Treasury Bond Excess Return Prediction: Machine Learning Insights

Yao Li[a] and Michael Rockinger[b]

December 2025

First Draft. Do not cite. Do not disseminate.

Abstract

We revisit Treasury excess-return predictability using monthly data, a broad macro-financial predictor set, and forecasting models ranging from linear regressions to hybrid machine-learning architectures. A data-driven selection procedure shows that only a few predictors—PMI, trend inflation, uncertainty, momentum, and the yield-curve slope—consistently matter. Predictability is strongly state-dependent: nonlinearities arise in low-PMI periods, where simple neural networks outperform linear benchmarks. A hybrid Elastic Net–MLP model delivers robust out-of-sample gains and economically meaningful improvements in portfolio performance. SHAP analyses reveal rich pairwise nonlinear interactions between PMI and other predictors. Machine learning methods thus provide insights for better investment decision-making than traditional linear methods.

JEL Classification Code: C45, C58, E43, E44, G12, G17

Keywords: Neural Networks, Switching, Real Activity, Treasury Bond Returns

[a]University of Lausanne and Swiss Finance Institute. HEC School of Management, Extranef. CH 1015 Lausanne, Switzerland. E-mail: yao.li@unil.ch.

[b]University of Lausanne and Swiss Finance Institute. HEC School of Management, Extranef. CH 1015 Lausanne, Switzerland. E-mail: michael.rockinger@unil.ch.

Prediction is very difficult, especially about the future.
Niels Bohr

1 Introduction

The predictability of the Treasury term structure is a central issue in fixed-income finance. Accurate forecasts of excess holding period returns (HPR) on government bonds matter for several reasons. First, they underpin portfolio allocation and the design of dynamic trading strategies for long-term investors. The Treasury market is sizable, with outstanding amounts in the trillions of dollars. Second, they are informative about the pricing of interest rate risk and the effectiveness of monetary policy. The behavior of the term structure constitutes a key indicator for the Central Banks. To understand how to forecast the term structure and measure deviations from forecasts can be relevant for a sound monetary governance.

In this paper we wish to contribute to this literature by introducing machine learning (ML) techniques and to further our understanding what could be a best model to predict future excess returns. ML techniques are inherently nonlinear and therefore answer the call by Conley et al. (1997) that nonlinear models should be used to model future returns instead of linear ones.

A large literature has documented that forward rates and yield-curve factors contain information about future bond returns, and has gradually extended the information set to include macroeconomic and financial variables.[1] More recently, Bianchi et al. (2021) and Ioannidis and Ka (2021) have applied regularized regressions and neural networks (NN) to bond return prediction, illustrating the potential of ML in this context. Those papers follow a trend that saw earlier the application of ML technique to stocks. The first to go in this direction are Gu et al. (2020), who introduced models of increasing complexity including the ordinary least squares (OLS); penalized linear regression methods such as LASSO, ridge regression (Ridge), and elastic net (ENet); dimension reduction techniques such as principal component analysis (PCA) and partial least square (PLS); random forests (RF); and feed-

[1]See, among many others, Campbell and Shiller (1991); Cochrane and Piazzesi (2005); Ludvigson and Ng (2007); Gargano et al. (2019); Sarno et al. (2016).

forward neural network (FFN). Later contributions are by Bali et al. (2020) who predict corporate bonds with such a battery of methods. See also Heaton et al. (2017).

However, existing work typically focuses on annual data, a limited set of predictors, or relatively rigid NN architectures. Moreover, the role of the state of the economy, as captured by the real-time business-cycle indicator: Purchasing Managers' Index (PMI), in conditioning the performance of linear versus nonlinear models, remains insufficiently explored.

This paper revisits the predictability of excess Treasury HPR using monthly data and a rich set of predictors that combine yield-curve information, macroeconomic factors, financial-market variables, and measures of uncertainty. Our starting point is the framework of Bianchi et al. (2021), who work with annual returns and McCracken and Ng (2016) macro factors, and the subsequent debate about the identification and robustness of such models at low frequencies (Bauer and Hamilton, 2018). Following Gargano et al. (2019), we move to a monthly frequency and broaden the predictor set in several dimensions. In addition to forward rates and term-structure factors, we incorporate the macroeconomic principal components from McCracken and Ng (2016), the stock-market predictors (Goyal et al., 2024), and recent variables that have attracted attention in the literature, such as the U.S. PMI index (Andreasen et al., 2021), trend inflation (Cieslak and Povala, 2015), and macro- and financial-uncertainty indices (Ludvigson et al., 2021).

Our empirical strategy proceeds in two steps. First, we search for a parsimonious yet informative subset of predictors. A large-scale exercise based on principal components and machine-learning factor extraction indicates that using all non-yield predictors mechanically would feed very noisy inputs into the NN. We therefore performed a preliminary exploratory analysis by regressing average excess returns on the full predictor set and retained variables that consistently exhibit strong predictive power. This procedure leads to a compact set of five key non-yield predictors: stock-market momentum (MOM), financial uncertainty (FINUNC), the U.S. PMI index (PMI), the trend inflation (CPITR), and a slope factor derived from the yield curve (YC2). In line with Boivin and Ng (2006), we confirm that "more is not necessarily better": the quality of predictors dominates sheer quantity. These retained data series have the further advantage that they are not revised. For this reason, our research is robust to the criticism that if one had used real-time data, then one would not find predictability, see Ghysels et al. (2018); Wan et al. (2022).

