Forecasting the Term Structure of Equity Betas: Implications for Valuation

Niklas Augustin* Kuntara Pukthuanthong†

We thank Michael O'Doherty and Adam Yore for helpful comments and suggestions. All errors are our own.

***** PRELIMINARY – PLEASE DO NOT CIRCULATE WITHOUT PERMISSION *****

This Version: May 5, 2025

Abstract

Estimating firm discount rates remains a fundamental challenge in asset pricing and corporate finance. We propose a novel approach that forecasts discrete elements of the beta term structure. Our results show that firm betas exhibit predictable variation across horizons and can be reliably forecasted for at least ten years. Incorporating expected betas into factor models reveals substantial dispersion in costs of equity, with an annual difference of approximately five percentage points between the top and bottom deciles of firms. Adjusting for these forward-looking costs of equity in discounted cash flow models reduces pricing errors by 10% relative to standard methods. Furthermore, accounting for changes in expected betas is necessary to construct truly market-neutral long-term portfolios. These findings highlight the importance of horizon-specific beta forecasts for improving the estimation of discount rates, valuation, and investment decisions.

Keywords: Beta estimation, machine learning, discounted cash flows

JEL Classification:

*Trulaske College of Business, University of Missouri. Email: nabvk@umsystem.edu
†Trulaske College of Business, University of Missouri. Email: pukthuanthongk@umsystem.edu

" When did our field stop being asset pricing and become asset expected returning? Why are betas exogenous?" – John Cochrane

Accurately measuring systematic risk has been a central focus of asset pricing research for decades. The most well-known model of risk and return is the Capital Asset Pricing Model (CAPM), developed by Lintner (1965) and Sharpe (1964). It stipulates a linear relationship between risk and return (or, in conditional versions, between expected risk and expected return), with a firm's market beta serving as its sole measure of risk. While there exists an extensive literature on computing backward-looking estimates of firm betas, less research has investigated forecasting future betas and the research that does, typically focuses on simple instrumental variable approaches in tests of the conditional CAPM, which are concerned only with short-term beta prediction.[1]

At the same time, it has been known for decades that betas are time-varying and thus that past beta estimates and future beta forecasts are different (Blume, 1971). The appropriate choice between these depends on the specific task at hand. For example, beta estimates are suitable for evaluating investment managers' performance or testing unconditional asset pricing models. Conversely, beta forecasts are more appropriate for forming hedging portfolios, making capital budgeting decisions, or valuing firms and projects, which is one of the key purposes of corporate finance.

This paper proposes and compares multiple methods to forecast future realized betas and tests their implications on corporate valuation. Prior studies have shown that 97% of analysts rely on the CAPM to estimate costs of capital, but that the discount rates used by analysts are constant (H Décaire and Graham, 2024). Hommel, Landier, and Thesmar (2023) find that discount rates obtained from factor models produce large pricing errors in discounted cash flow (DCF) valuations. One potential reason is that each valuation implicitly dictates a time horizon with which the beta forecast must align. For instance, if a cash flow is to be discounted for five years (e.g., in private equity), then a five-year beta forecast is

[1] Some notable studies include Boguth et al. (2011), Cosemans et al. (2016), Drobetz et al. (2023), Hollstein and Prokopczuk (2016), and Rosenberg and Guy (1976, 1995)

appropriate. Similarly, if a portfolio manager plans to capitalize on a mispricing over six months, the beta forecast should reflect that horizon.

To better reflect these windows, we explore whether it is possible to forecast an entire term structure of betas with a horizon of up to ten years and evaluate which forecasting methods yield the best out-of-sample predictions. Following the recent literature, we employ an extensive set of firm characteristics from Chen and Zimmermann (2021), diverse forecasting approaches including historical beta estimates, shrunk beta estimates, regression-based forecasts, and machine learning forecasts, and focus the analysis on the firm rather than the portfolio level.

We find that betas are highly predictable using even simple forecasting methods. At a one-year prediction horizon, all methods studied exhibit out-of-sample predictive power in the cross-section of betas.[2] However, the most widely adopted beta forecasts (i.e., rolling betas) make for the poorest predictors. Out-of-sample (OOS) R^2 values using rolling and shrunk beta estimates range from 10% to 20%, whereas regression-based methods achieve OOS R^2 values of 43%.

Consistent with the intuition that predicting betas further into the future becomes increasingly difficult, we find that prediction accuracy declines as the forecasting window lengthens. OOS R^2 values become negative when using recent and historical beta estimates for forecasts three years or more into the future, suggesting that assuming equal betas for all stocks performs better at such horizons. This finding aligns with recent research on differences in long-term expected returns. Keloharju, Linnainmaa, and Nyberg (2021) demonstrate that long-term expected return differences cannot be predicted, contributing to the widespread view that long-term costs of equity do not vary across firms (Keloharju, Linnainmaa, and Nyberg, 2022; Welch, 2021).

However, short-term OLS beta predictions, such as those used in conditional CAPM

[2]Previous studies have suggested that portfolio-level betas are predictable using macroeconomic state variables. Our study focuses on the understudied cross-sectional predictive power by demeaning all realized betas and beta forecasts.

research, yield positive R^2 values for horizons of up to eight years, suggesting that modern approaches can identify firms with distinct risk profiles. The best results are achieved using a term-structure-adjusted beta forecasting approach, which produces statistically significant predictive power in each year of the term structure. Similar findings are observed using other evaluation metrics, such as median absolute errors and the percentage of forecasts within a pre-specified range.

One challenge in evaluating beta predictions is the inherent unobservability of firm-level betas, which are measured with noise. Using a bootstrap analysis, we show that our results provide a lower bound on the prediction accuracy under mild assumptions about the correlation between forecasts and measurement errors in betas. Greater beta measurement error disproportionately reduces measured R^2 values relative to the R^2 between predicted and true betas. Additionally, we demonstrate that shrinking beta forecasts, while potentially increasing measured R^2 values, diminishes the true predictive power of the term-structure-adjusted model.

To provider further support for the usefulness of term-structure-adjusted beta forecasts, we test their performance in four applications. First, we show that implied discount rates vary more across firms than previously believed. By incorporating beta forecasts into the CAPM, one can construct a term structure of costs of equity.[3] We discount a fixed cash flow stream for each firm and compare the implied discount rates for the highest and lowest deciles of expected beta firms (as in Keloharju, Linnainmaa, and Nyberg (2021)). The lowest decile firms exhibit average annual discount rates approximately five percentage points lower than those of the highest decile, illustrating substantial variation in the cost of equity at the extremes. These improved discount rates can have real effects since discount rates impact investment decisions (Gormsen and Huber, 2023).

Second, we apply these cost of equity estimates to discounted cash flow (DCF) analyses.

[3]Recent studies have attempted to construct implied stock return term structures, but they often rely on dividend strips or black-box machine learning models, limiting their applicability to private or thinly traded assets and cases where interpretability is crucial Giglio, Kelly, and Kozak (2023).

Hommel, Landier, and Thesmar (2023) find that factor models like the CAPM perform poorly in discounting cash flows, and Gormsen and Huber (2024) show that 80% of deviations in a firm's perceived cost of equity result from managerial errors rather than economic differences. Using a sample of analyst price targets, observed prices, and earnings forecasts, we calculate implied firm valuations with back-of-the-envelope DCF models. Term-structure-adjusted beta forecasts consistently produce lower pricing errors compared to traditional methods, such as shrinking or using recent betas, across various assumptions about the equity risk premium and long-term growth rates. These findings highlight the potential to reduce misvaluations and capital distortions (Van Binsbergen and Opp, 2019).

Third, we illustrate the usefulness of term-structure-adjusted beta forecast in portfolio construction. Prior studies often consider the task of forming market neutral short-term portfolios. However, in practice, some investors may have holding horizons longer than just a month or a year. We thus form hedging portfolios with five year holding windows and show that term-structure-adjusted beta forecasts can be used to better calibrate portfolios' expected betas.

Fourth, we test whether term-structure-adjusted beta forecasts relate to the cross-section of short-term stock returns. Typically, asset pricing model tests use a single short-term expected beta (Hansen and Singleton, 1982). While theory does not clearly link long-term expected risk or expected changes in risk to returns, Clara (2017) provides some motivation. We find that the slope of the beta term structure carries a significant positive risk premium in Fama and MacBeth (1973) regressions.

Finally, we propose a unified term-structure model to predict multiple future betas using a single training process. Specifically, we employ a multi-output neural network and apply the Nelson and Siegel (1987) smoothing function to its beta forecasts, creating a smooth term structure. This approach contributes to the growing literature on using machine learning to predict betas (Alanis et al., 2024; Drobetz et al., 2023; Liu, Lin, and Gopal, 2024), as well as the traditional literature on beta estimation and prediction (Baesel, 1974; Blume, 1971,

1975; Rosenberg and Guy, 1976, 1995). We find that machine learning forecasts improve prediction accuracy and simplify training, as a single model can predict betas at all horizons.

A potential critique of this study is the CAPM's limited empirical validity. While some shortcomings may stem from inaccurate beta forecasts, others, such as the vast literature on anomalies unexplained by the CAPM, remain unaddressed. Our response is twofold. First, our procedure can be applied equally to estimating risk exposures in any other factor model, whose existence is guaranteed by the law of one price. Second, survey evidence shows that the CAPM remains widely used in practice, warranting attention to its main parameter (Graham, 2022; Graham and Harvey, 2001; Jacobs and Shivdasani, 2012).

The rest of the paper is structured as follows. Section I explains the data and summarizes the data sources. Section II describes five models for predicting firm-level betas. Section III presents the main empirical results, and Section IV discusses their statistical properties. Sections V and VI provide applications and extensions to beta term structure forecasts. Section VII concludes.

I Data and Sample Construction

Section I.A provides details on the sample, characteristic construction, and beta estimation. Section I.B reports summary statistics for estimated betas.

I.A Data Sources and Preparation

The sample includes all NYSE, Amex, and NASDAQ ordinary common stocks with return data available in the CRSP daily and monthly stock files and sufficient historical return data to compute annual beta estimates. We begin by estimating a separate CAPM regression at the firm-month level using one year rolling daily stock returns to obtain a time series of overlapping conditional CAPM betas. A firm must have 200 valid return observations over the prior 12 months to have a valid beta estimate. The regression specification follows

Hollstein, Prokopczuk, and Wese Simen (2017).

$$r^E_{i,t} = \alpha_{i,t-11\to t} + \beta_{i,t-11\to t} \times r^E_{Mkt,t} + \epsilon_{i,t}, \qquad t \in [t-11, t]. \tag{1}$$

When a firm is delisted from an exchange, we compound the delisting return and regular return if both are available and replace any missing returns with delisting returns.[4] Returns are always in excess of the risk-free rate. Data on the daily and monthly Fama and French (2015) factor returns, market returns, and risk-free rates are from Kenneth French's website.[5]

Firm characteristic data is obtained from Chen and Zimmermann (2021), who provide a comprehensive and clean dataset of 209 firm characteristics. The characteristics are signed such that higher values correspond to higher expected future returns, based on the findings of prior studies. The data are updated monthly, and when characteristics are updated less frequently (e.g., annually or quarterly), their most recent values are carried forward until the updated value becomes available. Annual accounting data is lagged by six months, and quarterly accounting data by one quarter, following the conventions in the literature. To mitigate the challenge of missing data and its associated selection bias, we follow Gu, Kelly, and Xiu (2020) and Freyberger, Neuhierl, and Weber (2020) and map each firm characteristic to the interval $[-0.5, 0.5]$ cross-sectionally and fill any missing values with zero.

The characteristic data is merged with monthly return data from CRSP as well as the manually estimated one-year betas. We require each characteristic to have a minimum of two unique monthly observations (i.e., to be non-missing for the entire sample). In total, we are left with 140 predictors covering around 3.5 million observations for 28,000 firms from January 1965 to December 2023. A detailed summary of all predictors obtained from Chen and Zimmermann (2021) is provided in table A.1. It should be noted that many predictors exhibit high correlations. Although not an immediate concern for prediction (Lewellen, 2014), high correlations complicate variable importance assessments.

[4] Shumway (1997) provides a discussion of the delisting bias.

[5] See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. we thank Kenneth French for making these data available.

I.B Empirical Properties of Beta Estimates

The fact that beta estimation and forecasting are different is predicated on the fact that betas are time-varying. Panel (a) of figure 1 plots the range of the middle 90% of estimated betas over time. In the cross-section, beta estimates vary between around 0 to 2. The average beta is 0.8, but the value-weighted average beta is 1.

The mean beta (thick black line) varies significantly over time, ranging between 0.5 in the 1980s and 1990s to 1 in the 1970s and 2010s. Prior studies have demonstrated that portfolio betas are predictable (Boguth et al., 2011; Cosemans et al., 2016; Shanken, 1990). Of course, there is no meaning to predicting overall market betas, which are by definition equal to one.

This study instead focuses on predicting the evolution of betas at the firm level. Panel (b) of figure 1 provides initial evidence that such predictions are possible, even for relatively long time horizons. Each month, all firms are sorted into quintile portfolios based on their most recently estimated beta ($\hat{\beta}_{i,t-11\to t}$). For all five portfolios, we calculate the average future realized betas for up to ten years ($\overline{\hat{\beta}}_{p,t+1\to t+12}, \overline{\hat{\beta}}_{p,t+2\to t+13}, \ldots, \overline{\hat{\beta}}_{p,t+109\to t+120}$). This exercise is repeated monthly and time series averages are reported in figure 1. The results illustrate that firm betas are persistent. Betas converge quickly in the first year, as predicted by prior papers who study beta estimation errors: top quintile betas tend to be overestimated and bottom quintile betas tend to be under-estimated (Blume, 1975; Boguth et al., 2011; Chan, 1988). After the first year, remaining differences in betas converge only slowly. Even after ten years, the average low-beta firm has a beta of 0.5 compared to a beta slightly above 1 for the highest quintile of firms.

II Forecasting Systematic Risk Over the Term Structure

The focus of this study is the predictability of beta term structures. We define beta term structures in section II.A and we summarize traditional approaches to forecasting future betas in section II.B. In section II.C, we introduce an extension to forecast betas along the term structure.

II.A Defining the Beta Term Structure

The beta term structure is defined analogously to term structures in fixed income, as a series of firm level betas calculated over varying future horizons—such as one-year, two-year, and three-year intervals. In this study, we focus on ten non-overlapping one-year periods following time t. Thus, the beta term structure of firm i at time t is defined as the set of betas over the first, second, ..., and tenth year following time t:

$$\mathcal{B}_{i,t} = \left\{\beta_{i,t+1\to t+12}, \beta_{i,t+13\to t+24}, \ldots, \beta_{i,t+109\to t+120}\right\}, \tag{2}$$

where $t+1 \to t+12$ indicates the time from beginning of month $t+1$ to end of month $t+12$. Consistent with the literature, all predictions are made and evaluated at the firm-month level (Drobetz et al., 2023).

True betas are never observed, and thus, ex-ante beta forecasts must be compared with ex-post beta estimates. Beta estimates are obtained using a regression of excess daily stock returns on excess daily market returns. Hollstein, Prokopczuk, and Wese Simen (2017) found this regression setup to be the optimal OLS estimation approach and its beta estimates are consistent estimates of the true beta under mild conditions (Andersen et al., 2006). Utilizing daily data also provides a reasonable balance between statistical efficiency and microstructure noise. The set of estimated betas $\{\hat{\beta}_{i,t+1\to t+12}, \ldots, \hat{\beta}_{i,t+109\to t+120}\}$ is denoted $\hat{\mathcal{B}}_{i,t}$ and makes

up the set of betas against which our forecasts errors are calculated.

II.B Traditional Forecasting Approaches

Traditional methods typically focus on predicting a single beta (typically the beta a month or year ahead). This practice is also standard in industry, where H Décaire and Graham (2024) document that "the discount rate used within a DCF model tends to be constant over the entire forecast horizon" (p. 4).

Because all our forecasts must have ten distinct elements to be compared against the ten ex-post beta realizations in the term structure ($\hat{\mathcal{B}}_{i,t}$), we denote constant forecasts as a set of ten elements that are all the same. In this section, we summarize the most widely used beta forecasting approaches, which will serve as benchmarks against term-structure-adjusted forecasts.[6]

The most common predictor of future betas is the most recently estimated beta, often referred to as a rolling beta. Rolling betas used in the literature employ varying levels of look-back windows (typically 1 year, 3 years, or 5 years) and data frequencies (daily vs. monthly). Hollstein, Prokopczuk, and Wese Simen (2017) find that a one-year window with daily data yields the lowest forecasting errors. We follow this practice and compute $\hat{\beta}_{i,t-11\to t}$ as a forecast of each element in the term structure. A comparison of different beta estimation methods is provided in the appendix. The forecasted term structure is denoted $\hat{\mathcal{B}}^{\text{Recent}}_{i,t} = \{\hat{\beta}_{i,t-11\to t}, \ldots, \hat{\beta}_{i,t-11\to t}\}$, where the superscript *Recent* indicates that the most recent beta estimate is used as a forecast.

An alternative measure is to use the average historical beta rather than the most recent one. We calculate the average of the five most recent non-overlapping one-year betas

$$\overline{\beta}_{i,t-59\to t} \equiv \frac{1}{5}\sum_{j=1}^{5}\hat{\beta}_{i,t-12j+1\to t-12(j-1)}.$$

[6] Refer to Hollstein and Prokopczuk (2016) and Hollstein, Prokopczuk, and Wese Simen (2017) for a more in-depth comparisons of beta forecasting methods.

This process more closely resembles industry practice, in which analysts typically use three to five years of data to estimate beta (H Décaire and Graham, 2024). As before, the beta prediction is then held constant for the next ten years.[7] We denote this term structure forecast $\hat{\mathcal{B}}_{i,t}^{\text{Historical}} \equiv \left\{\overline{\beta}_{i,t-59\to t}, \ldots, \overline{\beta}_{i,t-59\to t}\right\}$.

Another approach proposed to address the overestimation challenge in the most recent beta is to shrink betas towards unity, that is, to pull every firm's beta toward the market average beta of 1 (Blume, 1975; Vasicek, 1973). This is equivalent to what is taught in undergraduate and MBA textbooks. Bodie, Kane, and Marcus (2011), one of the most commonly used textbooks in investments, explain that "as an empirical matter, beta coefficients seem to move toward 1 over time" (p. 266). They go on to calculate an "adjusted beta" as $\beta_{Adj} = \frac{2}{3}\beta + \frac{1}{3}$. We follow their approach to create $\hat{\mathcal{B}}_{i,t}^{\text{Shrunk}} = \left\{\hat{\beta}_{i,t-11\to t} * 2/3 + 1/3, \ldots, \hat{\beta}_{i,t-11\to t} * 2/3 + 1/3\right\}$

The last traditional method of estimating future betas relies on ordinary least squares. Blume (1971) notes that regressing future betas on past betas is a method to empirically determine the optimal shrinkage towards 1. Forecasting beta at time t thus takes two steps. First, one estimates the most recent two betas ($\hat{\beta}_{i,t-11\to t}$ and $\hat{\beta}_{i,t-23\to t-12}$). Then, one estimates a cross sectional regression of the most recent beta on the beta from two periods ago:

$$\hat{\beta}_{i,t-11\to t} = \alpha_t + \gamma_t \hat{\beta}_{i,t-23\to t-12}. \tag{3}$$

Using the regression coefficient, next period's beta can then be predicted using the most recent beta and the regression coefficients

$$\hat{\beta}_{i,t+1\to t+12} = \alpha_t + \gamma_t \hat{\beta}_{i,t-11\to t}, \quad \forall i. \tag{4}$$

This approach is identical to shrinking the beta but optimizes the shrinkage coefficients.[8]

[7] If fewer than five historic betas are available, we replace the forecast with the average of the available betas.

[8] Ghysels and Jacquier (2006) show that block-estimated betas (i.e., non-overlapping betas) are valid instruments with uncorrelated error terms.

We follow Rosenberg and Guy (1976) and extend the approach to include another set of covariates: $\mathbf{X_{i,t-2}}$ in the first stage regression and $\mathbf{X_{i,t-1}}$ in the forecasting step. This set includes 139 firm characteristics that have been shown to predict returns. The first stage is estimated using a panel dataset covering the most recent 15 years of data. Figure 2 gives an example for the last observation in the sample ($t = 2013:12$) in panel (b). The top line shows the time span over which the independent variables are measured (grey indicates the training sample and blue the test sample). The second line plots the time period over which the betas (dependent variables) are measured. Equations 3 and 4 are calculated using only information available at time t. Forecasts using this approach are denoted $\hat{\mathcal{B}}_{i,t}^{\text{OLS-C}}$.

