Corporate Valuation: An Empirical Comparison of Discounting Methods*

Nicolas Hommel† Augustin Landier‡
David Thesmar§

July, 2025

Abstract

To evaluate projects or firms, the textbook recommendation is to discount expected cash-flows with CAPM-based returns. In this paper, we focus on public firms, for which we can construct expected cash-flows and measure market values. We empirically ask how different discounting methods do in terms of predicting observed market values. First, we find that variants of the CAPM do the worst. A simple flat 10% discounting rule nearly reduces the MSE of prediction by half. Second, we use "imputed IRRs" obtained from statistical models estimated on a separate training sample. We find that they drastically improve the quality of the prediction of firm values. Compared to using a flat discounting rule, our richest non-linear models can reduce the MSE by more than half. Third, we show that under standard assumptions about the production function, the value gain from using our richest statistical model of imputed IRR, compared to the CAPM, is large, of the order of 10%.

*We especially thank Hui Chen, Niels Gormsen and Jérôme Taillard for their thoughts, as well as numerous seminar and conference participants.

†Princeton University
‡HEC Paris
§MIT, NBER and CEPR

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

To estimate the value of a project or a firm, the classic textbook recommendation (e.g. Berk and DeMarzo (2006)) is to use a simplified version of the present value formula:

$$PV = \sum_{t \geq 1} \frac{E\pi_t}{(1+r)^t} \tag{1}$$

where $E\pi_t$ is the expected cash-flows at horizon t and r is the project's cost of capital. To calculate this cost of capital, the general recommendation is to use the CAPM. This approach is used pervasively by practitioners. Graham (2022) documents that, in 2021, 85% of large corporations use the CAPM to compute r. Jagannathan et al. (2016), and more recently, Gormsen and Huber (2025) show that firms' reported costs of capital are well explained by realized betas on the market, with a slope close to standard values of the equity premium. Hence, firms do indeed rely on the CAPM in order to compute the cost of capital used to discount future cash-flows.

However, this approach raises concerns that are widely recognized in the academic literature. 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, if at all, from low beta securities (see for instance Frazzini and Pedersen (2014), Hong and Sraer (2016) and the references therein).

In this paper, we empirically compare the performance of various discounting methods. Our testing ground consists of publicly listed firms, for which we can construct reliable measures of expected cash-flows and measure market values. For each discounting rule we study, we ask how well expected cash-flows, discounted using this rule, predict observed values out-of-sample. Our criterion is the Mean Squared Error (MSE) of prediction.

Our analysis generates three broad conclusions. We first benchmark our analysis with variants of the CAPM. We find that discounting expected cash-flows with CAPM-type rates predicts market values very poorly. In particular, we show that discounting expected cash-flows at a flat 10% rate does much better, with an MSE that is 40% lower than that obtained with CAPM discounting. We show that popular adjustments such as beta shrinkage, allowing for time-varying equity risk premium, or allowing for a richer set of factors, do not alter this conclusion: In fact, these alternatives are even worse.

Second, we explore the discount rates obtained from "imputed IRRs". To compute these imputed IRRs, we first estimate, in a training sample, statistical models of varying complexity to predict the cross-section of IRRs based on firm characteristics (features). We then apply these models to compute imputed IRRs in a separate test sample, which we then use as the cost of capital to discount expected cash flows. We show that this valuation method fits the cross-section of market values much better than valuations based on the CAPM, and better than the straight 10% rule. Our models differ in the number of variables they use (a suite of accounting variables, or only industry dummies) and the method to combine them (linear or non-linear). In particular, we establish that non-linear methods (Random Forest, Gradient Boosting, Neural Networks) do best, especially so with a larger number of features. In the end, the MSE of non-linear methods with the largest set of features is approximately 60% smaller than taking a flat discount rule based on the market-wide past rolling average (a rule not too different than a flat 10%).

Finally, we provide a way to quantitatively interpret the MSE of predictions through a *sufficient statistic* approach. Assume a given discounting rule R (CAPM, flat 10%, statistical model of IRR). Let $\text{MSE}_{\text{IRR}}(R)$ be the MSE of this rule in predicting the cross-section of IRRs. We show that the loss of market value incurred by a firm using rule R in capital budgeting instead of its actual IRR (the market-based discount rate) is given by:

$$\text{Mean value loss} \approx \left(\frac{1+r}{2(r+\delta)^2} \cdot \frac{\theta}{(1-\theta)^2}\right) \cdot \text{MSE}_{\text{IRR}}(R)$$

where r is the average discount rate of firms, δ the rate of capital obsolescence, and θ the elasticity of profits to the capital stock. This formula holds in a benchmark model of firm behavior, but we show that it is quantitatively robust to including more complex frictions. After calibration, we estimate that a firm going from the CAPM to the straight 10% rule for its capital allocation would increase its market value by about 8%. Going from 10% to a non-linear prediction based on our full array of accounting features would further increase market values by another 5.6%.

Our paper is primarily a contribution to the current discussion on capital budgeting: we explore which valuation method is best at predicting the market values of firms. We believe the conclusions can be useful beyond the domain of public firms, e.g. in private equity or for project-level valuation. Some of our statistical models restrict features to what would be available for project valuation (so, not detailed accounts): Even then, the gain over using the CAPM is very large.

We find that using the CAPM to discount cash-flows is strongly suboptimal compared to alternative rules and likely to destroy substantial firm value. This is in line with Dessaint et al. (2021), who show that acquisitions of low beta targets tend to be more value-destroying, indicating that the use of the CAPM destroys value through target overvaluation. Based on CFO surveys, Graham (2022) shows that DCF with the CAPM is among the most popular decision rules among large U.S. firms (about three-quarters of firms indicate they almost always use these rules for capital budgeting). This confirms results from earlier surveys (as Graham and Harvey (2001), Jagannathan et al. (2016)). Graham (2022) concludes that the "widespread reliance on the CAPM, even in the face of evidence of the model's empirical shortcomings, suggests that firms rely on simple and familiar techniques. The enduring popularity of the CAPM may also reflect teaching emphasis." Our contribution shows that discounting based on the CAPM can indeed be highly misleading as a method to recover firm or project values and is largely dominated by alternative methods, which can be simple to implement.

While surveys show that firms use the CAPM to compute their cost of capital, their decision rules are a bit more complex. 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 determine their minimum return expectations. Gormsen and Huber (2022) and Gormsen and Huber (2025) confirm this. They extract hurdle rates ("discount rates", i.e. the minimum IRR a project has to clear) and perceived cost of capital expressed by firms' managers in earnings calls. As expected from the surveys cited above, they confirm that the perceived cost of capital is highly correlated with firm-level CAPM-based WACC. But hurdle rates are quite distinct from the costs of capital (with a univariate R^2 of 0.17, see Gormsen and Huber (2025), Table 4). This confirms that, while they use the CAPM to discount cash-flows, firms do not blindly use the CAPM to decide whether to invest or not.

Similar evidence of distanciation from the CAPM is found in analyst reports. Ben-David and Chinco (2024) find in their sample of analyst reports that analysts apply multiples (specifically Trailing P/E Ratios) in about two-thirds of the cases. Their sample is limited (some 500 reports), but Bastianello et al. (2025) confirm in a much larger sample of reports that about 40% of analyst memos use DCF, and 65% use multiples. Décaire and Graham (2025) document that analysts' discount rates closely follow standard textbook prescriptions: nearly all employ a weighted average cost of capital (WACC) and (among those that report their method), 96.8% use the CAPM to determine the cost of equity. Overall, while a significant number of analysts are faithful to present value discounted by the CAPM, a much larger number uses costs of capital embedded in market valuation. Our paper suggests that this might yield better results than WACC-based DCF, as it is conceptually closer to a DCF using the IRR of comparables as the discount rate.

Finally, we share with Cohen et al. (2009) and Cho and Polk (2020) a focus on the cross-section of price levels (as opposed to returns, as in most asset pricing literature), but our goal and approach differ in two important ways. First, we do not use ex-post realized cash-flows to evaluate mispricing, but ask how to best predict current prices based on *ex-ante* characteristics. Our aim is to measure how close the resulting valuation is from observed prices *out of our estimation sample*, not whether prices do reflect fundamentals as measured in ex-post realizations. Another difference with these papers is that we test various discounting alternatives, while they focus on the cash-flow-based CAPM.

The paper is structured as follows: Section 2 describes the data and cash-flow forecasts that we use, and the construction of our discount rate candidates. Section 3 describes our main results, namely the comparison of various valuation models in terms of their out-of-sample fit. This section details the role of the different information sets used to condition discount rates. Using a simple model of firm investment, Section 4 shows how discounting errors translate into value loss from using the wrong discount rate, and Section 5 offers extensive robustness checks. Section 6 concludes.

2 Data and variable construction

2.1 Data

We combine data from S&P Global's Compustat, the Center for Research in Security Prices (CRSP), and IBES. We restrict ourselves to the period ranging from 1993 to 2022, to ensure that IBES forecasts are reasonably well-populated:

— Accounting data are from S&P Global's Compustat Fundamental Annuals. We focus on consolidated financial statements of industrial companies whose currency is the dollar. We further restrict the sample to statements in standardized data format.

— We take data on stock prices (`prc`), returns (`ret`), and shares outstanding (`shrout`) from the CRSP Monthly Stock Returns database. We restrict the sample to ordinary common equity shares of companies incorporated in the U.S. We merge these data with the financial statements from Compustat using the CRSP link table. We retrieve prices three months after the fiscal year-end.

— From the IBES Unadjusted Summary Files, we retrieve contemporaneous earnings per share (EPS) along with median forecasts of EPS up to two fiscal years ahead. Using unadjusted data avoids retroactive adjustments in stock shares which would not be reflected in CRSP price per share. We collect forecasts at the same date as stock prices (three months after fiscal year-end), ensuring analysts have incorporated recent financial data. We match the IBES data with the CRSP–Compustat sample based on the eight-digit CUSIP. If we find no match, we use the exchange ticker and the six-digit CUSIP.

Factor models require additional measures of risk premia and factor loadings:

— The safe rate of return R_t^f is the 10-year Treasury yield (FRED DGS10).

— Monthly returns on market (MKTRF_t), size (SMB_t), value (HML_t) and momentum (UMD_t) factors are from the WRDS Fama–French series. Risk premia are averages of these returns over five-year rolling windows.

— Stock factor loadings β_{it}^E come from the WRDS Beta Suite. For both the CAPM and the Fama–French–Carhart model, these betas are estimated on a rolling 252-day windows.[1] We then compute our monthly betas by averaging daily betas within each calendar month. When relevant, idiosyncratic volatility is the volatility of the residual from the Fama–French–Carhart model.

— For debt betas β_{it}^D, we use ratings from S&P Global. We assign a debt beta of 0 to safe firms (rating AAA), a debt beta of 0.2 to investment grade (AA+ to BBB-), and a debt beta of 0.4 to high yield and unrated (BB- to D and unrated).[2]

Finally, long-term GDP growth forecasts g_t are taken from the Survey of Professional Forecasters (SPF) maintained by the Federal Reserve of Philadelphia. 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.

2.2 Training and testing samples

To run our predictive exercises, we split our data in two equally-sized training and testing samples. Each year, we randomly assign half of the newly entering firms (identified with a `gvkey`) and allocate them to the training sample. Firms remain in their assigned group throughout the sample. This sampling technique prevents cross-sample contamination.

We repeat this train-test assignment procedure 100 times to ensure that our results are not a byproduct of a specific draw. We show the evolution of the sample size in Figure 1 for one train-test split. In a typical year, we have over 900 firms in each subsample for the entire period. By design, both subsamples are balanced and have the same size.

2.3 Timing and look-ahead bias

In valuing firms, avoiding look-ahead bias is critical. Using information not yet available to market participants would artificially inflate our apparent predictive power and undermine the validity of our analysis.

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

[2] These values are in line with the average debt betas computed by Schwert and Strebulaev (2014), with a coarser rating resolution.

To avoid look-ahead bias, we estimate all models annually on rolling five-year windows, using only the data from the training sample available up to and including year T. This procedure ensures that each model relies solely on information actually known in year T. All models should thus be indexed by T and the train-test draw but we will omit this to lighten notations.

One of our predictive features is the analyst consensus forecast of EPS, which may contain useful information about the future that is not subsumed by other features. To ensure that financial analysts have incorporated newly released financial statements, we use IBES forecasts released three months after the end of fiscal year t, that is, three months into fiscal year $t+1$. At that point, the financial statements for year t are public.

2.4 Expected free cash-flows

To value firms, we apply the standard present-value formula for enterprise value (EV), defined as equity plus net debt per share. Computing EV requires two ingredients: a forecast of free cash-flows and the discount rates applied to value them. This subsection details how we construct the expected free cash-flows; the next section explains our procedure for estimating discount rates.