Second, we compare a sequence of forecasting models that gradually increase in complexity, ranging from linear models to nonlinear neural networks. Among linear models, we consider OLS and standard regularized regressions (LASSO, Ridge, and Elastic Net). Among nonlinear models, we start with simple shallow Multilayer Perceptrons (MLP) with a single hidden layer and a few neurons, and then design hybrid architectures and regime-dependent models. Our benchmark NN design philosophy is deliberately conservative: instead of relying on very deep or opaque architectures, we favor simple models whose predictive gains can be interpreted and related to the economic environment. Preliminary experiments with more complex NN structures (multiple layers, larger hidden sizes, and various activation functions) also suggest that such complexity does not improve out-of-sample performance in a robust way. Ockham's razor applies here: simple NN, appropriately structured, tends to perform best.

A key empirical observation is that the predictive relationship between our predictors and excess Treasury HPR is highly state-dependent. Consistent with Gargano et al. (2019) and Andreasen et al. (2021), we find that bond return predictability is stronger in recessions or low-PMI states than in normal or boom periods. Ramsey RESET tests (Ramsey, 1969) applied to Elastic Net residuals indicate pronounced nonlinearities during normal- and low-PMI periods, whereas the linear specification appears reasonably well specified during high-PMI period. Pure MLP models, while not dominant overall, perform comparatively well in low-PMI periods, precisely where the linear model is most severely misspecified.

Motivated by these findings, we propose a family of hybrid linear–nonlinear models in which a neural network is used as a targeted correction to a strong linear backbone. Our main specification is an Elastic Net–MLP (ENet–MLP) hybrid: the Elastic Net provides the baseline linear forecast, and a shallow MLP learns a nonlinear correction to the residuals.The idea to predict residuals and to combine the forecasts of the residuals with the forecasts of the backbone is called in the ML literature boosting. Traditionally boosting is used in the context of tree based methods. We transport the idea to combinations of ENet and MLP.

We further exploit the information in PMI in two ways, since PMI data is available around the first and third business days of the month and is therefore valid for forecast excess Treasury HPR. First, we consider soft-switching models in which two independent MLPs produce alternative nonlinear forecasts, and PMI feeds into a third MLP that assigns state-dependent weights to these nonlinear forecasts. Second, we analyze hard-switch models that use the Elastic Net prediction in high-PMI (boom) periods and switch to the ENet–MLP hybrid model in normal- or low-PMI states. Finally, we compare these regime-switching frameworks to models that include PMI directly as a predictor, allowing the network to learn smooth nonlinear interactions between PMI and other variables.

Our main results can be summarized as follows.

1. Among linear models, regularized regressions substantially outperform OLS out of sample, with Elastic Net emerging as the best linear benchmark across maturities and predictor sets.

2. Simple MLPs alone yield limited gains overall, but their performance improves markedly in normal- and low-PMI periods, in line with the diagnostic test (Ramsey RESET test) that points to nonlinear misspecification in these regimes.

3. The ENet–MLP hybrid model delivers robust out-of-sample improvements in R^2_{OOS}, especially for short- and medium-maturity excess returns and during normal- and low-PMI periods. These gains are statistically significant when evaluated against the historical-average benchmark using the Clark and West (2007) test.

4. Using PMI purely as a switching variable is beneficial, but including PMI directly as a predictor in nonlinear models yields even stronger predictive performance. This suggests that PMI conveys richer information than a coarse classification into "recession" and "boom" states.

5. When translated into economic terms via certainty-equivalent returns and Sharpe ra-

tios under both power-utility and mean-variance preferences, the improvements from nonlinear and hybrid models are economically meaningful, particularly for longer-maturity Treasuries that are most sensitive to the business cycle.

Beyond forecasting performance, we also seek to interpret the mechanisms learned by the best-performing model. To this end, we use Shapley value (SHAP) decompositions to assess the time-varying importance of individual predictors and to quantify how their marginal effects interact with the state of the economy. We show that trend inflation (CPITR) consistently plays a dominant role, especially for long-maturity excess returns, while the importance of financial uncertainty and momentum has increased since the early 2000s. PMI and the yield-curve slope factor YC2 remain influential throughout the sample, reflecting their role as business-cycle indicators. Cross-elasticity measures further highlight rich nonlinear interactions between PMI and inflation, uncertainty, slope of yield curve, as well as momentum that would be difficult to be captured in standard linear models.

The paper is organized as follows. In Section 2, we follow Gu et al. (2020) and investigate the performance with OLS and regularized methods before switching to our ML methods. In Section 3, we discuss how we evaluate the models. In Section 4, we introduce the data. The paper culminates with Section 5 where we discuss the results of our estimation, the economic value of the model, and how ML techniques can lead to the detection of complex nonlinear patterns that traditional methods do not detect. Section 6 concludes with a summary of the main findings and their implications for Treasury return predictability.

