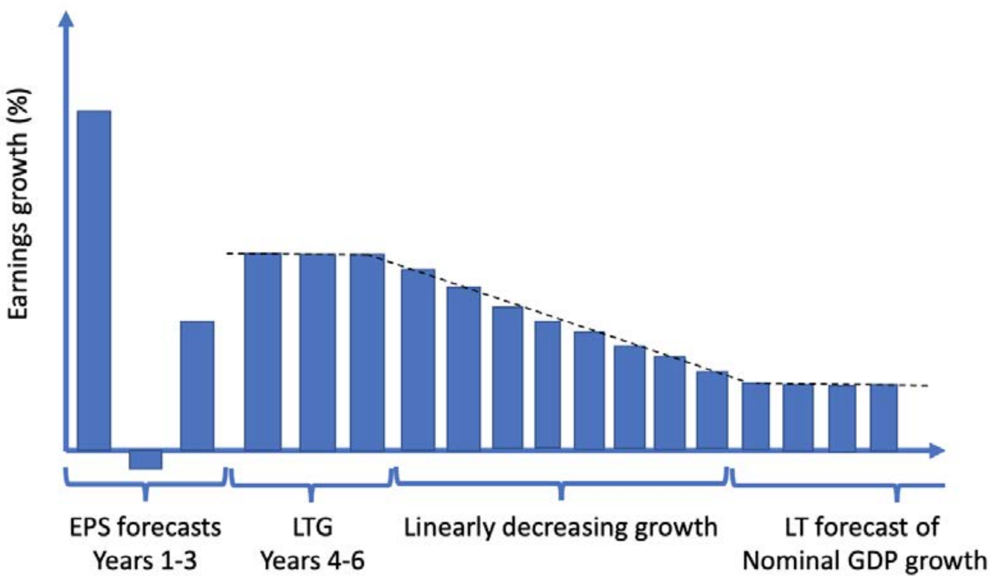
6.4 Robustness to lagging information

In interpreting our results, one question that arises naturally is whether the success of the ICC method vis-a-vis CAPM is due to its ability to use the most recent information about how the market prices firm characteristics, or whether it captures more persistent patterns in the cross-section of firm values which factor models do not capture. To test this, we perform the following test: We compare the performance of lagged valuation models, i.e. we predict prices using current firm characteristics but valuation models as estimated at various yearly lags. The results are reported in Figure A.5 to A.7. The picture that emerges is quite clear. While its performance slowly decays as we use lagged valuation formulas, the ICC method keeps performing largely better than the best version of the CAPM does (the flexible ERP one), even when one does not lag the CAPM parameter estimates. Indeed, for example, using valuation parameters that are five years old (applied to current firm characteristics), the "trivial ICC" delivers a log MSE of .18, which is still well ahead of the 0.55 log MSE of the "flexible ERP CAPM". This confirms that the CAPM is not dominated because it is more "slow-moving" than other methods, but rather because it fails to capture persistent differences in the cross-section of valuations that other models are able to catch.

6.5 The 4-stage growth model

In this Section, we explore the robustness of our findings with respect to a more sophisticated sequence of cash-flow forecasts. In our baseline specification, after the first three years of cash-flows forecasts made by analysts, the firm is expected to grow perpetually at the long-term growth rate of the economy. We now propose a different transition to this long-term growth rate. Instead, we construct the PV by assuming that, between year 4 and year 6, the firm's earnings are assumed to growth at a constant rate given by the consensus on "long-term growth" (LTG) from IBES Summary Files. After year 15, the firm is assumed to growth forever at the long-term nominal GDP growth rate embedded in the consensus of macro forecasters. Between years 6 and 15, growth is assumed to slow down linearly.

Figure 7: The 4-stage growth model



Note: We present here graphically the more involved model of cash-flow forecasts. After the first three years of analyst forecasts, growth during years 4-6 is assume to be equal to LTG (from IBES). GDP Growth after 15 years is equal to the consensus of macro forecasters from the Philly Fed. In between, the slowdown is linear.

Compared to the baseline model, on average across all firms, these alternative cash-flow forecasts tend to increase the duration of the stock: cash-flows between year 4 and year 15 tend to be larger, as LTG is typically larger than the GDP growth rate (with a median of 12% a year and a mean of 14%), which we implicitly used in our baseline approach. Thus, if a model is wrong about the discount rate, the MSE will be more impacted.

We report the comparative MSEs in Table 5, and find results similar to the baseline Table 3. The multiple model and forest-based models do not contain any assumption about the term structure of cash-flow forecasts, so their MSEs should be identical to the baseline, save for a sample selection effect. In the four-stage growth specification, models based on expected returns still do terribly. This is likely related to the larger duration of the four-stage growth set-up, which gives more weight to long-term cash-flows. The noise introduced by these models becomes more toxic as stock duration increases. Multiple, imputed ICC and loglinear models, however, perform well. The overall ranking of models is still preserved, with multiples being outperformed by the imputed ICC method. The term-structure allowed by the loglinear model still brings out additional predictive power.

Table 5: Performance of various models in predicting log prices in the four-stage growth setup

	Performance			
Model	Trivial	PE	Corp.acc.	Factor
CAPM				1.553
				(0.025)
CAPM (including a term-structure)				1.635
				(0.026)
CAPM (flexible ERP)				0.448
				(0.009)
CAPM (LW)				2.018
				(0.029)
Fama-French-Carhart				2.647
				(0.011)
Characteristics-based predicted returns				0.489
				(0.020)
Multiple	0.206	0.187	0.142	
	(0.006)	(0.005)	(0.003)	
ICC	0.163	0.139	0.124	
	(0.003)	(0.002)	(0.003)	
Loglinear model	0.148	0.126	0.111	
	(0.002)	(0.002)	(0.001)	
Random forest		0.100	0.089	
		(0.003)	(0.003)	
Local linear forest		0.099	0.086	
		(0.003)	(0.003)	

Note. This table reports the average MSE across models and information sets. Given a train-test split, we restrict our sample to observations for which the long-term growth IBES forecast is available. We further require that the log price prediction of every model be available. The MSE of model m is defined as

$$MSE^m = \frac{1}{N} \sum_{(i,t)} \left(\log P_{it} - \log \widehat{P}_{it}^m \right)^2$$

We repeat this procedure 100 times on different train-test splits and report the average MSE across all splits along with standard deviations. On average, there are 3,872 observations in this restriction of the validation sample.
Each line corresponds to a different model and each column corresponds to a different information set. Column (1) uses a constant only, column (2) uses industry dummies and log total assets ("PE" information set) and column (3) a large set of firm-level characteristics. In all models, expected cash-flows are given by analyst forecasts up to year 3 and then a perpetual growth assumption.

Also, we note as previously that giving the full term structure of forecasts to the multiple approach improves it dramatically (going from "PE" to "Corp. Acc."), making it quite competitive with imputed ICC.

6.6 Returns shrinkage for factor models

As an additional robustness check, we try to account for the fact that some predictions for factor models may yield $\widehat{r}_{it+1}^{m} < g_t$. Returns below the long-term growth forecast are incompatible with the NPV formula and cannot be used by an investor who wants to value a project using future cash-flows. Earlier, we winsorized these observations, replacing the estimated discount rates by $g_t + 1\%$.

In an alternative specification, we shrink returns for CAPM-based models and the four factor model towards 10% so that

$$\widehat{r}_{it+1}^{Shrinked} = 0.5 \times \widehat{r}_{it+1}^{m} + 0.5 \times 10\%$$

Shrinking returns this way significantly reduces the number of observations with predicted returns that are incompatible with the NPV-formula. We expect the performance of factor models to significantly improve since 10% is close to the empirical mean of the ICC, which already performs well as a single discount rate (see for instance Table 3: as we discussed previously, the imputed ICC approach with an average ICC for the entire sample – column "trivial" – already yields a reasonable MSE).