To compute the enterprise value, the relevant measure of cash-flows is free cash-flows to the firm (FCFF) defined, using Compustat Annual items, as:

$$\begin{aligned}\texttt{fcff}_{it} &= (1-\tau_{it})\,\texttt{ebit}_{it} + \texttt{dp}_{it} - \Delta\texttt{nwc}_{it} - \texttt{capx}_{it},\\ \Delta\texttt{nwc}_{it} &= \Delta\texttt{rect}_{it} + \Delta\texttt{invt}_{it} - \Delta\texttt{ap}_{it}.\end{aligned}$$

These classic formulas connect the economic concept of free cash-flows to accounting numbers: We begin with operating profit before interest (`ebit`), reduce it by corporate income tax $(1-\tau_{it})$, and add back the non-cash depreciation charge (`dp`). We then subtract the increase in net working capital (Δ`nwc`), which captures the extra cash locked into day-to-day operations when receivables (`rect`) or inventories (`invt`) rise faster than payables (`ap`). Finally, we deduct capital expenditures (`capx`). We use the same statutory tax rate for all firms, which we fix at 35% before 2018 and 21% starting 2018. In Table 1, we report variable definitions.

To estimate *expected* future free cash-flows per share, we rely on a statistical model, trained on 5-year rolling windows in the training sample. For each horizon $h \in \{1, 2, \ldots, 5\}$, we note this model as f_h where:

$$\frac{\texttt{fcff}_{it+h}}{\texttt{shrout}_{it}} = f_h(X_{it}) + \epsilon_{it,h},$$

where X_{it} is a set of characteristics. Dividing free cash-flows by the number of shares outstanding from CRSP (`shrout`) puts the free cash-flow to the firm (`fcff`) on a per-share basis, ready to be discounted to obtain a per-share enterprise value. We retrieve shares outstanding three months after the fiscal year end, like prices. This avoids look-ahead bias and ensures that share numbers are consistent with stock prices.

Our baseline characteristics: current free cash-flows to the firm, return on assets, tangibility ratio, asset growth, cash ratio, age, idiosyncratic volatility, log of total assets, IBES forecast of EPS at $t+1$ and $t+2$, and IBES actuals EPS, as well as fixed effects for 2-digit SIC sectoral codes. We interact all variables with an indicator for negative current free cash-flows. Table 1 summarizes the construction of each variable. Note that, since the functions $(f_h)_{h\in\{1,..,5\}}$ are estimated on 5-year rolling windows in the training sample, they should be indexed by t and the train-test draw. We omit this index to lighten notations.

We assume that the functions $(f_h)_{h\in\{1,..,5\}}$ are linear and estimate them using lasso. More precisely, for each horizon h, we regress the cash-flow per share on observable characteristics X_{is} available at date s. Because we are working on 5-year rolling windows, we require $s+h$ to be between $t-5$ and t. The parameter for the lasso penalty is chosen through three-fold cross-validation. Importantly, the estimation of this model uses only data from the training sample.

In Table 2, we report information on the quality of our forecasts in the *test* sample. We consider two sets of characteristics: all characteristics described in the previous paragraph, including EPS forecasts (first 4 columns), and all characteristics *except* EPS forecasts (last 4 columns)[3]. For each horizon h, we report the average MSE across 100 train/test draws. Models are estimated on the training samples and MSE are computed on the testing samples. We also provide the decomposition of MSE into variance and bias (MSE equals variance plus squared bias).

Three important facts emerge from Table 2. First, the MSE of our cash-flow forecasting model increases with horizon, as expected. Second, the MSE is almost entirely coming from the variance component. Lasso is of course not unbiased, and given our short time period, even an unbiased estimator could appear ex post biased. However, this bias has negligible impact on performance. Third, including IBES EPS forecasts as a feature does not reduce the MSE very much (though it helps slightly more at longer horizons). This is in contrast with the findings in de Silva and Thesmar (2024): In their paper, EPS forecasts did improve forecasting of future EPS (even at longer horizons). In this paper, the difference is that we are predicting a different object: Free cash-flows to the firm.

[3]This guarantees that our forecasts do not contain price information, which could happen if analysts adjust their forecasts to fit prices.

2.5 True and estimated value of the firm

Our goal is to estimate the enterprise value per share of a firm, and compare this estimate with the true value of the firm. In line with standard valuation practices, we define the true value as being the enterprise value per share, which we compute using our data through the formula, for firm i at date t:

$$\text{EV}_{it} = \underbrace{\texttt{prc}_{it}}_{\text{stock price}} + \underbrace{\frac{\texttt{dltt}_{it} + \texttt{dlc}_{it} - \texttt{che}_{it}}{\texttt{shrout}_{it}}}_{\text{net debt per share}},$$

where enterprise value is defined, as is standard, net of cash holdings.

To estimate the enterprise value per share, we follow standard practice of computing the present value of future cash-flows. We explicitly forecast future cash-flows up to five years, and then assume perpetual growth. The choice of the 5-year cutoff is somewhat arbitrary but in line with analyst reports analyzed by Décaire and Graham (2025). Earlier versions of this paper had a three year cutoff and the results were similar. We conjecture that using a forecasting horizon up to 10 years would not change our results dramatically. Let r be the discount rate (we will explore many alternative ways of computing it), and g_t be the perpetual growth after year 5 (defined as the long-term nominal GDP growth rate from the Survey of Professional Forecasters, as described in Section 2.1).

Then, for a given cost of capital r, the present value per share of firm i at date t is given by:

$$\text{PV}(r; X_{it}, g_t) = \frac{f_1(X_{it})}{1+r} + \frac{f_2(X_{it})}{(1+r)^2} + \cdots + \frac{f_5(X_{it})}{(1+r)^5} + \underbrace{\frac{f_5(X_{it})}{(1+r)^5} \cdot \frac{1+g_t}{r-g_t}}_{\text{Terminal Value}}, \tag{2}$$

Note that we implicitly assume a flat term structure of discount rates, as is common in corporate valuation.

We now discuss how we estimate the cost of capital r. We use two main approaches, factor models and internal rate of returns of comparable stocks.

2.6 Cost of capital: Factor models

For factor models, we compute the cost of capital using the Weighted Average Cost of Capital (WACC) formula:

$$r_{it}^{\text{WACC}} = r_{it}^{E} \times e_{it} + r_{it}^{D} \times (1 - e_{it}) \times (1 - \tau_{it}), \tag{3}$$

where e_{it} is the (market) equity ratio defined as:

$$e_{it} = \frac{\texttt{prc}_{it} \times \texttt{shrout}_{it}}{\texttt{prc}_{it} \times \texttt{shrout}_{it} + \texttt{dltt}_{it} + \texttt{dlc}_{it} - \texttt{che}_{it}},$$

and r_{it}^{E} and r_{it}^{D} are the cost of equity and debt. We winsorize the equity ratio at ±3 to limit the influence of outliers. To compute the cost of debt, we always use the CAPM and write: $r_{it}^{D} = R_t^f + \beta_{it}^{D} \times 5\%$ where, as described in Section 2.1, R_t^f is the safe rate of return and β_{it}^{D} is the CAPM beta of debt computed using ratings.

To compute the cost of equity, we use a variety of methods based on factor models:

Flat 10% This is in order to benchmark the different factor models. We use $r_{it} = 10\%$ for all firms, every period.

Straight CAPM. Given the emphasis placed on CAPM in introductory finance classes, it is natural to use it as a benchmark. In this implementation of the CAPM, we compute the cost of equity as:

$$r_{it}^{E,\text{CAPM}} = R_t^f + \beta_{it}^{E} \times 5\%,$$

where we set the ERP at 5%, close to the recent consensus of practitioners collected by Fernandez et al. (2024).

LW CAPM. Since β_{it}^{E} is measured with noise, there is a concern that CAPM-based expected equity returns may be too imprecise. To deal with this, we shrink the empirical beta in the above formula towards size-dependent target values, using the procedure recommended by Levi and Welch (2017). We consider two shrinkage intensities. Under "low shrinkage," we assign a weight of 0.65 to the empirical CAPM beta and 0.35 to the target beta, while "high shrinkage" applies equal weighting to both betas. The target beta values are assigned based on market capitalization: stocks in the lowest tercile receive a target beta of 0.5, those in the middle tercile receive 0.7, and those in the highest tercile receive 0.9.

Flexible ERP. The previous two implementations of the CAPM impose a constant equity risk-premium of 5%. To obtain an upper bound for the CAPM's explanatory power, we let the data estimate an equity risk-premium that gives the CAPM the best possible fit. To do this, we compute the cost of equity for each firm as:

$$r_{it}^{\text{WACC,flex ERP}} = R_t^f + \left(e_{it}\beta_{it}^{E} + (1 - e_{it})(1 - \tau_{it})\beta_{it}^{D}\right) \times \text{ERP}_t^{\text{flex}}$$

where, every year, we estimate $\text{ERP}_t^{\text{flex}}$ to minimize, in the training sample, the average squared difference between $r_{it}^{\text{WACC,flex ERP}}$ and the implied internal rate of return. Of course, we estimate $\text{ERP}_t^{\text{flex}}$ using only information from the training sample and using a 5-year rolling window, so as to avoid any form of overfitting. We show the evolution of this fitted ERP in Appendix Figure A.2. Unsurprisingly, it comoves with valuations.

Fama–French–Carhart. Finally, we try a richer multifactor model. Richer models are better candidates for measuring the cost of capital, since they have been shown to explain the cross-section of expected returns better than the CAPM. We settle here for the 4-factor model of Fama and French (1993) and Carhart (1997) which includes four risk factors: market (MKTRF_t), size (SMB_t), value (HML_t), and momentum (UMD_t), which we pull directly from WRDS, as explained in Section 2.1. For firm i at date t, we thus calculate the cost of equity as:

$$r_{it}^{E,\text{FF}} = R_t^f + \beta_{it}^{\text{MKTRF}} \times \text{MKTRF}_{it} + \beta_{it}^{\text{SMB}} \times \text{SMB}_t + \beta_{it}^{\text{HML}} \times \text{HML}_t + \beta_{it}^{\text{UMD}} \times \text{UMD}_t.$$

For each one of these alternative factor-based approaches, we then plug the cost of equity into equation (3).

2.7 Cost of capital: Predicted IRR

This paper proposes an alternative way of computing the cost of capital in the test sample: the internal rate of return (IRR) of comparable firms in the training sample. To calculate it on the testing sample, we proceed in three steps.

First, in the training sample, we define the IRR as the positive solution of:

$$\text{PV}_{it}(\text{IRR}_{it}) = \text{EV}_{it}.$$

We solve this equation only if the terminal cash-flow, $f_5(X_{it})$ is strictly positive, and we restrict the resulting IRR to lie between 0 and 30%. As shown in Figure 2 the cross-section and time-series of these IRRs are well-behaved. The top two panels concern the IRR obtained using the cash-flow model $f_h(X_{it})$ estimated *without* IBES EPS forecasts. In the bottom two panels, we add the IBES EPS forecasts to the list of features X_{it}. As expected, including IBES forecasts does not make much of a difference. Most observations have an IRR between 5 and 15%, with a mean slightly higher than 10%. In the past 30 years, IRRs have been steadily decreasing, in line with a secular downward trend in interest rates. Between 1992 and 2020, the IRR decreased by about 5ppt, a pace similar to 10 year treasury yields (Graham (2022) finds a similar trend for self-reported costs of capital used by firms). Since 2020, the mean IRR has rebounded by about 2ppt, consistent with the recent rise of interest rates.

Second, still in the training sample, we regress the IRR on a linear combination of observables (which may or may not overlap with X_{it} that are used to predict free cash-flows). We estimate via OLS:

$$\text{IRR}_{it} = g(W_{it}) + u_{it}, \tag{4}$$

where the functional form g and the set of variables W_{it} varies across applications. The above equation is estimated on 5-year rolling windows in the training sample, which includes all firm-year observations between $t-5$ and t. Technically, the function g should have a subscript t that we omit in our notation.

In the spirit of the recent literature which applies statistical learning to finance, we try various parameterizations for g. For instance, Gu et al. (2020) explore the ability of various methods to predict future *realized* returns using a wealth of classic characteristics. We instead focus here on IRRs as measures of *expected* returns, but also try various standard machine learning methods: (1) plain non-penalized OLS (which may overfit when many characteristics are included), (2) penalized linear models (lasso and ridge), (3) random forest, (4) extreme gradient boosting (XGBoost), and (5) a simple feed-forward neural network. The first two methods are linear, but the other three are not. Implementation details, including hyperparameter tuning, are provided in Appendix B.

We explore several sets of variable W_{it}, going from the simplest (only one constant) to the richest (essentially, all variables used to forecast free cash-flows). The idea behind this approach is that such "comparable IRR" approach requires data, which may be abundant when evaluating public companies (in which case accounting data, as well as various analyst forecasts, are available), but much sparser when evaluating new firms or internal projects (in which case only size and industry may be available). Thus, we explore the following six sets of variables (see Table 1 for a summary):

1. **Rolling IRR.** As a benchmark, we first reduce W_{it} to a constant. So the predicted value is given by the average IRR taken over the 5-year rolling window in the training sample.

2. **Private project.** This small set of variables is for non-listed projects and private firms that do not yet have existing accounting variables. Characteristics W_{it} are restricted to firm age and 2-digit SIC industry and year dummies.

3. **Private project + leverage.** We then separately add book leverage, computed as the ratio of long-term plus short-term debt to book value. This could also be part of the set of variable used for start-ups or capital budgeting. We add leverage separately in order to evaluate its explanatory power on valuation. In the classic WACC valuation approach, for given risk, leverage reduces taxes, the cost of capital and boosts valuation (see for instance Berk and DeMarzo, 2006). We use book

leverage instead of market leverage in order to exclude firm market value from our predictive features.