Almost none of the traditional methods to forecast betas take an explicit stance on the term structure of betas. One notable exception is Clara (2017). However, his paper uses options data to infer a term structure and is thus restricted to (1) the short end of the term structure (less than two years) and (2) large firms with liquid option markets. Hollstein, Prokopczuk, and Wese Simen (2017) also show that option implied betas can suffer from serious forecasting errors. We generalize the OLS approach to predict a full beta term structure in the next section and propose a new prediction method in section VI.

II.C A New Predictive Framework

The approach to predict term-structure-adjusted betas closely resembles the OLS-C approach described above, but iteratively predicts ten unique betas. Thus, in stage 1, we run ten unique panel regressions:

$$\hat{\beta}_{i,t+1+12\tau \to t+12+12\tau} = \alpha_\tau + \gamma_\tau \hat{\beta}_{i,t-11\to t} + \Omega X_t, \quad \forall i, t = t-59, t-58, \ldots, t-12, \tau = 0, 1, \ldots, 9.$$

At first glance, it might appear that there is a contamination issue in this procedure. For example, when $\tau = 9$, the training sample ends at $t-12$ and the dependent variable is thus $\hat{\beta}_{i,t+97\to t108}$, a future value unknown at time t. However, the beta to be predicted in the test

sample ($\beta_{i,t+109\to120}$) does not overlap with the training data and no part of the test data will be in the training set. A remaining concern, however, is that this training could not have been done historically. The overconditioning can lead to biases if the characteristic-beta relationship is strongly time varying. Most prior research using instrumental variables to test the conditional CAPM assumes that this relationship is not time-varying at all, and we follow suite. One would need to increment the gap between the training and testing samples to ten years. Given the low number of non-overlapping ten-year periods we can study, and the fact that this paper is not focused on real-time portfolio construction, we choose not to include this gap.

Stage 2 of the procedure is identical to that of OLS-C, with the exception that the α and γ components are horizon-varying. The ten elements of $\hat{\mathcal{B}}_{i,t}^{\text{OLS-TS}}$ are thus not equal to one another.

III Evaluating Beta Forecasts

In this section we compare the predictions generated using the forecasting methods outlined in section II. Section III.A summarizes the criteria used to compare predictions. Main results are reported in section III.B and an analysis of the drivers of forecasting power is provided in section III.C.

III.A Performance Criteria for Predictive Accuracy

The out-of-sample period begins in January 1980 and ends in December 2013 to allow us to compute ten-year realized betas for all test observations. Predicted betas are compared with ex-post realized betas based on three criteria: the median absolute error (MAE), the cross-sectional R-Squared (R^2), and a proportional goodness of fit indicator (% Within). Models are trained on a rolling basis using a 15-year rolling window as outlined in figure 2.

The out-of-sample MAE is calculated as the median absolute difference between predicted

betas and future estimated betas

$$MAE^{\text{Model}}_{OOO,\tau} = \frac{1}{T}\sum_{t=1980-01}^{2023-12} Med\Big(\hat{\beta}_{i,t+1+12\tau\to t+11+12\tau} - \hat{\beta}^{Model}_{i,t+1+12\tau\to t+11+12\tau}\Big).$$

The MAE is calculated cross-sectionally for each year-month combination and separately for all ten forecasting windows (i.e., $\tau = 0, \ldots, 9$). Throughout the paper, we report time-series averages of the monthly cross-sectional MAEs.

The second criteria is the OOS R^2. Since the focus of the study is the cross-section of expected returns, we focus on the cross-sectional OOS R^2. Results using the raw R^2 are qualitatively similar. Again, R^2s are calculated for each year-month combination and then averaged across the 408 months. The cross-sectional out-of-sample R^2 for horizon τ is defined as

$$R^{2,\text{Model}}_{OOS,\tau} = 1 - \frac{1}{T}\sum_{t=1980-01}^{2023-12} \frac{\sum_{i=1}^{N}(\hat{\tilde{\beta}}_{i,t+1+12\tau\to t+11+12\tau} - \hat{\tilde{\beta}}^{\text{Model}}_{i,t+1+12\tau\to t+11+12\tau})^2}{\sum_{i=1}^{N}(\hat{\tilde{\beta}}_{i,t+1+12\tau\to t+11+12\tau})^2},$$

where the tilde in $\hat{\tilde{\beta}}$ denotes the cross-sectionally demeaned beta forecast or estimate.

The last evaluation metric employed in this paper (% Within) is an index variable that takes the value one if the forecasted and estimated betas are within a threshold κ, which is set equal to 0.3. The average of this indicator is calculated cross-sectionally and then averaged across the year-month combinations as with the two other performance criteria

$$\%\text{Within}^{\text{Model}}_{OOO,\tau} = \frac{1}{T}\sum_{t=1980-01}^{2023-12} \frac{\sum_{i=1}^{N} I_{-}\Big\{|\hat{\beta} - \hat{\beta}^{Model}| \leq 0.3\Big\}}{\sum_{i=1}^{N} 1}.$$

III.B Evaluation of Forecast Performance

Table 1 reports the performance of the models for all ten prediction horizons. Panel (a) reports median absolute errors, scaled $100x$ for readability.

As implied by economic theory, MAEs increase monotonically in the prediction horizon, no matter which forecasting approach is used. The further out into the future one predicts

a firm's beta, the more noisy the forecasts become.

Our main result, however, is that there are large differences in accuracy between the models. At all horizons, OLS beta forecasts outperform rolling betas, average historical betas, and shrunk betas. Between the two OLS approaches, the horizon-adjusted approach is more accurate at each horizon, except for the one-year horizon at which both methods make the same prediction.

More importantly, the differences across the models increase with the prediction window. When predicting one-year betas, OLS forecast median errors are 25.47 compared to an MAE of 26.63 using rolling beta forecasts. However, at the ten-year window, the same MAEs are 31.81 and 40.6, an almost 10x increase in the difference between the models.

For practitioners who prefer not to explicitly model future betas, the optimal choice between shrinking betas and not shrinking them depends on the prediction window. For the first three years, MAEs using simple rolling betas are lower than those using shrunk betas. However, in years four to ten, shrunk betas produce lower MAEs than rolling betas.

In addition to the MAE, the table also reports the cross-sectional R^2 performance of the models in panel (b). The R^2 has a unique interpretability, because it inherently contrast a model's forecasting power to a null model. In the cross-sectional R^2, the null model is to assign the same beta forecast to each firm. For horizons multiple years into the future, recent research has suggested that this might be the optimal forecast of expected returns (Keloharju, Linnainmaa, and Nyberg, 2021).

Using traditional beta prediction approaches, we find that this recommendation is justified. In the long-run, rolling betas, historical betas, shrunk betas, and short-term beta predictions all fail to have any explanatory power compared to assuming equal betas across firms, as indicated by negative R^2 values. Rolling betas have no explanatory power after two years, although a five-year look-back window improves this to three years. Shrunk rolling betas have their first negative R^2 seven years after prediction and short-term OLS forecasts produce negative R^2 beginning in year eight. The assertion that "[l]ong-term discount rates

do not vary across firms" (Keloharju, Linnainmaa, and Nyberg, 2021) thus seems plausible.

However, we find that the term-structure-adjusted OLS forecasts produce positive R^2 values in the cross-section for all ten years. The R^2s fall gradually from 43.16 to 14.42, indicating that it is possible to predict cross-sectional differences in betas for at least ten years.

Finally, panel (c) reports the percent of predictions that fall within a range of 0.3 of the ex-post realized beta. This evaluation metric provides an important check of economic significance. An error in the beta forecast of 1 leads to an error in the cost of equity estimate equal to the equity risk premium, which in the United States is currently around 4.6%.[9] Beta estimates within ±0.3 thus translate into cost of equity estimates within ±1.38 of the true cost of equity, making the 0.3 threshold a good rule of thumb for beta accuracy. Performance with other cutoffs is discussed in the appendix. As before, OLS forecasts outperformed forecasts using historical betas for all time horizons and the horizon-adjusted forecasts outperformed the constant OLS forecasts at each time horizon (except for the one-year horizon at which both are equal). While only 40% of all forecasts are within 0.3 of the future estimate when using recent betas, 50% of all forecasts achieve the threshold using term structure forecasts.

Significance levels of the prediction criteria are estimated using Newey and West (1986) standard errors. We compare each model's forecasts to those obtained using rolling betas and report the significance of the performance difference in table 2.

First, we find that, especially at short horizons, one-year rolling beta and average five-year betas do not perform significantly differently. This implies that the choice of look-back window when estimating betas may be of secondary importance. OLS predictions, on the other hand, provide statistically significant improvements over using recent betas. Horizon-adjusted OLS forecasts improve over recent beta forecasts for all horizons at a one percent significance level using R^2 as the performance criteria. The results are the same using MAE and %Within, except for at the one-year window.

[9] See https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/ctryprem.html

Panel (d) of table 2 assesses the performance difference between the short-term OLS forecasts and the term structure specific OLS forecasts. For all years and evaluation metrics, the term structure specific forecasts outperform the forecasts assuming a flat term structure of systematic risk. This finding highlights the importance of explicitly calibrating forecasting methods to forecasting horizons and of accounting for expected changes in systematic risk when valuing firms.

It turns out that the average predictive performance hides significant time variation in the accuracy of beta forecasts. Typically, time variation in forecast accuracy is of secondary importance because investors with short prediction horizons can rely on the law of large numbers. However, in our setting with long prediction horizons, investors realistically only participate in the markets for up to six non-overlapping ten-year periods. Forecasts of firm risk turning out worse than expected can thus have significant impacts on lifetime utility.

In figure 4, we therefore plot the time series of the out-of-sample R^2 ($R^{2,\text{Model}}_{OOS,\tau,t}$) of the one-year and ten-year beta forecasts. Panel (a) plots the R^2 when using recent beta forecasts and panel (b) plots the R^2 using the term-structure OLS forecasts. The figure illustrates that using term-structure adjusted forecasts has a secondary advantage over competitive methods: the accuracy of the predictions is more predictable. In fact, there was only a single occurrence in which ten-year OLS forecasts had a negative R^2. Comparatively, during 2007, 1997, and 1986, even one-year beta forecasts using recent betas had negative R^2s ranging as low as -50%.

III.C Determinants of Forecasting Power

We conclude this section by analyzing how variable importance and coefficient estimates differ across one-year and ten-year predictions. Prior research has uncovered strong linkages between firm characteristics and future betas. For example, increases in a firm's financial leverage or variability of a firm's earnings might contribute to systematic risk, while increases in the payout ratio might reduce a firm's systematic risk (Rosenberg and Guy, 1976).

The term-structure setup in this paper allows us to further analyze which firm characteristics predict *changes* in future betas. However, these results should be analyzed cautiously, given the high correlations across predictors. For example, the manually calculated rolling one-year betas correlate with the betas from Frazzini and Pedersen (2014) (*BetaFP*) with a correlation coefficient of 0.6. However, in almost all regressions, the relationship between future betas and current values of *BetaFP* is negative. This does not indicate that future and past betas are inversely correlated, but is merely a consequence of *BetaFP*'s high correlation with the manually calculated betas.

Nonetheless, in figure 5, we plot the 30 most important predictor variables along with their coefficients. The variable importances are calculated as an equal-weighted average of the coefficients in each monthly regression.[10]

The strongest predictor of high future betas are high current betas. This persistence in firm beta is widely recognized at the short-term horizon (Blume, 1975; Boguth et al., 2011; Vasicek, 1973). Adding to the this research, we document that betas are persistent for very long time periods: The 1-year and 10-year coefficients on recent betas are almost identical (around 0.75), mirroring the results of figure 1 panel (b). This result also provides evidence for the hypothesis that traditional auto-correlation coefficients estimated by regressing betas on lagged betas are biased downward due to measurement error.

Other strong predictors of high future short-term betas are high momentum values (e.g., *Momentum in High Volume Stocks, Momentum (6 month), Momentum (12 month)*). Boguth et al. (2011) have shown that appropriately conditioning for high expected future beta using historical beta estimates as instruments can significantly reduce the momentum anomaly. Our finding that momentum stocks tend to have higher short-term betas resonate with theirs.

Coefficients for long-term regressions differ significantly from those of short-term regres-

[10] All variables are transformed onto an interval from -0.5 to 0.5 to allow for comparison in coefficient magnitudes, with the exception of recent betas, which are not transformed. In the data, all variables signs were adjusted such that high value are correlated with high expected returns. We undo this transformation to simplify interpretability.

sions. Momentum, for example, does not predict high long-term betas. Out of all predictors, the FF3 and CAPM idiosyncratic risk observe the highest increase in coefficient from the one-year to the ten-year window. Over longer periods, idiosyncratic risk may reflect deeper structural vulnerabilities or operational inefficiencies which increase a firm's systematic risks.[11]

Characteristics that correlate with small future betas tend to be more persistent. They include high volume, dividend initiation, and high valuation ratios (e.g., high sales-to-price or book-to-market ratios). These variables all indicate perceived safeness. Low-beta predictors are fundamental firm characteristics, while high-beta predictors are often related to firm's trading activities. In unreported results, we find that high-beta firms tend to experience larger changes in beta than low-beta firms, in line with the finding that the characteristics that predict low betas next year tend to also do so in the long-term. In the appendix, we directly report the characteristics whose coefficients change the most over the ten-year window (figure A.1).

IV Bias in Performance Evaluation and Shrinkage

The prior section showed that the term structure of betas can be predicted with substantial accuracy. Section IV.A shows that the estimated forecast accuracy from the prior section is biased downward. Section IV.B tests whether additional improvements can be obtained by shrinking OLS-TS forecasts.

IV.A Bias in Performance Metrics

Like Hollstein and Prokopczuk (2016), the prior section compared ex ante forecasts of betas to ex post realized betas. However, realized betas are measured with noise and thus the true prediction accuracy is unknown. It is thus important to analyze whether the performance

[11] These results do not hold in univariate regressions.

measures are biased and, if so, in which way.

We first establish in the appendix that, under mild conditions, beta forecasts obtained from OLS-TS are unbiased. Empirically, this claim is supported by the fact that the relationship between predicted betas and realized betas in figure 3 is close to 1.[12]

Given that the beta forecasts are unbiased, we next perform a simulation analysis to illustrate the resulting bias in performance measurement. The benefit of conducting a simulation is that population parameters and measurement errors are observable.

The modeled economy consists of 100 firms, each having a beta determined by three state variables X_i following the linear function: $Y_i = \beta' X_i$, where $X_i \sim N_3(0, \sigma_X \times I)$. The economy lasts for two periods: $t = 1, 2$. For each firm, we draw a realization of the state variables for the two periods: $X_{i,1}$ and $X_{i,2}$ and then compute the true firm betas $Y_{i,1}$ and $Y_{i,2}$. The econometrician only observes a noisy estimate of the true betas: $\hat{\hat{Y}}$, which is given by $\hat{\hat{Y}} = Y + \mu$, where the measurement error (μ) follows a normal distribution with mean 0 and variance σ_μ. Given the observed information tuple $(\hat{\hat{Y}}_{i,1}, X_{i,1})$, the econometrician can estimate the relationship between Y and X and then use X_2 to make a prediction of $Y_{i,2}$.[13]

We estimate the 100 forecasts of period 2 betas ($\hat{Y}_{i,2}$) and calculate the true R^2 (between $Y_{i,2}$ and $\hat{Y}_{i,2}$) as well as the measured R^2 (between $\hat{\hat{Y}}_{i,2}$ and $\hat{Y}_{i,2}$). This simulation is ran $10,000$ times for various levels of measurement errors (σ_μ) and variances of the state variables (σ_X). We then plot the true R^2 (panel a), measured R^2 (panel b), and bias in R^2 (panel c), which is defined as the difference between the R^2 using true betas and the R^2 using measured betas. A positive bias hence implies that the measured R^2 is too low.

The results are depicted in figure 6. Panel (a) indicates that the true R^2s only drop when measurement error variance far outweighs the variance of the independent variables. The relative levels of measurement error and variance of X determine the precision of the

[12]The same pattern can also be produced if beta forecasts predict only the measurement errors of future betas.

[13]This setup assumes that X is observable before $\hat{\hat{Y}}$. Alternatively, one can think of an out-of-sample forecast and replace X_1 and X_2 with X_0 and X_1.

parameters estimated in the first stage, which is the only determinant of the true R^2.[14] In principle, given enough observations, the true R^2 would be equal to one no matter the measurement error.

On the contrary, the measured R^2 falls when measurement error is introduced. A high enough level of measurement error can make forecasting power appear almost non-existent if the measurement error makes up most of the variation in betas. Most importantly the simulation shows that for any reasonable range of measurement error and variance in predictors, the true R^2 is always higher than the measured R^2. Thus, the performance measures from section III are biased downward and actual beta predictability is even stronger than what the prior results suggest.

IV.B Optimal Level of Forecast Shrinkage

The fact that true betas are unobservable also makes it difficult to determine whether forecasts of betas are too extreme and should be shrunk as suggested by Vasicek (1973). We adjust the prior simulation and assume that firm characteristics and true betas are equal in both periods. We then repeat the analysis and create two separate forecasts of beta: the period 1 measured beta and the OLS forecasted beta.

Panel (a) of figure 7 plots the R^2 between period 1 measured betas and period 2 observed betas for varying levels of shrinkage and measurement error. The variance in the characteristics is fixed. The results reveal the traditional pattern, that shrinking beta forecasts in the presence of measurement error improves accuracy. However, as illustrated by Blume (1975), the regression approach of section II directly estimates the optimal shrinkage level. Panel (b) illustrates this point by plotting the R^2 between true betas and beta predictions from regressions. For all levels of measurement error, employing a shrinkage to the forecasts

[14] The first stage estimates are unbiased estimators whose variance depends positively on the measurement error and inversely on the variance of the independent variables. The true R^2 is determined by the error in the first stage coefficient estimate. Because the estimate is unbiased, the expected magnitude of the measurement error is increasing in the variance of the coefficient estimate and inherits the same properties in relation to the measurement error and variance.

reduces forecast accuracy.

Most importantly, the simulation indicates that the regression approach always outperforms the recent beta forecasts. Panel (c) plots the difference between the R^2 obtained from the regression approach in panel (b) and the recent beta approach in panel (a). For all combinations of measurement error and shrinkage, the difference is positive. In addition, the difference is increasing in measurement error, indicating that employing direct predictive models becomes more important as measurement errors increase. Generally, prediction models have more information to obtain reasonable forecasts. In addition, the set of potential forecasts evaluated in prediction models encompasses the set of using simple historical averages and thus can only rarely underperform.

To give empirical credence to this analysis, we compare the range of beta forecasts implied by OLS-TS predictions to the range of beta estimates used by analysts in practice. H Décaire and Graham (2024) show that subjective betas used by financial analysts have a mean of 1.1 with an interquartile range (IQR) of 0.4. Since analysts typically don't utilize term-structure adjusted betas and instead assume a constant beta, we compute the arithmetic average beta along the term structure for each observation. In line with the assumptions used by practitioners, our analysis also reveals an IQR of 0.4 in average forecasted betas, suggesting that the range of OLS-TS betas is similar to that of analyst betas used in practice. We end with the caveat that rather than applying a uniform shrinkage factor, a potentially fruitful avenue may be to shrink beta estimates for which a strong prior exists or to winsorize beta estimates that appear to be extreme. Such an analysis is conduced in the appendix.

V Applications to Asset Pricing and Investment Analysis

This section analyzes four applications of beta term structure predictions. In section V.A, we show that costs of equity vary in the cross-section at levels higher than previously believed.

In section V.B, we show that beta term structures yield superior estimates of costs of equity in discounted cash flow (DCF) valuations. We then, in section V.C, illustrate the usefulness of term-structure adjusted betas in forming market neutral portfolios. Finally, in section V.D, we show that forecasted betas are related to the cross-section of stock returns.

V.A Variation in the Cost of Equity Across Firms

A consensus is emerging that expected returns across firms converge quickly over the term structure (Keloharju, Linnainmaa, and Nyberg, 2021), implying that firms have very similar costs of equity. **fama2008dissecting** conclude that "Given the imprecision of beta estimates for individual stocks, we judge that little is lost in omitting them from the cross-section regressions (in effect assuming they are all equal to 1.0)". Ivo Welch notes that "the CAPM is worse than useless". He states that "[t]he primary disagreement which remains among finance professors is whether it is merely worse than useless or statistically significantly worse than useless" (Welch, 2021, p. 189) and prescribes that using the CAPM with identical betas of 1 for each firm seems to be at least as good as using the CAPM as traditionally used. In this section, we explicitly model the dispersion in costs of equities across firms and compare our results to the existing evidence.