Results are reported in Table A.6 of the Appendix. As expected, the performance of factor models is significantly improved. This is especially true for the CAPM, whose MSE is divided by three. However, the imputed ICC and the loglinear models continue to outperform both the multiple model and factor models, even with a trivial information set. As before, these models reach near-completeness when a rich set of corporate accounts variable is used for prediction.

7 Conclusion

We compare how good various valuation techniques are at explaining the cross-section of equity prices. Using forecasts of future payouts to equity and characteristics of firms observed at a given point in time (excluding the price itself), we predict equity values with several techniques, most of them involving the present value formula that is widely taught in the classroom. We compare these predicted values to effective prices, and rank valuation techniques based on their ability to reduce the cross-sectional mean square error. We find that computing a discount rate by using the CAPM or other models of expected returns is highly detrimental, when it comes to predict the cross-section of price levels. Even a simple constant multiple of forecasted earnings does better as a valuation tool. Discounting with a simple horizon-independent imputed ICC, function of firm characteristics, obtains better results in our validation sample. Allowing for time-varying expected returns provides a marginal improvement vis-a-vis this method, and in some instances, an a-structural approach, using machine learning does even better. Overall, we recommend an alternative valuation method to what is taught in the classroom. While CAPM-based valuation is normatively grounded (under heavy model assumptions), it is a very poor predictor of the cross-section of firm prices. We propose, as an implementable alternative, to use discount rates obtained on comparables – e.g. by industry. Such discount rates are very easy to use. They yield an MSE that is about 5 times smaller than CAPM-implied discount rate. They provide more accurate valuation figures, are very easy to implement and can thus be easily taught in class.

8 Bibliography

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Ang, Andrew, and Jun Liu, 2004, How to discount cashflows with time-varying expected returns, *The Journal of Finance* 59, 2745–2783.

Athey, Susan, Julie Tibshirani, and Stefan Wager, 2019, Generalized random forests, *The Annals of Statistics* 47, 1148 – 1178.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, 2013, Cross-country differences in productivity: The role of allocation and selection, *American Economic Review* 103, 305–34.

Berk, Jonathan, and Peter DeMarzo, 2006, *Corporate Finance*.

Biau, Gérard, and Erwan Scornet, 2016, A random forest guided tour, *TEST* 25, 197–227.

Breiman, Leo, 2001, Random forests, *Machine Learning* 45, 5–32.

Brennan, Michael J, 1997, The term structure of discount rates, *Financial Management* 81–90.

Broda, Christian, and David E. Weinstein, 2006, Globalization and the Gains From Variety*, *The Quarterly Journal of Economics* 121, 541–585.

Campbell, John, 2017, *Financial Decisions and Markets: A Course in Asset Pricing*.

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *The Review of Financial Studies* 1, 195–228.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *Journal of finance* 52, 57–82.

Catherine, Sylvain, Thomas Chaney, Zongbo Huang, David Sraer, and David Thesmar, 2022, Quantifying Reduced-Form Evidence on Collateral Constraints, *Journal of Finance* .

Cochrane, John, 2011, Discount rates, *Journal of Finance* 64.

Cohen, Randolph, Christopher Polk, and Tuomo Vuolteenaho, 2009, The price is (almost) right, *Journal of Finance* 64, 2739–2782.

De La O, Ricardo, Xiao Han, and Sean Myers, 2022, The cross-section of subjective expectations: Understanding prices and anomalies.

De la O, Ricardo, and Sean Myers, 2021, Subjective cash flow and discount rate expectations, *The Journal of Finance* 76, 1339–1387.

Dean, Joel, 1951, *Capital Budgeting* (Columbia University Press).

Décaire, Paul, 2021, Capital budgeting and idiosyncratic risk.

DeSilva, Tim, and David Thesmar, 2021, Noise in expectations: Evidence from analyst forecasts, Technical Report 28963, NBER.

Dessaint, Olivier, Jacques Olivier, Clemens Otto, and David Thesmar, 2021, Capm-based comapny (mis)valuation, *Review of Financial Studies* 34.

Dewing, Arthur, 1934, *The Financial Policy of Corporations* (The Ronald Press Company).

Fama, Eugene F., and Robert R. Bliss, 1987, The information in long-maturity forward rates, *The American Economic Review* 77, 680–692.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Farre-Mensa, Joan, Roni Michaely, and Martin Schmalz, 2014, Payout policy, *Annual Review of Economics* 6.

Frazzini, Andrea, and Lasse Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Friedberg, Rina, Julie Tibshirani, Susan Athey, and Stefan Wager, 2021, Local linear forests, *Journal of Computational and Graphical Statistics* 30, 503–517.

Fudenberg, Drew, Jon Kleinberg, Annie Liang, and Sendhil Mullainathan, 2022, Measuring the completeness of economic models, *Journal of Political Economy* 130, 956–990.

Gebhardt, William R., Charles M. C. Lee, and Bhaskaran Swaminathan, 2001, Toward an implied cost of capital, *Journal of Accounting Research* 39, 135–176.

Gollier, Christian, 2020, The welfare cost of ignoring the beta, Working paper.

Gormsen, Niels, and Kilian Huber, 2022a, Corporate discount rates, Working paper.

Gormsen, Niels, and Kilian Huber, 2022b, Equity factors and firms' perceived cost of capital, Working paper.

Graham, John, 2022, Corporate finance and reality, *Journal of Finance* .

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Reiew of Financial Studies* 33, 2223–2273.

Gupta, Arpit, and Stijn VanNieuwerburgh, 2021, Valuing private equity strip by strip, *Journal of Finance forthcoming* .

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, 2009, *The elements of statistical learning: data mining, inference and prediction*, second edition (Springer).

Hennessy, Christopher A., and Toni M. Whited, 2005, Debt Dynamics, *Journal of Finance* 60, 1129–1165.

Hong, Harrison, and David Sraer, 2016, Speculative betas, *Journal of Finance* 71, 2095–2144.

Hughes, John, Jing Liu, and Jun Liu, 2009, On the relation between expected returns and implied cost of capital, *Review of Accounting Studies* 14, 246–259.

Jacobs, Michael, and Anil Shivdasani, 2012, Do you know your cost of capital?, *Harvard Business Review* .

Jagannathan, Ravi, David Matsa, Iwan Meier, and Vefa Tarhan, 2016, Why do firms use high discount rates?, *Journal of Financial Economics* 120, 445–463.

Keloharju, Matti, Juhani Linnainmaia, and Peter Nyberg, 2020, Long-term discount rates do not vary across firms, *Forthcoming Journal of Financial Economics* .

Levi, Yaron, and Ivo Welch, 2017, Best practice for cost-of-capital estimates, *Journal of Financial and Quantitative Analysis* 52, 427–463.

Midrigan, Virgiliu, and Daniel Xu, 2014, Finance and misallocation: Evidence from plant-level data, *American Economic Review* .

Nembrini, Stefano, Inke R König, and Marvin N Wright, 2018, The revival of the Gini importance?, *Bioinformatics* 34, 3711–3718.

So, Eric, 2013, A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts?, *Journal of Financial Economics* 108, 615–640.

Stein, Jeremy, 1996, Rational capital budgeting in an irrational world, *Journal of Business* 69, 429–455.

Tauchen, George, 1986, Finite state markov-chain approximations to univariate and vector autoregressions, *Economics Letters* 20, 177–181.

van Binsbergen, Jules H., Xiao Han, and Alejandro Lopez-Lira, 2020, Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases, *Working Paper* .

Welch, Ivo, 2021, The cost of capital: If not the capm, then what?

Wright, Marvin N., and Andreas Ziegler, 2017, ranger: A fast implementation of random forests for high dimensional data in C++ and R, *Journal of Statistical Software, Articles* 77, 1–17.