4. **Private project + FCFF forecasts.** Here, to the previous set of variables (age, industry), we add the five free cash-flow forecasts $(f_h(X_{it}))_{h\in\{1,..,5\}}$ and the latest available EPS at t. The idea is that these forecasts are anyways produced in the valuation exercise, and can therefore be used to build comparable IRRs. While this is somewhat atypical to use cash-flow expectations to compute the cost of capital, these may contain relevant information about risk.

5. **Private firm for research.** We add to "Private project variables" information that could be easily available for economic researchers working on Census-like data. We thus include: 2-digit SIC industry dummies, age, year, but also physical capital, sales, and book value.

6. **Public firm.** Our last set of variables is the most comprehensive and assumes the person doing the valuation has access to information from Compustat and IBES, but naturally not prices. We include all variables used in "Private project + FCFF forecasts." To those, we add: book leverage, return on assets, tangibility ratio, cash holdings, total assets, annual growth of total assets between $t-1$ and t, and idiosyncratic volatility as defined in Section 2.1.

Note that we only keep observations at the intersection of all six information sets. In particular, we drop firms for which we do not have two years of IBES EPS forecasts.

Third, once we have estimated $\widehat{g}$ on rolling windows and for each one of the 6 sets of variables on the training sample, we construct the predicted IRR $\hat{g}(W_{it})$ on the testing sample. These values are the predicted IRRs we will use in valuation. The rest of the paper is devoted to evaluating the quality of factor-based and predictive IRRs.

3 Main results

We are now set to estimate the enterprise value per share of firms in our test sample, using the present value formula (2) and the various measures of the cost of capital described in sections (2.6-2.7).

3.1 Computing mean-squared errors

Our ultimate goal is to evaluate which cost of capital does best at predicting the observed enterprise value. After computing, for each firm in the testing sample, $\text{PV}(r; X_{it}, g_t)$ for given discount rate r, and then compare it to the observed value per share EV_{it}. For each method, we use the Mean Squared Error as our measure of fit. So for each cost of capital method $R = (r_{it})_{(i,t)}$, we define MSE as:

$$\text{MSE}_{\text{EV}}(R) = \frac{1}{N_{\text{Test}}} \sum_{(i,t)\in\mathbb{T}} \left(\log\left(\text{PV}_{it}(r_{it})\right) - \log\left(\text{EV}_{it}\right)\right)^2$$

where $\mathbb{T}$ is our notation for the test sample, and N_{Test} is the number of firm-year observations in $\mathbb{T}$. We use the same cash-flows in all scenarios, so any differences among models arise solely from the choice of discount rate. Importantly, both the cash-flow and discount-rate models are estimated in the training sample. There is one small abuse of notation in the above formula. Remember that the results reported in what follows are taken *on average* across 100 train/test splits, i.e., across 100 different randomly drawn test samples.

3.2 Factor models

We start with cost of capital coming from factor models as described in Section 2.6. We show the mean squared-error in Figure 3. Three salient features emerge. The left panel uses cash-flows forecasts predicted without IBES EPS forecasts, while the right panel uses these forecasts. Conclusions are similar in both panels, consistent with evidence from Table 2. First, the flat 10% discount rate performs best: The out-of-sample MSE of the best factor model is over 50% higher. Second, the CAPM performs significantly worse than the flat discount rate, with an MSE approximately 75% larger. Adjusting the CAPM by shrinking the beta towards 1 reduces the variance component somewhat (CAPM-induced valuation are less dependent on raw beta) but increases the bias so that the resulting MSE is, in the end, larger. Adjusting the ERP every year to equate CAPM discount rates with true IRRs reduces the bias to zero, but the MSE remains significantly larger than the flat discount approach. This is because it does not help predict the cross-section of value quite as much as it is used by the CAPM formula. Finally, adding more factors does not help. If anything, it makes corporate valuation sensitive to factor loadings, which do not seem to be predictive in the cross-section.

3.3 Predicted IRR

We now evaluate how discounting expected cash-flows with *predicted IRRs* fits observed enterprise values. In some instances, the predicted values take extreme values that are implausible but have a disproportionate impact on the MSE. To handle these extreme cases, we remove these extreme values using the following rules. First, if the IRR is missing or lower than g_t, we replace the observation with the average IRR in the training sample for that year. Second, we winsorize EV per share predictions at \$5,000. We drop negative EV per share predictions.

The results are summarized in Table 3, which reports the predictive results for the different predictive techniques and the different sets of variables described in Section 2.7. We focus here on the version where IBES EPS forecasts are used to predict future cash-flows. Several results emerge. First, the MSE obtained by using the rolling average IRR in the training sample is 0.70, approximately 20% lower than the MSE obtained using a flat 10% discount rate across the board, the best of the "factor models" (see right panel, Figure 3). This is because rolling IRRs capture time-varying valuations of future cash-flows more closely. Second, adding industry dummies further improves the MSE down to 0.52 – a reduction of about 15%. Thus, a simple version of this methodology, which amounts to computing rolling IRRs for each industry, is already quite precise. Third, in all methodologies, controlling for leverage alone does not improve the fit. This is in contrast with the typical WACC methodology, where leverage reduces the cost of capital through the effect of the tax shield. In our data, we do not find any cross-section evidence of this.

The fourth fact is that adding more variables improves predictions when using non-linear methods. As seen in the first 3 rows of the table, the fit of penalized and non-penalized regressions improves by about 20% when moving from the first set of variables (industries only, "private project") to the last one (a slew of Compustat-driven variables "Public firms"). Looking at the last three rows, the non-linear methods experience a decrease in MSE of 40 to 50% after including all the possible variables. This is consistent with the general notion that non-linear methods do better with a larger number of variables. In the end, the best outcome is obtained from the three non-linear methods and the full set of variables ("public firm"), and the MSE is 0.3, over three times smaller than taking a flat 10%, and five times smaller than CAPM-based valuation.

4 Predicting discount rates

In this section, we analyze the MSE of IRRs – instead of the MSE of values that we focused on so far. It writes as:

$$\mathrm{MSE}_{\mathrm{IRR}}(R) = \frac{1}{N_{\mathrm{Test}}} \sum_{(i,t)\in\mathbb{T}} (r_{it} - \mathrm{IRR}_{it})^2.$$

It measures the extent to which the IRR model $R = (r_{it})_{i,t}$ fits the data-generated IRR on the test sample. It is related to the MSE of log prices, since a model R that fits observed IRR well, will also ensure that the R-implied valuation fits observed values well. By definition of duration, given $\mathrm{MSE}_{\mathrm{IRR}}$ leads to a bigger valuation error $\mathrm{MSE}_{\mathrm{EV}}(R)$ for longer duration firms (firms whose expected cash-flows grow faster).

An advantage of looking at MSE_{IRR} is that it is quantitatively interpretable. As we show in this Section, under a standard model of firm behavior, the MSE of IRRs is proportional to the present value loss of not using the exact discount rate of cash-flows. Thus, the difference between MSEs of IRRs of two different models allows us to measure the % value gain for a hypothetical firm from moving to the less precise to the more precise prediction model.

4.1 Interpreting the MSE of IRRs

We describe now the model under which such MSE of IRRs can be interpreted. Assume the firm is infinitely-lived with productivity z_t and true discount rate r. The firm only employs one input, capital k_t, chosen in period t to be productive in period $t+1$. The firms' EBITDA at $t+1$ is $F(k_t)$, and δ is the rate of depreciation.

When choosing capital, the firm is *not using* the right discount rate but instead $r^* \neq r$. Its present value is thus given by:

$$V(z_0; r; r^*) = \mathbb{E}\left(\sum_{t\geq 0} \frac{1}{(1+r)^t}\left(-k_t + \frac{1}{1+r}\left((1-\delta)k_t + z_{t+1}F(k_t)\right)\right)\middle| z_0\right)$$

The following proposition computes the value loss from using r^* instead of r:

Proposition 1. *The value loss of using r^* instead of r is given by the approximate relation:*

$$\frac{V(z_0; r; r) - V(z_0; r; r^*)}{V(z_0; r; r)} \approx \left(\frac{1+r}{2} \cdot \frac{\epsilon_F}{1-\epsilon_F} \cdot \frac{1}{-\epsilon_{F'}}\right) \cdot \left(\frac{r^* - r}{r+\delta}\right)^2$$

where ϵ_F is the elasticity of F and $\epsilon_{F'}$ the elasticity of F'.

Proof. See Appendix C. □

The above formula simply states that, if the value loss is bigger when (1) the discount rate wedge $(r - r^*)$ is larger and (2) the production function F is closer to linear. This is because more linear production functions lead to a bigger response of investment to discount rates, and therefore more mistakes. Note also, because of the envelope theorem, the value loss is a quadratic function of the discount rate wedge: The value impact is second order because firms are close to their unconstrained optimal investment policy.

To apply this formula, assume $F(k) \propto k^\theta$. Further, assume that there is a cross-section of firms which differ in the discount rate wedges $(r - r^*)$ for all sorts of reasons. In this case, the average value loss across all these firms writes as:

$$\text{Mean value loss} \approx \left(\frac{1+r}{2(r+\delta)^2} \cdot \frac{\theta}{(1-\theta)^2} \right) \cdot \text{MSE}_{\text{IRR}}(R) \tag{5}$$

where $\text{MSE}_{\text{IRR}}(R)$ is the mean squared difference between actual discount rates r and used discount rates r^*.

This is the formula we will apply in our MSE interpretation. The true (i.e. market-based) discount rates are represented by the true IRR, while r^* comes from a given model of IRRs. $\text{MSE}_{\text{IRR}}(R)$ is the MSE of this model. Then, assuming the discount rate of all projects in a firm is well represented by the IRR, the formula gives us the value lost, on average, coming form firms not using their IRR, but the model instead. Of course, this interpretation of the model MSE relies on the assumption that all firms have the same production function F. It is easy, but beyond the scope of the present paper, to extend this framework to account for industry heterogeneity.

We also explore in Appendix D whether the conclusions drawn from this analysis are sensitive to the simple model we used. We use a model with additional frictions in the tradition of the modern structural corporate finance literature (e.g. Hennessy and Whited (2005) and many others). The frictions we add are: costly equity issuance, collateral constraint and capital adjustment costs. We then solve the model numerically to obtain the sensitivity of the value loss from using the wrong cost of capital on the size of the discount rate wedge $(r - r^*)$. We find that, for reasonable errors, adjustment costs have little impact on the value loss. Financing constraints have more effect – they dampen the value loss of cost of capital mistakes – but the estimated loss has a similar magnitude.

4.2 MSEs of IRR models: the Data

Table 4 shows the IRR MSEs for different predictive methods. It is the equivalent of Table 3, but for IRRs. To help with readability, these tables use IRRs and their predicted values in *basis points* – so that the true values should be divided by 10,000. Similar patterns emerge: using industry comparables significantly reduces the prediction errors for all methods. Adding further variables also improves, but especially so for non-linear methods. Figure 4 compares the performance of IRR models with factor models.

In Appendix Figures A.3 and A.4, we also show that, while the MSE of IRRs tends to increase since the 1990s, for all models g and all features sets W_{it}, this is because the cross-sectional variance of IRRs is also trending upwards. As a fraction of cross-sectional variance, the share that is predictable is stable over time, as high as 60% for non-linear models and the largest set of features, and as low as some (time-varying) 20% for "private project + leverage" (i.e. industry dummies and book leverage). As in the Table, we see in Figure A.4 that two groups of models arise: linear (with or without penalization) and non-linear, and that non-linear models outperform more strongly in feature-rich environments.

4.3 Using the MSEs of IRRs to estimate the value gains from using the right discount rates

Without additional modeling, the numbers in Table 4 are not easy to interpret. But in our simple economic framework we can do this. To apply formula 5, we set $\delta = 5\%$, $r = 10\%$ and $\theta = 0.54$. Our calibration of θ is consistent with mark-ups of 25%, constant returns in production, and a Cobb-Douglas capital share of 0.3 (see, for instance, Catherine et al. (2022)).[4] With these numbers, the loss of value from not using the actual discount rate is 62 times the IRR MSE.

Under this calibration, the estimates in Figure 4 suggest that taking a simple past average IRR achieves considerable value gains over the CAPM rule. Given the formula, using a rolling average of the IRR instead of the actual IRR yields a value loss of 7.8% (62×12.5bps). The CAPM results in substantially larger value losses around 19% (62×30bps). Overall, going from the CAPM (a value loss of 19%) to the rolling average of past IRRs (a value loss of 7.8%) yields a value gain of approximately 11%, a very large difference.

Increasing model complexity achieves substantial additional gains. Already, over the rolling average IRR, *industry-specific* rolling IRRs reduces the value loss from 7.8 to 6.1% (62×9.9bps), unlocking 1.7% in market value. Now, taking the average of the three non-linear models, and the largest possible number of features, the value loss further drops to 3.4% (62×5.5bps).[5] So using non-linear models of IRR would help create an additional 2.7% of value over industry-specific rolling IRRs.

5 Robustness

In this last section, we explore two robustness checks.