We model the cost of equity analogously to equation 5.[15] Each month, we sort firms into deciles based on their average forecasted cost of equity for the next ten years. Table 3 reports the cross-sectional average costs of equity for the top and bottom deciles for each year in columns 2 and 3. For example, firms with the highest predicted betas have average one-year costs of equity of 11.12%.[16] On the other hand, the lowest beta firms have an average cost of equity of just 0.36% indicating an expected beta of almost zero. The process is then repeated for year 2 to year ten.

Even after ten years, the lowest decile firms still have average costs of equity of 2.4% compared to 9% for firms in the top decile. This result indicates that implied costs of capital

[15] We let the risk-free rate be equal to zero since it doesn't impact the dispersion in costs of equity.
[16] This is analogous to the firms having an expected beta of $11.12/8 = 1.39$.

vary significantly across firms and that expected exposure to systematic risk persists for long periods of time. We follow Keloharju, Linnainmaa, and Nyberg (2021) and estimate a single implied cost of equity. To do so, we first extrapolate costs of equity for years 11 to 100 following

$$\hat{r}_{E,i,t+\tau} = \hat{r}_{E,i,t+\tau-1} * 0.9 + 0.8\%.$$

By year 98, the one-year costs of equity of the top and bottom decile firms have converged as can be seen in columns 2 and 3. We then calculate two separate cash flow streams, each of which pays \$1 in year one and grows at a constant rate g, which is set equal to 1% and 5%, representing a value and a growth company. The cash flows are reported in the last two columns. Finally, the present value of each cash flow stream is calculated using the two series of costs of equity. The bottom panel of table 3 reports the four present values. The implied constant discount rate is imputed using the Gordon growth formula: $P = \frac{CF_1}{r-g}$.

Our results show that differences in systematic risk exposure result in large persistent differences in expected returns. The result is particularly pronounced for low-growth firms, whose costs of equity range from 4.55% to 9.3%. Fluctuations in even short-term expected returns can have large impacts on firms with low duration–whose values come from mostly short-term cash flows. Even a dividend stream growing at 5%, however, will have very different prices under the different discount rates. Implied discount rates range from 6.4% to 9.1%, resulting in the first cash flow stream having a price three times as high as the second.

V.B Implications for Discounted Cash Flow Models

DCF models are widely used in corporate valuation and taught in most business schools (Berk and DeMarzo, 2011; Kaplan and Ruback, 1995). To study whether term structure beta estimates improve DCF valuations, we conduct back-of-the-envelope DCF analyses following Hommel, Landier, and Thesmar (2023).

Beta term structure predictions are merged with mean analyst price targets and earnings

forecasts from IBES. In addition, we also collect actual prices from CRSP. For each firm-month, we then estimate ten costs of equity

$$\hat{r}^{e,\text{Model}}_{i,t,\tau} = r_{F,t} + \hat{\beta}^{\text{Model}}_{i,t+1+12*\tau \to t+12+12*\tau} * E(r_E), \quad \tau = 0, 1, \ldots, 9 \tag{5}$$

We then explicitly evaluate the following two-stage growth model

$$\hat{PV}^{M}_{i,t} = \left[\sum_{j=1}^{10} \frac{\hat{E}_t\left[CF_{t+12*j}\right]}{\prod_{\tau=1}^{j}(1+\hat{r}^{M}_{i,t,12*\tau})} \right] + \frac{1}{\prod_{\tau=1}^{10}(1+\hat{r}^{M}_{i,t,12*\tau})} \times \frac{\hat{E}_t\left[CF_{t+132}\right]}{r_E - g}. \tag{6}$$

Expected cash flows are estimated using average analyst earnings forecasts. Each test is repeated for a range of terminal growth rates ranging from 0% to 2%.[17] Terminal costs of equity are set equal to the equity risk premium of 8%.

Predicted prices are compared to analyst price forecasts and real prices analogously to prior tests. We compute the median absolute error, the R^2 between predicted and observed prices, and the percent of times that the predicted price is within 25% of the true price. Results are presented in table 4 for analyst price targets and in table 5 for real prices.

We find that DCF models incorporating term-structure-adjusted costs of equity exhibit significantly stronger correlations between estimated prices and analyst price forecasts compared to DCF models that rely on historical beta estimates or assume constant future betas. Specifically, the highest attainable R^2 between analyst price targets and DCF-based price estimates reaches 67.8% when using term-structure-adjusted discount rates, whereas the highest R^2 achieved using historical beta estimates is 62.45%. These results are consistent when evaluated using MAEs, with minimum MAEs of 10.18 and 11.24 for term-structure-adjusted and historical beta-based estimates, respectively. Furthermore, when measuring the proportion of observations within 25% of the actual price, the term-structure-adjusted approach outperforms the historical beta method, yielding 37.5% compared to 34.8%.

[17] H Décaire and Guenzel (2023) shows that the average long-term growth (LTG) expectation by analysts is around 2%. The paper also shows that IBES LTG forecasts bare little resemblance to terminal growth rate expectations.

Costs of equity obtained from shrunk beta estimates provide results that are qualitatively very close to those obtained using term-structure-adjusted beta forecasts. For example, the R^2 value using shrunk beta inputs is 66.51 and for both MAEs and the % Within measures, costs of equity based on shrunk betas perform second best, implying that shrunk betas do estimate costs of capital well. This finding creates a puzzle, given shrunk beta's poor predictive power of future realized systematic risk.

We next examine an alternative explanation of the findings. Given the widely accepted practice of shrinking betas, we compare the performance of DCF models against realized prices, rather than analyst forecasts. In efficient markets, market prices should closely correspond to fundamental firm value derived from future cash flows and discounted at an appropriate cost of capital. As before, discount rates derived from term-structure-adjusted beta forecasts perform the best among all three predictors. The gap in performance between using term-structure-adjusted beta expectations and shrunk betas increases (86.1% vs. 63.6%) and short-term OLS beta forecasts attain even higher R^2 values than shrunk betas.

Overall, our results indicate that term-structure-adjusted beta forecasts are better suited as inputs into firm valuation models. Costs of equity derived from these betas provide price estimates more closely aligned with analyst price targets and observed market prices.

V.C Using Beta Forecasts for Portfolio Hedging

We now test how forecasted betas can be used in constructing portfolios. Prior research has investigated the role of short-term beta forecasts in constructing market-neutral short-term portfolios (Cosemans et al., 2016). However, in some circumstances investors may plan to hold their portfolio for an extended period of time. For example, private equity funds typically invest in 12 to 15 firms and have a holdings period of about five years (Ivashina, 2023). We employ a bootstrap analysis to infer how various beta forecasts perform in constructing minimum-variance market-neutral quasi-PE portfolios. The quasi-PE portfolios are portfo-

lios of 15 public firms held for five years. The use of public firms allows for the calculation of realized systematic risk in the post-formation period.

Following Ghysels and Jacquier (2006), the objective is to construct a market-neutral portfolio with minimum variance:

$$\begin{aligned} \min_{\{\mathbf{w}\}} \quad & \sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}} \\ \text{s.t.} \quad & \mathbf{w}^\top \hat{\boldsymbol{\beta}}_{i,t+1 \to t+60} = 0. \end{aligned}$$

Details on the weight optimization are provided in the appendix. Each month, we form 20 potentially overlapping portfolios of 15 random firms. Firms must have non-missing return observations for the two years prior to sorting and five years after sorting.[18] Optimal market-neural weights are then calculated for each portfolio based on recent betas and average predicted betas. Finally, the portfolio's return is tracked for five years.

Figure 8 plots the distribution of ex-post realized portfolio betas. The blue line plots the distribution of ex-post betas using the OLS-TS beta forecasts whereas the red line plots the distribution using recent betas forecasts. The results show that portfolio's formed on past betas are not truly market neutral. On average, future betas of these portfolios are positive. Once adjusted for expected changes in systematic risk, we are able to construct portfolios that remain market neutral over five years. In practice, short-selling constraints and the lack of firms with negative expected betas makes it unreasonable for private equity firms to obtain true market neutrality. However, the analysis reveals that investment managers are better served incorporating expected changes in beta into their portfolio formation.

V.D Beta Term Structures and the Cross-Section of Returns

One might argue that the CAPM has been shown not to predict returns. Thus, there is no point in forecasting betas and measuring betas' implications on costs of capital. However,

[18] This requirement induces a survivorship bias. In practice, a firm dropping out of the sample would create a change in the expected beta.

there seems to be consensus in the literature that (a) the CAPM is widely used (Graham, 2022) and that (b) the CAPM measures systematic risk well (Jensen, 2023). Furthermore, the lack of a correlation between market betas and realized returns may also be due to mispricing (for example through an overconfidence bias as shown in Jensen (2023)).[19] In that case, costs of equity across firms would still differ. Finally, beta forecasts are useful in other applications such as the DCF models shown above or in creating market neutral portfolios.

Still, in this section we analyze how the term structure of expected betas are related to the returns. To do so, we decompose the term structure of systematic risk exposures into two parts: a level and a slope, akin to the fixed income literature and Clara (2017). We deviate slightly from the literature and use the one-year predicted beta as the level. The reason is that the one-year expected beta should be directly related to the one-year expected return. Following the literature, we calculate the slope as the difference between the ten-year and the one-year forecasted beta. We calculate one-year rolling daily exposures to the factors in Fama and French (2015) using daily returns (except for the market factor). We then run monthly Fama and MacBeth (1973) regressions of stock returns on the lagged factor exposures, forecasted one-year beta, and forecasted term structure slope. Results are reported in panel (a) of table 6.

Our main finding is that the predicted beta level is insignificantly related to returns. However, the coefficient on the slope of the term structure is positive and statistically significant (except for in column 6). The coefficient is between 1.5 and 2, indicating that firms whose betas are expected to further increase in the future, have higher returns. This result is also the opposite of what is predicted by the theory (Clara, 2017).

The remaining factors are insignificant, which can be attributed to (1) generally weak performance of factor models and (2) an errors-in-variable bias that is particularly strong with individual securities.

[19] For a summary of the empirical evidence of the CAPM, view Fama and French (2004).

We conclude this section my forming double-sported portfolios on size and term structure slopes. Firms are sorted into 25 portfolios each January and monthly returns are calculated from 1980 : 01 to 2013 : 13. A high slope minus low slope portfolio performs strongest for very large stocks with an average monthly return of 0.9 and a t-statistic of 3.97. Overall, these results indicate that forecasted betas are related to expected returns, although a deeper analysis of the mechanism is outside the scope of this paper.

VI Machine Learning for Beta Prediction

Using OLS to forecast the term structure of betas is simple and transparent, but significantly constraining. OLS can only model linear relationships and is unable to detect interactions between predictor variables (Breiman, 2001). In addition, the model must be trained separately for each prediction horizon (a new OLS model must be fit to forecast the one-year, two-year, etc. betas). We propose a more direct prediction method in this section.

We extend the baseline OLS model in two ways. First, we employ a feed-forward neural network with a multi-output layer. This structure allows us to use a single model to forecast each element of the term structure. Second, we use the Nelson and Siegel (1987) model to fit a smoothed term structure curve, reducing the noise in each estimate.

Neural networks are ideally suited for this task as they are able to model many outputs simultaneously and can approximate any well behaved function arbitrarily well (Scarselli and Tsoi, 1998). They are also at the forefront of modern machine learning applications such as transformers in large language processing, computer vision, and many other fields.

Figure 9 illustrates the structure of the network. We use a feed-forward architecture, as it is the most standard type of neural network, but the application of more advanced models is certainly possible. The baseline model utilizes a single hidden layer with 32 hidden nodes and the output layer forecasts each predicted beta as a linear function of these hidden node values. Details on the training method and hyperparameters are provided in the appendix.

Table 7 reports the performance of the raw neural network and the Nelson Siegel adjusted neural network. Even the single neural net is able to forecast ten-year betas with a higher R^2, lower MAE, albeit a slightly lower percent of firms whose betas are predicted within |0.3|. Once applying the NS-transformation, the R^2 jumps from 14.27 to 18.5 percent, a almost 30 percent improvement. In unreported results, we extend the neural network model and apply other algorithms such as regression trees and obtain results significantly higher than the ones reported here. A review of optimal forecasting methods is left for future research.

VII Conclusion

The estimation of discount rates has been a debated topic for decades. Much focus has been put on predicting short term returns, with success. However, little work has been done on predicting term structures of equity returns, in particular using interpretable methods. On the contrary, most recent research proposes to not bother predicting costs of equities, expected returns, or systematic risk exposures in the first place, culminating in the view that long-term expected returns do not vary across firms.

We attempt to challenge this claim by providing the first comprehensive analysis of the predictability of firm's CAPM betas along the term structure. We show that betas are highly predictable for horizons of up to ten years using both simple (OLS) methods and machine learning methods. Importantly, expected betas are not constant over time, and models that allow betas to vary over the term structure obtain a higher accuracy than those that do not.

Given our forecasts of firm betas, and assuming the CAPM, we show that costs of equity, and thus discount rates, vary significantly more across firms than prior research would suggest. We apply these costs of equity estimates to firm valuation using discounted cash flow analyses and compare implied firm values to analyst price targets and observed market prices. Our results indicate that term-structure-adjusted betas yield costs of equity that result in better pricing power, particularly compared to real prices.

Overall, we recommend the exact opposite of what prior studies have concluded. Rather than to stop predicting betas overall, we recommend predicting future betas on a more granular basis, at the firm level and across different horizons in the term structure.

References

Alanis, E., Lesseig, V., Payne, J. D., & Quijano, M. (2024). Can machine learning methods predict beta? *Applied Economics*, 1–15.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Wu, G. (2006). Realized beta: Persistence and predictability. In *Econometric analysis of financial and economic time series* (pp. 1–39). Emerald Group Publishing Limited.

Baesel, J. B. (1974). On the assessment of risk: Some further considerations. *The Journal of Finance*, *29*(5), 1491–1494.

Berk, J., & DeMarzo, P. (2011). Corporate finance, global ed. *Essex: Person Education Limited*.

Blume, M. E. (1971). On the assessment of risk. *The Journal of Finance*, *26*(1), 1–10.

Blume, M. E. (1975). Betas and their regression tendencies. *The Journal of Finance*, *30*(3), 785–795.

Bodie, Z., Kane, A., & Marcus, A. J. (2011). *Investments*. McGraw-hill.

Boguth, O., Carlson, M., Fisher, A., & Simutin, M. (2011). Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics*, *102*(2), 363–389.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, *16*(3), 199–231.

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

Chan, K. (1988). On the contrarian investment strategy. *Journal of business*, 147–163.

Chen, A. Y., & Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review, Forthcoming*.

Clara, N. (2017). The term-structure of systematic risk.

Cosemans, M., Frehen, R., Schotman, P. C., & Bauer, R. (2016). Estimating security betas using prior information based on firm fundamentals. *The Review of Financial Studies*, *29*(4), 1072–1112.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of financial economics*, *7*(2), 197–226.

Drobetz, W., Hollstein, F., Otto, T., & Prokopczuk, M. (2023). Estimating stock market betas via machine learning. *Journal of Financial and Quantitative Analysis*, 1–56.

Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, *18*(3), 25–46.

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

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, *81*(3), 607–636.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of financial economics*, *111*(1), 1–25.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, *33*(5), 2326–2377.

Ghysels, E., & Jacquier, E. (2006). Market beta dynamics and portfolio efficiency. *Available at SSRN 711942*.

Giglio, S., Kelly, B. T., & Kozak, S. (2023). *Equity term structures without dividend strips data* (tech. rep.). National Bureau of Economic Research.

Gormsen, N. J., & Huber, K. (2023). *Corporate discount rates* (tech. rep.). National Bureau of Economic Research.

Gormsen, N. J., & Huber, K. (2024). *Firms' perceived cost of capital* (tech. rep.). National Bureau of Economic Research.

Graham, J. R. (2022). Presidential address: Corporate finance and reality. *The Journal of Finance*, *77*(4), 1975–2049.

Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of financial economics*, *60*(2-3), 187–243.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

H Décaire, P., & Graham, J. R. (2024). Valuation fundamentals. *John Robert, Valuation Fundamentals (September 09, 2024)*.

H Décaire, P., & Guenzel, M. (2023). What drives very long-run cash flow expectations? *Marius, What Drives Very Long-Run Cash Flow Expectations*.

Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, 1269–1286.

Hollstein, F., & Prokopczuk, M. (2016). Estimating beta. *Journal of Financial and Quantitative Analysis*, *51*(4), 1437–1466.

Hollstein, F., Prokopczuk, M., & Wese Simen, C. (2017). *How to estimate beta?* (Tech. rep.). Hannover Economic Papers (HEP).

Hommel, N., Landier, A., & Thesmar, D. (2023). *Corporate valuation: An empirical comparison of discounting methods* (tech. rep.). National Bureau of Economic Research.

Ivashina, V. (2023). Private equity financing. *Handbook of the Economics of Corporate Finance: Private Equity and Entrepreneurial Finance*, 139.

Jacobs, M. T., & Shivdasani, A. (2012). Do you know your cost of capital? *Harvard business review*, *90*(7/8), 118–124.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance*, *56*(2), 699–720.

Jensen, T. I. (2023). Subjective risk and return. *Available at SSRN 4276760*.

Kaplan, S. N., & Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis. *The journal of Finance*, *50*(4), 1059–1093.

Keloharju, M., Linnainmaa, J., & Nyberg, P. (2022). Do you really know your cost of capital? *Journal of Applied Corporate Finance*, *34*(3), 116–128.

Keloharju, M., Linnainmaa, J. T., & Nyberg, P. (2021). Long-term discount rates do not vary across firms. *Journal of Financial Economics*, *141*(3), 946–967.

Lewellen, J. (2014). The cross section of expected stock returns. *Forthcoming in Critical Finance Review, Tuck School of Business Working Paper*, (2511246).

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, *20*(4), 587–615.

Liu, Y., Lin, J., & Gopal, A. (2024). Neuralbeta: Estimating beta using deep learning. *arXiv preprint arXiv:2408.01387*.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, 473–489.

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Roberts, M. R., & Whited, T. M. (2013). Endogeneity in empirical corporate finance1. In *Handbook of the economics of finance* (pp. 493–572, Vol. 2). Elsevier.

Rosenberg, B., & Guy, J. (1976). Prediction of beta from investment fundamentals: Part one. *Financial Analysts Journal*, *32*(3), 60–72.

Rosenberg, B., & Guy, J. (1995). Prediction of beta from investment fundamentals. *Financial Analysts Journal*, *51*(1), 101–112.

Scarselli, F., & Tsoi, A. C. (1998). Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results. *Neural networks*, *11*(1), 15–37.

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, *45*(1-2), 99–120.

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

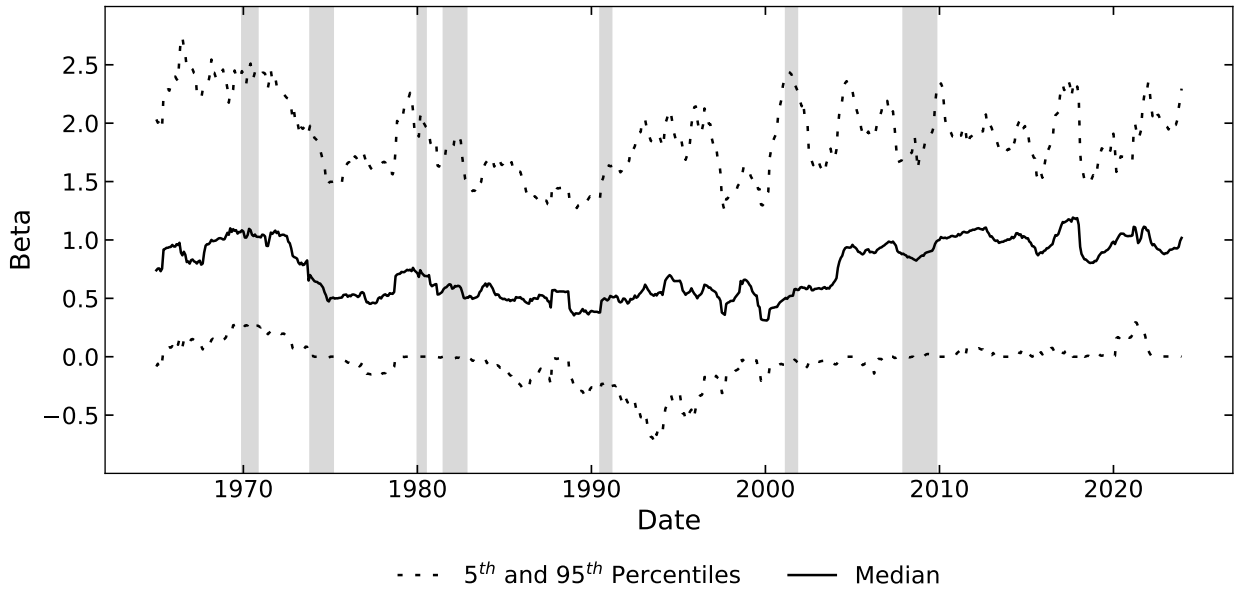
Shumway, T. (1997). The delisting bias in crsp data. *The Journal of Finance*, *52*(1), 327–340.