APPENDIX – FOR ONLINE PUBLICATION

A Appendix Material for Section 2

The firm-level optimization problem is a standard dynamic programming one. While it admits no closed form solution, it can be solved numerically by discretizing the state space and using value function iteration. This gives us a numerical estimate for the optimal investment-borrowing policy function. Given, we simulate a large panel of firms whose cash-flows we discount at a rate $r^\star$. This gives us estimates for the value of a firm using the wrong discount rate, which we use to compute the value us.

Calibration. Table A.1 summarizes our calibration. The one parameter that is important is the elasticity of demand φ, which we set to 6.7 in our baseline calibration as in Midrigan and Xu (2014). We also try a larger value 10 which corresponds to some estimates in Broda and Weinstein (2006) but is admittedly in the upper part of the acceptable range. The other parameters take standard values.[6]

Discretization. We work with a discretized version of the state space. For the productivity process, we implement the method in Tauchen (1986) to create a grid with 30 points for log TFP. We use bounds of -3 and 3 standard deviations. We then create a 50×30 grid for capital and debt. For capital, we form an equally spaced grid from $\log k_{\min}$ to $\log k_{\max}$, where

$$k_{\min} = \frac{1}{2}\left(\frac{(1-\tau)\theta e^{\rho z_{\min}+\frac{\sigma}{2}}}{r_f(1-\tau M)+(1-\tau)\delta}\right)^{\frac{1}{1-\theta}} \quad ; \quad k_{\max} = \left(\frac{(1-\tau)\theta e^{\rho z_{\max}+\frac{\sigma}{2}}}{r_f(1-\tau M)+(1-\tau)\delta}\right)^{\frac{1}{1-\theta}}.$$

For debt, we use an equally spaced grid for the debt-to-capital ratio ranging from $-m = -0.25$ to $M = 0.4$. Finally, we use an equally spaced grid with 41 points for discount rates ranging from 0.05 to 0.15.

Solution method. For each r in the grid, we solve for the value function, policy function, and stationary distribution.

1. Starting from a guess for the value function $\widehat{V}_0$, iterate on the Bellman equation until convergence. To do so, we use a grid search to solve, for every $n > 0$,

$$\left(\widehat{\kappa}_n(z,k,d;r), \widehat{\Delta}_n(z,k,d;r)\right) = \arg\max_{k',d'} \Pi(z,k,d,k',d') + \frac{1}{1+r}\mathbb{E}\left[\widehat{V}_{n-1}(z,k,d;r) \mid z\right],$$

   and update the value function accordingly. To speed-up convergence, we use 30 steps of policy function iteration for each step of value function iteration. We assess convergence by checking

$$\max\left|\log \widehat{V}_n(z,k,d;r) - \log \widehat{V}_{n-1}(z,k,d;r)\right| \le 10^{-5}.$$

   Denote $\widehat{V}(\cdot;r)$, $\widehat{\kappa}(\cdot;r)$, and $\widehat{\Delta}(\cdot;r)$ the value and policy functions at the end of the procedure.

2. Simulate N firms over T periods using the estimated policy $\widehat{\kappa}(\cdot;r)$ and $\widehat{\Delta}(\cdot;r)$ starting from an arbitrary state. Discard the first T_0 periods for each firm, and define

$$\widehat{\mu}(S;r) = \frac{1}{N(T-T_0+1)}\sum_{i=1}^{N}\sum_{t=T_0}^{T}\mathbf{1}\{S_{it}=S\},$$

   where $S=(z,k,d)$, i denotes a firm, and t denotes time. We use $N=10^5$, $T=500$, and $T_0=300$. We ensure that those values are sufficient to attain convergence by checking that we obtain results

[6] For the constraints on leverage and cash, we roughly use the third quartile for the book leverage ratio and cash ratio in our sample. Moreover, the parameters ρ and σ reported the scaled log TFP process z.

similar to the 10^{-3} using different seeds and uniformly drawn initial states.

To estimate the valuation loss from discounting at the wrong loss, we use the following method.

1. For each $r^\star$ in the grid, draw initial states from the estimated stationary distribution $\widehat{\mu}(\cdot; r^\star)$.
2. Using these initial states, for each r in the grid, simulate N firms over T periods using the estimated policy functions $\widehat{\kappa}(\cdot; r)$ and $\widehat{\Delta}(\cdot; r)$. We use $N = 10^5$ and $T = 300$.
3. Using the simulated cash-flows compute the present value

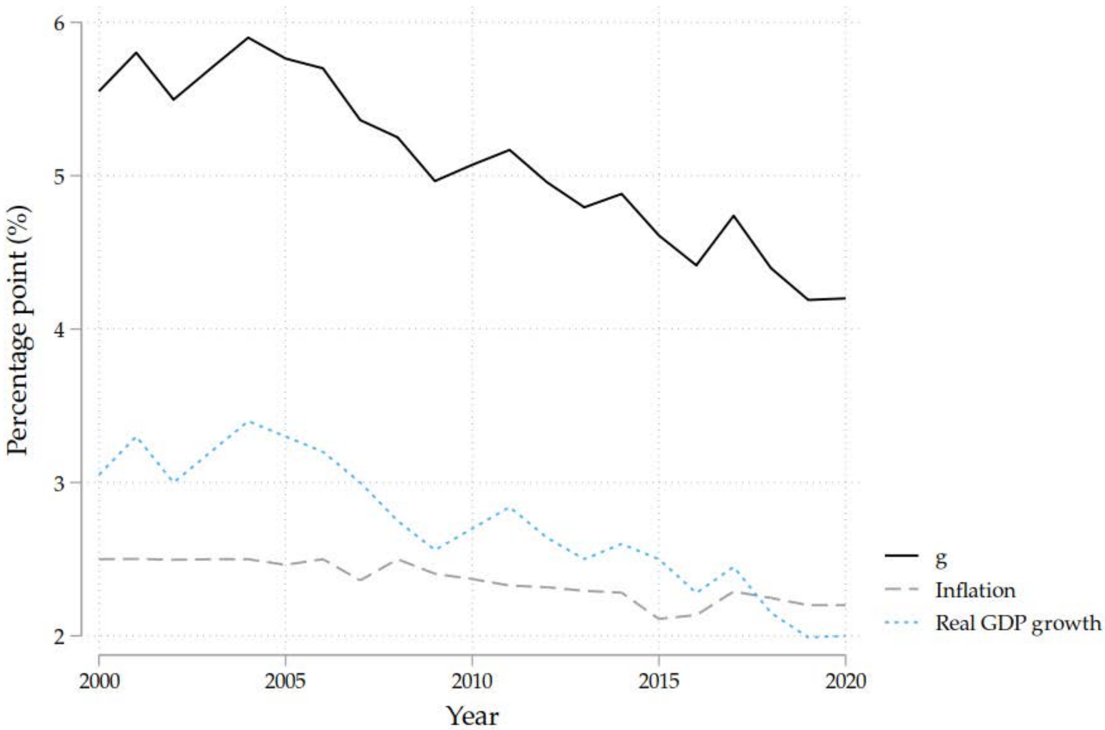
$$\widehat{W}(r, r^\star) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=0}^{T-1} \frac{\Pi\left(z_{it}, k_{it}, d_{it}, \widehat{\kappa}(z_{it}, k_{it}, d_{it}; r), \widehat{\Delta}(z_{it}, k_{it}, d_{it}; r)\right)}{(1 + r^\star)^t}.$$

   The estimated value loss is

$$\frac{\widehat{W}(r^\star, r^\star) - \widehat{W}(r, r^\star)}{\widehat{W}(r^\star, r^\star)}.$$