[4]These assumptions correspond to a firm revenue given by $R(k,l) = (k^\alpha l^{1-\alpha})^{1-1/\varphi}$, where $\alpha = .3$. Given constant price elasticity of demand φ, the mark-up is $\frac{1}{\varphi-1}$, so that $\varphi = 1 + 1/.25 = 5$. Assuming l is chosen without friction, then, the firm's EBITDA is given by:

$$F(k) = \max_l \left[R(k,l) - wl\right] \propto k^{\frac{(1-1/\varphi)\alpha}{1-(1-1/\varphi)(1-\alpha)}}.$$

In this model, the elasticity of EBITDA to k is thus a function of $\theta = \frac{1-1/\varphi\alpha}{1-(1-1/\varphi)(1-\alpha)}$. Plug in $\alpha = .3$ and $\phi = 5$ and find $\theta = .54$

[5]Note that this value differs slightly from the one obtained by averaging the values in Table 4, Column (6). This is because the MSEs in Figure 4 are computed for firms with valid factor discount rates. This subsample is different from the one in Table 4.

5.1 Price information in IBES

Since our results use IBES earnings forecasts to predict future free cash-flows, a concern is that current valuations are already included by analyst forecasts in our present value calculation. In other words, we are indirectly using the value to be predicted as one feature. A first defense of our approach is that this should bias all methods, whether using factor models or any model of IRR prediction. It is not obvious how this would affect the very clear ranking of methods that we find. A second defense is that our method to forecast cash-flows is evaluated on its ability to actually forecast cash-flows, not prices. So we are trying to zoom in on the part of IBES forecasts which is useful to predict cash-flows.

We can, however, propose a stronger test: We just replicate our methodology *without* using IBES forecasts to predict future cash-flows or IRRs. When evaluating factor models, we saw in Figure 3 that the use of IBES barely affected the predictive power of the different models (the right and left panels are similar). When evaluating predictive models of the IRR, we show in Table 5 that not using IBES forecasts does not change the quantitative and qualitative conclusions of our analysis on log prices. Appendix Table A.1 reproduces the results on IRR MSEs without using IBES forecasts, and also finds no significant difference with Table 4.

5.2 Long-term versus short-term value

Another concern with our analysis is that it uses information that is "too fresh" to guide capital budgeting decisions. The intuition of this concern is that IRRs may vary a lot across years – as valuations do – so that investment would be too dependent on short-term fluctuations of the market. A justification of our baseline approach is, however, that IRRs tend to move at low frequencies, as shown in Figure 2. Thus, we expect IRRs predicted using older information to also perform quite well in explaining the cross-section of values.

Figure 5 clarifies this defense. In this Figure, we report MSE of IRRs for different methods and different lags. For each firm (i, t), we use the model we estimated in our baseline approach, but lag the predicted value by h year, i.e. take $g(W_{it-h})$ instead of $g(W_{it})$ as our measure of predicted IRR. We then compute the MSE of this lagged prediction:

$$\text{MSE}_{\text{IRR}}^{h} = \frac{1}{N_{\text{Test}}} \sum_{(i,t)\in\mathbb{T}} \left(g(W_{it-h}) - \text{IRR}_{it}\right)^2$$

over the test sample. To maximize readability, we focus on the model that is most likely to overfit the training sample, so choose g as the linear fit with OLS.

The key takeaway of Figure 5 is that, even with a five-year lag, the MSE does not increase much. With the full set of Compustat variables for W_{it} ("Public firm"), the MSE of IRR increases from 8 to 16, so it has the same order of magnitude. This trend is, however, less pronounced with fewer variables, and totally absent in the "Rolling IRR" case where we set W_{it} to a constant. The modest increase of MSE we observe are thus presumably because OLS "overfits" the training sample.

6 Conclusion

This paper revisits perhaps the most fundamental step in valuation—the choice of discount rate—and delivers three empirical results. First, we find that textbook factor-based approaches, from the simple CAPM to richer Fama-French-Carhart specifications, poorly explain the cross-section of market values. Indeed, discounting all cash-flows at a flat 10 percent consistently outperforms any version of the CAPM in terms of mean-squared-log-error. Second, we show that an "imputed IRR" derived from comparable firms substantially reduces pricing errors, even when estimated using simple linear methods and basic industry identifiers. When non-linear learners utilize a full set of balance-sheet and analyst variables, pricing errors fall to approximately one-fifth of those achieved by the CAPM benchmark. Finally, employing a standard investment model, we demonstrate that using CAPM-based hurdle rates can destroy firm value by about 7–10 percent compared to a data-driven IRR rule. Our findings have significant practical and pedagogical implications, as the method we propose for computing IRRs from comparable firms can readily be applied in valuation practice. Future research may build upon the machine-learning framework presented here to explore variations in discount rates across geographies, private capital markets, and project-level data.

7 Bibliography

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.

Bastianello, Francesca, Paul Décaire, and Marius Guenzel, 2025, Mental models and financial forecasts.

Ben-David, Itzhak, and Alex Chinco, 2024, Expected eps× trailing p/e and asset pricing .

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

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

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

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* .

Cho, Thummim, and Christopher Polk, 2020, Asset pricing with price levels, *Journal of Finance* .

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

de Silva, Tim, and David Thesmar, 2024, Noise in expectations: Evidence from analyst forecasts, *Review of Financial Studies* .

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

Décaire, Paul, and John Graham, 2025, Valuation fundamentals.

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

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.

Fernandez, Pablo, Diego Garcia de la Garza, and Lucià Fernàndez Acìn, 2024, Survey: Market risk premium and risk-free rate used for 96 countries in 2024.

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

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

Gormsen, Niels Joachim, and Kilian Huber, 2025, Corporate discount rates, *American Economic Review* 115, 2001–2049.

Graham, John R, 2022, Presidential address: Corporate finance and reality, *The Journal of Finance* 77, 1975–2049.

Graham, John R, and Campbell R Harvey, 2001, The theory and practice of corporate finance: Evidence from the field, *Journal of financial economics* 60, 187–243.

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

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.

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

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* .

Schwert, Michael, and Ilya A. Strebulaev, 2014, Capital structure and systematic risk, Technical report, Stanford University Graduate School of Business; AQR Arbitrage, LLC; National Bureau of Economic Research, Rock Center for Corporate Governance at Stanford University Working Paper No. 178. 34 pages. Posted: 8 Apr 2014. Last revised: 14 Apr 2014.

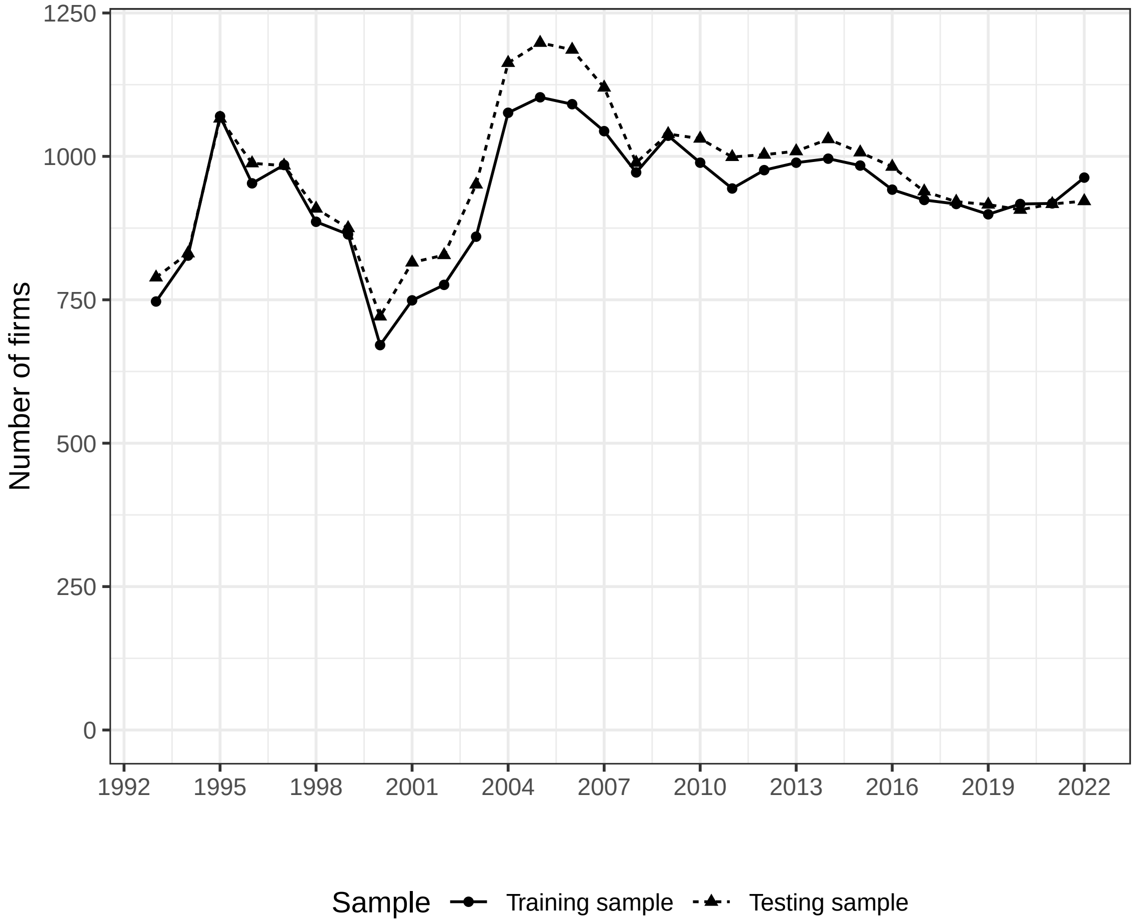
Tauchen, George, 1986, Finite state markov-chain approximations to univariate and vector autoregressions, *Economics Letters* 20, 177–181.

Welch, Ivo, 2021, The cost of capital: If not the capm, then what?, *Management and Business Review* 1.

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.