2 Theoretical Models

We start by introducing some notations. All financial variables are assumed to be annualized. The basic ingredient is $y_t^{(n)}$, the continuous compounded yield (log-yield) at time t of a pure discount bond paying 1 unit of currency at time $t + n$. In Appendix A, we remind more formally how this yield is related to bond prices and show that the excess holding period return (henceforth HPR) between times t and $t + h$ of a pure discount bond, maturing n

time units after, and denoted by $xr^{(n)}_{t,t+h}$ is:

$$xr^{(n)}_{t,t+h} = \frac{n-h}{h}\left[y^{(n)}_t - y^{(n-h)}_{t+h}\right] + y^{(n)}_t - y^{(h)}_t.$$

The objective of this research is to predict the future HPRs with a state of the art methodology. Thus, we seek an estimate of $\mathbb{E}[xr^{(n)}_{t,t+h}|F_t]$, where F_t is the set of all available information up to time t. In special cases, a conditional expectation will be linear. More generally, we expect a nonlinear relation so that

$$xr^{(n)}_{t,t+h} = \Phi(X_t, t) + u_t,$$

where X_t, the vector of explanatory variables, is in the information set at time t. We assume that u_t is some random noise orthogonal to the space generated by the X_{t-1}. Φ is a potentially nonlinear function which can also depend on time.

In practice, we will consider non-overlapping monthly data, say from 0 to h, from h to $2h$, etc. To simplify the notations for the empirical work, it is convenient to change notation and denote by t the index. We will drop t since we will use the same time horizon all over.

2.1 Linear Models

For estimation purposes, it is convenient to gather HPRs of bonds with different maturities, $n = 1, \cdots, N$ into an $1 \times N$ row vector, $y_t = [xr^{(1)}_{t+h}, \cdots, xr^{(N)}_{t+h}]$. It is also convenient to stack all those vectors y_t, for $t = 1, \cdots, T$ into a matrix Y. Similarly, we include K explanatory variables in a row vector, x_t, and this for $t = 1, \cdots, T$. All the x_t are assumed to be temporally aligned with y_t and stacked in another matrix $\tilde{X}$. Formally

$$\underset{T\times M}{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \quad \underset{T\times K}{\tilde{X}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix}.$$

We can now turn to the estimation methods, increasing the complexity, ranging from the linear to the nonlinear.

2.1.1 Ordinary least squares

The OLS predictors can all be estimated simultaneously. Define $X = [\mathbf{e}, \tilde{X}]$ the matrix with a constant vector of ones, $\mathbf{e}$ in the first column. Set $Y = XB + U$, where B is a matrix of constant parameters. OLS seeks to minimize the sum of squared column-wise errors. The textbook estimator of B is $\hat{B} = (X'X)X'Y$.

The $T \times N$ matrix of residuals U is defined as $\hat{U} = Y - XB$. Those residuals can be further examined for the goodness of (in sample) fit. Given additional information, presented as a $1 \times K$ vector x_{t+h}, it is possible to predict the vector of future excess returns y_{t+h} using $\hat{y}_{t+h} = [1, x_{t+h}]\,\hat{B}$. The estimation that we just described will be performed on an increasing sample. We keep a minimal set of data to initialize the estimations and then add in each step one month.

2.1.2 Regularized regression models

While the Ordinary Least Squares (OLS) estimator is well-known to be the Best Linear Unbiased Estimator (BLUE) under the Gauss-Markov assumptions, it is also known to suffer from high variance, particularly in settings with a large number of predictors or significant multicollinearity, Hastie et al. (2009). This high variance often leads to poor predictive performance on new data, a key motivation for the use of regularized methods like Ridge Regression or the LASSO regression. In the context of predicting excess treasury HPR, a discussion of such models can be found in Bianchi et al. (2021) or Kynigakis and Panopoulou (2022).

The purpose of regularization is to reduce the number of parameters and to shrink the parameters. Whereas the OLS problem consisted a set of estimates which minimizes the sum of squared residuals, the regularized problem introduces a penalty function denoted $P(B)$

so that the error minimization becomes

$$\text{argmin}_B \, ||Y - \tilde{X}B||_2 + P(B)$$
$$P(B) = \lambda \sum_{b \in B^*} \left[\alpha |b| + (1-\alpha) b^2 \right].$$

Since the constant in B centers the variables, normally, those elements are not included in the penalty. So, B^* is the matrix of parameters except for those on the constant (in the multivariate case this is B without the first row).

The parameter λ is a tuning parameter and will indicate how binding the penalty should be. The α is a parameter that controls the trade off between the L_1 and L_2 norms. When $\alpha = 0$, the only term that matters in the penalty is the L_2 norm. In this case, one shrinks the parameters, and one obtains the ridge regression of Hoerl and Kennard (1970). This method is known to stabilize the estimation.

When $\alpha = 1$ then the emphasis is on the L_1 norm of the parameters. This yields the least absolute shrinkage and selection LASSO penalty which was initially introduced by Tibshirani (1997). This method is interesting because it forces some of the parameters to zero. By doing so, one eliminates multicollinearity which further stabilizes the estimates.

Based on those insights Zou and Hastie (2005) combine both the L_1 and L_2 component by taking intermediate values of $\alpha \in (0, 1)$. The resulting estimation procedure is named elastic net. It encourages a grouping effect, where strongly correlated predictors tend to be in or out of the model together. The elastic net is particularly useful when the number of predictors is much bigger than the number of observations where the LASSO is not a very satisfactory variable selection method.

In this paper, all estimations are based on increased windows. The first set of models are estimated in December 1989. Subsequently, parameters are re-estimated every 2 years using an expanding window that incorporates all the available data up to each re-estimation date. There exists an academic discussion of which method is better. Stock and Watson (2003) and Pesaran and Timmermann (2007). This discussion is important if one suspects structural breaks. Our choice of increasing window is driven by the belief that if there is a structural break in the data, somehow the model should capture this break. This is the reason why we allow for switching between components. The second is of practical nature: estimation of nonlinear models requires a relatively large sample size. The choice of a rolling window might lead to the spurious finding that there is no nonlinearity when there is one because of a lack of observations.