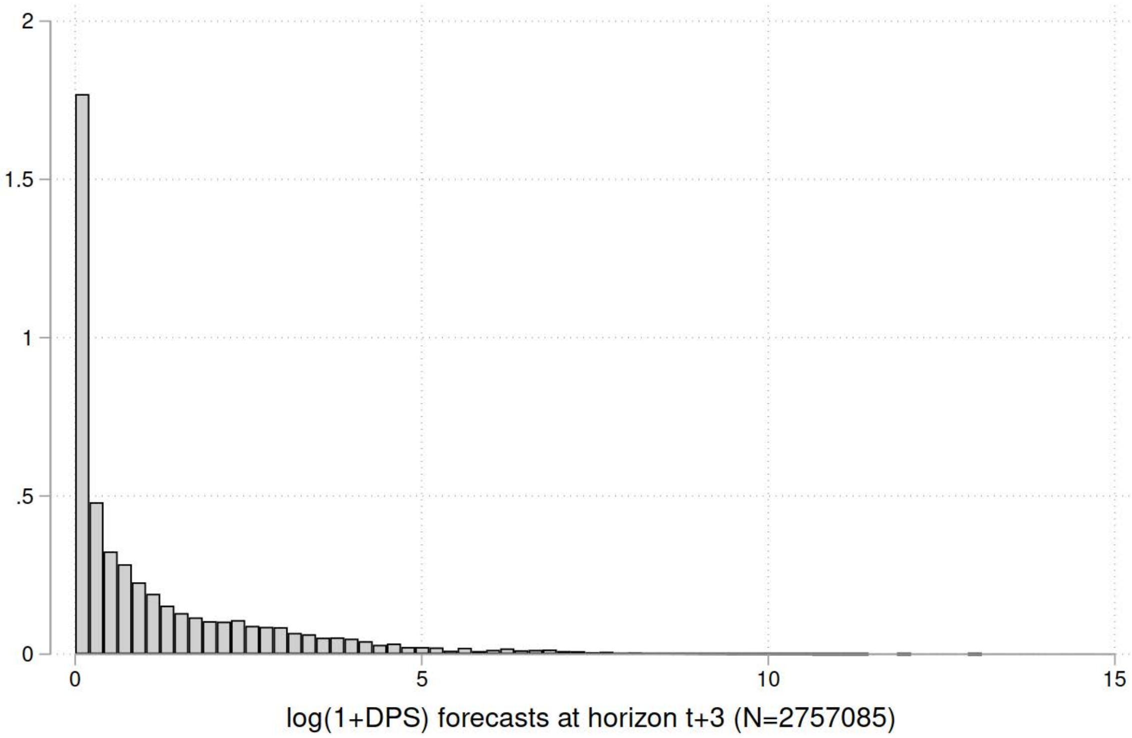
B Appendix Figures

Figure A.1: Time series of long-run nominal GDP forecasts



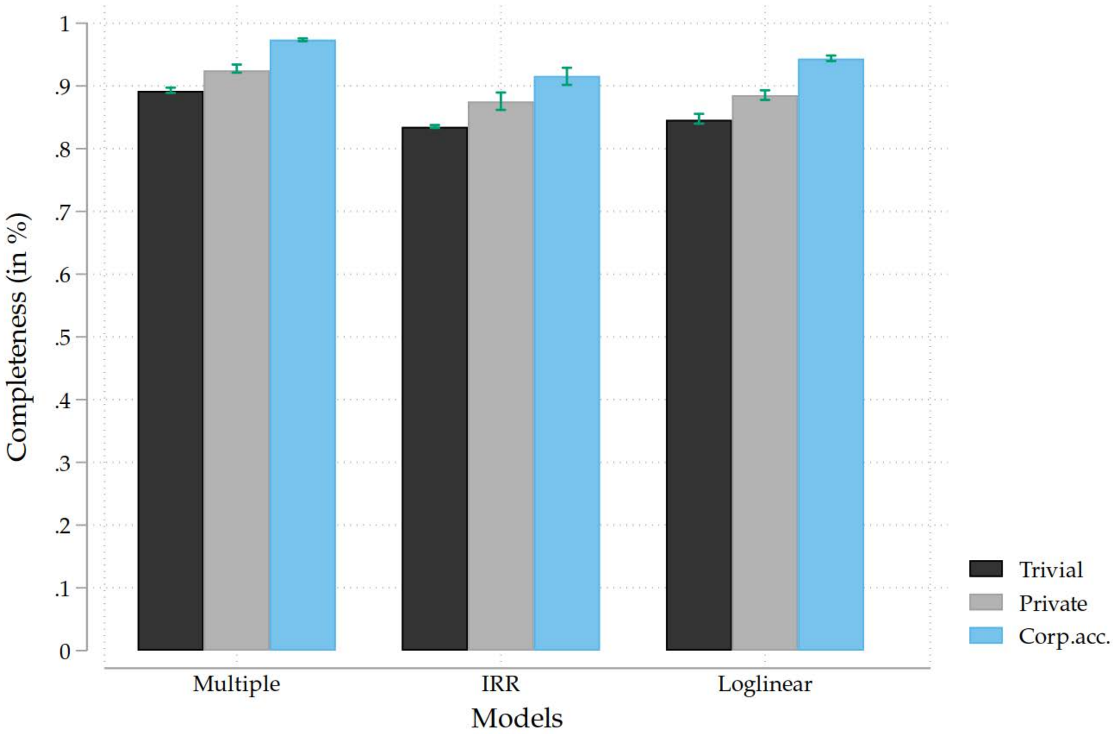
Note. We add 10-year real GDP growth forecasts to the 10-year horizon inflation forecasts. For both variables, we use the median consensus forecast from the Survey of Professional Forecasters maintained by the Philadelphia Federal Reserve.

Figure A.2: Distribution of (log) DPS forecasts



Note. We use the median consensus firm-level 3 years ahead DPS forecast from the IBES Summary files. Out of the 2,700,000 observations, over 500,000 (about 19%) are exactly zero.

FIGURE A.3: Completeness of predictive models using statistical forecasts

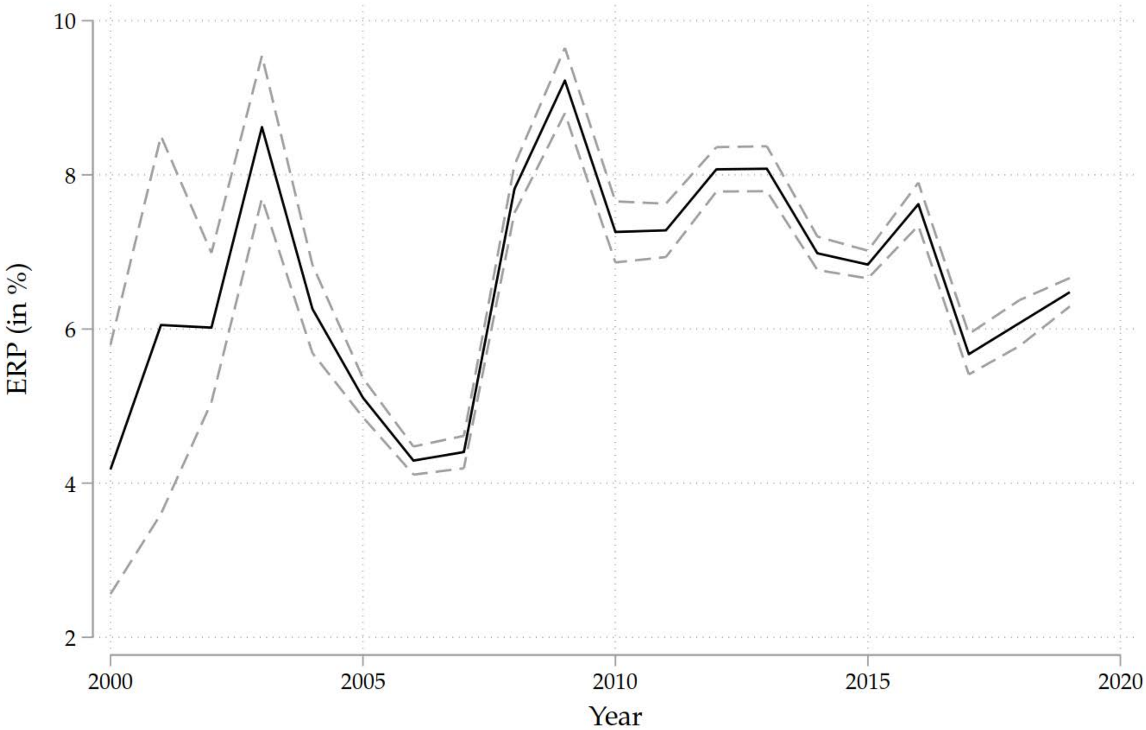


Note. This figure reports the median of the completeness metric of Fudenberg et al. (2022) over 100 train-test splits using statistical forecasts. Confidence intervals at the 90% level are displayed. Completeness is defined as

$$\frac{\text{MSE}_{\text{benchmark}} - \text{MSE}_{\text{model}}}{\text{MSE}_{\text{benchmark}} - \text{MSE}_{\text{irreducible}}}$$