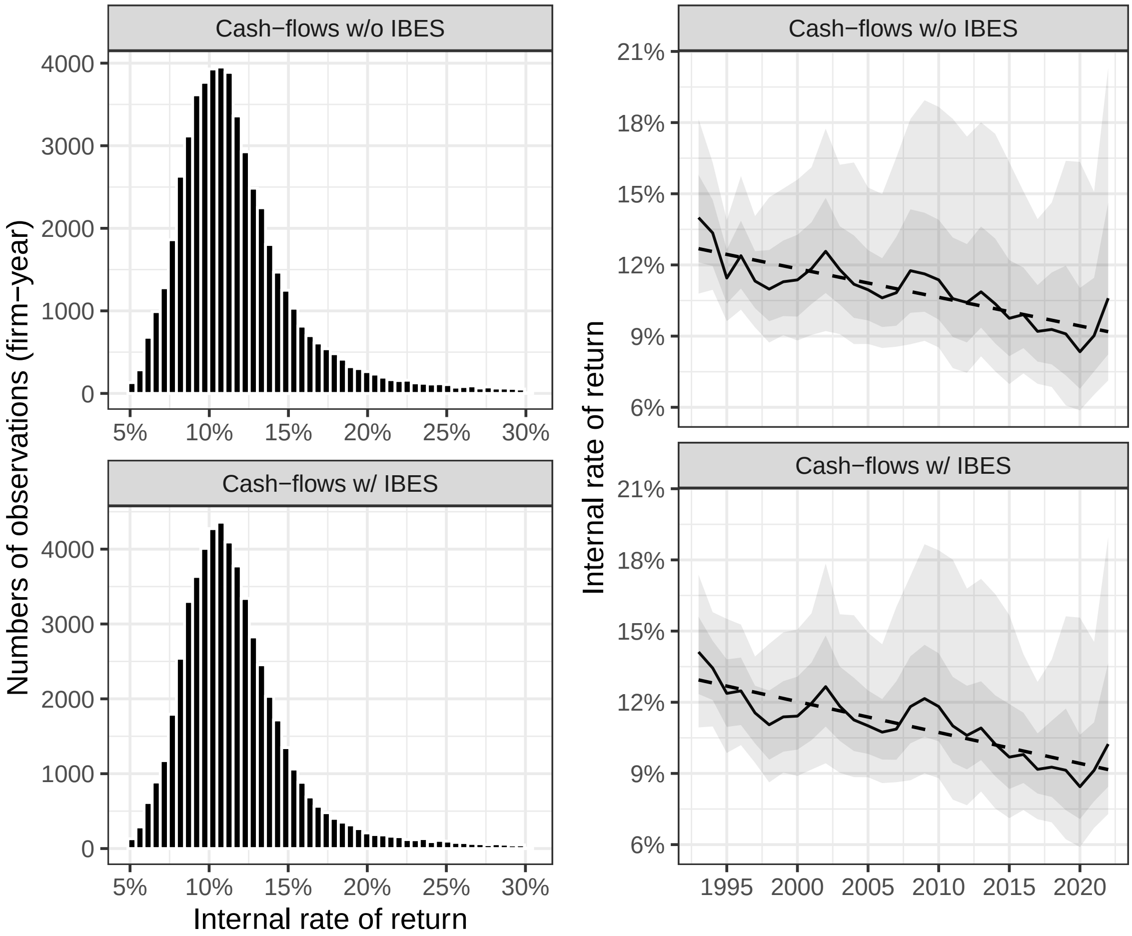
Figures

Figure 1: Sample size of the training and testing sample



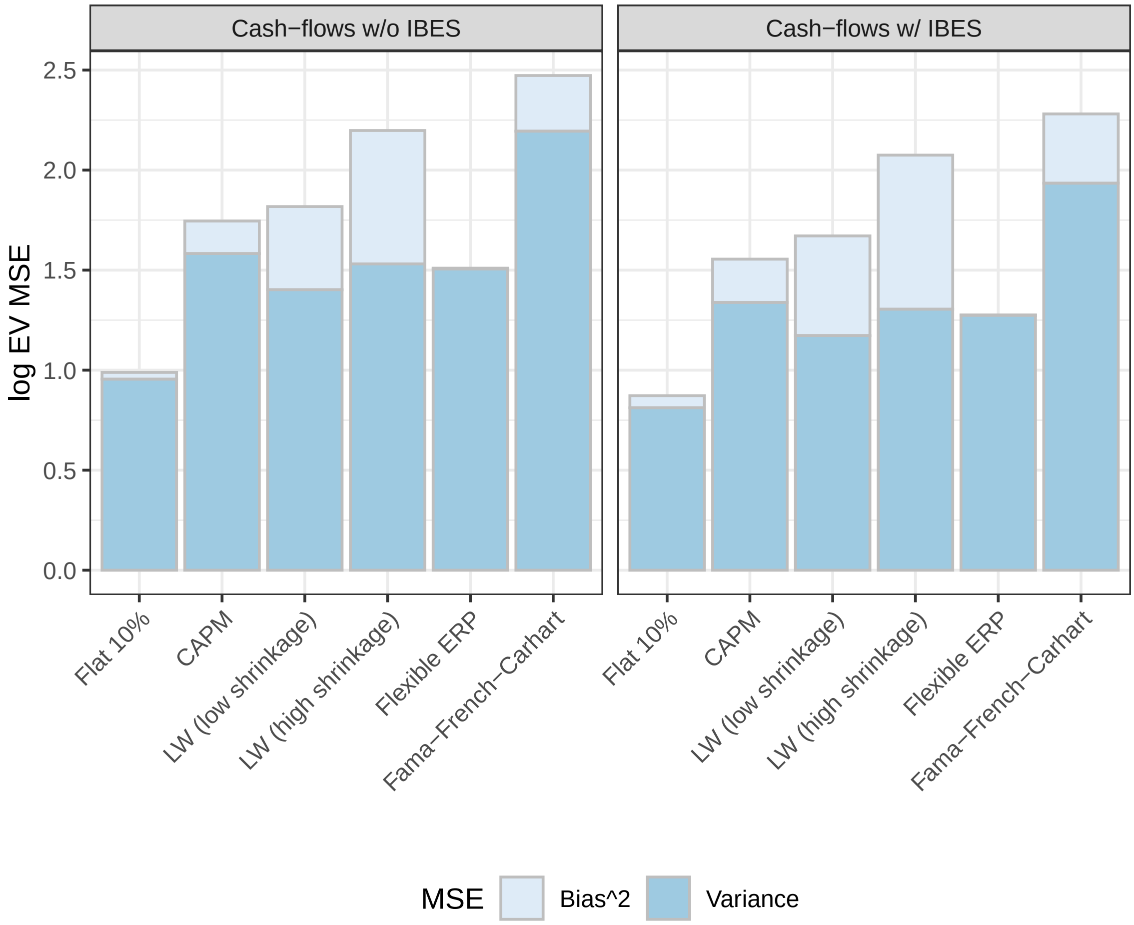
Note. This figure shows the number of firms in our sample in every year for a single train-test split. The construction of the two subsamples (training and testing) is explained in the text. Go back to main text.

Figure 2: Cross-section and time-series of IRRs



Note. This figure displays the distribution (left panels) and time-series evolution (right panels) of the internal rate of return (IRR) for a single train-test split. The top panels show IRRs calculated using cash flow forecasts that exclude IBES data, while the bottom panels use forecasts that include IBES data. In the time-series plots, the solid line shows the median IRR, while the dark and light gray shaded areas indicate the interquartile and interdecile ranges. The dashed line shows the regression line. IRRs are calculated as the discount rate that equates the present value of forecasted free cash flows to the observed enterprise value. Go back to main text.

Figure 3: MSE of values using variants of the CAPM as discount rates

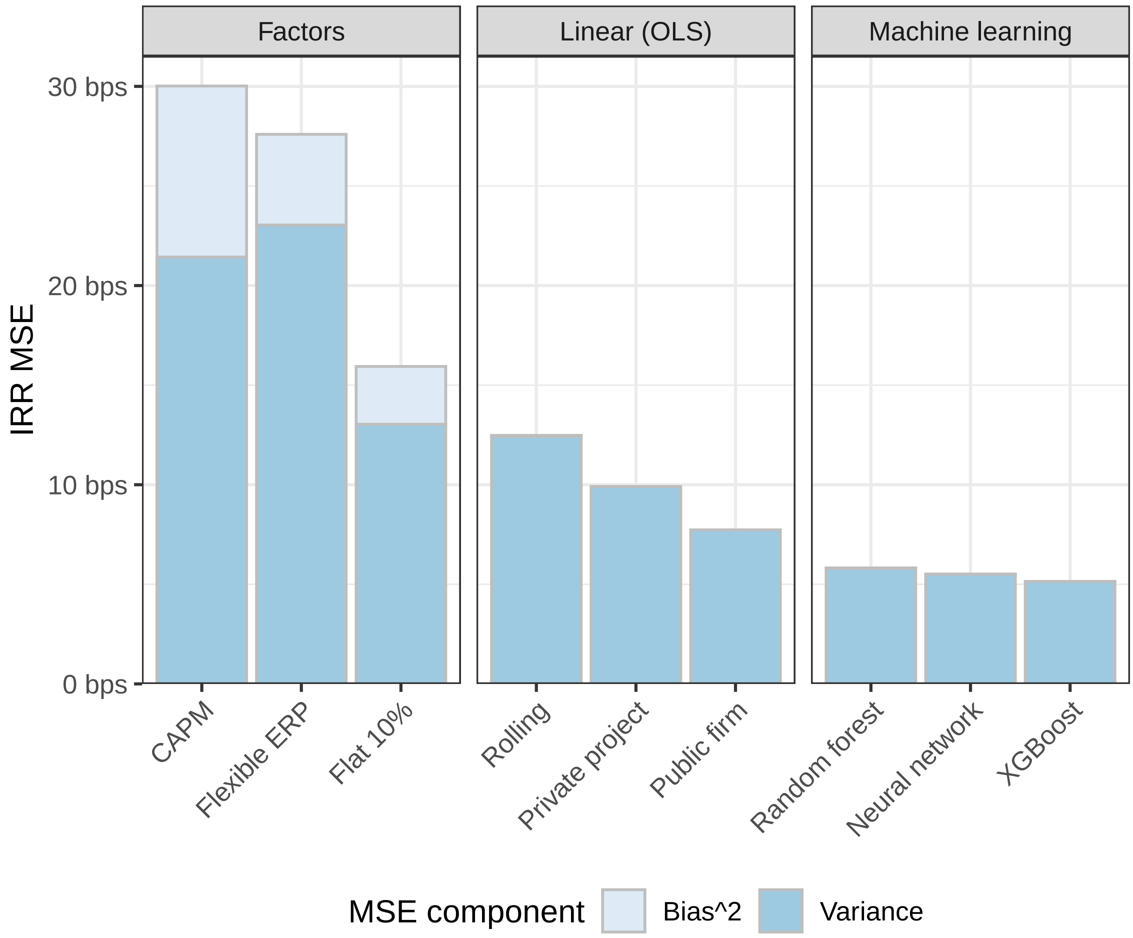


Note. This figure displays the out-of-sample performance of factor models in predicting enterprise values. Our metric of performance for model R is the mean squared log error

$$\text{MSE}_{\text{EV}}(R) = \frac{1}{N_{\text{Test}}} \sum_{(i,t)\in\mathbb{T}} \left(\log \text{PV}_{it}(r_{it}) - \log \text{EV}_{it} \right)^2,$$

where EV_{it} is the enterprise value and PV_{it} is the present value of future cash flows discounted at the rate r_{it} predicted by model R. The bars decompose the total MSE into its squared bias (light blue) and variance (dark blue) components. Future cash flows are forecasted using a lasso model estimated in the training sample with IBES as predictors (right panel) and without (left panel). The factor models included are the CAPM with and without Levi–Welch shrinkage, the Fama–French–Carhart four factors model, and a flexible ERP CAPM. We include a flat 10% discount rate as a benchmark. Results are averaged over 100 train-test splits." Go back to main text

Figure 4: MSE of IRRs using variants of the CAPM and statistical models as discount rates



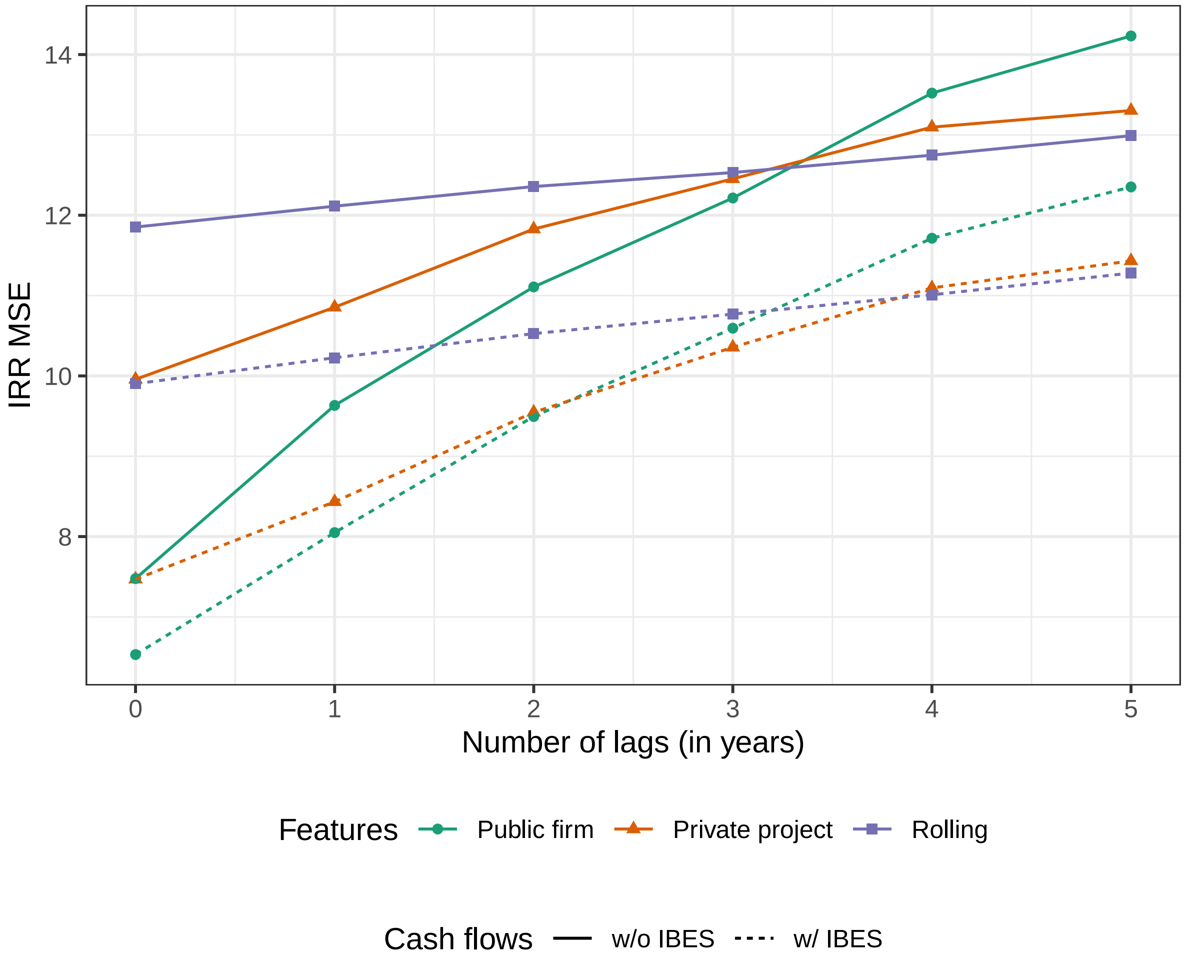
Note. This figure compares the out-of-sample performance of factor models and statistical models in predicting discount rates. Our metric of performance for model R is the mean squared error

$$\text{MSE}_{\text{IRR}}(R) = \frac{1}{N_{\text{Test}}} \sum_{(i,t)\in\mathbb{T}} (r_{it} - \text{IRR}_{it})^2 \,,$$

where, r_{it} denotes the discount rate predicted by model R. We report values multiplied by 10,000 for readability (i.e., expressed in basis points). The IRR is defined as the discount rate that equates the enterprise value and the present value of future cash flows. Future cash flows are forecasted using a lasso model estimated in the training sample.