2.2 nonlinear Models

Our quest for an optimal linear model was driven by several principles. First, the explanatory variables should include all potentially interesting variables that have been used in the recent academic literature. Second, it should involve NN architectures that are not just out of the box but potentially new. Third, it should be possible to interpret the findings that result from the estimation of the NN.

Concerning the first point, it is well known, see Gargano et al. (2019) that prediction of excess returns is better in recession periods than during booms. This has led Andersen et al. (2024) to estimate an OLS model where during recessions a dummy variable enters the equation.Their switching is induced not by a latent variable model in the spirit of a hidden Markov chain but by directly conditioning on the PMI index which is a real-time variable. We will further address the state of the literature when we discuss the choices of explanatory variables.

Concerning the second point, we started by using simple NN varying the number of neurons and the number of layers in those NN from 1 to 3 with various architectures such as, say [8-4-8] meaning a first layer with 8 neurons, a second one with 4 and a third one with 8 neurons. We also experimented with inverted structures such as [32-16-8] and this for different numbers of neurons. We also experimented with the choice of activation function (Tanh vs ReLu) etc. Those experiments led us to conclude that simpler models are better: Ockham's razor also applies here. For this reason, we focused mostly on single layer NN with few neurons and enriched the simple models by adding more structure to them.

Our choice of model is therefore based on a significant amount of preliminary experiments with different architectures.

As we will show later, in Table 1, the Ramsey RESET Test results show that the Elastic Net residuals exhibit nonlinearity, especially during the normal- and low-PMI periods. This implies that the linear regression used to predict excess treasury HPR is misspecified. To address this misspecification, we further explore whether a neural network, which is capable of learning flexible nonlinear interactions, can improve the predictive power. In this section, we first test a basic MLP model. We find that although the overall performance is poor, the MLP performs well during the low PMI period. This pattern is consistent with the results of the Ramsey RESET Test. Therefore, we go a step further and build a linear-plus-nonlinear hybrid model, where a neural network is used only to correct the residuals from the linear Elastic Net model. In addition, we have two in-sample training strategies. First, the in-sample validation uses the full validation set. After this initial assessment, we refine the model by performing in-sample validation only on the subset of validation data corresponding to the normal- and low-PMI periods.

2.2.1 Multilayer Perceptron models

We use standard Multilayer Perceptron (MLP) models, denoted as $NN^{[i]}$ for $i \in \{3, 5, 7\}$, to test whether a simple nonlinear model can improve the prediction of excess Treasury HPR.[2] Specifically, we use a model with one hidden layer, and we experiment with three different sizes for the hidden layer: 3, 5, and 7 neurons. We keep the network shallow to explore the contribution of basic nonlinearity rather than the nonlinear model complexity.

All models use the Tanh activation function in the hidden layer to introduce the nonlinearity. To prevent overfitting, we apply dropout after the hidden layer. The network is trained using conventional backpropagation with AdamW optimizer. Parameters are updated based on minimizing the in-sample mean squared errors (MSE) on the training dataset, and are tested on the in-sample validation dataset. We also apply L1 and L2 regulariza-

[2] We denote the number of neurons or an architecture of neurons within square brackets. [3] is a single layer NN with 3 neurons. $[16-4]$ would be a two layer NN with 16 neurons in the first layer and 4 in the second layer. We denote with $[3, 5, 7]$ when we estimate alternatively NN with either 3, 5, or 7 neurons in a single layer.

tion to the MLP parameters, which is inspired by the multicollinearity of predictors and the outperformance of regularized linear models. Empirically, applying regularization on MLP parameters helps to improve significantly the out-of-sample prediction performance. The validation losses are used to monitor the learning process. If the validation loss is not updated after a patience of 50 iterations, training is terminated by early stopping. We also employ a learning rate scheduler based on the validation loss. When the validation loss plateaus after 10 iterations, the scheduler will reduce the learning rate by 30% to help the model approach the local optimum.

To ensure that the predictive results of the MLP models are not driven by random initialization, at each estimation window, each MLP is trained under 100 distinct random seeds. This choice reflects two considerations. First, neural networks used here are shallow, which makes the computational cost of multiple restarts negligible relative to the benefits of obtaining stable results. Second, preliminary experiments revealed variation in validation performance across initializations. Using multiple seeds mitigates this instability and reduces the variance of the final forecasts. After training MLP models independently with 100 random seeds, we save the best-performing 10 models according to a moving-average validation loss. The moving-average criterion assigns equal weight to the validation loss from the current estimation window and from the previous window. This smoothing prevents overreacting to temporary fluctuations in the validation loss. The predicted excess treasury HPR, following the next 2 years of each estimation window, is the exponentially weighted average of forecasts from these top 10 models.