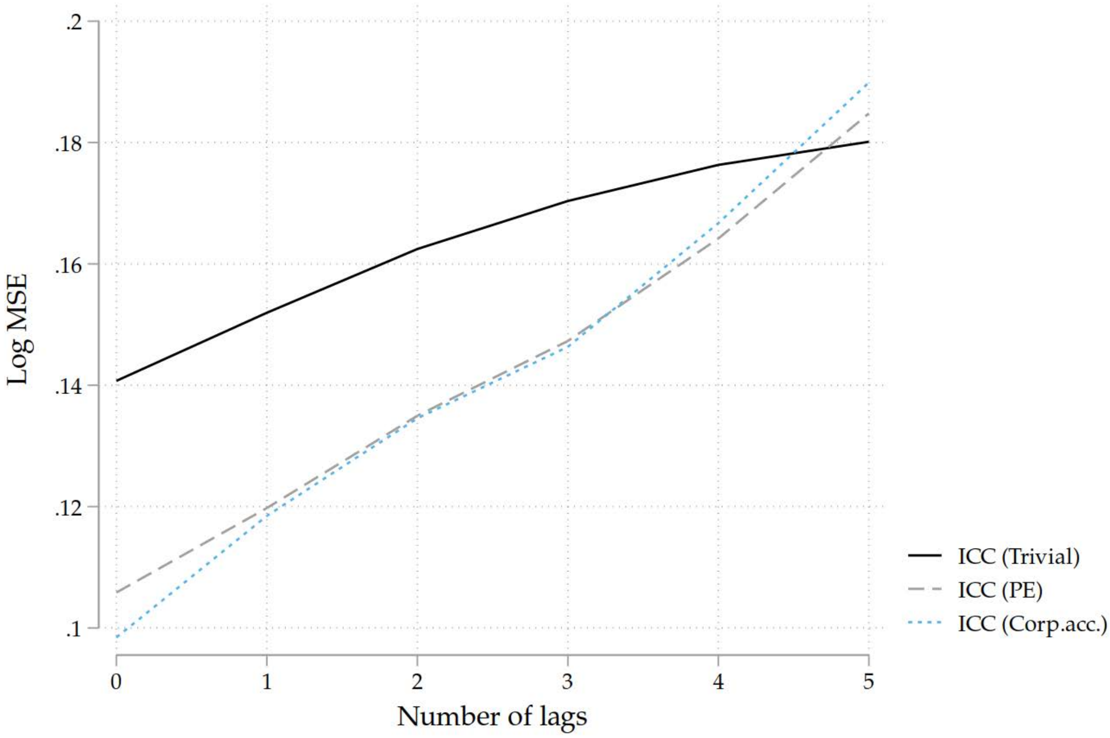
where MSE_m denotes the mean-squared error of model m. The irreducible error is proxied by the error of the local linear forest model trained using corporate accounts information. The benchmark model is the CAPM.

Figure A.4: Optimal equity risk-premium for the CAPM



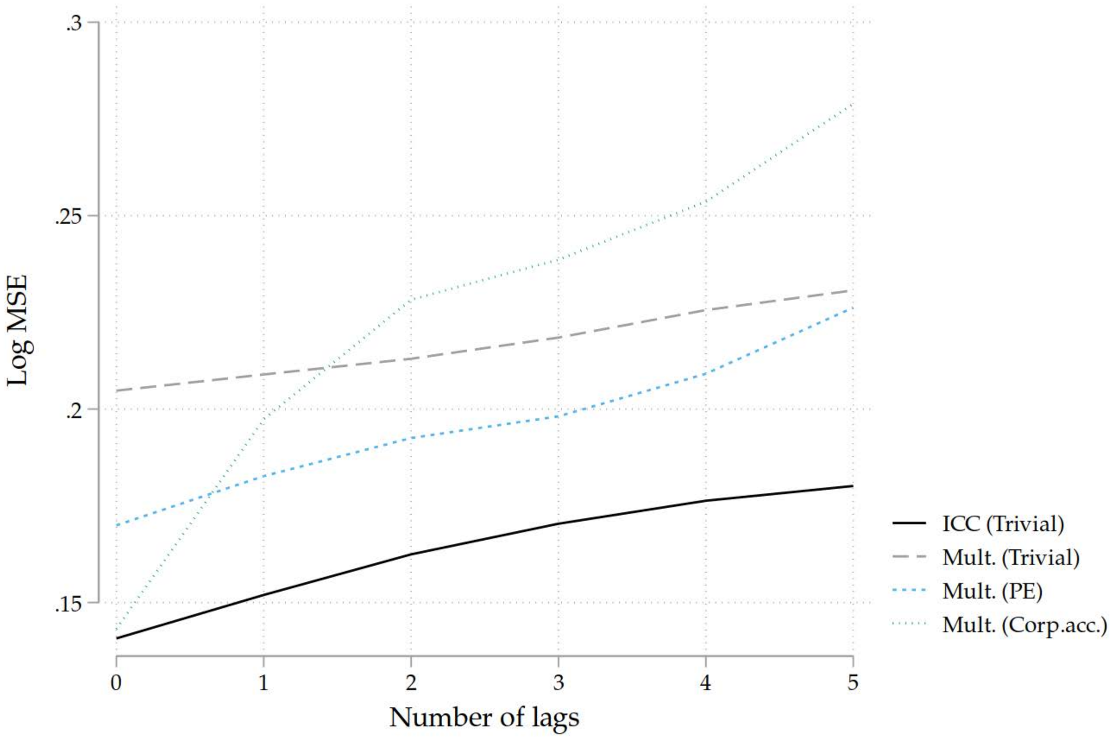
Note. For each year, we estimate the risk-premium that minimizes the in-sample mean-squared error of the CAPM using nonlinear least squares. The dotted line shows pointwise bootstrapped confidence intervals at the 95% level.