The left panel shows factor models. The middle panel shows linear (OLS) models using progressively richer sets of firm characteristics (from 'Rolling' to 'Public firm'). The right panel shows various machine learning models using the most comprehensive 'Public firm' feature set. All models are trained on the training sample and evaluated on the test sample, with results averaged over 100 splits. Go back to main text

FIGURE 5: Using lagged IRR for prediction



Note. This figure displays the out-of-sample performance of lagged models of IRR as predictors of discount rates. Our metric of performance is the mean-squared error

$$\text{MSE}_{\text{IRR}}(h) = \frac{1}{N_{\text{Test}}} \sum_{(i,t)\in\mathbb{T}} \left(r_{it-h} - \text{IRR}_{it}\right)^2,$$

where r_{it-h} is predicted IRR lagged by h years. We report values multiplied by 10,000 for readability (i.e., expressed in basis points). The IRR is defined as the discount rate that equates the enterprise value and the present value of future cash flows. Future cash flows are forecasted using a lasso model estimated in the training sample including IBES forecasts as predictors (dotted lines) or not (solid lines). Predicted IRRs are computed using a linear model estimated in the training sample and lagged by h years. As described in the main text, we consider increasingly informative characteristics from "Rolling" to "Public firms." See text for details. Go back to main text

Tables

TABLE 1: Main variables

Panel A. Variables used in the computation of free cash flow to the firm (FCFF)

Name	Description
`ebit`	Earnings before interest and taxes
`dp`	Depreciation and amortization expense (added back because non-cash)
`capx`	Capital expenditures (cash outflow for fixed assets)
`rect`	Accounts receivable (gross)
`invt`	Inventories
`ap`	Accounts payable
`shrout`	Common shares outstanding[(a)]
τ	Effective corporate tax rate, computed as the statutory U.S. federal rate

Panel B. Predictive variables

Name	Description	FCFF	(1)	(2)	(3)	(4)	(5)	(6)
Industry	First two digits of `sich`	✓	–	✓	✓	✓	✓	✓
Year	`fyear`	–	–	✓	✓	✓	✓	✓
Cash ratio	`che/at`	✓	–	–	–	–	–	✓
ROA	`ebitda/at`	✓	–	–	–	–	–	✓
Book leverage	$(\texttt{dltt} + \texttt{dlc})/\texttt{at}$	✓	–	✓	–	–	–	✓
Tangibility	`ppent/at`	✓	–	–	–	–	–	✓
Asset growth	$\texttt{at}_t/\texttt{at}_{t-1}$	✓	–	–	–	–	–	✓
Assets	log(`at`)	✓	–	–	–	–	✓	✓
Sales	log(`sale`)	–	–	–	–	–	✓	–
PPE	log(`ppent`)	–	–	–	–	–	✓	–
$\texttt{EPS}_t$	Actual EPS from IBES	✓	–	–	–	✓	–	✓
$\texttt{F}_t\texttt{EPS}_{t+1}$	Median forecast from IBES	✓	–	–	–	–	–	–
$\texttt{F}_t\texttt{EPS}_{t+2}$	Median forecast from IBES	✓	–	–	–	–	–	–
$\texttt{FCFF}_t$	Realized free cash flow to firm	✓	–	–	–	–	–	–
$\texttt{FCFF}_{t+1}$	Fitted FCFF one year ahead	–	–	–	–	✓	–	✓
$\texttt{FCFF}_{t+2}$	Fitted FCFF two years ahead	–	–	–	–	✓	–	✓
$\texttt{FCFF}_{t+3}$	Fitted FCFF three years ahead	–	–	–	–	✓	–	✓
$\texttt{FCFF}_{t+4}$	Fitted FCFF four years ahead	–	–	–	–	✓	–	✓
$\texttt{FCFF}_{t+5}$	Fitted FCFF five years ahead	–	–	–	–	✓	–	✓
Age	$\texttt{fyear} - \min(\texttt{fyear})$	✓	✓	✓	✓	✓	–	✓
IVOL	Idiosyncratic volatility (FF4)	✓	–	–	–	–	–	✓

[(a)] We use the CRSP-adjusted share count to stay consistent with market prices.
Note. In Panel B, Column (1) describes the rolling IRR, column (2) private project, (3) private project with leverage, (4) private project with free cash flows forecast, (5) private firm for research, and (6) public firm.

Table 2: Performance of our predictive model of FCFF

	Cash-flows with IBES				Cash-flows w/o IBES				
h	Bias	Var	MSE	R^2	Bias	Var	MSE	R^2	N
1	0.31	12.26	12.36	0.23	0.33	12.58	12.69	0.21	30,302
2	0.38	14.55	14.70	0.16	0.44	15.01	15.20	0.13	26,449
3	0.44	15.89	16.08	0.16	0.51	16.36	16.62	0.13	23,416
4	0.42	17.97	18.15	0.14	0.52	18.49	18.76	0.11	20,948
5	0.35	20.86	20.99	0.11	0.46	21.46	21.67	0.09	18,819

Note. This table shows the out-of-sample MSE of our predictive model for free cash-flows to the firm, as defined in the main text. The left panel reports MSE for when the median IBES EPS forecasts and actual EPS are included as predictors, and the right panel reports the MSE when they are not. We compute MSEs over 100 train-test splits and report the average. Go back to main text

Table 3: MSE for log EV deviations using cash-flows predicted with IBES

Variables used (W_{it})	Rolling IRR	Private project	Private project + leverage	Private project + FCFF forecasts	Private firm for research	Public firm
	(1)	(2)	(3)	(4)	(5)	(6)
OLS	0.7	0.53	0.52	0.47	0.49	0.44
Ridge	–	0.54	0.54	0.47	0.50	0.43
Lasso	–	0.52	0.52	0.46	0.48	0.43
RF	–	0.50	0.49	0.37	0.46	0.31
XGBoost	–	0.49	0.49	0.35	0.46	0.26
NN1	–	0.48	0.48	0.37	0.44	0.31

Note. This table reports the out-of-sample performance across model and predictive variables. Our metric of performance for model R is the mean-squared log error

$$\mathrm{MSE}_{\mathrm{EV}}(R) = \frac{1}{N_{\mathrm{Test}}} \sum_{(i,t)\in\mathbb{T}} \Big(\log \mathrm{PV}_{it}(r_{it}) - \log \mathrm{EV}_{it} \Big)^2,$$

where EV_{it} is the enterprise value and PV_{it} is the present value of future cash flows discounted at the rate r_{it} predicted by model R. Given a train-test split, we estimate in the training sample a lasso model for future cash flows including IBES forecasts as predictors. We compute the IRR as the discount rate that equates the enterprise value and the present value of future cash flows. We train statistical models to predict the IRR in the training sample using various sets of firm characteristics as predictors, as described in the main text. We then predict cash flows and IRRs, and the corresponding present values, out-of-sample. We reproduce this exercise for 100 train-test splits and report the average performance. Each line corresponds to a different statistical, and each column to a different set of characteristics. Go back to main text

Table 4: MSE for IRR deviations using cash-flows predicted with IBES

Variables used (W_{it})	Rolling IRR	Private project	Private project + leverage	Private project + FCFF forecasts	Private firm for research	Public firm
	(1)	(2)	(3)	(4)	(5)	(6)
OLS	11.15	8.82	8.79	8.27	8.37	6.98
Ridge	–	8.79	8.78	8.47	8.35	7.06
Lasso	–	8.77	8.74	8.22	8.30	6.93
RF	–	7.93	7.70	6.60	7.38	5.32
XGBoost	–	7.69	7.53	6.39	7.16	4.73
NN1	–	7.67	7.54	6.37	7.02	5.04

Note. This table reports the average out-of-sample performance across models and sets of predictive variables. Our metric of performance for model R is the mean-squared error

$$\text{MSE}_{\text{IRR}}(R) = \frac{1}{N_{\text{Test}}} \sum_{(i,t)\in\mathbb{T}} (r_{it} - \text{IRR}_{it})^2 ,$$

where r_{it} denotes the discount rate predicted by model R. We report values multiplied by 10,000 for readability (i.e., expressed in basis points). Given a train-test split, we estimate in the training sample a lasso model for future cash flows including IBES forecasts as predictors. We compute the IRR as the discount rate that equates the enterprise value and the present value of future cash flows. We train statistical models to predict the IRR in the training sample using various sets of firm characteristics as predictors, as described in the main text. We then predict cash flows and IRRs out-of-sample. We reproduce this exercise for 100 train-test splits and report the average performance. Each line corresponds to a different statistical, and each column to a different set of characteristics. Go back to main text

Table 5: MSE for log price deviations using cash-flows predicted without IBES

Variables used (W_{it})	Rolling IRR	Private project	Private project + leverage	Private project + FCFF forecasts	Private firm for research	Public firm
	(1)	(2)	(3)	(4)	(5)	(6)
OLS	0.89	0.70	0.70	0.60	0.67	0.55
Ridge	–	0.71	0.72	0.60	0.68	0.53
Lasso	–	0.70	0.70	0.59	0.66	0.53
RF	–	0.67	0.66	0.47	0.65	0.36
XGBoost	–	0.67	0.67	0.46	0.64	0.31
NN1	–	0.66	0.66	0.50	0.62	0.39

Note. This table reports the out-of-sample performance across model and predictive variables. Our metric of performance for model R is the mean-squared log error

$$\text{MSE}_{\text{EV}}(R) = \frac{1}{N_{\text{Test}}} \sum_{(i,t)\in\mathbb{T}} \left(\log \text{PV}_{it}(r_{it}) - \log \text{EV}_{it} \right)^2,$$

where EV_{it} is the enterprise value and PV_{it} is the present value of future cash flows discounted at the rate r_{it} predicted by model R. Given a train-test split, we estimate in the training sample a lasso model for future cash flows excluding IBES forecasts as predictors. We compute the IRR as the discount rate that equates the enterprise value and the present value of future cash flows. We train statistical models to predict the IRR in the training sample using various sets of firm characteristics as predictors, as described in the main text. We then predict cash flows and IRRs, and the corresponding present values, out-of-sample. We reproduce this exercise for 100 train-test splits and report the average performance. Each line corresponds to a different statistical, and each column to a different set of characteristics. Go back to main text

Online Appendix to "Corporate Valuation: An Empirical Comparison of Discounting Methods"

Nicolas Hommel Augustin Landier
David Thesmar

This Appendix contains additional material:

— Appendix A: Appendix Tables and Figures

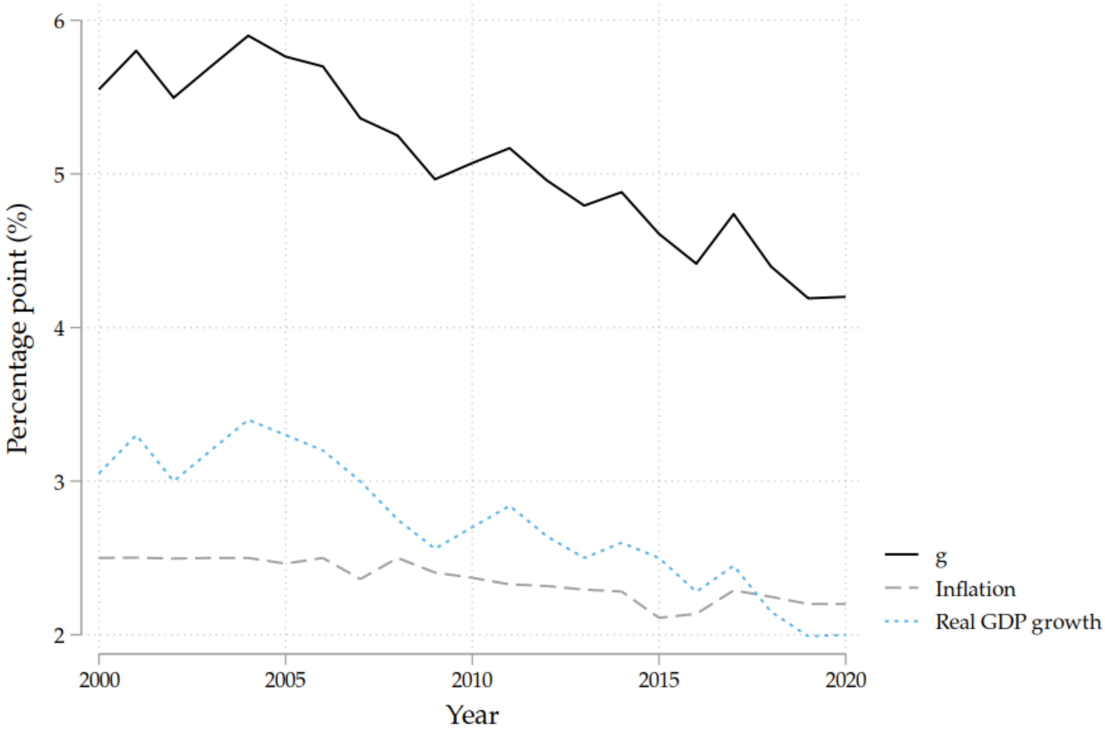
— Appendix B: Details on ML implementations

— Appendix C: Proof of sufficient statistic interpretation of MSE of IRRs

— Appendix D: Numerical evaluation of value loss of using the wrong cost of capital with classic frictions