Van Binsbergen, J. H., & Opp, C. C. (2019). Real anomalies. *The Journal of Finance*, *74*(4), 1659–1706.

Vasicek, O. A. (1973). A note on using cross-sectional information in bayesian estimation of security betas. *The Journal of Finance*, *28*(5), 1233–1239.

Welch, I. (2021). The cost of capital: If not the capm, then what? *Management and Business Review*, *1*(1), 187–194.

(a) Betas Over Time



(b) Cross-Sectional Convergence of Betas

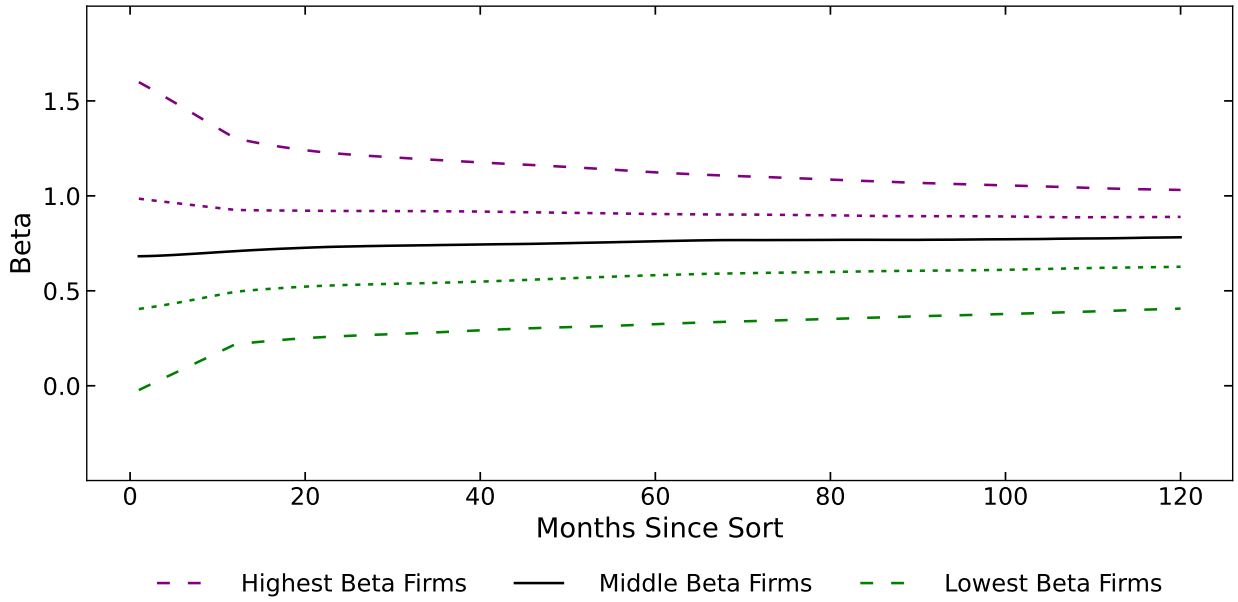
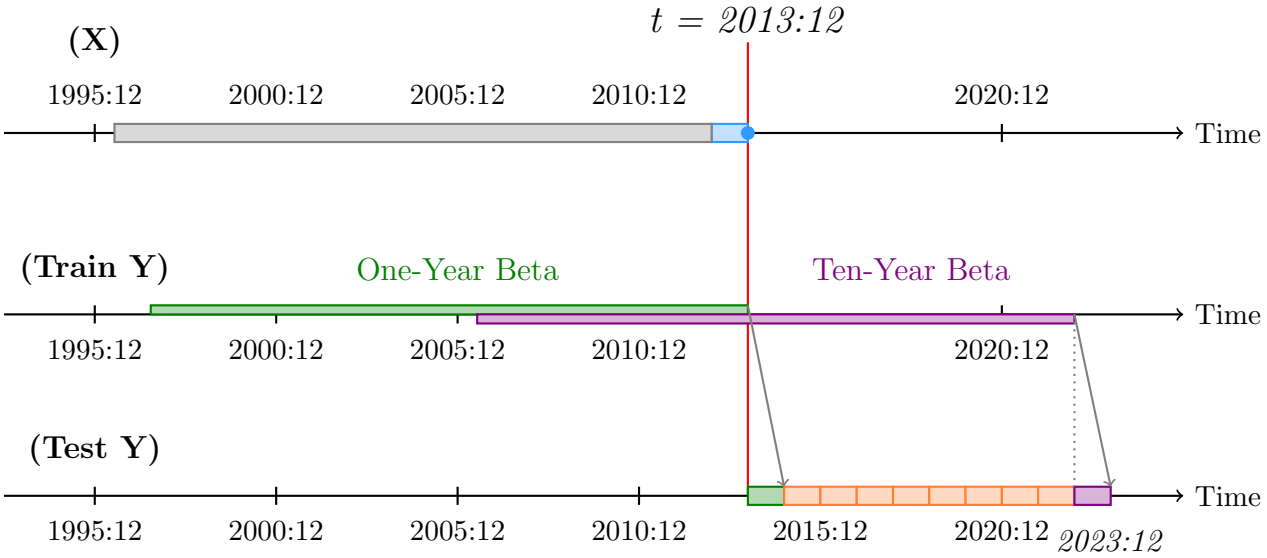


Figure 1: Beta Summary Statistics

This figure illustrates the behavior of equity betas estimated via one-year rolling regressions of daily excess stock returns on excess market returns. Panel (a) displays the median beta (solid line) and the 90% confidence interval (dotted lines) over time. Panel (b) tracks the average beta of quintile portfolios formed monthly based on rolling beta estimates. Portfolio betas are averaged across firms within each quintile and plotted for ten years post-formation.

(a) Train Test Split Example



(b) Example Observation

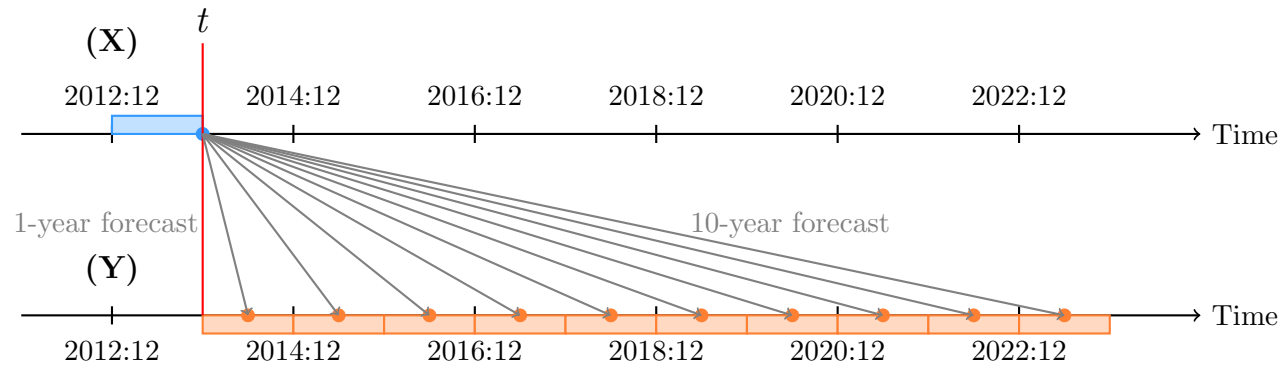


Figure 2: Representative Data Structure
This figure illustrates the methodology for splitting data into training and testing sets and an example of beta estimation along the term structure. Panel (a) depicts the train-test split for the final observation ($t = 2013:12$). Predictors used for training are measured from 1997:02 to 2012:12, while testing predictors are measured from 2013:01 to 2013:12. One-year betas ($\hat{\beta}_{i,t+1\to t+12}$) used as dependent variables in the training set are measured between 1998:02 and 2013:12, and testing betas are measured from 2014:01 to 2014:12. For each subsequent beta along the term structure, measurement years are lagged by one year. Panel (b) provides a detailed example of the measurement periods for the observations at $t = 2013:12$. It depicts the predictor measurement window in blue and the beta estimation windows across the term structure in orange.

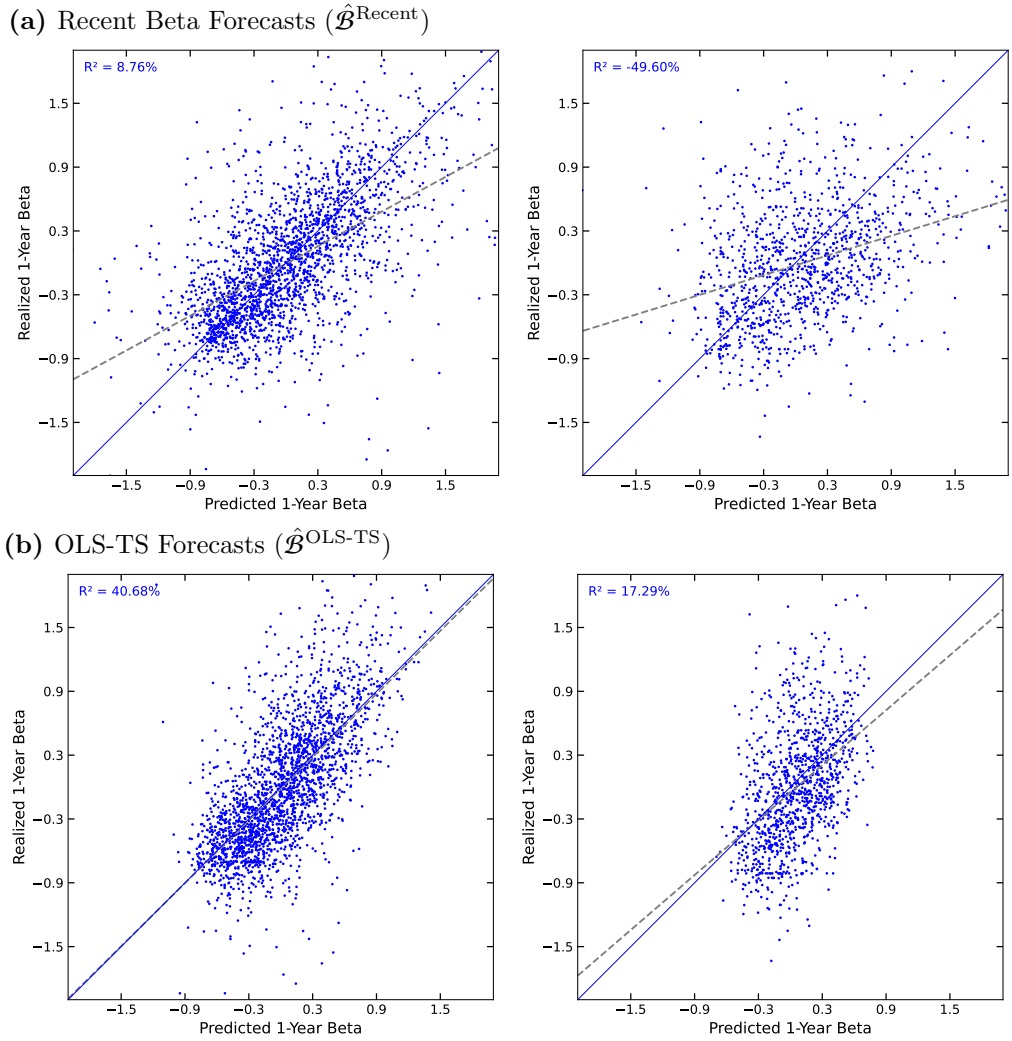
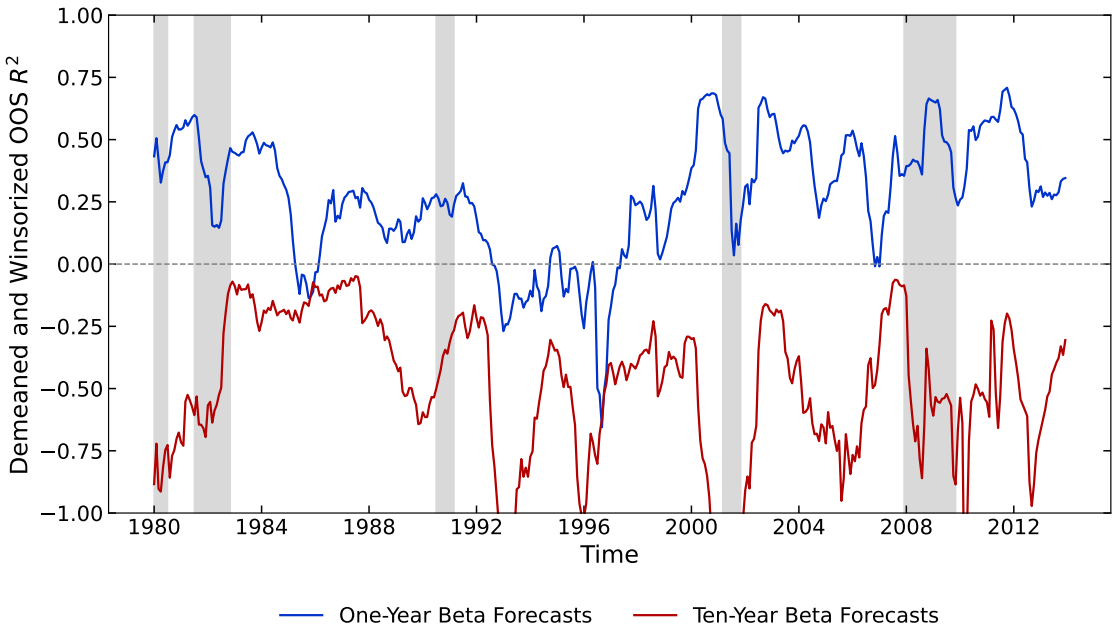


Figure 3: Predicted vs. Estimated Firm Betas

This figure compares predicted beta forecasts to realized beta estimates. The left column shows scatterplots of beta forecasts versus realized beta estimates for the subsequent year, while the right column focuses on realized beta estimates from the tenth year after the forecast. Panel (a) uses the most recently observed betas as forecasts, while panel (b) employs beta estimates derived from the term-structure-adjusted OLS method detailed in Section II.C. The solid blue line represents the 45-degree line with slope = 1 between predicted and realized betas. The dotted gray line depicts a regression line of realized betas on predictions. For clarity, a random sample of 2,500 observations is shown.

(a) Performance of Recent Beta Forecasts



(b) Performance of OLS Beta Forecasts

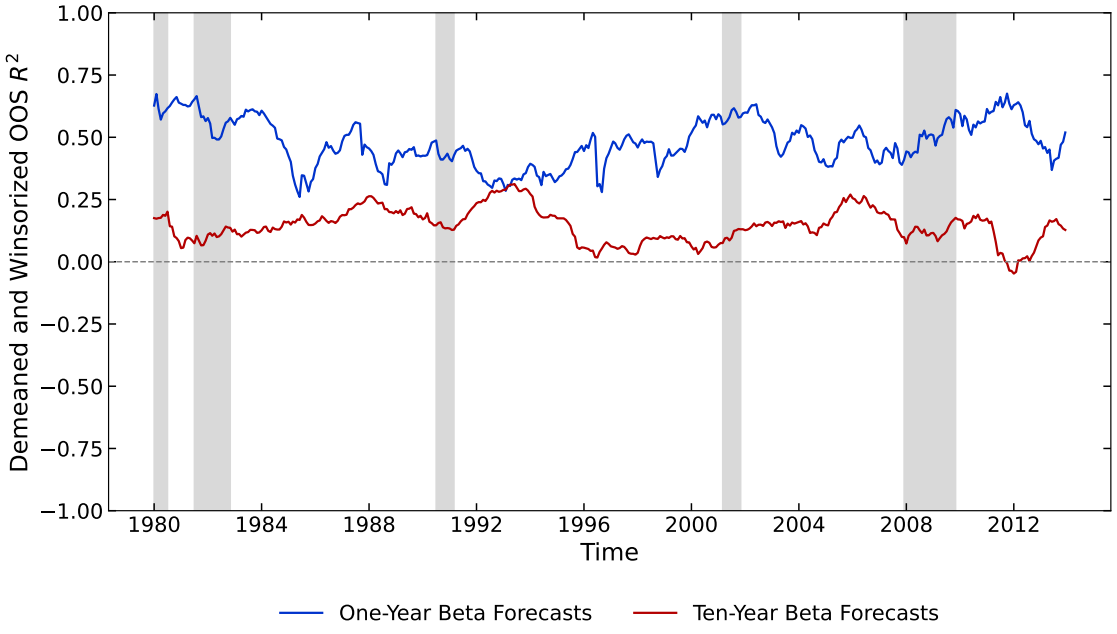


Figure 4: Out-Of-Sample R^2 Over Time
This figure compares the prediction accuracy of one-year and ten-year beta forecasts using out-of-sample (OOS) R^2 values. Monthly OOS R^2 values are calculated following section III.A. Panel (a) presents the performance using recent betas as predictors, while panel (b) shows the performance using term-structure-adjusted OLS forecasts. NBER recessions are shaded in gray in both panels.

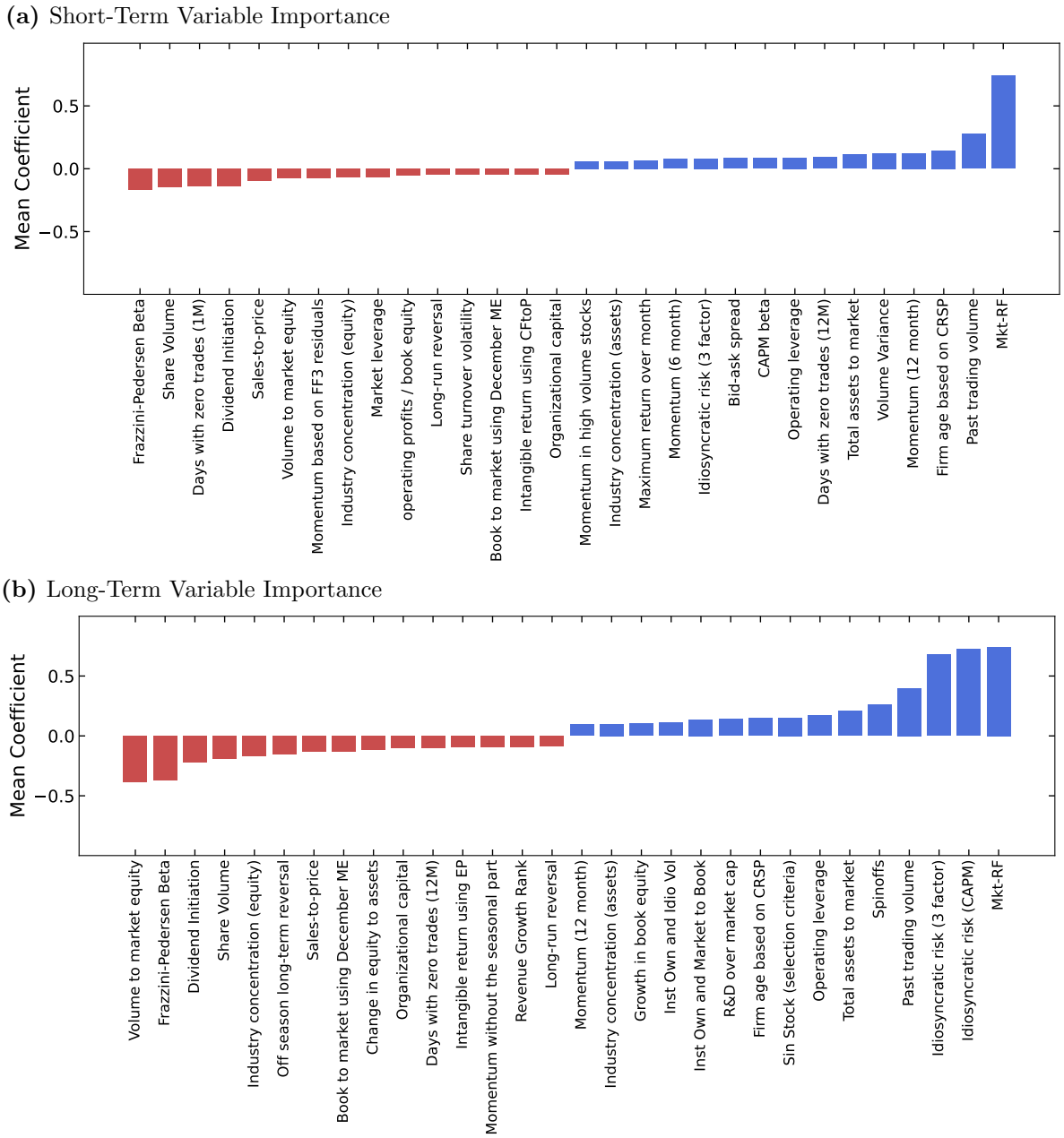


Figure 5: Variable Importances

This figure illustrates the short- and long-term relationships between firm characteristics and future beta estimates. Each month, regressions of future betas on past characteristics are ran as described in section II.C. Time-series averages of regression coefficients are calculated for short-term regressions (panel (a)), where the dependent variable is $\hat{\beta}i, t+1 \rightarrow t+12$, and of long-term regressions (panel (b)), where the dependent variable is $\hat{\beta}i, t+109 \rightarrow t+120$. The 30 strongest predictors of future betas are shown. Positive coefficients plotted in blue and negative coefficients in red.