FIGURE A.5: Performance of the IRR when using lagged information



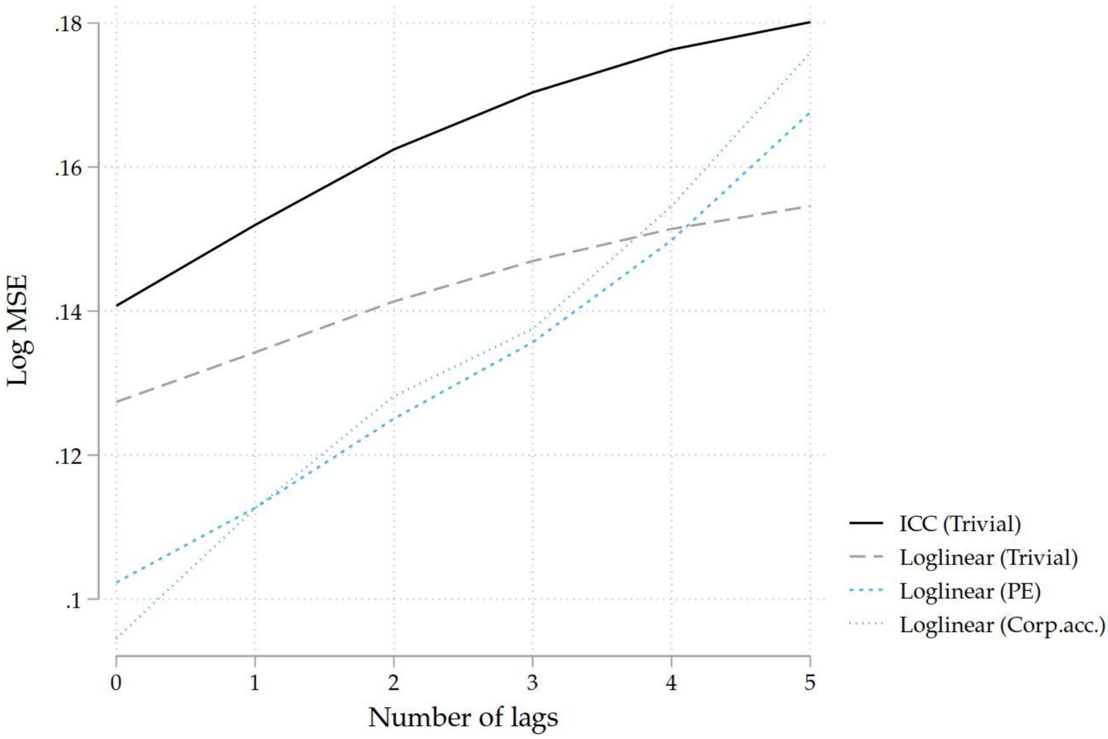
Note. This figure plots the out-of-sample MSE obtained by using the lagged discount rate $\widehat{\text{ICC}}_{it-h}$ to discount the cash-flows of firm i at time t. We compute the MSE using only those observations for which all three models (ICC, multiple, and loglinear) are available for all lags h.

Figure A.6: Performance of the multiple when using lagged information



Note. This figure plots the out-of-sample MSE obtained by using the predicted multiple $\widehat{\text{Multiple}}_{it-h}$ lagged by h years to compute the price of firm i at time t. We also plot the corresponding MSE using a lagged discount rate predicted using a trivial information set for comparison. We compute the MSE using only those observations for which all three models (ICC, multiple, and loglinear) are available for all lags h.

Figure A.7: Performance of the loglinear model when using lagged information



Note. This figure plots the out-of-sample MSE obtained by using the predicted loglinear discount rate term $\widehat{\text{DR}}_{it-h}$ lagged by h years to compute the price of firm i at time t. We also plot the corresponding MSE using a lagged discount rate predicted using a trivial information set for comparison. We compute the MSE using only those observations for which all three models (ICC, multiple, and loglinear) are available for all lags h.

C Appendix Tables

Table A.1: Calibration summary

Parameter	Description	Value	Source
α	Capital expenditure share	0.33	Bartelsman et al. (2013)
φ	Demand elasticity	6.7	Broda and Weinstein (2006)
θ	Production curvature	0.65	$\frac{\alpha(\varphi-1)}{1+\alpha(\varphi-1)}$
ρ	Persistence of TFP	0.85	Catherine et al. (2022)
σ	Volatility of TFP innovations	0.30	–
γ	Investment costs	0.10	Catherine et al. (2022)
δ	Depreciation rate	0.06	Midrigan and Xu (2014)
r_f	Risk-free borrowing rate	0.03	Catherine et al. (2022)
M	Maximum leverage	0.40	Compustat
m	Maximum cash	0.25	Compustat
τ	Tax rate	0.33	Statutory tax rate
λ	Equity issuance cost	0.10	Catherine et al. (2022)

Table A.2: List of predictive variables

Variable	(1)	(2)	(3)	(4)	Construction
Industry		✓	✓		`gind`
Size (assets)		✓	✓		Log `at`
EPS_t			✓		From IBES Actuals
F_tEPS_{t+1}			✓		Median EPS forecast from IBES
F_tEPS_{t+2}			✓		Median EPS forecast from IBES
F_tEPS_{t+3}			✓		Median EPS forecast from IBES
ROA			✓		`ebitda` divided by `at`
Book leverage			✓		Total debt (`dltt` plus `dlc`) over `at`
Age			✓		2019 minus year of entry in Compustat
Cash ratio			✓		`che` over `at`
Assets growth			✓	✓	`at` growth rate
IVOL			✓	✓	Idiosyncratic volatility from FFC model
Tangibility			✓	✓	`ppent` over `at`
CAPEX				✓	`capx`
Size (market cap.)				✓	Log market cap
Net issue				✓	Log difference of `shrout`
Dividend payout				✓	`dvc` over market cap
Book to market				✓	`ceq` over market cap
Amihud ratio				✓	See Appendix D
Market leverage				✓	Debt plus market cap over market cap

Note. Column (1) describes the trivial information set, column (2) the "PE" information set, column (3) the corporate accounts information set, and column (4) the variables used for predicting returns.

Table A.3: Summary statistics

	Final Sample ($N = 19416$)				
	Mean	S.d.	Med.	Min	Max
Variable	(1)	(2)	(3)	(4)	(5)
Age	31.72	16.90	27.00	2.00	69.00
Amihud ratio	–7.43	1.99	–7.66	–10.31	3.65
Assets growth	1.12	0.27	1.06	0.60	2.76
Book leverage	0.23	0.17	0.22	0.00	0.87
Book to market	4.83	3.83	3.91	0.39	36.36
Capital expenditures	4.38	1.94	4.42	–6.21	8.00
Cash ratio	0.14	0.16	0.09	0.00	0.99
Dividend payout ratio	0.01	0.02	0.01	0.00	0.90
EPS	2.24	1.99	1.78	–2.87	8.69
EPS forecast at $t + 1$	1.57	1.42	1.20	0.03	9.77
EPS forecast at $t + 2$	1.80	1.55	1.39	0.09	10.63
EPS forecast at $t + 3$	2.02	1.71	1.57	0.13	11.88
Idiosyncratic volatility	1.85	0.93	1.65	0.45	12.75
Log market capitalization	14.95	1.63	14.90	8.07	20.98
Log sale	7.64	1.65	7.62	0.85	13.15
Net issue	–0.04	0.15	–0.00	–0.84	0.18
ROA	0.13	0.08	0.12	–0.66	0.39
Tangibility ratio	0.24	0.23	0.16	0.00	0.88

Note. Columns (1) to (5) report summary statistics for the final sample.

Table A.4: In-sample predictive regressions of models

	(1) Mult.	(2) ICC	(3) DR	(4) Mult.	(5) ICC	(6) DR	(7) Mult.	(8) ICC	(9) DR
Log assets				-0.09	0.08	-0.03	-0.05	0.11	-0.02
				(0.01)	(0.03)	(0.01)	(0.01)	(0.04)	(0.01)
EPS for. $t+1$							-0.32	0.14	-0.01
							(0.05)	(0.16)	(0.03)
EPS for. $t+2$							0.15	-0.88	0.15
							(0.05)	(0.22)	(0.04)
EPS for. $t+3$							0.16	0.57	-0.09
							(0.03)	(0.14)	(0.03)
EPS							-0.08	0.27	-0.07
							(0.02)	(0.06)	(0.01)
ROA							-1.56	0.20	-0.28
							(0.19)	(0.60)	(0.14)
IVOL							-0.07	0.80	-0.13
							(0.02)	(0.10)	(0.02)
Tangibility							0.37	-0.75	0.21
							(0.08)	(0.25)	(0.06)
Assets growth							0.01	-0.22	0.06
							(0.03)	(0.12)	(0.03)
Book lvg.							-0.34	1.85	-0.35
							(0.06)	(0.26)	(0.05)
Age							0.00	0.00	-0.00
							(0.00)	(0.00)	(0.00)
Cash							0.89	-2.23	0.63
							(0.09)	(0.31)	(0.07)
Constant	3.43	9.52	-1.71						
	(0.03)	(0.17)	(0.03)						
Observations	9,863	9,863	9,863	9,863	9,863	9,863	9,863	9,863	9,863
R^2	-0.00	0.00	-0.00	0.19	0.22	0.22	0.38	0.34	0.32
Fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Note. This table shows in-sample predicting regressions for one particular train-test split. We show estimates using the entirety of the training sample although predicting regressions are estimated on rolling windows using available information only. Columns (1) to (3) reports estimates for the "trivial" information set, while columns (4) to (6) report estimates for the "PE" information set and columns (7) to (9) for the "corporate accounts" information set. When included, fixed effects are year and industry dummies. Standard errors are two-way clustered by firm and year.