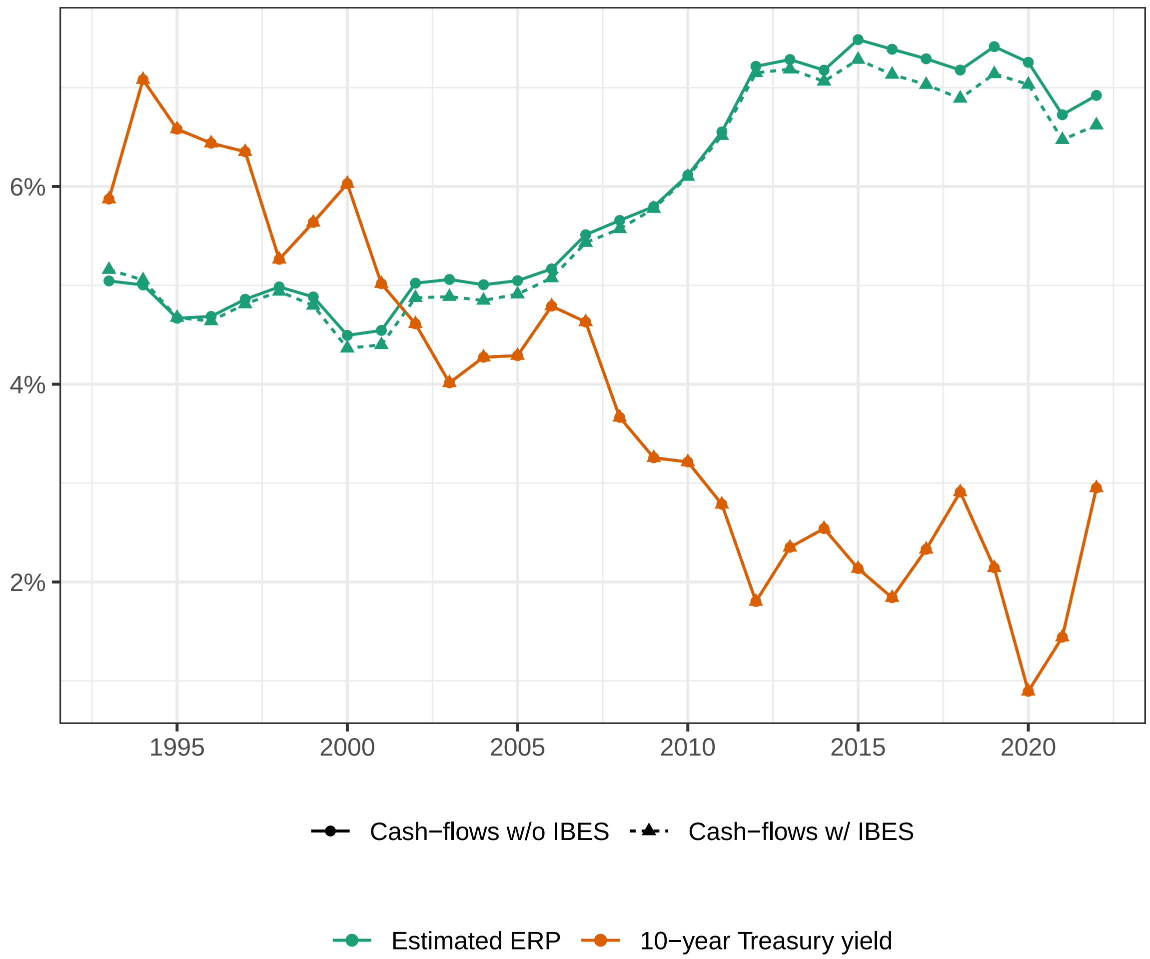
A Appendix Figures and Tables

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.Go back to main text

Figure A.2: Time series of ERP estimates



Note. This figure displays the time-series evolution of the fitted equity risk premium (ERP) and the 1o-year Treasury yield (FRED series DGS10). We estimate the ERP to minimize the distance between the CAPM WACC and the observed IRR in the training sample, as detailed in the main text. We show two versions of the estimated ERP: one where IRRs are derived from cash flows forecasted with IBES data (dashed line) and one without (solid line). Go back to main text.

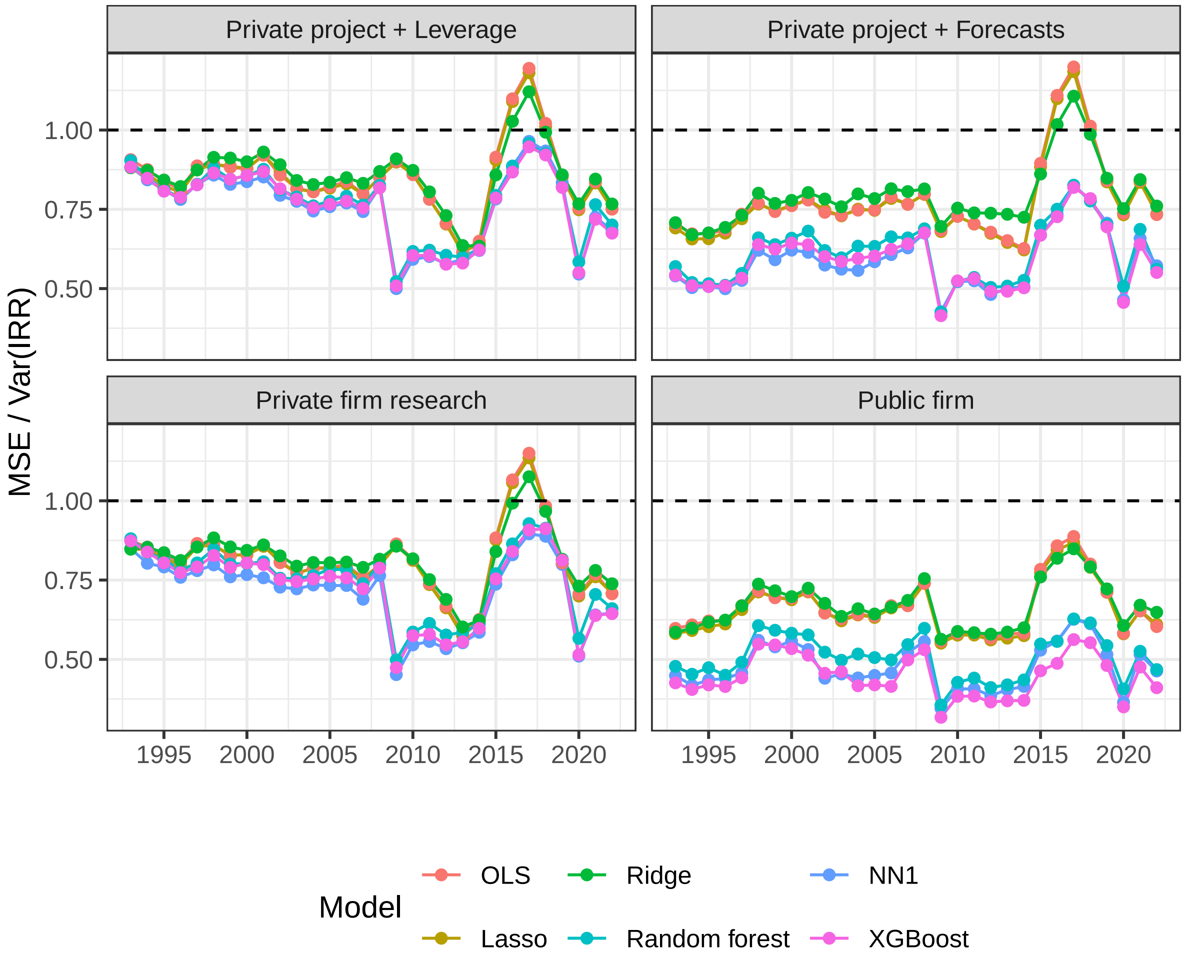
Figure A.3: Trends in RMSE of IRRs

Note. This figure shows the time-series evolution of the out-of-sample root mean squared error (RMSE) of IRR predictions for different models g and sets of features W. The RMSE is defined as

$$\text{RMSE}_{\text{IRR}}(g, W, t) = \sqrt{\frac{1}{N_{\text{Test}}(t)} \sum_{i \in \mathbb{T}_t} (g(W_{it}) - \text{IRR}_{it})^2}.$$

Each panel corresponds to a model g, and each line corresponds to a set of features W. All models are trained on the training sample and evaluated on the test sample, with results averaged over 100 splits. Go back to main text

Figure A.4: Trends in MSE of IRRs, relative to the variance of IRRs



Note. This figure shows the time-series evolution of the out-of-sample mean squared error (MSE) of IRR predictions scaled by the out-of-sample IRR variance. The scaled MSE is defined as

$$\frac{\text{MSE}_{\text{IRR}}(g, W, t)}{\text{Var}_t \text{IRR}_{it}} = \frac{1}{\text{Var}_t \text{IRR}_{it}} \frac{1}{N_{\text{Test}}(t)} \sum_{i \in \mathbb{T}_t} (g(W_{it}) - \text{IRR}_{it})^2.$$

Each panel corresponds to a set of features W, and each line corresponds to a model g. All models are trained on the training sample and evaluated on the test sample, with results averaged over 100 splits.
Go back to main text

Table A.1: MSE for IRR deviations using cash-flows predicted without IBES

Variables used (W_{it})	Rolling IRR	Private project	Private project + leverage	Private project + FCFF forecasts	Private firm for research	Public firm
	(1)	(2)	(3)	(4)	(5)	(6)
OLS	13.31	11.19	11.16	10.37	10.85	8.05
Ridge	–	11.12	11.11	10.41	10.77	8.08
Lasso	–	11.12	11.08	10.28	10.76	7.97
RF	–	10.21	10.02	8.79	9.86	6.43
XGBoost	–	9.98	9.85	8.77	9.63	5.65
NN1	–	9.96	9.86	8.82	9.47	6.17

Note. This table reports the average out-of-sample performance across models and sets of predictive variables. Our metric of performance for model R is the mean-squared error

$$\text{MSE}_{\text{IRR}}(R) = \frac{1}{N_{\text{Test}}} \sum_{(i,t)\in\mathbb{T}} (r_{it} - \text{IRR}_{it})^2 ,$$

where r_{it} denotes the discount rate predicted by model R. We report values multiplied by 10,000 for readability (i.e., expressed in basis points). Given a train-test split, we estimate in the training sample a lasso model for future cash flows excluding IBES forecasts as a predictor. We compute the IRR as the discount rate that equates the enterprise value and the present value of future cash flows. We train statistical models to predict the IRR in the training sample using various sets of firm characteristics as predictors, as described in the main text. We then predict cash flows and IRRs out-of-sample. We reproduce this exercise for 100 train-test splits and report the average performance. Each line corresponds to a different statistical, and each column to a different set of characteristics.Go back to main text

B ML methods implementations

B.1 Models

Linear models. Our first set of estimators are linear models, which minimize objectives of the form

$$\mathcal{L}(\beta, \lambda_1, \lambda_2) = \sum_{it} (\text{IRR}_{it} - \beta^\top X_{it})^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2,$$

where dependent variable is the IRR, and X_{it} is a set of predictors. We consider three special cases: OLS ($\lambda_1 = \lambda_2 = 0$), ridge ($\lambda_1 = 0$), and lasso ($\lambda_2 = 0$). For ridge and lasso, we tune the hyperparameter using cross-validation, as described below.

Random forests. We consider two tree-based algorithms. The first is random forests. At a high level, random forests are an ensemble method that combines predictions from many regression trees using bootstrap aggregation (bagging).

Regression trees partition the data into "leaves," with the goal of minimizing mean squared error. Within each leaf, the outcome of interest is approximated by its average. Because determining a globally optimal partition is computationally prohibitive, we proceed recursively using 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. However, they usually are weak predictors and tend to be unstable due to their recursive-choice construction.

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 a number of 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 a point is the average prediction over all trees.

For our random forest implementation, we use the R package `ranger` by Wright and Ziegler (2017), which implements the algorithm of Breiman (2001) along with measures of variable importance. We tune the parameters using cross-validation on the number minimum leaf size ℓ and the number of variables drawn at each split m, as detailed below.

XGboost. Our second tree-based algorithm is Extreme Gradient Boosting (XGBoost). XGBoost is an ensemble learning method that builds a sequence of B shallow trees to minimize mean squared error plus regularization. Although each individual tree is a weak learner, boosting proceeds iteratively, fitting each new tree on the residuals of the current ensemble. Intuitively, later trees therefore focus on dimensions that the current ensemble misses.

To avoid overfitting and minimize compute, XGBoost combines several strategies. First, the contribution of each new tree is scaled via a learning rate η. Second, only randomly selected a subset of columns and observations are used to grow each new tree. Third, training stops early if the loss function does not decrease for a predetermined number of steps in a holdout sample.

We train the histogram-based algorithm with a loss-guided (leaf-wise) growth policy, which bins continuous features. Although this implementation introduces approximations which may harm prediction accuracy, gains in computation time allow us to explore a finer hyperparameter grid that more than compensates for any loss in predictive accuracy.

Neural network. Our final model is a fully connected feed-forward neural network, which are flexible and well-suited to capture nonlinear interactions missed by simpler models. We explore architectures with one to three hidden layers, each with 128 or 258 units. We use ReLU activations for the hidden layers. We use the Adam optimizer, which combines momentum and adaptive learning rates to accelerate convergence and improve stability.

Given their flexibility, neural networks are prone to overfitting, especially given the moderate size of our dataset. We therefore apply three standard regularization techniques. First, we penalize large weights using L^2 weight decay. Second, we use dropout, which randomly turns off a subset of units during training. Third, we use early stopping based on validation loss: Training halts if performance does not improve after a fixed number of epochs, thereby reducing the risk of overfitting and speeding up training.