(a) True R^2 **(b)** Measured R^2 **(c)** Bias in R^2

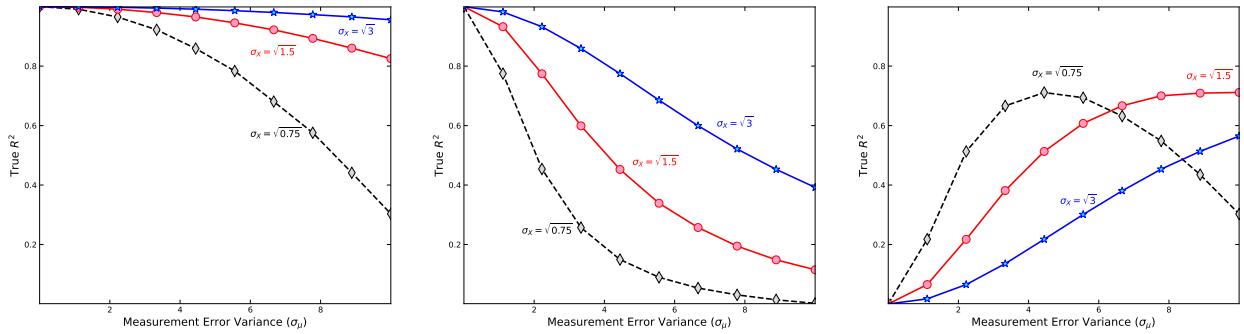


Figure 6: Bias is Performance Measures

This figure shows the results of a simulation analysis that estimates how measurement errors in betas can lead to a downward bias in performance measures. The simulated two-period economy consists of 100 firms, whose betas depend on three firm characteristics: X_i in a predetermined linear fashion:

$$Y_i = 2X_{1,i} - 2X_{2,i} + 0.3X_{3,i}.$$

The characteristics are i.i.d. normal with variance σ_X and are independently drawn for each firm in periods one and two. True betas are unobserved and measured with error μ, which is also normal and has variance σ_μ. In period one, a cross-sectional regression of betas on characteristics is run. Given period-two characteristics, period-two betas are predicted. The cross-sectional R^2 between the true firm betas and predicted betas (True R^2) as well as the measured firm betas and predicted betas (measured R^2) are calculated. This simulation is repeated $10,000$ times. Average results across all simulations are plotted for various levels of σ_μ and σ_X. Panel (a) plots the true R^2 and panel (b) plots the measured R^2. Panel (c) plots the biased R^2, which is defined as the difference between the true and the measured R^2. A positive value in panel (c) indicates that the measured R^2 obtained in prior tests is lower than the R^2 between the population parameters.

(a) R^2 Using Shrinkage Predictions Without OLS

(b) R^2 Using Shrinkage Predictions With OLS

(c) R^2 Improvements With OLS By Shrinkage

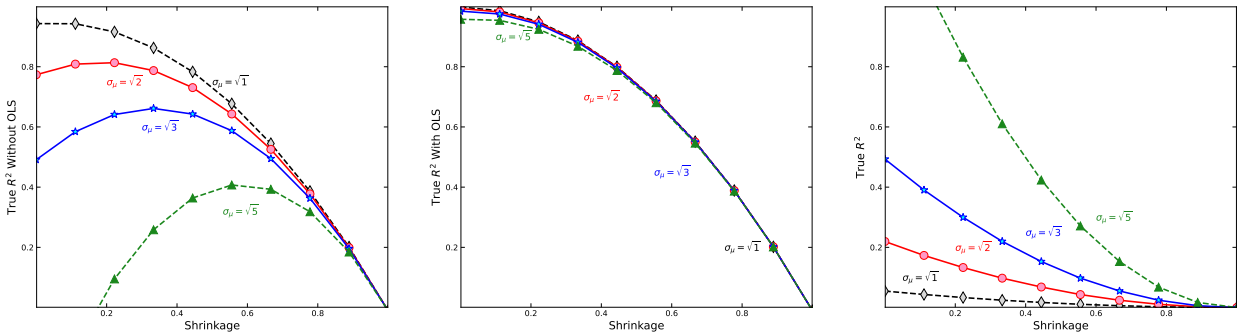


Figure 7: Optimal Shrinkage Given Measurement Error

This figure shows the results of a simulation analysis to estimate the optimal shrinkage level of the OLS-forecasted betas. The simulated two-period economy consists of 100 firms, whose betas depend on three firm characteristics: X_i in a predetermined linear fashion:

$$Y_i = 2X_{1,i} - 2X_{2,i} + 0.3X_{3,i}.$$

The characteristics are i.i.d. normal with variance $\sigma_X = 1.5$ and are independently drawn for each firm. This analysis assumes that betas are constant across the two periods. True betas are unobserved and measured with error μ, which is also normal and has variance σ_μ. In period one, a cross-sectional regression of betas on characteristics is estimated and given period-two characteristics, a beta of each firm is predicted. The cross-section R^2 between the true firm betas and period one measured betas (True R^2 Without OLS) as well as the OLS predicted betas (True R^2 With OLS) are calculated. Betas are multiplied by a shrinkage factor between zero and one to simulate the practice of shrinking betas towards their cross-sectional means. This simulation is repeated $10,000$ times. Average results across all simulations are plotted for various levels of σ_μ and shrinkage. Panel (a) plots the true R^2 using betas measured in period 1. Panel (b) plots the true R^2 using betas predicted using OLS. Panel (c) plots the difference between the observations in panels (a) and (b). A positive value in panel (c) indicates that the OLS predictions outperformed the predictions using period one betas.

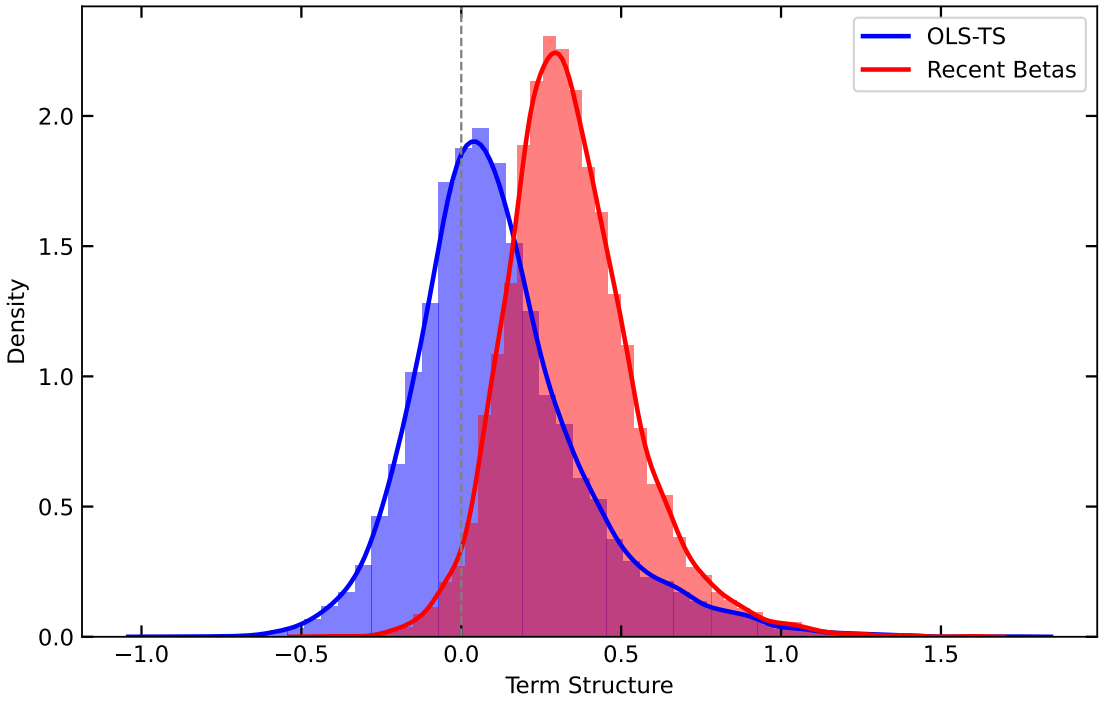


Figure 8: 5 Year Buy-And-Hold Portfolio

This figure shows the distribution of ex-post realized betas for market-neutral hedged portfolios. Each month, 20 portfolios with 15 random firms are formed. The weights in the portfolio are optimized to minimize expected variance subject to being market neutral (i.e., having an expected beta of 0). The market of a monthly rebalanced portfolio are calculated for the following five years. The figure plots the distribution of ex-post portfolio betas for portfolios formed on expected betas based on OLS-TS in blue and for portfolios formed on recent betas in red. More details on the optimization process for the weights are in the Appendix.

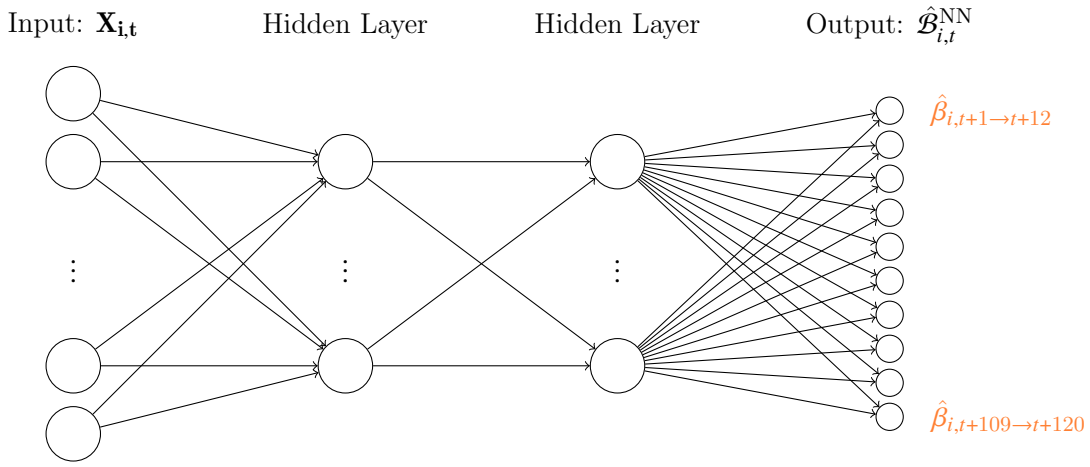


Figure 9: Neural Network Architecture
This figure illustrates the architecture of the multi-output feed-forward neural network (MLP). The network consists of an input layer with 140 nodes, corresponding to the 139 firm characteristics and $\hat{\beta}_{i,t-11\to t}$. The model uses a single hidden layer with 32 nodes, but any number of hidden layers can be employed as depicted in the illustration. The output layer consists of 10 nodes representing the ten dependent variables in the term structure ($\hat{\mathcal{B}}_{i,t}$). Each layer is fully connected and the Adam algorithm is used for weight training. The activation function applied for each node is the sigmoid function.

Table 1: Prediction Performance

This table reports the beta prediction accuracy of the five forecasting models described in sections II. For each model and each prediction horizon, forecasts are evaluated cross-sectionally based on out-of-sample R^2 (panel (a)), median absolute forecast errors (panel (b)), and the percent of observations within 0.3 of the observed future beta estimate (panel (c)) as decried in section III.A. The predictive model used to make the forecasts is denoted in the column *Forecast Type*.

Forecast Type	Years since Prediction									
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year	7 Year	8 Year	9 Year	10 Year
Panel (a): Cross-Sectional Out-Of-Sample R^2										
Recent	17.58	-0.86	-10.77	-19.74	-28.37	-34.77	-40.62	-42.82	-46.49	-49.59
Historical Average	9.44	4.32	-0.21	-4.09	-8.24	-11.76	-14.18	-16.65	-19.16	-21.42
Shrunk	34.71	23.31	17.25	11.6	6.29	2.8	-0.89	-2.84	-5.28	-7.31
OLS Constant	43.16	31.57	24.58	19.09	13.7	9.27	5.46	2.44	-0.42	-2.92
OLS Term Structure	43.16	33.48	28.82	25.2	22.03	20.14	18.49	16.96	15.46	14.42
Panel (b): Median Absolute Error ($\times 100$)										
Recent	26.63	30.2	32.38	34.27	35.76	36.76	37.91	38.68	40.01	40.6
Historical Average	32.28	33.26	34.28	35.5	36.31	36.88	37.44	38.04	38.6	38.93
Shrunk	29.12	31.16	32.27	33.36	33.99	34.35	34.88	35.17	36.03	36.07
OLS Constant	25.47	28.02	29.57	30.95	31.92	32.62	33.29	34.04	34.9	35.36
OLS Term Structure	25.47	27.68	28.81	29.88	30.47	30.81	30.92	31.13	31.56	31.81
Panel (c): Percent Within $\lvert 0.3 \rvert$										
Recent	55.35	50.34	47.68	45.68	44.37	43.38	42.45	41.62	40.5	40.0
Historical Average	47.67	46.68	45.71	44.39	43.65	43.17	42.69	42.11	41.56	41.27
Shrunk	51.66	48.28	46.65	45.33	44.9	44.54	44.07	43.66	42.74	42.8
OLS Constant	57.41	53.41	51.24	49.49	48.34	47.55	46.83	46.09	45.08	44.58
OLS Term Structure	57.41	54.01	52.4	51.01	50.27	49.87	49.94	49.71	49.1	48.86

Table 2: Prediction Performance - Statistical Significance

This table evaluates the statistical significance of the predictive performance of the five forecasting models described in section II. Each month, forecasts are evaluated cross-sectionally based on out-of-sample R^2 (panel (a)), median absolute forecast errors (panel (b)), and the percent of observations within 0.3 of the observed future beta estimate (panel (c)) as described in section III.A. The table reports the statistical significance of the monthly difference between the evaluation metric using the most recent betas as forecasts and the model reported in column *Forecast Type*. Panel (d) compares the term-structure-adjusted OLS forecasts to the constant OLS forecasts. Statistical significance is estimated used Newey and West (1986) standard errors with 12 lags. Each column denotes a separate prediction window.

Forecast Type	Years since Prediction									
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year	7 Year	8 Year	9 Year	10 Year
Panel (a): Cross-Sectional Out-Of-Sample R^2										
Historical Average	-1.93	1.13	2.12	2.69	3.0	3.67	4.28	4.95	5.57	5.69
Shrunk	7.99	8.26	8.28	7.76	8.14	9.0	9.98	12.15	13.19	14.57
OLS Constant	7.74	7.29	7.04	6.66	6.87	7.58	8.17	9.57	10.16	10.54
OLS Term Structure	7.74	7.5	7.43	7.07	7.4	8.45	9.32	11.37	12.4	13.53
Panel (b): Median Absolute Error (×100)										
Historical Average	-6.4	-3.66	-2.19	-1.43	-0.58	-0.12	0.46	0.64	1.36	1.66
Shrunk	-4.62	-1.67	0.18	1.52	3.0	4.17	4.64	4.81	4.87	5.22
OLS Constant	1.67	3.28	4.6	5.73	5.94	5.88	6.08	6.34	6.83	6.96
OLS Term Structure	1.67	3.58	5.45	7.19	7.82	8.43	8.95	9.71	10.2	10.8
Panel (c): Percent Within |0.3|										
Historical Average	-7.1	-4.05	-2.12	-1.51	-0.83	-0.24	0.28	0.58	1.23	1.46
Shrunk	-4.96	-2.57	-1.32	-0.53	0.92	2.21	2.81	3.08	3.02	3.7
OLS Constant	2.21	3.69	4.62	5.58	6.14	6.17	6.79	6.99	7.27	7.89
OLS Term Structure	2.21	4.18	5.54	6.73	7.65	8.76	10.51	11.49	11.97	14.06
Panel (d): Term-Structure-Adjusted OLS vs. OLS-C										
R^2	-	5.35	6.26	6.08	6.4	6.92	7.47	8.48	9.96	12.34
MAE	-	-3.34	-4.75	-5.21	-5.72	-6.4	-7.48	-8.49	-8.97	-7.93
Pct. Within	-	4.34	5.46	5.53	5.91	6.55	8.0	8.87	9.35	8.41

Table 3: Implied long-term discount rates

This table reports the implied discount rates for high- and low-beta firms based on the term-structure-adjusted OLS model. Each month, firms are ranked by their average predicted beta, and the top and bottom decile averages are calculated. These averages are converted into costs of equity using a risk-free rate of 0 and an expected market return of 8%, following the CAPM. After year 10, costs of equity evolve as $\hat{r}_{E,i,t+\tau} = \hat{r}_{E,i,t+\tau-1} \cdot 0.9 + 0.8\%$. Columns (1) and (2) show costs of equity by year, while columns (3) and (4) present geometric average costs of equity from year 1 to τ, where τ is indicated in column *Year t*. We also simulate the growth of a \$1 payout at 1% and 5% growth rates (columns *Cash Flows*), and discount these streams using the derived rates. Using the discounted values, growth rates, and first payments, implied constant discount rates are calculated for both growth scenarios and deciles using the Gordon Growth equation.

Year t	Discount Rates				Cash Flows	
	Year $t-1$ to t		Year 0 to t		Growth rate, g	
	Bottom Decile	Top Decile	Bottom Decile	Top Decile	1%	5%
1	0.36	11.12	0.36	11.12	1.0	1.0
2	0.78	10.55	0.57	10.84	1.01	1.05
3	1.02	10.26	0.72	10.64	1.02	1.1
4	1.22	10.03	0.85	10.49	1.03	1.16
5	1.41	9.81	0.96	10.35	1.04	1.22
6	1.59	9.62	1.06	10.23	1.05	1.28
7	1.74	9.39	1.16	10.11	1.06	1.34
8	1.94	9.28	1.26	10.01	1.07	1.41
9	2.13	9.13	1.35	9.91	1.08	1.48
10	2.35	9.02	1.45	9.82	1.09	1.55
⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	8.0	8.0	6.79	8.28	2.63	113.6
99	8.0	8.0	6.8	8.27	2.65	119.28
100	8.0	8.0	6.82	8.27	2.68	125.24

Firm Valuation and Implied Discount Rates

Growth Rate, g	Firm	Price	Discount Rate
1%	Bottom Decile	28.42	4.55
1%	Top Decile	12.17	9.30
5%	Bottom Decile	75.12	6.40
5%	Top Decile	25.72	9.08

Table 4: Analyst Target Prices Predictions

This table reports out-of-sample R^2 (panel (a)), median absolute forecast errors (panel (b)), and the percent of observations within 25% of the observed future beta estimate (panel (c)) between prices estimated using DCF models with varying discount rates and analyst price targets. For each observation, we compute the ten-year term structure of predicted betas using different forecasting methods and convert them into discount rates with a risk-free rate of 0 and market risk premiums from 8% to 12%. The discounted cash flow model used is:

$$\hat{PV}^M_{i,t} = \left[\sum_{j=1}^{10} \frac{\hat{E}_t\left[CF_{t+12*j}\right]}{\prod_{\tau=1}^{j}(1+\hat{r}^M_{i,t\tau})} \right] + \frac{1}{\prod_{\tau=1}^{10}(1+\hat{r}^M_{i,t,\tau})} \times \frac{\hat{E}_t\left[CF_{t+132}\right]}{r_E - g}. \tag{7}$$

Cash flows are explicitly discounted for ten years, with a terminal value in year ten based on one of three discount rates and long-term growth rates from 0% to 2%. Post-five-year cash flows are imputed using the IBES long-term growth rates.