Table A.5: In-sample predictive regressions of the ICC

	(1) ICC	(2) ICC	(3) ICC	(4) ICC	(5) ICC	(6) ICC	(7) ICC	(8) ICC
Beta	0.24	-0.10	0.07	-0.04	0.26	-0.31	-0.04	0.16
	(0.17)	(0.15)	(0.13)	(0.11)	(0.22)	(0.20)	(0.21)	(0.17)
IVOL		0.57	0.37	0.20		0.77	0.57	0.32
		(0.10)	(0.05)	(0.06)		(0.20)	(0.10)	(0.11)
Dividend yield			1.72	3.65			0.19	5.37
			(1.65)	(1.80)			(3.12)	(4.58)
Net issue			-0.02	0.04			0.16	-0.03
			(0.23)	(0.16)			(0.25)	(0.15)
Book-to-market			0.13	0.11			0.16	0.18
			(0.02)	(0.01)			(0.03)	(0.03)
ROA			3.44	2.43			2.69	2.06
			(0.60)	(0.72)			(0.60)	(0.69)
Amihud ratio			0.11	0.08			0.30	0.38
			(0.05)	(0.04)			(0.11)	(0.12)
Market leverage			0.40	0.52			0.55	0.71
			(0.07)	(0.08)			(0.14)	(0.13)
Assets growth			0.16	0.11			-0.65	-0.71
			(0.08)	(0.08)			(0.13)	(0.13)
Capex			-0.53	-0.45			-0.56	-0.41
			(0.08)	(0.07)			(0.09)	(0.07)
Size			0.35	0.34			0.18	0.02
			(0.10)	(0.09)			(0.14)	(0.14)
Tangibility			-2.49	-2.08			-2.48	-1.89
			(0.29)	(0.31)			(0.44)	(0.41)
Observations	9,553	9,553	9,553	9,553	4,671	4,671	4,671	4,670
R^2	0.00	0.07	0.29	0.40	0.00	0.08	0.31	0.44
Fixed effects	No	No	No	Yes	No	No	No	Yes

Note. Note. This table shows in-sample predicting regressions for one particular train-test split. In columns (1) to (4), we use the ICC computed from IBES forecasts. In columns (5) to (8), we use the ICC computed from statistical forecasts. When included, fixed effects are year and industry dummies. Standard errors are two-way clustered by firm and year.

Table A.6: Performance of various models in predicting log prices with shrinkage of factor models towards 10%

	Performance			
Model	Trivial	PE	Corp.acc.	Factor
CAPM				0.245
				(0.008)
CAPM (including a term-structure)				0.245
				(0.008)
CAPM (flexible ERP)				0.164
				(0.003)
CAPM (LW)				0.219
				(0.006)
Fama-French-Carhart				0.480
				(0.004)
Characteristics-based predicted returns				0.676
				(0.012)
Multiple	0.206	0.187	0.142	
	(0.006)	(0.005)	(0.003)	
ICC	0.126	0.105	0.097	
	(0.004)	(0.007)	(0.005)	
Loglinear model	0.121	0.104	0.092	
	(0.003)	(0.005)	(0.003)	
Local linear forest		0.099	0.086	
		(0.003)	(0.003)	

Note. This table reports the average MSE across models and information sets. We shrink the returns predicted by factor models towards 10%. Given a train-test split, we restrict our sample to observations for which the log price prediction of every model is available. The MSE of model m is defined as

$$MSE^m = \frac{1}{N} \sum_{(i,t)} \left(\log P_{it} - \log \widehat{P}_{it}^m \right)^2$$

We repeat this procedure 100 times on different train-test splits and report the average MSE across all splits along with standard deviations. On average, there are 3,872 observations in this restriction of the validation sample.

Each line corresponds to a different model and each column corresponds to a different information set. Column (1) uses a constant only, column (2) uses industry dummies and log total assets ("PE" information set) and column (3) a large set of firm-level characteristics. In all models, expected cash-flows are given by analyst forecasts up to year 3 and then a perpetual growth assumption.

TABLE A.7: Performance of various models in predicting log prices with and without leverage

	Subjective		Statistical		Four-stage	
Model	PE	LVG	PE	LVG	PE	LVG
Multiple	0.187	0.186	0.140	0.136	0.187	0.186
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)
ICC	0.105	0.105	0.176	0.176	0.139	0.138
	(0.007)	(0.005)	(0.014)	(0.016)	(0.002)	(0.003)
Loglinear model	0.104	0.102	0.169	0.166	0.126	0.125
	(0.005)	(0.004)	(0.009)	(0.009)	(0.002)	(0.002)

Note. This table reports the average MSE across models and information sets. Given a train-test split, we restrict our sample to observations for which the log price prediction of every model is available. The MSE of model m is defined as

$$MSE^m = \frac{1}{N} \sum_{(i,t)} \left(\log P_{it} - \log \widehat{P}_{it}^m \right)^2$$

We repeat this procedure 100 times on different train-test splits and report the average MSE across all splits along with standard deviations. On average, there are 3,872 observations in this restriction of the validation sample.
Each line corresponds to a different model and each column corresponds to a different information set. The "PE" columns use industry and year dummies, along with log total assets ("PE" information set). The "LVG" column also includes book leverage.
We compare the performance of models across three sets of forecast: subjective IBES forecasts, statistical forecasts, and forecasts constructed according to a four-stage growth model.

Table A.8: Predicting FCFF and FCFE using a linear model

Horizon	$t+1$	$t+2$	$t+3$
Panel A: Mean forecast error			
Free cash-flows to the firm	-0.05	-0.08	-0.01
Free cash-flows to equity	-0.06	0.05	0.22
Panel B: s.d. of forecast error			
Free cash-flows to the firm	2.12	2.31	2.37
Free cash-flows to equity	4.25	4.54	4.74
Panel C: R^2_{OOS}			
Free cash-flows to the firm	0.44	0.36	0.36
Free cash-flows to equity	0.80	0.77	0.77
Observations	5471	5471	5471

Note. We compute the out-of-sample forecasting error $\mathrm{FCF}_{it+h} - \widehat{\mathrm{FCF}}_{it+h}$ for free cash-flows to the firm and free cash-flows to equity, where $\widehat{\mathrm{FCF}}_{it+h}$ is computed using a linear model trained in-sample. Panels A and B report the means and s.d. of these errors, by rule and horizon. Panel C shows the OOS R^2 measured as:

$$R^2_{OOS} = 1 - \frac{\sum_{it}\left(\mathrm{FCF}_{it+h} - \widehat{\mathrm{FCF}}_{it+h}\right)^2}{\sum_{it}\left(\mathrm{FCF}_{it+h} - \overline{\mathrm{FCF}}_{t+h}\right)^2}$$

where $\overline{\mathrm{FCF}}_{t+h}$ is the constant forecast, so the statistic R^2_{OOS} represents the MSE gain of the forecasting rule relative to the average of past realizations.