2.2.2 Soft-switch on MLPs

As the empirical results show that a single MLP model doesn't improve predictive power overall but shows improvements during the low-PMI period, which is also consistent with the Ramsey RESET test, we propose a soft switching model (we denote those by $NN_{ss}^{[i]}$ for $i = 3, 5, 7$) between two independent nonlinear MLPs, with the weighting for each MLP's forecasts given by the current PMI level. As shown in Figure 1, the inputs $\tilde{X}_t \in \{\mathbb{R}^{10}, \mathbb{R}^{14}, \mathbb{R}^{15}\}$ are passed simultaneously into two independent MLPs to give two forecasts $x_{t+h}^{(n),\{1\}}$ and

$x_{t+h}^{(n),\{2\}}$. The predictor vector $\tilde{X}_t^i$ belongs to one of three sets: $\mathbb{R}^{10}$, containing the ten forward rates $\{f_t^{(j)}\}_{j=1}^{10}$; $\mathbb{R}^{14}$, augmenting these forward rates with four additional predictors:MOM, FINUNC, CPITR, YC2; or $\mathbb{R}^{15}$, further including the PMI. The current PMI level is passed to a third MLP to give weights ($w_{t+h}^{\{1\}}$ and $w_{t+h}^{\{2\}}$) for each forecast. The final forecast $x_{t+h}^{(n)}$ is a weighted combination of the two nonlinear predictions. The training strategy for the MLPs is the same as described in Section 2.2.1.

2.2.3 Elastic Net-MLP hybrid model

The Ramsey RESET test results indicate that the Elastic Net residuals exhibit significant nonlinearity, especially during the normal- and low-PMI periods. However, the pure MLP models show weak overall performance and only deliver improvements during low-PMI periods, precisely the periods where the Ramsey RESET test indicates stronger misspecification. This pattern suggests that a pure nonlinear model is unnecessary, but that a targeted nonlinear adjustment could be beneficial. Based on this insight, we propose an Elastic Net-MLP (ENet-MLP) hybrid model. We keep the Elastic Net as the backbone of the forecasting model and use a simple one-layer MLP to correct only the residuals of the linear prediction.

As shown in Figure 2, the inputs $\tilde{X}_t^{(n)} \in \{\mathbb{R}^{10}, \mathbb{R}^{14}, \mathbb{R}^{15}\}$ are passed simultaneously into two components: the Elastic Net produces the linear prediction $\hat{\ell}_{t+h}^{(n)} \in \mathbb{R}^7$, and a one-hidden-layer MLP (with 3, 5, or 7 neurons) generates the nonlinear residual correction $\hat{\varepsilon}_{t+h}^{(n)} \in \mathbb{R}^7$. The final excess treasury HPR forecast $\hat{x}_{t+h}^{(n)}$ combines both parts. The training strategy for the MLP is the same as described in Section 2.2.1.

Further, we design different training and evaluation procedures to better target the nonlinearities identified by the Ramsey RESET test. First, when training the MLP residual correction component, the in-sample validation is performed using the full validation set, denoted as $LN^{[i]}$ for $i \in \{3, 5, 7\}$, which includes all PMI regimes. Next, we further refine the model by conducting an in-sample validation only on the subset of validation observations that fall within the normal- and low-PMI periods, denoted as $LN^{[i],s}$ for $i \in \{3, 5, 7\}$, in which the Ramsey RESET Test also suggests stronger misspecification of the linear model.

2.2.4 Hard-switch Models

In addition to relying on validation-based, in-sample adjustments for observations that fall into the normal- and low-PMI regimes, we also introduce a set of hard-switch models, denoted as $LN_{hs}^{[i]}$ and $LN_{hs}^{[i],s}$ for $i \in \{3, 5, 7\}$, which differ in the validation strategy as described in Section 2.2.3. These models use the value of PMI_t to decide which forecasting method should be applied to predict $xr_{t+h}^{(n)}$.

When PMI_t falls into the high-PMI period, we use the Elastic Net model only to forecast $xr_{t+h}^{(n)}$. This choice is supported by the Ramsey RESET test, which does not reveal meaningful misspecification for the linear model in this regime. In contrast, when PMI_t falls into the normal- or low-PMI period, we use the ENet-MLP hybrid models $LN^{[i]}$ or $LN^{[i],s}$ for $i \in \{3, 5, 7\}$ to capture the nonlinear dynamics.

3 Model Evaluation and Performance

Now that we have introduced the estimation methods, we turn to evaluating model quality. Two approaches are used. The first approach compares each model's prediction errors with those of a benchmark, which we discuss in the next subsection. Because strong predictive accuracy does not necessarily translate into good economic performance, we also assess the models in a portfolio-allocation setting, using either power-utility or mean-variance optimization.

3.1 Relative Model Evaluation

Our main criterion for model evaluation is Campbell and Thompson (2008). They propose the use of an out of sample R^2. With this measure, it is possible to compare the performance of a given model $\mathcal{M}$ with some benchmark model $\mathcal{M}_B$. We denote by $t = t_0, \cdots, T$ the sample of available data and by $\hat{xr}_{t|t-1}^{(n)}(\mathcal{M})$ the prediction using model $\mathcal{M}$. All the excess returns are for a given holding period h. For convenience, we omit the h in the following.

$$R^2_{\text{OOS}} \equiv 1 - \frac{\sum_{t=t_0}^{T} \sum_{m=1}^{M} \left(\hat{xr}^{(n)}_{t|t-1}(\mathcal{M}) - xr^{(n)}_t \right)^2}{\sum_{t=t_0}^{T} \sum_{m=1}^{M} \left(\hat{xr}^{(n)}_{t|t-1}(\mathcal{M}_B) - xr^{(n)}_t \right)^2}.$$

In the above equation, t_0 denotes the forecast origin. A sample $1, \cdots, t_0$ is typically required to initialize the estimation and determine hyper-parameters. Trivially, R^2_{OOS} can be used also in the situation of a single series.