	Long Term Growth								
	0%			1%			2%		
r_E	6%	8%	10%	6%	8%	10%	6%	8%	10%
Panel (a): R^2									
Recent	5.88	50.61	62.45	-43.31	37.88	59.03	-144.4	14.93	52.8
Historical	12.6	55.02	65.4	-33.28	43.93	62.91	-128.15	23.56	58.02
Shrunk	15.52	56.92	66.51	-28.43	46.87	64.64	-119.52	28.21	60.68
OLS-C	9.71	53.13	64.0	-37.06	41.73	61.39	-133.43	20.93	56.35
OLS	22.17	60.03	67.82	-17.92	51.47	66.61	-101.61	35.23	63.62
Panel (b): Median Absolute Error									
Recent	15.76	11.36	11.35	19.95	12.51	11.24	27.1	14.69	11.55
Historical	16.65	11.67	11.21	21.25	13.11	11.26	28.74	15.65	11.7
Shrunk	16.02	10.96	10.79	20.51	12.29	10.62	27.83	14.75	10.87
OLS-C	16.02	11.33	10.99	20.36	12.57	10.88	27.6	14.78	11.31
OLS	15.44	10.57	10.68	19.81	11.8	10.18	26.99	14.04	10.36
Panel (c): Pct. Within 25%									
Recent	25.22	34.83	33.34	20.77	31.98	34.33	15.4	27.18	33.84
Historical	23.6	32.52	35.5	19.64	29.2	35.08	15.68	25.57	33.19
Shrunk	24.61	34.0	36.13	20.46	31.07	36.8	16.19	26.8	35.01
OLS-C	24.65	34.04	36.06	20.38	31.18	35.37	15.51	26.28	34.22
OLS	25.79	36.53	37.24	20.84	32.93	37.23	15.74	27.9	37.53

Table 5: Share Price Prediction

This table reports out-of-sample R^2 (panel (a)), median absolute forecast errors (panel (b)), and the percent of observations within 25% of the observed future beta estimate (panel (c)) between prices estimated using DCF models with varying discount rates and observed prices from CRSP. For each observation, we compute the ten-year term structure of predicted betas using different forecasting methods and convert them into discount rates with a risk-free rate of 0 and market risk premiums from 8% to 12%. The discounted cash flow model used is:

$$\hat{PV}^M_{i,t} = \left[\sum_{j=1}^{10} \frac{\hat{E}_t\left[CF_{t+12*j}\right]}{\prod_{\tau=1}^{j}(1+\hat{r}^M_{i,t\tau})} \right] + \frac{1}{\prod_{\tau=1}^{10}(1+\hat{r}^M_{i,t,\tau})} \times \frac{\hat{E}_t\left[CF_{t+132}\right]}{r_E - g}. \tag{8}$$

Cash flows are explicitly discounted for ten years, with a terminal value in year ten based on one of three discount rates and long-term growth rates from 0% to 2%. Post-five-year cash flows are imputed using the IBES long-term growth rates.

	Long Term Growth								
	0%			1%			2%		
r_E	6%	8%	10%	6%	8%	10%	6%	8%	10%
Panel (a): R^2									
Recent	2.98	52.14	65.29	-50.2	38.34	61.51	-159.16	13.56	54.71
Historical	-15.01	42.64	60.08	-76.78	25.57	54.86	-201.74	-4.32	46.01
Shrunk	-5.59	48.41	63.59	-62.13	33.85	59.75	-177.12	7.99	52.93
OLS-C	3.23	52.64	65.6	-49.02	39.59	62.36	-156.0	16.09	56.39
OLS	9.98	56.74	68.06	-38.49	45.47	65.78	-138.24	24.84	61.22
Panel (b): Median Absolute Error									
Recent	15.58	9.94	8.57	20.02	11.6	8.91	27.44	14.03	9.66
Historical	17.08	10.88	8.96	22.24	12.84	9.55	30.64	15.87	10.48
Shrunk	16.43	10.23	8.38	21.56	12.04	8.81	29.76	14.94	9.68
OLS-C	16.23	10.28	8.51	21.05	11.91	8.92	28.78	14.61	9.75
OLS	15.74	9.64	7.98	20.58	11.3	8.24	28.33	14.03	8.95
Panel (c): Pct. Within 25%									
Recent	20.46	30.99	33.32	15.9	26.81	32.98	11.86	22.32	31.16
Historical	19.08	27.1	33.64	15.23	23.95	30.14	11.18	20.23	28.12
Shrunk	19.76	29.58	35.84	15.48	25.94	33.39	11.18	21.66	30.67
OLS-C	19.69	29.81	33.83	15.5	26.3	32.94	11.36	21.96	30.78
OLS	20.15	31.79	35.94	15.47	27.87	35.53	11.35	22.46	33.55

Table 6: Expected Betas and Expected Returns

This table presents the results of Fama and MacBeth (1973) regressions of stock returns on predicted stock betas. Each month, we run a cross-sectional regression of excess one-year stock returns on firms' average predicted beta:

$$r_{i,t,\tau} = \alpha_{t,\tau} + \gamma_{t,\tau}\hat{\beta}^{TS}_{i,t-1,1} + \hat{S}^{TS}_{i,t-1,10,1} + \epsilon_{i,t,\tau}.$$

The term structure slope, $\hat{S}^{TS}_{i,t-1,10,1}$, is calculated as the difference between predicted ten-year and one-year betas. The table reports the average coefficient $\overline{\gamma}$, the total number of observations, and the number of months with cross-sectional regressions. Panel (b) provides average portfolio returns following the portfolio sort methodology of Jegadeesh and Titman (2001). Significance levels account for overlapping returns using Newey-West standard errors (Newey and West, 1986).

Panel A: Fama Mac-Beth Regressions

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Mkt-RF	-0.0		-0.18	-0.17		-0.2
	(-1.22)		(1.49)	(-0.56)		(-0.76)
Slope		0.02	1.81	2.0	1.98	1.58
		(1.97)	(-0.54)	(1.69)	(2.03)	(1.44)
SMB				0.09	0.05	0.08
				(1.09)	(0.46)	(0.77)
HML				0.05	0.15	0.11
				(0.82)	(1.85)	(1.54)
CMA					0.03	0.03
					(0.54)	(0.6)
RMW					0.0	-0.02
					(0.01)	(-0.24)
Num. Obs.	1,042,071	1,042,071	1,042,071	1,042,071	1,042,071	1,042,071
R^2 (in %)	2.19	1.14	2.45	3.52	3.72	4.5

Panel B: Portfolio Sorts

	Mean Return						t-statistic					
	Low	2	3	4	High	H-L	Low	2	3	4	High	H-L
Small	1.05	0.97	1.16	1.16	1.4	0.35	2.73	3.23	4.55	5.23	5.91	1.41
2	0.86	1.0	0.96	0.93	1.1	0.24	2.4	3.56	3.97	4.69	5.12	1.05
3	1.05	1.03	1.09	1.14	1.27	0.22	2.89	3.54	4.03	4.54	5.0	1.11
4	0.96	1.12	1.19	1.17	1.52	0.56	2.5	3.44	4.12	3.96	5.01	3.0
Large	0.9	1.33	1.3	1.56	1.77	0.87	2.04	3.32	3.39	4.12	4.4	3.97

Table 7: Machine Learning Performance

This table reports the predictive performance of a plain neural network and the Nelson and Siegel (1987) adjusted neural network. The neural network architecture is described in figure 9 and the Nelson and Siegel (1987) adjustment is described in section VI. For each model and each prediction horizon, the out-of-sample R^2 (R^2), median absolute error (MAE), and percent of beta forecasts within |0.3| ($Pct.Within$) are calculated following section III.A. Each column denotes a separate prediction horizon. Panel (a) reports results for the plain neural network and panel (b) reports results including the adjustment.

Evaluation Type	Years since Prediction									
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year	7 Year	8 Year	9 Year	10 Year
Panel (a): Plain Neural Network Performance										
R^2	41.95	33.44	28.97	25.52	22.59	20.81	19.11	17.27	15.64	14.27
MAE	25.32	27.45	28.66	29.7	30.27	30.65	30.7	31.03	31.49	31.64
PCT. Within	57.67	54.11	52.3	50.87	50.12	49.69	49.71	49.36	48.76	48.64
Panel (b): Nelson-Siegel Adjusted Neural Network										
R^2	47.18	38.11	33.93	31.1	28.48	26.49	24.59	22.51	20.55	18.5
MAE	24.43	27.2	28.44	29.39	29.91	30.26	30.46	30.87	31.55	31.88
PCT. Within	59.02	54.6	52.67	51.43	50.73	50.26	50.05	49.64	48.75	48.36

Appendix

Comparison of Beta Estimation Approaches

The main analysis in this paper relies on ex-post realized beta estimates obtained from one-year rolling regressions. In practice, many alternative methods to estimate betas have also been proposed and are contained in the set of independent variables. In this section, we compare the similarity between the different beta estimation approaches.

All approaches are applied to a uniform dataset that merges daily return data from the Center for Research in Security Prices (CRSP) with factor data from Kenneth French's website. Daily returns and delisting returns from CRSP are compounded if both are available. If either is missing, the return is imputed from the other. Excess daily returns are calculated using the risk-free rate from Kenneth French's data. All betas are computed at the firm-month level. Observations with returns less than negative one are excluded. We now describe the beta estimation methods considered.

Daily One-Year Rolling Regression

Following the main section of the paper, daily one-year rolling regression estimates are formed by estimating the regression parameter $\beta^{1\text{-Yr_Rolling}}_{i,t-11\to t}$ in the regression

$$r^E_{i,t} = \alpha_{i,t-11\to t} + \beta^{1\text{-Yr_Rolling}}_{i,t-11\to t} \times r^E_{Mkt,t} + \epsilon_{i,t}.$$

To ensure a reliable estimate, a firm must have at least 200 non-missing return observations over the prior year. Other look-back windows are also widely used in practice. For robustness, we also estimate betas using three- and five-year windows, requiring at least 600 and 1000 non-missing observations, respectively. These betas are denoted $\beta^{3\text{-Yr_Rolling}}$ and $\beta^{5\text{-Yr_Rolling}}$.

Partial Market Beta

In addition to univariate market betas, we estimate the market beta in regressions that include the Fama and French (2015) five factors

$$r^E_{i,t} = \alpha_{i,t-11\to t} + \beta^{\text{Partial_FF5}}_{i,t-11\to t} \times r^E_{Mkt,t} + \gamma' \text{FF5}_{-Mkt-RF} + \epsilon_{i,t}.$$

Regressions are again ran using a one year rolling window and require at least 200 non-missing return observations. If the factors were entirely orthogonal, the rolling one-year betas would remain unchanged regardless of additional factors. In practice, the market factor is negatively correlated with all four of the other Fama and French (2015) factors.

Frazzini and Pedersen Beta

Frazzini and Pedersen (2014) decompose beta into its three components: one correlation and two standard deviations

$$\beta^{F\&P}_{i,t} = \rho_{i,Mkt} \frac{\sigma_i}{\sigma_{Mkt}}.$$

In addition, they use daily log returns to estimate standard deviations and three-day overlapping log returns for correlations, whereas our regression based implementations use raw daily returns.[20]

Second, Frazzini and Pedersen (2014) use a one-year window to estimate standard deviations and a five-year window to estimate the correlation since correlations appear to be more persistent than volatility. Implementing the longer look-back window into our tests complicates our analysis because betas measured up to five years apart rely on overlapping data. Thus, we estimate two versions of the Frazzini and Pedersen beta. First, we use a one-year look-back window to estimate both the correlations and the standard deviations

$$\beta_{i,t}^{F\&P_1_Yr} = \rho_{i,Mkt,t-11\to t}\frac{\sigma_{i,t-11\to t}}{\sigma_{Mkt,t-11\to t}}.$$

Second, we replicate the original betas using the five-year window to estimate correlations.

$$\beta_{i,t}^{F\&P_5_Yr} = \rho_{i,Mkt,t-59\to t}\frac{\sigma_{i,t-11\to t}}{\sigma_{Mkt,t-11\to t}}.$$

Log Return Regressions

Logarithmic returns are commonly used due to their ability to handle compounding. For returns close to zero, the logarithmic return $log(R) \approx R - 1$ approximates the raw returns well. Given the presence of large outliers in returns, betas obtained from regressions using raw returns and log returns might differ. We thus also estimate log return betas using a regression of the form

$$\log\big(R_{i,t}\big) - \log\big(R_{F,t}\big) = \alpha_{i,t-11\to t} + \beta_{i,t-11\to t}^{1_Yr_Log} \times \Big(\log\big(R_{Mkt,t}\big) - \log\big(R_{F,t}\big)\Big) + \epsilon_{i,t}.$$

Dimson's Beta

Finally, we implement Dimson (1979)'s beta adjustment for infrequent trading. The regression specification is:

$$r_{i,t}^{E} = \alpha_{i,t} + \beta_{i,t,1} r_{Mkt,t}^{E} + \beta_{i,t,2} r_{Mkt,t-1}^{E} + \beta_{i,t,3}\left(\frac{r_{Mkt,t-2}^{E} + r_{Mkt,t-3}^{E} + r_{Mkt,t-4}^{E}}{3}\right) + \epsilon_{i,t},$$

and the rolling beta estimate is:

$$\hat{\beta}_{i,t-11\to t}^{1_Yr_Dimson} = \hat{\beta}_{i,t,1} + \hat{\beta}_{i,t,2} + \hat{\beta}_{i,t,3}.$$

This adjustment accounts for the fact that infrequently traded stocks may not immediately reflect market movements, leading to a downward bias in traditional beta estimates. By including lagged market returns, Dimson's approach captures the delayed response of such stocks to market-wide shocks, providing a more accurate measure of systematic risk.

[20] Excess log returns are calculated as the difference between the natural logarithm of the stock return and the natural logarithm of the risk-free rate: $\log\left(R_{i,t}^{e}\right) \equiv \log\big(R_{i,t}\big) - \log\big(R_{F,t}\big)$.

Comparison of Beta Estimation

Table A.2 provides a comparison of the different beta estimates following Hollstein and Prokopczuk (2016). In panel (a), we report summary statistics for the beta estimates. Summary statistics are calculated at the monthly level and the table reports the mean of the summary statistic time series. Theory predicts that value-weighted average betas must be close to 1. Hollstein and Prokopczuk (2016), however, show that many beta estimation methods (e.g., option implied methods) yield beta estimates that are substantially different from 1. It can be seen that the rolling beta estimates do not suffer from this problem, with the exception of those estimated using Frazzini and Pedersen (2014), which have a mean of 1.07 when calculated using a five-year window for the correlation estimate and of 1.05 using the one-year window.

The value-weighted mean and arithmetic mean differ substantially for rolling regression based approaches (e.g., between 1.00 and 0.82 using the one-year rolling regression), implying that a few large firms have high betas. This result is persistent across all estimation methods, but less pronounced using Frazzini and Pedersen (2014) betas and betas using Dimson's illiquidity adjustment. Finally, the 98 percentile range of betas contains some extreme values of betas below zero and above 3. The minimum and maximum betas are unreported but range from -24 to 51, suggesting that outliers in beta estimates are present.

Panel (b) of the table reports the correlation between the beta estimates. The correlation matrix shows that beta estimates using different approaches can differ substantially, with some correlations being as low as 0.5. The five-year Frazzini and Pedersen, 2014 and the liquidity-adjusted Dimson (1979) betas have the lowest correlation with the standard one-year rolling beta of 0.69. In addition, a firm's beta changes substantially once controlling for other factors. The Fama and French (2015) five factor market beta has a correlation of 0.73 with the one-factor beta.

Robustness in Training Process

This section provides robustness analysis using alternative training methods and adjustments in the beta forecasting methods. First, we apply weighted least squares instead of ordinary least squares in section VII. In section VII, we re-train the models using winsorized betas to control for unreasonable beta estimates and outliers in beta estimates.

Weighted Least Squares

Both the dependent and independent variables are measured with error. Each firm's recent and future betas are estimated and regressions are ran of measured future betas on measured past betas (see equation 1). Since betas are obtained from regressions, the uncertainty (i.e., the standard error) in beta estimates is known. We update the training process by using a weighted least square (WLS) algorithm to estimate betas rather than OLS.

For each observation, the standard error of the beta estimate is obtained. The model of section II is augmented by using WLS to find the beta estimates. The weights are equal to the inverse of the sum of the standard error and a constant $\lambda = 0.01$ to mitigate the exploding weights issue for standard errors close to zero. Stage two in the forecasting process is the same.

We report prediction accuracy of WLS beta forecasts in table A.3. For completeness, the performance metrics using recent betas, historical betas, and shrunk betas are repeated, but are identical to those in table 1. Overall, the results close to those obtained using OLS, which indicates that the noise in the beta estimates did not negatively impact the second-stage estimates. This finding can be explained using the simulation results in figure 6. When provided with enough observations, precise and unbiased estimates of the true parameters can still be obtained.

Winsorizing Betas

In section VII, we mentioned that extreme beta estimates as low as -24 and as high as 51 exist in the data. These betas are almost surely a result of measurement error, rather than true market risk. Outliers influence the main analysis of the paper in two ways. First, they can impact the reliability of beta forecasts by changing the characteristic-beta estimates from equation 3. However, the prior section showed that this is unlikely the case, as the extreme values also would tend to be underweight using WLS. A remaining concern is that beta outliers impact the performance metrics. The cross-sectional R^2, for example, is known to be sensitive to outliers. To address this concern we redo the prediction analysis winsorizing all beta forecasts and estimates using a lower bound of 0.5 and an upper bound of 2.5.

The prediction performance using winsorized betas is shown in table A.4. As expected, the performance metrics are improved relative to those of table 1. For example, the MAE of ten-year beta forecasts using the term-structure-adjusted OLS methodology drops to 0.24 compared to 0.32 and the fraction of ex-post betas within 0.3 of the predicted beta increases from 49% to 62%. However, one needs to exercise caution when interpreting these results. The simulation study in section IV shows that the performance measures become biased downward as measurement error is introduced in the variables. Winsorizing the beta

estimates corrects for some of the measurement error and it is thus expected that the performance metrics improve. In the extreme case of winsorizing the variables completely, average errors would be come 0, for example.

Robustness of %*Within*

The main analysis reports the percent of beta forecasts that fall within a range of 0.3 from the estimated ex-post betas. To provide an idea of the distribution of absolute beta forecasting errors, we also provide the percentages based on cutoffs of 0.1, 0.5, and 1.0. These results are reported in table A.6.

Across all cutoff levels, we find that the term structure specific beta forecasts provide the best long-term forecasts of beta. At the 1-year horizon the recent beta is slightly more frequently within 0.1 of the true future beta than the OLS forecasts, but the trend is opposite at the 0.5 and 1.0 cutoff. Long-term forecasts using the term structure specific OLS predictions always outperform traditional methods. Almost 95% of forecasts are within one unit of the estimated beta. In addition, we find that the OLS forecasts are particularly useful to get precise long-term beta estimates. OLS forecasts outperform the recent beta forecasts by around 18% in the tight range of 0.1 versus by only 8% for the wide range of 1.0.

Total R^2

The R^2 values reported in the paper always refer to cross-sectional R^2 values. This performance metric isolates the true across firm predictive power of the model, but is less frequently used in the asset pricing literature. For comparability with other studies, we thus also report the traditional out-of-sample R^2 calculated as

$$R^{2,\text{Model}}_{OOS,\tau} = 1 - \frac{1}{T}\sum_{t=1980-01}^{2023-12} \frac{\sum_{i=1}^{N}(\hat{\beta}_{i,t+1+12\tau\to t+11+12\tau} - \hat{\beta}^{\text{Model}}_{i,t+1+12\tau\to t+11+12\tau})^2}{\sum_{i=1}^{N}(\hat{\beta}_{i,t+1+12\tau\to t+11+12\tau} - 1)^2}, \quad \tau = 0,\ldots,9.$$

The difference between the cross-sectional and regular R^2 is that the cross-sectional R^2 first de-means both the predicted and true betas. The disadvantage of using the regular R^2 is that a null-model must first be chosen (typically a constant). The null model in the cross-sectional R^2 is simply to predict the mean beta for each firm. Since the market beta and the value-weighted average beta in the data are 1, we use 1 as the null forecasts.

Table A.5 reports the raw R^2 values. The results using this analysis are largely unchanged to the main results in the paper. Recent betas and historical average betas make poor predictors of long-term betas. Short-term beta predictions also perform poorly, albeit with a slightly positive compared to negative R^2 one in table 1. By far the best performing model is still the term structure adjusted OLS forecast.