Table A.9: Performance of various models in predicting log prices using free cash-flows

	Performance			
Model	Trivial	PE	Corp.acc.	Factor
Panel A – Analyst forecasts				
CAPM				0.919
				(0.015)
CAPM (including a term-structure)				0.967
				(0.016)
CAPM (LW)				1.093
				(0.015)
Fama-French-Carhart				1.555
				(0.016)
ICC	0.178	0.154	0.138	
	(0.005)	(0.005)	(0.005)	
Panel B – FCFF forecasts				
CAPM				2.113
				(0.060)
CAPM (including a term-structure)				2.548
				(0.068)
CAPM (LW)				2.438
				(0.066)
Fama-French-Carhart				2.581
				(0.068)
ICC	0.508	0.417	0.327	
	(0.029)	(0.025)	(0.017)	
Panel C – FCFE forecasts				
CAPM				1.520
				(0.059)
CAPM (including a term-structure)				1.584
				(0.061)
CAPM (LW)				1.929
				(0.073)
Fama-French-Carhart				2.451
				(0.092)
ICC	0.311	0.279	0.251	
	(0.018)	(0.017)	(0.016)	

Note. This table reports the average MSE across models and information sets. Given a train-test split, we restrict our sample to observations for which the log price prediction of every model is available. The MSE of model m is defined as

$$MSE^m = \frac{1}{N} \sum_{(i,t)} \left(\log P_{it} - \log \widehat{P}_{it}^m \right)^2$$

We repeat this procedure 100 times on different train-test splits and report the average MSE across all splits along with standard deviations.
Panel A is taken from Table 3, which is computed on a different subsample as Panels B and C. This is only to simplify presentation, as computing all MSEs on the sample subsample would not alter the results.

Table A.10: Comparison of model performance in the training sample and the testing sample

	Performance			
Model	Trivial	PE	Corp.acc.	Factor
CAPM				–0.9
				(–2.7)
CAPM (including a term-structure)				–0.9
				(–2.6)
CAPM (flexible ERP)				0.6
				(1.6)
CAPM (LW)				–0.8
				(–3.0)
Fama-French-Carhart				–0.3
				(–1.5)
Characteristics-based predicted returns				–0.1
				(–0.4)
Multiple	–1.0	12.4	16.4	
	(–1.5)	(17.0)	(21.0)	
ICC	–0.2	14.4	15.9	
	(–0.4)	(20.6)	(21.4)	
Loglinear model	–0.3	15.7	18.4	
	(–0.5)	(23.0)	(24.0)	
Random forest		142.3	158.4	
		(95.3)	(100.7)	
Local linear forest		71.1	254.1	
		(65.5)	(107.9)	

Note. This table reports the performance loss coming from out-of-sample prediction, measured as

$$\left(\frac{\text{MSE}_{\text{In-sample}}}{\text{MSE}_{\text{Out-of-sample}}} - 1\right) \times 100.$$

We compute this measure for 100 different train-test splits and report its average across all splits. The t-statistic for the test that the average performance loss is zero is in parentheses.

D Data appendix

In this section, we describe our methodology in more details. Variable definitions are summarized in Table A.2, which also contains which variable is used in which predictive regression. We report summary statistics for the final dataset in Table A.3.

Sample selection steps. In our baseline results, we use the following procedure:

1. keep observations between year 1980 and 2019 (for predictive regressions) and between 2000 and 2019 (for the final sample).
2. keep observations whose share code (`shrcd` in CRSP) is 10 or 11, and whose exchange code (`exchcd` in CRSP) is 1, 2, or 3.
3. drop observations whose price per share (`prc` in CRSP) is below 1 or above 5,000.
4. drop observations for which any of the following variable is negative: long-term debt (`dltt`), debt in current liabilities (`dlc`), total assets (`at`), shares outstanding (`shrout`)
5. drop observations when the `permno` appears twice in the same fiscal year in Compustat, or when the IBES ticker appears twice in the same fiscal year in the IBES Unadjusted Summary Files.
6. drop observations which cannot be merged across all datasets: CRSP, Compustat, CAPM loadings and Fama–French–Carhart loadings from the WRDS beta suite, Fama–Bliss discount bonds, IBES Unadjusted Summary Files, and IBES Actuals.
7. drop observations for which any variable used in predictive regressions, or used to construct such variable, is missing (see Table A.2 for a complete list).
8. drop observations for which EPS forecasts at horizons $t+1$, $t+2$, or $t+3$ are not available.
9. drop observations for which EPS forecasts are negative.
10. drop observations for which the predicted log price of any model is below zero or above seven.

For robustness results only, we add the following steps

9a drop observations for which the long-term growth forecast from IBES is missing.

9b drop observations for which actual earnings at time $t+1$, $t+2$, or $t+3$ are missing.

9c drop observations for which earnings as predicted by the statistical model are negative.

For the results using free cash-flows only, we add the following steps

9a drop observations for which FCFF or FCFE are missing.

9b drop observations for which the forecasted FCFF or FCFE at horizon $t+3$ is negative.

Definition of variables.

- CRSP–Compustat variables.
  - Returns. We use the CRSP Monthly Stock File to compute returns over the year

$$\text{Future returns}_m = \prod_{k=0}^{11} \texttt{ret}_{m-k},$$

where m denotes a monthly time period.

- Market capitalization is defined as the product of `prc` and `shrout`.
- Illiquidity measure of Amihud (2002). Using daily data from CRSP, we compute the absolute value of returns (`ret`) divided by the volume traded in dollars (`prc` times `vol`). We drop penny stocks (`prc` < 1) and observations for which the volume traded is negative. We compute the average of this variable by month, and keep those observations for which there are at least 7 daily trading days in a month. Then, we winsorize the illiquidity measure at the 1% and 99% levels. Finally, we rescale the it and take logs.
- Asset growth is $\mathtt{at}_t/\mathtt{at}_{t-1}$.
- Book leverage is debt (see below) over `at`
- Book to market is `ceq` over market capitalization.
- Capital expenditures (capex) are `capx`.
- Cash is defined as `che` over equity `ceq`.
- Debt is the sum of long-term debt `dltt` and current liabilities `dlc`.
- Dividend payout is `dvc` over market capitalization.
- Industry-level payout ratio. We compute the firm-level payout ratio including stock repurchases, which we define as `dvc` plus `prstkc` minus `pstkrv` over `ni`. We winsorize this ratio at 0 and 1. We then compute its mean at the sector level (`gsec`) on a 5-year rolling window.
- Free cash flows to the firm (FCFF) are `oancf` minus `capx` over `shrout`.
- Free cash flows to equity (FCFE) are FCFF $-(1-\tau)\cdot$ `intpn` $+\,\Delta$Debt over `shrout`, where τ is the corporate tax rate. We use a tax rate of 35% throughout the sample.
- Industry is `gind`.
- Market leverage is debt plus market capitalization over market capitalization.
- Returns on assets (ROA) are `ebitda` over `at`.
- Returns on equity (ROE) are net income `ni` over equity `ceq`.
- Sector is `gsec`.
- Size (assets) is defined as $\log$ `at`, and size (market cap) is defined as $\log$ `prc` $+ \log$ `shrout`.
- Tangibility is `ppent` over `at`.

• Macroeconomic variables.

- Long-run inflation forecasts are taken from the SPF (median CPI10 series).
- Long-run real GDP growth forecasts are taken from the SPF (median RGDP10 series).
- Nominal GDP growth forecasts are the sum of long-run inflation and real GDP forecasts.
- 5-year treasury bonds are taken from FRED (DGS5 series at the monthly frequency).
- AA corporate yields are taken from FRED (BAMLC0A2CAAEY series at the monthly frequency).

• WRDS Factor Suite variables.

- Loadings for the CAPM and the Fama–French–Carhart models are estimated at the stock level using daily returns. We take our estimates from the WRDS Beta Suite, and we use the monthly average of those daily betas.

Winsorization steps. We winsorize variables at the 1% and 99% levels to minimize the influence of outliers, except CRSP variables (`prc`, `shrout`, and `ret`). We implement this step before constructing new variables. After constructing a variable, we also winsorize it at the 1% and 99% levels.