A standard benchmark is the expectation hypothesis (EH), which states that the best predictor of future returns is the historical average:

$$\hat{xr}^{(n)}_{t|t-1}(EH) = \frac{1}{t-1} \sum_{u=1}^{t-1} xr^{(n)}_u.$$

As the sample grows, this estimator becomes more precise because it averages over more observations. We use the test procedure developed by Clark and West (2007) to test if one model is better than another. Below, we will present results when we consider the EH model as the benchmark. In the Technical Appendix, we gathered the results when one compares models cross-wise.

3.2 Portfolio allocation

In this section we discuss the economic value of the predictions. For this purpose we consider a investor, either with power utility or a second order approximation thereof, the mean-variance utility. This investor determines the optimal allocation between the 1-month risk-free return and an allocation which can consist of one specific bond or all Treasury bonds. One can then determine the relevance of a model using the certainty equivalent criterion.

3.2.1 Power utility allocation

We consider first the univariate setting and then move to the portfolio allocation for the multivariate setting. Following Gargano et al. (2019), the objective is to maximize the utility function

$$U(w_t, xr^{(n)}_{t+h}) = \frac{\left[(1-w_t)\exp(\tilde{y}_t) + w_t \exp(\tilde{y}_t + xr^{(n)}_{t,t+h})\right]^{1-\gamma}}{1-\gamma}, \quad \gamma > 0.$$

Here $\tilde{y}$ is the return of the risk-free asset, e.g. $\tilde{y} = y^{(h)}_t$. This rate of return is known at time t and applies for an h period investment. To do so, we follow the approach of Brandt and Santa-Clara (2006) and approximate expected utility by the average utility which we maximize. Hence in the univariate problem we seek for month t an allocation given by

$$w^{(n)}_t \in \text{argmax}_w \sum_{s=1}^{t-1} \frac{\left[(1-w)\exp(\tilde{y}_s) + w \exp(\tilde{y}_s + \hat{xr}^{(n)}_{s,s+1})\right]^{1-\gamma}}{1-\gamma}.$$

The solution can be easily found since the problem is scalar. A typical choice of parameter of risk aversion is $\gamma = 5$. Once the weights $w^{(n)}_t$ have been obtained, one can determine the gross return of a portfolio.[3]

Presently, we are going to discuss the multivariate case. For this purpose we consider the $M \times 1$ vector $xr_{t+h} = [xr^{(1)}_{t+h}, \cdots, xr^{(n)}_{t+h}]'$. To obtain the weights that should be allocated to the M assets in this vector, Campbell and Viceira (1999) assume joint normality of the excess returns, and by doing so are able to generate a closed form solution for the optimal portfolio allocation. Before presenting their allocation, we need to introduce some notations. Denote by $\Sigma_{t+h|t}$ the conditional covariance matrix of xr_{t+h}. Also denote by $\sigma^2_{t+h|t}$ the diagonal elements of the matrix $\Sigma_{t+h|t}$. In this case, the optimal weights on the bonds are in the vector

$$w_t = \frac{1}{\gamma}\left(\Sigma_{t+h|t}\right)^{-1}\left[E_t[xr_{t,t+h}] + \frac{1}{2}\sigma^2_{t+h|t}\right],$$

where $E_t[xr_{t,t+h}]$ is the conditional expectation of the future excess return.

To render this expression operational, the conditional expectation $E_t[xr_{t,t+h}]$ is replaced

[3] An important practical observation is that to obtain sensible values in 1 both the risk-free rate and the excess return need to be annualized. If those conventions are not respected, the coefficient of risk aversion ($\gamma = 5$) needs to be adjusted. For the gross return, $R_{p,t} = 1 + hR_{f,t} + hw_t' R_{t+h}$, which corresponds to an investment of one month, the risk-free rate and excess return need to be multiplied by the holding period h. This is the same as dividing by 12.

by the best predictor $\hat{xr}_{t,t+h}$ generated by some model. Then there is the choice of the covariance matrix. This choice affects the allocation on a first order. The inspiration of how to estimate this matrix comes originally from Foster and Nelson (1996). It has been further promoted, for instance in Fleming et al. (2001).

The idea is to estimate $\Sigma_{t+h|t}$ using an exponentially weighted average of past innovations as follows. Since $\Sigma_{t+h|t}$ corresponds to a variance of deviations of returns from their best forecast, define ε_t, the forecast error. This forecast error can be obtained from $\varepsilon_t \equiv xr_{t|t-1} - \hat{xr}_{t|t-1}$. It is then possible to estimate the covariance matrix using:

$$\hat{\Sigma}_{t+h|t} = \sum_{l=0}^{\infty} A_{t-l} \odot \varepsilon_{t-l}\varepsilon'_{t-l}.$$

We let $A_{t-l} = \alpha \exp(-\alpha l)\mathbf{11'}$ be a matrix, with the same dimension as $\Sigma_{t+h|t}$ with constant elements. In practice, the infinite sum needs to be truncated (we choose to keep 120 months in a rolling window). This requires adjusting the weights so that their sum equals 1. The decay rate, α is chosen by Fleming et al. (2001) to be 0.063. Bianchi et al. (2021) use 0.05, and so do we. Some robustness analysis verified that the results do not depend on the exact choice of decay rate.

Inspection of the weights revealed that they are mostly reasonable. However, to avoid unrealistic allocation, we winsorize all the weights by the constraints $-1 \leq w_i \leq 2$, see Fleming et al. (2001), or Bianchi et al. (2021). Alternative winsorizations by restricting the weights to long-only strategies, $0 \leq w_i \leq 1$, did not fundamentally alter the findings.