Bias in Beta Predictions

The conditions under which the OLS forecasts are unbiased follow Roberts and Whited (2013)'s setting including measurement error. The intertemporal nature of the prediction imposes one additional constraint. Let β_{t-1} be the true beta of a firm and let $\mathbf{X}_{t-1}$ represent the state variables (e.g., firm characteristics) that determine beta. Assume that there exists a relationship between β_{t-1} and $\mathbf{X}_{t-1}$ in the population

$$\beta_{t-1} = \mathbf{X}'_{t-1}\gamma_{t-1} + \epsilon.$$

However, a firm's rolling beta $\hat{\beta} = \beta + \mu$ contains measurement error. When regressing $\hat{\beta}_{t-1}$ on $\mathbf{X}_{t-1}$, the estimate $\hat{\gamma}_{t-1}$ takes the well-known form $\hat{\gamma}_{t-1} = (\mathbf{X}'_{t-1}\mathbf{X}_{t-1})^{-1}(\mathbf{X}'_{t-1}\beta^*_{t-1})$. Substitution leads to the two classic orthogonality conditions.

$$\hat{\gamma}_{t-1} = \gamma_{t-1} + (\mathbf{X}'_{t-1}\mathbf{X}_{t-1})^{-1}\mathbf{X}'_{t-1}(\epsilon + \mu).$$

For $\hat{\gamma}_{t-1}$ to be a consistent estimate of γ_{t-1}, the error term and the measurement error must be orthogonal to the independent variables. This condition would be violated in our setting if firm characteristics (e.g., trading volume) are correlated with the measurement error.

The intertemporal structure of the prediction process imposes an additional constraint. The predicted firm beta in period two is obtained as

$$\hat{\beta}_{t-2} = \mathbf{X}'_{t-1}\hat{\gamma}_{t-1}.$$

Thus, even if $\hat{\gamma}_{t-1}$ is unbiased, beta predictions can be biased if $\gamma_{t-1} \neq E(\gamma_{t-2})$. This additional assumption is imposed in any out-of-sample predictive setting and is typically ignored and we will follow this trend in this paper.

Details on Hedging Portfolio Construction

The hedging portfolio is constructed by solving an optimization problem that aims to minimize the expected portfolio variance while ensuring an expected beta of zero. In addition, we enforce the constraints that the weights add to one and that they be in the range $w_i \in [-0.3, 0.3]$, which ensures that no single holding drives the portfolio performance. The full minimization problem is

$$\begin{aligned} \min_{\{\mathbf{w}\}} \quad & \sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}} \\ \text{s.t.} \quad & \mathbf{w}^\top \hat{\boldsymbol{\beta}}_{\boldsymbol{i,t+1 \to t+60}} = 0 \\ & \mathbf{w}^\top \mathbf{1} = 1 \\ & w_i \in [-0.3, 0.3] \quad \forall i. \end{aligned}$$

The covariance matrix $\boldsymbol{\Sigma}$ is estimated using a two-year rolling window with daily returns, $\mathbf{w}$ denotes the portfolio weights and $\hat{\boldsymbol{\beta}}_{\boldsymbol{i,t+1 \to t+60}} = \frac{1}{5} \sum_{\tau=0}^{4} \hat{\boldsymbol{\beta}}_{\boldsymbol{i,t+1+12\tau \to t+12+12\tau}}$ is the vector of expected five-year returns. To solve this optimization problem, we employ the Sequential Least SQuares Programming (SLSQP) method from the `scipy.optimize` library.

Empirical Test of Beta Unbiasedness

We also analyze the relationship between ex-ante predicted betas and ex-post realized betas by running regressions of the form

$$\hat{\beta}_{i,t+1+12\tau\to t+12+12\tau} = \gamma \times \hat{\beta}^{OLS-TS}_{i,t+1+12\tau\to t+12+12\tau} + \alpha_t + \omega_i + \epsilon_{i,t}, \tag{9}$$

where α_t are time fixed effects and ω_i are firm fixed effects. We also test the null hypothesis that $\hat{\beta}_{i,t+1+12\tau\to t+12+12\tau} = 1$ using a t-test.

Regression results are reported in table A.7. In the univariate regressions and the regressions with time fixed effects, there is an almost one-for-one relationship between predicted beta and ex-post realized beta. This relationship flattens over the prediction horizon, with a coefficient of 0.96 for one-year regressions and 0.81 for ten-year regressions.

One possible explanation for the flattening of the line is that high beta firms are more likely to experience financial distress (Campbell, Hilscher, and Szilagyi, 2008). At the ten-year window, these high beta firms have fallen out of the sample, and the regression is estimated only using the surviving firm. In this case, ex-ante beta predictions are unbiased even if they appear biased in the ex-post regressions.

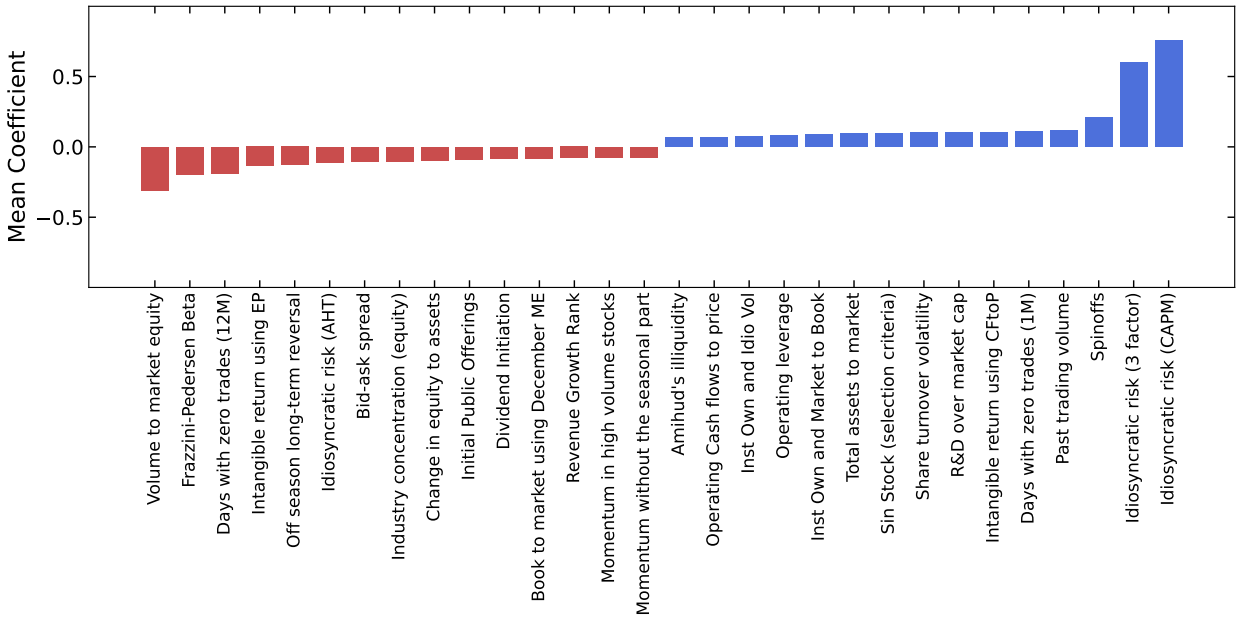


Figure A.1: Variable Importance Changes

This figure visualizes the difference between short- and long-term relationships between firm characteristics and future betas. Each month, regressions of future betas on past characteristics are ran as described in section II.C. Time-series averages of regression coefficients are calculated for short-term regressions, where the dependent variable is $\hat{\beta}i, t+1 \rightarrow t+12$, and of long-term regressions, where the dependent variable is $\hat{\beta}i, t+109 \rightarrow t+120$. For each characteristic, the difference between the average long-term and average short-term coefficient is then calculated. The differences between the long-term and short-term coefficients are reported for the 15 characteristics whose coefficients fell the most (red) and whose coefficient rose the most (blue).

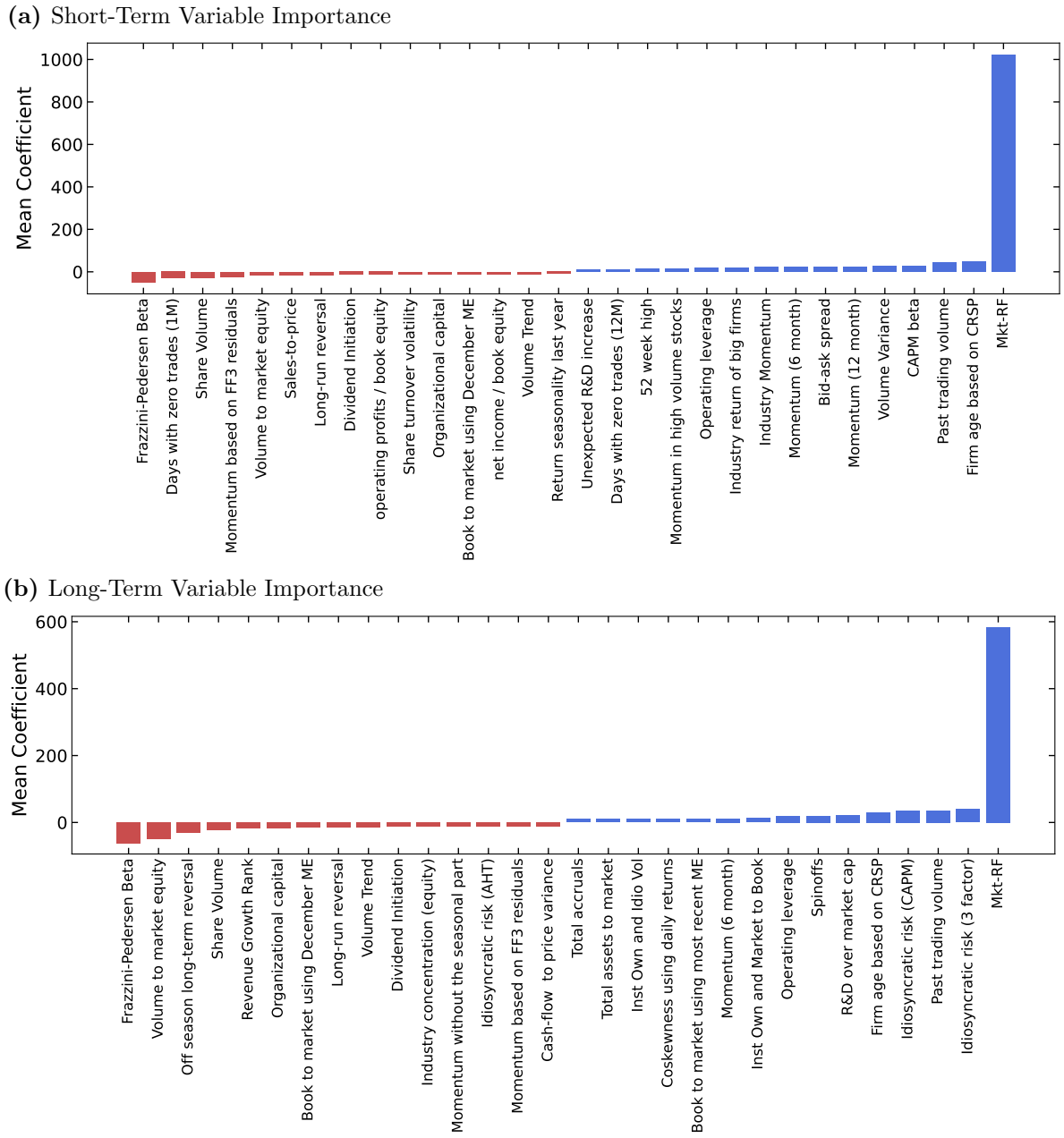


Figure A.2: Variable Importance Significances

This figure illustrates the short- and long-term relationships between firm characteristics and future beta estimates. Each month, regressions of future betas on past characteristics are ran as described in section II.C. Time-series averages of the t-statistics are calculated for short-term regressions (panel (a)), where the dependent variable is $\hat{\beta}i, t+1 \rightarrow t+12$, and of long-term regressions (panel (b)), where the dependent variable is $\hat{\beta}i, t+109 \rightarrow t+120$. The average t-statistic of the 15 strongest negative and positive beta predictors are plotted in red and blue, respectively.

Table A.1: Characteristics This table summarizes the predictor variables from Chen and Zimmermann (2021). Out of the 209 predictors, only 139 with at least two unique values in each month were selected as predictors in $\mathbf{X}_t$.

No.	Acronym	Author	Year	Explanation
1	abnormalaccruals	Xie	2001	Abnormal Accruals
2	accruals	Sloan	1996	Accruals
3	accrualsbm	Bartov and Kim	2004	Book-to-market and accruals
4	activism1	Cremers and Nair	2005	Takeover vulnerability
5	activism2	Cremers and Nair	2005	Active shareholders
6	adexp	Chan, Lakonishok and Sougiannis	2001	Advertising Expense
7	ageipo	Ritter	1991	IPO and age
8	am	Fama and French	1992	Total assets to market
9	analystrevision	Hawkins, Chamberlin, Daniel	1984	EPS forecast revision
10	analystvalue	Frankel and Lee	1998	Analyst Value
11	announcementreturn	Chan, Jegadeesh and Lakonishok	1996	Earnings announcement return
12	aop	Frankel and Lee	1998	Analyst Optimism
13	assetgrowth	Cooper, Gulen and Schill	2008	Asset growth
14	beta	Fama and MacBeth	1973	CAPM beta
15	betafp	Frazzini and Pedersen	2014	Frazzini-Pedersen Beta
16	betaliquidityps	Pastor and Stambaugh	2003	Pastor-Stambaugh liquidity beta
17	betatailrisk	Kelly and Jiang	2014	Tail risk beta
18	betavix	Ang et al.	2006	Systematic volatility
19	bidaskspread	Amihud and Mendelsohn	1986	Bid-ask spread
20	bm	Rosenberg, Reid, and Lanstein	1985	Book to market using most recent ME
21	bmdec	Fama and French	1992	Book to market using December ME
22	bookleverage	Fama and French	1992	Book leverage (annual)
23	bpebm	Penman, Richardson and Tuna	2007	Leverage component of BM
24	brandinvest	Belo, Lin and Vitorino	2014	Brand capital investment
25	cash	Palazzo	2012	Cash to assets
26	cashprod	Chandrashekar and Rao	2009	Cash Productivity
27	cboperprof	Ball et al.	2016	Cash-based operating profitability
28	cf	Lakonishok, Shleifer, Vishny	1994	Cash flow to market
29	cfp	Desai, Rajgopal, Venkatachalam	2004	Operating Cash flows to price
30	changeinrecommendation	Jegadeesh et al.	2004	Change in recommendation
31	chassetturnover	Soliman	2008	Change in Asset Turnover
32	cheq	Lockwood and Prombutr	2010	Growth in book equity
33	chforecastaccrual	Barth and Hutton	2004	Change in Forecast and Accrual
34	chinv	Thomas and Zhang	2002	Inventory Growth
35	chinvia	Abarbanell and Bushee	1998	Change in capital inv (ind adj)
36	chnanalyst	Scherbina	2008	Decline in Analyst Coverage
37	chnncoa	Soliman	2008	Change in Net Noncurrent Op Assets
38	chnwc	Soliman	2008	Change in Net Working Capital
39	chtax	Thomas and Zhang	2011	Change in Taxes
40	citationsrd	Hirschleifer, Hsu and Li	2013	Citations to RD expenses
41	compequiss	Daniel and Titman	2006	Composite equity issuance
42	compositedebtissuance	Lyandres, Sun and Zhang	2008	Composite debt issuance
43	consrecomm	Barber et al.	2002	Consensus Recommendation

Table A.1 Continuation.

No.	Acronym	Author	Year	Explanation
44	convdebt	Valta	2016	Convertible debt indicator
45	coskewacx	Ang, Chen and Xing	2006	Coskewness using daily returns
46	coskewness	Harvey and Siddique	2000	Coskewness
47	cpvolspread			Call minus put volaility
48	credratdg	Dichev and Piotroski	2001	Credit Rating Downgrade
49	customermomentum	Cohen and Frazzini	2008	Customer momentum
50	dcpvolspread			Change in put volatility minus change in call volatility
51	debtissuance	Spiess and Affleck-Graves	1999	Debt Issuance
52	delbreadth	Chen, Hong and Stein	2002	Breadth of ownership
53	delcoa	Richardson et al.	2005	Change in current operating assets
54	delcol	Richardson et al.	2005	Change in current operating liabilities
55	deldrc	Prakash and Sinha	2012	Deferred Revenue
56	delequ	Richardson et al.	2005	Change in equity to assets
57	delfinl	Richardson et al.	2005	Change in financial liabilities
58	dellti	Richardson et al.	2005	Change in long-term investment
59	delnetfin	Richardson et al.	2005	Change in net financial assets
60	divinit	Michaely, Thaler and Womack	1995	Dividend Initiation
61	divomit	Michaely, Thaler and Womack	1995	Dividend Omission
62	divseason	Hartzmark and Salomon	2013	Dividend seasonality
63	divyieldst	Litzenberger and Ramaswamy	1979	Predicted div yield next month
64	dnoa	Hirshleifer, Hou, Teoh, Zhang	2004	change in net operating assets
65	dolvol	Brennan, Chordia, Subrahmanyam	1998	Past trading volume
66	downrecomm	Barber et al.	2002	Down forecast EPS
67	dvolcall			Change in call vol
68	dvolput			change in put vol
69	earningsconsistency	Alwathainani	2009	Earnings consistency
70	earningsforecastdisparity	Da and Warachka	2011	Long-vs-short EPS forecasts
71	earningsstreak	Loh and Warachka	2012	Earnings surprise streak
72	earningssurprise	Foster, Olsen and Shevlin	1984	Earnings Surprise
73	earnsupbig	Hou	2007	Earnings surprise of big firms
74	ebm	Penman, Richardson and Tuna	2007	Enterprise component of BM
75	entmult	Loughran and Wellman	2011	Enterprise Multiple
76	ep	Basu	1977	Earnings-to-Price Ratio
77	equityduration	Dechow, Sloan and Soliman	2004	Equity Duration
78	exchswitch	Dharan and Ikenberry	1995	Exchange Switch
79	exclexp	Doyle, Lundholm and Soliman	2003	Excluded Expenses
80	feps	Cen, Wei, and Zhang	2006	Analyst earnings per share
81	fgr5yrlag	La Porta	1996	Long-term EPS forecast
82	firmage	Barry and Brown	1984	Firm age based on CRSP
83	firmagemom	Zhang	2004	Firm Age - Momentum
84	forecastdispersion	Diether, Malloy and Scherbina	2002	EPS Forecast Dispersion
85	fr	Franzoni and Marin	2006	Pension Funding Status
86	frontier	Nguyen and Swanson	2009	Efficient frontier index

Table A.1 Continuation.

No.	Acronym	Author	Year	Explanation
87	governance	Gompers, Ishii and Metrick	2003	Governance Index
88	gp	Novy-Marx	2013	gross profits / total assets
89	gradexp	Lou	2014	Growth in advertising expenses
90	grcapx	Anderson and Garcia-Feijoo	2006	Change in capex (two years)
91	grcapx3y	Anderson and Garcia-Feijoo	2006	Change in capex (three years)
92	grltnoa	Fairfield, Whisenant and Yohn	2003	Growth in long term operating assets
93	grsaletogrinv	Abarbanell and Bushee	1998	Sales growth over inventory growth
94	grsaletogroverhead	Abarbanell and Bushee	1998	Sales growth over overhead growth
95	herf	Hou and Robinson	2006	Industry concentration (sales)
96	herfasset	Hou and Robinson	2006	Industry concentration (assets)
97	herfbe	Hou and Robinson	2006	Industry concentration (equity)
98	high52	George and Hwang	2004	52 week high
99	hire	Bazdresch, Belo and Lin	2014	Employment growth
100	idiovol3f	Ang et al.	2006	Idiosyncratic risk (3 factor)
101	idiovolaht	Ali, Hwang, and Trombley	2003	Idiosyncratic risk (AHT)
102	illiquidity	Amihud	2002	Amihud's illiquidity
103	indipo	Ritter	1991	Initial Public Offerings
104	indmom	Grinblatt and Moskowitz	1999	Industry Momentum
105	indretbig	Hou	2007	Industry return of big firms
106	intanbm	Daniel and Titman	2006	Intangible return using BM
107	intancfp	Daniel and Titman	2006	Intangible return using CFtoP
108	intanep	Daniel and Titman	2006	Intangible return using EP
109	intansp	Daniel and Titman	2006	Intangible return using Sale2P
110	intmom	Novy-Marx	2012	Intermediate Momentum
111	investment	Titman, Wei and Xie	2004	Investment to revenue
112	investppeinv	Lyandres, Sun and Zhang	2008	change in ppe and inv/assets
113	invgrowth	Belo and Lin	2012	Inventory Growth
114	io_shortinterest	Asquith Pathak and Ritter	2005	Inst own among high short interest
115	iomom_cust	Menzly and Ozbas	2010	Customers momentum
116	iomom_supp	Menzly and Ozbas	2010	Suppliers momentum
117	leverage	Bhandari	1988	Market leverage
118	lrreversal	De Bondt and Thaler	1985	Long-run reversal
119	maxret	Bali, Cakici, and Whitelaw	2010	Maximum return over month
120	meanrankrevgrowth	Lakonishok, Shleifer, Vishny	1994	Revenue Growth Rank
121	mom12m	Jegadeesh and Titman	1993	Momentum (12 month)
122	mom12moffseason	Heston and Sadka	2008	Momentum without the seasonal part
123	mom6m	Jegadeesh and Titman	1993	Momentum (6 month)
124	mom6mjunk	Avramov et al	2007	Junk Stock Momentum
125	momoffseason	Heston and Sadka	2008	Off season long-term reversal
126	momoffseason06yrplus	Heston and Sadka	2008	Off season reversal years 6 to 10
127	momoffseason11yrplus	Heston and Sadka	2008	Off season reversal years 11 to 15
128	momoffseason16yrplus	Heston and Sadka	2008	Off season reversal years 16 to 20
129	momrev	Chan and Ko	2006	Momentum and LT Reversal
130	momseason	Heston and Sadka	2008	Return seasonality years 2 to 5
131	momseason06yrplus	Heston and Sadka	2008	Return seasonality years 6 to 10
132	momseason11yrplus	Heston and Sadka	2008	Return seasonality years 11 to 15
133	momseason16yrplus	Heston and Sadka	2008	Return seasonality years 16 to 20
134	momseasonshort	Heston and Sadka	2008	Return seasonality last year
135	momvol	Lee and Swaminathan	2000	Momentum in high volume stocks
136	mrreversal	De Bondt and Thaler	1985	Medium-run reversal

Table A.1 Continuation.