B.2 Implementation details

Linear models. We pick the penalization rate using 3-fold cross validation, blocked by `gvkey`. We use the base grid in the `glmnet` implementation in R. Once the optimal hyperparameters are selected, we refit the penalized model to the entire training sample.

Random forest. We use the R package `ranger`. We tune hyperparameters using 3-fold cross-validation blocked by `gvkey`. We use the following values

— `mtry` $\in \left\{1, \lfloor\sqrt{p}/2\rfloor, \lfloor\sqrt{p}\rfloor\right\}$, where p is the number of features,

— `min.node.size` $\in \{1, 5\}$,

— `ntree` $= 1,000$,

— `sample.fraction` $= 0.50$.

Once the optimal hyperparameters are selected, we refit the random forest to the entire training sample.

XGBoost. We use the R package `xgboost`. We tune hyperparameters using 3-fold cross-validation blocked by `gvkey`. We tune the following parameters:

— `max_depth` $\in \{2, 4, 6\}$,

— `eta` $\in \{0.01, 0.1\}$,

— `colsample_by_tree` $\in \{0.30, 0.65, 1.00\}$,

— `subsampling` $= 0.50$,

— `nrounds` $= 2,000$,

— `early_stopping_rounds` $= 20$.

We cap the number of boosting rounds at $2,000$ and use early stopping after 20 rounds of no out-of-fold improvement. Once the optimal hyperparameters are selected, we refit XGBoost to the entire training sample.

Neural network. We use the R package `keras`, and `kerastuneR` for tuning. At the beginning of each training, we set aside 20% of `gvkey` from the training sample for validation hold-out. We rescale the IRR to be within $[0, 1]$. We use the following parameter grid:

— number of layers is in $\{1, 2, 3\}$,

— `units` $\in \{128, 256\}$,

— `dropout_rate` $\in \{0.3, 0.5\}$,

— `l2_strength` $\in \{10^{-6}, 10^{-5}, 10^{-4}\}$.

To minimize compute, we use a random search which explores five draws from this hyperparameter space. Each candidate model is trained for up to 200 epochs with batch-size 64. We use early stopping with a patience of 15, monitoring the MSE on the validation hold-out set.

C Proof of Proposition 1

1. Start with a couple of notations. Call $\pi(k, z)$ the cash flow:

$$\pi(k, z; r) = -k + \frac{1}{1+r}\left(zF(k) + (1-\delta)k\right)$$

and $\kappa(z_0; r)$ the solution of the FOC of profit maximization:

$$E(z|z_0)F'(\kappa(z_0; r)) = r + \delta.$$

Note that this capital stock is dynamically optimal in this model with capital adjustment cost nor financing constraint.

The value of a firm with true cost of capital r but using cost of capital r^* is given by:

$$V(z; r; r^*) = \mathbb{E}_0\left(\sum_{s\geq 1}\frac{1}{(1+r)^s}\mathbb{E}_s\left(\pi(\kappa(z_s; r^*), z_{s+1}; r\right)\right)$$

where $\kappa(z_s; r^*)$ is the capital stock chosen at date s by a firm who believes its cost of capital is r^*. We are interested in computing the value loss of using r^* instead of r:

$$\Delta V = V(z; r; r^*) - V(z; r; r)$$

Now, to compute ΔV, we rely on the second order approximation $V(z; r; r^*)$ in the neighborhood of $r^* = r$:

$$V(z; r; r^*) \approx V(z; r; r) + \frac{\partial V}{\partial r^*}(z; r; r)(r^* - r) + \frac{1}{2}\frac{\partial^2 V}{(\partial r^*)^2}(z; r; r)(r^* - r)^2$$

2. Let us start with the first derivative. By the envelope theorem, the first term is zero:

$$\begin{aligned}\frac{\partial V}{\partial r^*}(z; r; r) &= \mathbb{E}_0\left(\sum_{s\geq 0}\frac{1}{(1+r)^s}\frac{\partial \mathbb{E}_s\pi(z_{s+1}, \kappa(z_s; r); r)}{\partial r^*}\right)\\ &= \mathbb{E}_0\left(\sum_{s\geq 0}\frac{1}{(1+r)^s}\cdot\frac{\partial \mathbb{E}_s\pi}{\partial k}(z_s, \kappa(z_s; r); r)\cdot\frac{\partial \kappa}{\partial r^*}(z_s; r)\right)\end{aligned}$$

By definition of $\kappa(z_s; r)$

$$\frac{\partial \mathbb{E}_s\pi}{\partial k}(z_s, \kappa(z_s; r); r) = 0, \tag{C.1}$$

so that

$$\frac{\partial V}{\partial r^*}(z; r; r) = 0.$$

3. Let us now compute the second derivative:

$$\frac{\partial^2 V}{(\partial r^*)^2}(z;r;r) = \mathbb{E}_0\left(\sum_{s\geq 0}\frac{1}{(1+r)^s}\left(\frac{\partial^2 \mathbb{E}_s\pi}{\partial k^2}(z_s,\kappa(z_s;r);r).\left(\frac{\partial\kappa}{\partial r^*}(z_s;r)\right)^2\right)\right)$$
$$+\mathbb{E}_0\left(\sum_{s\geq 0}\frac{1}{(1+r)^s}\left(\frac{\partial \mathbb{E}_s\pi}{\partial k}(z_s,\kappa(z_s;r);r).\frac{\partial^2\kappa}{(\partial r^*)^2}(z_s;r)\right)\right)$$

From the definition of κ in Equation (C.1), the second term is zero for $r^* = r$, so we only need to compute the first term. From the definitions of π and κ, we know that:

$$\frac{\partial^2 \mathbb{E}_s\pi}{\partial k^2}(z_s,\kappa(z_s;r);r) = \mathbb{E}(z_{s+1}|z_s)F''\left(\kappa(z_s,r)\right)$$

and, from the differentiation of the FOC:

$$\frac{\partial\kappa}{\partial r^*}(z_s;r) = \frac{1}{\mathbb{E}(z_{s+1}|z_s)F''\left(\kappa(z_s,r)\right)}$$

so that:

$$\frac{\partial^2 \mathbb{E}_s\pi}{\partial k^2}(z_s,\kappa(z_s;r^*);r).\left(\frac{\partial\kappa}{\partial r^*}(z_s;r^*)\right)^2 = \frac{1}{\mathbb{E}(z_{s+1}|z_s)F''\left(\kappa(z_s,r)\right)}$$

Now, we factor out expected cash flows:

$$\begin{aligned}\frac{\partial^2 \mathbb{E}_s\pi}{\partial k^2}(z_s,\kappa(z_s;r);r)\cdot\left(\frac{\partial\kappa}{\partial r^*}(z_s;r)\right)^2 &= \frac{1}{-\kappa(z_s;r)+\frac{1}{1+r}\left((1-\delta)k_s + E(z_{s+1}|z_s)F(\kappa(z_s;r))\right)}\\ &\times\frac{1}{\mathbb{E}(z_{s+1}|z_s)F''(\kappa(z_s;r))}\times\mathbb{E}_s\pi\\ &= (1+r)\frac{1}{E(z_{s+1}|z_s)F(\kappa(z_s;r))-(r+\delta)\kappa(z_s;r)}\\ &\times\frac{1}{\mathbb{E}(z_{s+1}|z_s)F''(\kappa(z_s;r))}\times\mathbb{E}_s\pi.\end{aligned}$$

Then, making use of the fact that

$$E(z_{s+1}|z_s) = \frac{1}{F'(\kappa(z_s;r))}(r+\delta),$$

we obtain that:

$$\begin{aligned}\frac{\partial^2 \mathbb{E}_s\pi}{\partial k^2}(z_s,\kappa(z_s;r);r)\cdot\left(\frac{\partial\kappa}{\partial r^*}(z_s;r)\right)^2 &= \frac{1+r}{(r+\delta)}\frac{1}{F/F'-\kappa}\\ &\times\frac{1}{(r+\delta)F''/F'}\times\mathbb{E}_s\pi\\ &= \frac{1+r}{(r+\delta)^2}.\frac{F'}{F-F'\kappa}\cdot\frac{F'}{F''}\times\mathbb{E}_s\pi\\ &= \frac{1+r}{(r+\delta)^2}.\frac{\epsilon_F}{1-\epsilon_F}\cdot\frac{1}{\epsilon_{F'}}\times\mathbb{E}_s\pi\end{aligned}$$

This gives us the second derivative of the value function:

$$\frac{\partial^2 V}{(\partial r^*)^2}(z;r;r) = \frac{1+r}{(r+\delta)^2} \cdot \frac{\epsilon_F}{1-\epsilon_F} \cdot \frac{1}{\epsilon_{F'}} \times \underbrace{\mathbb{E}_0\left(\sum_{s\geq 0} \frac{1}{(1+r)^s}\pi_s(\kappa(z_s;r), z_{s+1};r)\right)}_{=V(z,r;r)}$$

where ϵ_F is the elasticity of output and $\epsilon_{F'}$ the elasticity of MRPK.

4. So, the value loss from taking the wrong discount rate r^* is given by the following approximate relationship:

$$\frac{V(z;r;r^*) - V(z;r;r)}{V(z;r;r)} \approx \left(\frac{1+r}{2} \cdot \frac{\epsilon_F}{1-\epsilon_F} \cdot \frac{1}{\epsilon_{F'}}\right) \cdot \left(\frac{r^*-r}{r+\delta}\right)^2$$

QED.

D Value Loss from cost of capital error error: Taking frictions into account

In this Appendix, we establish that frictions do not significantly affect the conclusion drawn from the frictionless model of Section 4.1. We do this by using a model including financing constraints and adjustment costs, and show that the loss coming from adopting the wrong discount rate is similar for calibrated parameter.

D.1 Set-Up

The extended 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.

D.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:

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

$$k_{t+1} = \kappa(z_t, k_t, d_t; r)$$

$$d_{t+1} = \Delta(z_t, k_t, d_t; r)$$

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 $\frac{W(r^\star,r^\star)-W(r,r^\star)}{W(r^\star,r^\star)} > 0$.

D.3 Solution procedure

This suggestion provides details on the implementation procedure, but can be skipped by readers only interested in the outcome.

This 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.

Table D.2: 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)

Calibration. Table D.2 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

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

   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

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

   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 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

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

[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.

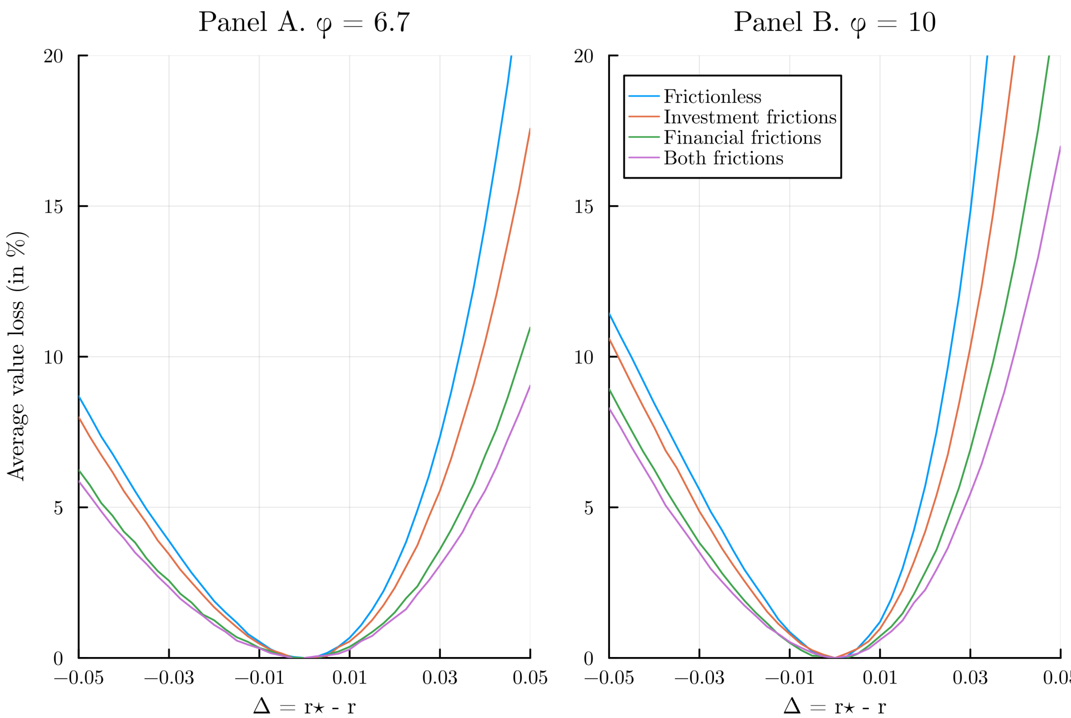
The estimated value loss is

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

D.4 Results

In Figure D.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 D.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 ($\gamma = 0.1$, $\lambda = 0.1$, $M = 0.4$, $m = 0.25$).

Figure D.2 confirms the analysis of the frictionless model that costs 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 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.

The second message from Figure D.2 is that frictions dampen the value loss a bit, but they remain of a similar magnitude. Thus, the approximated formula of proposition 1 is a good approximation of more constrained models, although it is of course not perfect. Financial frictions, more than adjustment costs, affect the value loss.