3.2.2 Mean-variance allocation

The mean-variance allocation of Markowitz suggests that an investor with a coefficient of risk aversion of γ will solve $\max_w w'E_t[xr_{t,t+h}] - \gamma w'_t \Sigma_{t+h|t} w_t/2$. The solution of this optimization is well-known to be $w_t = \Sigma^{-1}_{t+h|t} E_t[xr_{t,t+h}]/\gamma$. The variance-covariance matrix $\Sigma_{t+h|t}$ can be estimated as indicated in the previous power-utility allocation.

3.2.3 Certainty Equivalent

When we wish to compare two dynamic trading strategies, say S_1 and S_2, with corresponding allocations $w_t(S_1)$ and $w_t(S_2)$ for all t, it is convenient to evaluate the certainty equivalent for the investor to switch from strategy S_1 to S_2. This is the amount of return that the investor receives or loses by adhering to S_1 instead of S_2. Denote this certainty equivalent by θ. To obtain it, define the utility for a given period and allocation and the average utility over all periods as:

$$\begin{aligned} U_u(w_u(S), xr_u) &= \frac{\left[(1 - \sum_{m=1}^{M} w_u^{(n)}(S) \exp(\tilde{y}_u) + \sum_{m=1}^{M} w_u^{(n)}(S)) \exp(\tilde{y}_u + xr_{u,u-1}^{(n)})\right]^{1-\gamma}}{1-\gamma} \\ \bar{U}(w(S), xr) &= \sum_{u=1}^{t-1} U_s(w_u(S), xr_u). \end{aligned}$$

The certainty equivalent is then defined by

$$U(w(S_1), xr) = U(w(S_2), xr + \theta).$$

The certainty equivalent is, thus, the amount of performance, measured in annualized returns, one should give to the investor with the suboptimal strategy S_2 to render this investor indifferent with the investor using the optimal strategy S_1. Alternatively, one can interpret θ as the maximal amount one could remove from the S_1 strategy yet having the investor to be better off than under the S_2 strategy. This interpretation is useful if there are transactions costs in which case θ can be compared to those.

To test if the certainty equivalent is statistically significant, we perform the test if the per period utilities for the different allocations are the same. Since it can happen that a strategy S_1 is suboptimal to another strategy S_2, we test for equality of the two utilities associated with those strategies. Formally, we put all the per-period utilities $U_u(w_u(S_1), xr_u) - U_u(w_u(S_2), xr_u)$ into a single vector and then regress the differences on a constant. We test after determining the t-statistics with an HAC standard error using 12 lags.

4 Data Used

In this section we document the origin of the treasury term structure data (yield data) as well as of the other non-yield variables. We will refer in the following to non-yield variables as being either constructs obtained with yield variables, examples being forward rates or functions thereof (like the Cochrane and Piazzesi (2005) index CP) or additional macro or finance variables. In the following, we discuss the term structure data used, i.e. our fundamental data.

4.1 Term Structure Data

As mentioned in the introduction, we downloaded three sets of term structure data. These datasets are the one referred to as GSW from Gürkaynak et al. (2007). The one referred to as LW from Ludvigson and Ng (2009) and last the one by Filipović et al. (2022a,b) to which we will refer to as FPY. It turned out that the moments of the three datasets as well as some more complicated model estimations were very similar. For this reason, we decided to use throughout the GSW data which has the advantage of yielding results that are comparable with the literature. Most of the works in the literature use the GSW dataset. An exception is the paper by Bianchi et al. (2021) and Wan et al. (2022) who use the LW dataset. Using alternative datasets than GSW, entails the risk that the available sample covers a smaller dataset.[4] Once we had found a satisfactory model we also estimated it on the other datasets to ascertain that the results are not driven by a peculiar dataset.

One of the well known drawbacks of the GSW dataset is that it does not involve T-Bill rates in its construction. For this reason, for the one month maturity, i.e. the reference risk-free rate, we always use the T-Bill rate.[5] The GSW is also constructed using off-the-run bonds, which are less liquid and for which one would expect somewhat higher yields than if one had used highly liquid bonds.[6] This dataset goes from 14 June 1961 to 12 Sept 2025

[4] In the online appendix, we give more details concerning LW and FPY datasets and also present moments and perform comparisons with the GSW dataset.

[5] Interestingly, the T-Bill rates are also not equal to corresponding yields in the othe datasets.

[6] The dataset we used in this paper has been downloaded on 23 Sept 2025 from `https://www.federalreserve.gov/data/nominal-yield-curve.htm`

and is at daily frequency. It contains the daily Svensson (1994) parameters. We remind that Svensson's formula determines the yield for a maturity τ with

$$y(\tau) = \beta_0 + \beta_1 \frac{1 - e^{-\tau/\lambda_1}}{\tau/\lambda_1} + \beta_2 \left(\frac{1 - e^{-\tau/\lambda_1}}{\tau/\lambda_1} - e^{-\tau/\lambda_1} \right) + \beta_3 \left(\frac{1 - e^{-\tau/\lambda_2}}{\tau/\lambda_2} - e^{-\tau/\lambda_2} \right). \quad (1)$$

All the parameters $\beta_0, \cdots, \beta_3$ as well as λ_1 and λ_2 are in the database. In the earlier years running from 14 June 1961 to 1972, the GSW parameters could only be estimated using bonds with maturities up to 7 years. We follow the literature and also use the estimates to extrapolate up to the 10 year horizon. So, by using (1) it is easy to construct daily yield curves. We do this by evaluating $y(\tau)$ on a monthly grid covering the 1 to 10 year maturity.