No.	Acronym	Author	Year	Explanation
137	ms	Mohanram	2005	Mohanram G-score
138	netdebtfinance	Bradshaw, Richardson, Sloan	2006	Net debt financing
139	netdebtprice	Penman, Richardson and Tuna	2007	Net debt to price
140	netequityfinance	Bradshaw, Richardson, Sloan	2006	Net equity financing
141	netpayoutyield	Boudoukh et al.	2007	Net Payout Yield
142	noa	Hirshleifer et al.	2004	Net Operating Assets
143	numearnincrease	Loh and Warachka	2012	Earnings streak length
144	operprof	Fama and French	2006	operating profits / book equity
145	operprofrd	Ball et al.	2016	Operating profitability R&D adjusted
146	opleverage	Novy-Marx	2010	Operating leverage
147	optionvolume1	Johnson and So	2012	Option to stock volume
148	optionvolume2	Johnson and So	2012	Option volume to average
149	orderbacklog	Rajgopal, Shevlin, Venkatachalam	2003	Order backlog
150	orderbacklogchg	Baik and Ahn	2007	Change in order backlog
151	orgcap	Eisfeldt and Papanikolaou	2013	Organizational capital
152	oscore	Dichev	1998	O Score
153	patentsrd	Hirschleifer, Hsu and Li	2013	Patents to RD expenses
154	payoutyield	Boudoukh et al.	2007	Payout Yield
155	pctacc	Hafzalla, Lundholm, Van Winkle	2011	Percent Operating Accruals
156	pcttotacc	Hafzalla, Lundholm, Van Winkle	2011	Percent Total Accruals
157	predictedfe	Frankel and Lee	1998	Predicted Analyst forecast error
158	pricedelayrsq	Hou and Moskowitz	2005	Price delay r square
159	pricedelayslope	Hou and Moskowitz	2005	Price delay coeff
160	pricedelaytstat	Hou and Moskowitz	2005	Price delay SE adjusted
161	probinformedtrading	Easley, Hvidkjaer and O'Hara	2002	Probability of Informed Trading
162	ps	Piotroski	2000	Piotroski F-score
163	rd	Chan, Lakonishok and Sougiannis	2001	R&D over market cap
164	rdability	Cohen, Diether and Malloy	2013	R&D ability
165	rdcap	Li	2011	R&D capital-to-assets
166	rdipo	Gou, Lev and Shi	2006	IPO and no R&D spending
167	rds	Landsman et al.	2011	Real dirty surplus
168	realestate	Tuzel	2010	Real estate holdings
169	realizedvol			Realized vol
170	recomm_shortinterest	Drake, Rees and Swanson	2011	Analyst Recommendations and Short-Interest
171	residualmomentum	Blitz, Huij and Martens	2011	Momentum based on FF3 residuals
172	retconglomerate	Cohen and Lou	2012	Conglomerate return
173	returnskew	Bali, Engle and Murray	2015	Return skewness
174	returnskew3f	Bali, Engle and Murray	2015	Idiosyncratic skewness (3F model)
175	rev6	Chan, Jegadeesh and Lakonishok	1996	Earnings forecast revisions
176	revenuesurprise	Jegadeesh and Livnat	2006	Revenue Surprise
177	rio_disp	Nagel	2005	Inst Own and Forecast Dispersion
178	rio_mb	Nagel	2005	Inst Own and Market to Book
179	rio_turnover	Nagel	2005	Inst Own and Turnover
180	rio_volatility	Nagel	2005	Inst Own and Idio Vol

Table A.1 Continuation.

No.	Acronym	Author	Year	Explanation
181	rivolspread			Realized minus implied vol
182	roaq	Balakrishnan, Bartov and Faurel	2010	Return on assets (qtrly)
183	roe	Haugen and Baker	1996	net income / book equity
184	sfe	Elgers, Lo and Pfeiffer	2001	Earnings Forecast to price
185	shareiss1y	Pontiff and Woodgate	2008	Share issuance (1 year)
186	shareiss5y	Daniel and Titman	2006	Share issuance (5 year)
187	sharerepurchase	Ikenberry, Lakonishok, Vermaelen	1995	Share repurchases
188	sharevol	Datar, Naik and Radcliffe	1998	Share Volume
189	shortinterest	Dechow et al.	2001	Short Interest
190	sinalgo	Hong and Kacperczyk	2009	Sin Stock (selection criteria)
191	skew1	Xing, Zhang and Zhao	2010	Volatility smirk near the money
192	smileslope	Yan	2011	Put volatility minus call volatility
193	sp	Barbee, Mukherji and Raines	1996	Sales-to-price
194	spinoff	Cusatis, Miles and Woolridge	1993	Spinoffs
195	std_turn	Chordia, Subra, Anshuman	2001	Share turnover volatility
196	surpriserd	Eberhart, Maxwell and Siddique	2004	Unexpected R&D increase
197	tang	Hahn and Lee	2009	Tangibility
198	tax	Lev and Nissim	2004	Taxable income to income
199	totalaccruals	Richardson et al.	2005	Total accruals
200	trendfactor	Han, Zhou, Zhu	2016	Trend Factor
201	uprecomm	Barber et al.	2002	Up Forecast
202	varcf	Haugen and Baker	1996	Cash-flow to price variance
203	volmkt	Haugen and Baker	1996	Volume to market equity
204	volsd	Chordia, Subra, Anshuman	2001	Volume Variance
205	volumetrend	Haugen and Baker	1996	Volume Trend
206	xfin	Bradshaw, Richardson, Sloan	2006	Net external financing
207	zerotrade12m	Liu	2006	Days with zero trades
208	zerotrade1m	Liu	2006	Days with zero trades
209	zerotrade6m	Liu	2006	Days with zero trades

Table A.2: Comparison of Beta Estimation Methods

This table compares eight approaches for estimating a market betas: a one-year, three-year, and five-year rolling betas, two versions of the Frazzini and Pedersen (2014) beta, a one-year rolling beta based on log returns rather than raw returns, a partial beta adjusted for the four non-market factors in Fama and French (2015), and a beta employing Dimson (1979) adjustment for infrequent trading. Each month, betas are estimated at the firm-month level. Panel (a) reports summary statistics on the betas and panel (b) presents a correlation matrix. More detailed explanations of the beta estimation methods are provided in section VII.

Panel (a): Summary Statistics

	N. Obs.	Mean_VW	Mean	Std. Dev.	1st Pct.	25th Pct.	Median	75th Pct.	99th Pct.
$\beta^{1_Yr_Rolling}$	3,318,656	1.00	0.82	0.6	-0.38	0.41	0.76	1.17	2.4
$\beta^{F\&P_1_Yr}$	3,318,656	1.05	0.96	0.67	-0.44	0.51	0.91	1.37	2.73
$\beta^{F\&P_5_Yr}$	3,318,656	1.07	1.01	0.6	-0.07	0.59	0.95	1.36	2.69
$\beta^{1_Yr_Log}$	3,318,656	1.00	0.82	0.59	-0.36	0.41	0.76	1.17	2.39
$\beta^{3_Yr_Rolling}$	2,837,098	1.01	0.82	0.5	-0.1	0.46	0.78	1.14	2.14
$\beta^{5_Yr_Rolling}$	2,433,463	1.00	0.82	0.47	-0.02	0.48	0.78	1.12	2.03
$\beta^{Partial_FF5}$	3,318,656	1.00	0.85	0.63	-0.61	0.45	0.84	1.23	2.45
$\beta^{1_Yr_Dimson}$	3,318,656	1.01	1.04	0.8	-0.63	0.53	0.97	1.48	3.25

Panel (b): Correlation Matrix

	$\beta^{1_Yr_Rolling}$	$\beta^{F\&P_1_Yr}$	$\beta^{F\&P_5_Yr}$	$\beta^{1_Yr_Log}$	$\beta^{3_Yr_Rolling}$	$\beta^{5_Yr_Rolling}$	$\beta^{Partial_FF5}$	$\beta^{1_Yr_Dimson}$
$\beta^{1_Yr_Rolling}$	1.0	0.87	0.69	0.99	0.83	0.76	0.73	0.69
$\beta^{F\&P_1_Yr}$	0.87	1.0	0.73	0.88	0.73	0.67	0.66	0.73
$\beta^{F\&P_5_Yr}$	0.69	0.73	1.0	0.69	0.76	0.79	0.53	0.56
$\beta^{1_Yr_Log}$	0.99	0.88	0.69	1.0	0.83	0.76	0.72	0.68
$\beta^{3_Yr_Rolling}$	0.83	0.73	0.76	0.83	1.0	0.94	0.59	0.56
$\beta^{5_Yr_Rolling}$	0.76	0.67	0.79	0.76	0.94	1.0	0.55	0.52
$\beta^{Partial_FF5}$	0.73	0.66	0.53	0.72	0.59	0.55	1.0	0.54
$\beta^{1_Yr_Dimson}$	0.69	0.73	0.56	0.68	0.56	0.52	0.54	1.0

Table A.3: Prediction Performance - Weighted Least Squares

This table compares the predictive performance of the five forecasting models described in sections II.B and II.C. For each model and each prediction horizon, forecasts are evaluated cross-sectionally based on out-of-sample R^2 (panel (a)), median absolute forecast errors (panel (b)), and the percent of observations within 0.3 of the observed future beta estimate (panel (c)) as decried in section III.A. Each column denotes a separate prediction horizon. Panels (a), (b), and (c) report results using the R^2, MAE, and $Pct.Within$ respectively.

Forecast Type	Years since Prediction									
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year	7 Year	8 Year	9 Year	10 Year
Panel (a): Cross-Sectional Out-Of-Sample R^2										
Recent	17.58	-0.86	-10.77	-19.74	-28.37	-34.77	-40.62	-42.82	-46.49	-49.59
Historical Average	9.44	4.32	-0.21	-4.09	-8.24	-11.76	-14.18	-16.65	-19.16	-21.42
Shrunk	20.41	17.26	14.46	11.87	9.48	7.37	5.82	4.17	2.52	1.26
OLS Constant	39.82	26.27	18.55	12.06	5.71	0.81	-3.44	-6.6	-10.02	-12.61
OLS Term Structure	39.82	30.15	26.04	22.34	19.01	17.11	15.77	13.64	11.84	10.75
Panel (b): Median Absolute Error (×100)										
Recent	26.63	30.2	32.38	34.27	35.76	36.76	37.91	38.68	40.01	40.6
Historical Average	32.28	33.26	34.28	35.5	36.31	36.88	37.44	38.04	38.6	38.93
Shrunk	35.05	35.33	35.57	36.17	36.39	36.44	36.5	36.65	36.81	36.72
OLS Constant	24.1	26.92	28.79	30.48	31.62	32.57	33.55	34.36	35.58	36.18
OLS Term Structure	24.1	26.66	28.23	29.63	30.33	30.75	30.96	31.36	32.15	32.3
Panel (c): Percent Within $\|0.3\|$										
Recent	55.35	50.34	47.68	45.68	44.37	43.38	42.45	41.62	40.5	40.0
Historical Average	47.67	46.68	45.71	44.39	43.65	43.17	42.69	42.11	41.56	41.27
Shrunk	43.34	43.18	43.09	42.51	42.42	42.45	42.43	42.3	42.14	42.38
OLS Constant	59.46	54.86	52.19	50.12	48.8	47.71	46.85	45.99	44.75	44.22
OLS Term Structure	59.46	55.76	53.19	51.23	50.25	49.72	49.71	49.2	48.32	48.21

Table A.4: Prediction Performance - Winsorized Betas

This table compares the predictive performance of the five forecasting models described in sections II.B and II.C. For each model and each prediction horizon, forecasts are evaluated cross-sectionally based on out-of-sample R^2 (panel (a)), median absolute forecast errors (panel (b)), and the percent of observations within 0.3 of the observed future beta estimate (panel (c)) as decried in section III.A. Each column denotes a separate prediction horizon. Panels (a), (b), and (c) report results using the R^2, MAE, and $Pct.Within$ respectively.

Forecast Type	Years since Prediction									
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year	7 Year	8 Year	9 Year	10 Year
Panel (a): Cross-Sectional Out-Of-Sample R^2										
Recent	15.49	-6.01	-19.37	-27.86	-37.56	-44.54	-50.75	-51.73	-56.44	-56.56
Historical Average	-28.86	-36.36	-42.84	-47.27	-52.92	-57.77	-59.23	-61.35	-63.65	-64.8
Shrunk	8.7	4.19	0.27	-2.87	-6.12	-9.01	-10.38	-12.11	-13.9	-14.98
OLS Constant	37.96	21.37	11.15	4.16	-3.12	-8.67	-12.83	-15.26	-19.12	-20.15
OLS Term Structure	37.96	25.86	20.43	16.72	13.0	16.95	15.52	14.27	12.87	11.85
Panel (b): Median Absolute Error (×100)										
Recent	12.65	15.57	17.56	19.34	20.86	21.95	23.34	24.44	25.79	26.67
Historical Average	30.25	30.86	31.54	32.43	32.98	33.48	33.98	34.5	35.01	35.36
Shrunk	21.98	22.51	23.05	23.89	24.43	24.92	25.38	25.85	26.26	26.58
OLS Constant	11.03	13.47	15.25	16.93	18.17	19.28	20.53	21.61	22.89	23.83
OLS Term Structure	11.03	13.26	14.81	16.34	17.28	22.65	22.84	23.01	23.45	23.82
Panel (c): Percent Within $\|0.3\|$										
Recent	70.18	65.7	63.18	61.18	59.78	58.65	57.21	56.12	54.81	54.02
Historical Average	50.08	49.4	48.65	47.52	46.97	46.57	46.07	45.5	44.97	44.62
Shrunk	63.56	62.62	61.62	60.22	59.52	58.92	58.24	57.52	56.84	56.37
OLS Constant	73.49	69.48	66.97	64.93	63.66	62.44	61.1	60.01	58.62	57.75
OLS Term Structure	73.49	70.39	68.67	67.04	66.15	63.95	63.68	63.42	62.65	61.95

Table A.5: Prediction Performance - Total R^2

This table compares the predictive performance of the five forecasting models described in sections II.B and II.C. For each model and each prediction horizon, forecasts are evaluated based on the total out-of-sample R^2. The computing of the R^2 is explained in section VII. Each column denotes a separate prediction horizon. Panels (a), (b), and (c) report results using the R^2, MAE, and $Pct.Within$ respectively.

Forecast Type	Years since Prediction									
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year	7 Year	8 Year	9 Year	10 Year
Recent	60.5	54.68	51.3	48.68	46.13	45.46	44.2	45.18	44.81	46.47
Historical Average	58.66	57.02	55.87	54.79	54.13	53.76	54.24	54.2	54.16	54.85
Shrunk	60.6	59.74	59.31	58.86	58.77	58.97	59.71	59.97	60.24	61.35
OLS Constant	72.5	68.29	65.74	64.04	62.33	61.43	60.8	60.43	59.72	60.34
OLS Term Structure	72.5	69.07	67.6	66.76	66.08	66.27	66.6	66.85	66.94	67.79

Table A.6: Prediction Performance - Cutoff Robustness

This table provides robustness analyses on the beta prediction performance tested in table 1. Five beta forecasting approaches are used and described in sections II.B and II.C. For each model and each prediction horizon, forecasts are evaluated cross-sectionally based on the percent of observations within various cutoffs of the observed future beta estimate. In panel (a), the cutoff is 0.1. In panel (b), the cutoff is 0.5, and in panel (c) the cutoff is 1. The calculation of the performance measure is explained in section III.A. Each column denotes a separate prediction horizon. Panels (a), (b), and (c) report results using the R^2, MAE, and $Pct.Within$ respectively.

Forecast Type	Years since Prediction									
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year	7 Year	8 Year	9 Year	10 Year
Panel (a): Percent Within |0.1|										
Recent	22.84	20.11	18.58	17.58	16.88	16.4	15.98	15.58	15.05	14.79
Historical Average	18.06	17.57	17.12	16.49	16.13	15.82	15.58	15.27	14.99	14.85
Shrunk	18.41	17.0	16.37	15.94	15.83	15.57	15.49	15.21	14.81	14.83
OLS Constant	22.24	20.23	19.22	18.35	17.87	17.46	17.23	16.87	16.34	15.98
OLS Term Structure	22.24	20.45	19.56	18.82	18.43	18.12	18.16	18.01	17.68	17.51
Panel (b): Percent Within |0.5|										
Recent	74.36	69.87	67.49	65.35	63.91	62.9	61.75	61.09	59.84	59.6
Historical Average	68.06	66.9	65.89	64.59	63.93	63.48	62.93	62.37	61.86	61.58
Shrunk	75.99	72.48	70.66	68.83	67.88	67.36	66.7	66.3	65.43	65.57
OLS Constant	78.74	75.13	72.92	71.08	69.88	68.97	68.16	67.4	66.55	66.06
OLS Term Structure	78.74	75.43	73.8	72.46	71.77	71.64	71.52	71.38	71.02	70.99
Panel (c): Percent Within |1.0|										
Recent	92.88	91.13	90.1	89.13	88.24	87.74	87.13	86.78	86.28	86.21
Historical Average	91.6	91.09	90.55	89.88	89.56	89.33	89.13	88.87	88.74	88.66
Shrunk	94.68	93.61	93.01	92.39	91.84	91.6	91.3	91.16	90.89	91.09
OLS Constant	95.51	94.29	93.45	92.71	92.22	91.82	91.44	91.06	90.74	90.64
OLS Term Structure	95.51	94.46	93.96	93.45	93.25	93.25	93.19	93.13	93.1	93.18

Table A.7: Bias in Beta Predictions Test

This table presents the results of regressing ex-post realized betas on ex-ante predicted betas. Pooled panel regressions are ran separately for each of the ten year horizons (columns 2 through 11). Coefficients are estimated excluding fixed effects, including time-fixed effects, including firm-fixed effects, and including both time and firm-fixed effects. The regression specification is

$$\hat{\beta}_{i,t+1+12\tau \to t+12+12\tau} = \gamma \times \hat{\beta}^{OLS-TS}_{i,t+1+12\tau \to t+12+12\tau} + \alpha_t + \omega_i + \epsilon_{i,t}, \tag{10}$$

The table reports the slope coefficients (γ) and the associated t-statistic when testing the null hypothesis that the coefficient is equal to one.

Fixed Effects	Years since Prediction									
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year	7 Year	8 Year	9 Year	10 Year
Linear Model	0.96	0.94	0.93	0.93	0.90	0.88	0.87	0.86	0.83	0.81
	(-47)	(-58)	(-63)	(-58)	(-75)	(-84)	(-84)	(-88)	(-97)	(-105)
Time FE	0.94	0.92	0.91	0.90	0.89	0.87	0.86	0.84	0.83	0.82
	(-71)	(-77)	(-76)	(-76)	(-82)	(-85)	(-84)	(-89)	(-90)	(-90)
Firm FE	0.61	0.40	0.31	0.24	0.17	0.11	0.06	0.03	-0.00	-0.02
	(-314)	(-384)	(-395)	(-390)	(-393)	(-394)	(-382)	(-377)	(-368)	(-358)
Time + Firm FE	0.63	0.45	0.36	0.32	0.24	0.18	0.14	0.12	0.07	0.06
	(-304)	(-369)	(-385)	(-373)	(-389)	(-399)	(-391)	(-386)	(-388)	(-384)