The dataset provided by GSW does not only contain the parameter estimates but also a set of forward rates. We used those forward rates and compared them with forward rates constructed using formula A.4. We obtained exactly identical forward rates.

We construct annualized excess holding period returns over a time horizon h using the expression

$$xr_{t,t+h}^{(n)} = \frac{n-h}{h} \left[y_t^{(n)} - y_{t+h}^{(n-h)} \right] + y_t^{(n)} - y_t^{(h)}.$$

In this formula t is the time index. With n we denote the maturity of a bond. This formula states that the excess holding period yield has several components. A first component captures the consequence of buying an m maturity bond and selling it h time units after. The second component contains the term spread, i.e. the difference between the m maturity bond and the short term bond. This paper follows Gargano et al. (2019) and we consider a one-month holding period return $h = 1/12$. It is recommended to express all time periods in an annual frequency: one month is 1/12. In our formulation, the excess returns are annualized. For comparison purposes, Gargano et al. (2019) express their returns per holding period. When we consider exactly their estimation period we match closely their moments. In a similar vein, when we compare our moments with those reported in Ioannidis and Ka (2021) and Wan et al. (2022) (who rely on the Liu and Wu (2021) dataset, and align the time horizons of the data samples, our construction obtains identical values down to all decimal places.

4.2 Predictors

So far, our focus has been on constructing holding-period excess returns from the term structure. We now introduce the predictors. As mentioned in the introduction, we rely on an extended set of predictors that has recently attracted attention in the related literature. In this sense, we extend Bianchi et al. (2021) by incorporating additional variables.

Our first set of estimators consists of the historical average of excess returns. This variable is the benchmark and corresponds to the Expectation Hypothesis according to which all the information required to predict future yield curves is already contained in the existing yield curves. See Campbell and Shiller (1991).

Next, forward rates are known to be key predictors that have attracted lots of attention. See Fama and Bliss (1987), Campbell and Shiller (1991), and Cochrane and Piazzesi (2005). The study by Ludvigson and Ng (2007) confirms the findings of Cochrane and Piazzesi (2005) before proposing an additional predictor based on macro-economic variables.

We define the annualized continuously compounded forward rate, $f_t^{(n-k,n)}$, as the rate agreed at time t, for a loan between $t+n-k$ and $t+n$. We remind in Appendix A.ii that the annualized forward rate, with tenor k, is

$$f_t^{(n-k,n)} = \frac{1}{k}\left[n \cdot y_t^{(n)} - (n-k) \cdot y_t^{(n-k)}\right]. \tag{2}$$

Having constructed forward rates, it is easy to derive the Fama-Bliss forward spreads by subtracting from (2) the risk-free rate that holds over the allocation period. It is similarly easy to obtain the Cochrane and Piazzesi (2005), CP, predictor for holding period h obtained by regressing the average of future excess returns on current forward rates, and then to use the parameter estimates and known forward rates to predict future excess returns.

We construct the first three principal components (YC1, YC2, and YC3) of monthly GSW dataset. The interpretation of these components follows the extensive term-structure literature stating that the first three PCA factors correspond closely to the traditional level, slope, and curvature movements of the yield curve, see Litterman (1991). Table 2 reports the correlations between these 3 PCA factors and level, slope, and curvature movements constructed directly from the GSW dataset.

We wanted to directly compare those principal components with the level, spread, and convexity playing an important role in the interpretation of the yield of Diebold and Li (2006). To do so, we construct the level, spread, and convexity as follows. The level factor is defined as the cross-sectional mean of yields across all maturities. The slope factor is constructed as the spread between the short (3-month) and long (10-year) ends of the yield curve, following the convention in the literature that uses the 3-month Treasury yield as the proxy for the short rate, see Fama and Bliss (1987); Diebold et al. (2006). The curvature factor is computed by comparing the medium-term yield at 2 years to the short- and long-term yields at 3 months and 10 years, respectively, where the medium-term sector of the yield curve (approximately 2–3 years) best captures the peak curvature loading in the Nelson–Siegel model, see Nelson and Siegel (1987); Diebold et al. (2006).

The correlations indicate that YC1 is effectively a level factor, YC2 closely corresponds to slope, and YC3 captures curvature. In this light, we decided to perform our preliminary analysis of potential predictors using the PCA components.

Following Ludvigson and Ng (2009), additional macroeconomic variables describing the state of the economy should also be considered as predictors. This led us to download the McCracken and Ng (2016) data.[7] They also provide on their website Matlab codes which detect missing values, remove outliers, and propose transforms to render their variables stationary. They also propose a grouping of the variables into 8 categories. The logic of using PCA of macro-economic variables as predictors initially goes back to Ludvigson and Ng (2007). We call the resulting factors $\mathrm{PCA}_1, \cdots, \mathrm{PCA}_8$. We did not go beyond the 8th factor. They construct a predictor, called LN, which results as a linear combination of PCA

[7]This data was sourced on 23 Sep 2025 from `https://www.stlouisfed.org/research/economists/mccracken/fred-databases`. This dataset covers 1999-08 to 2014-12 and the period from 2015 to 2024-12. There is also monthly data for 2025 (01 till 08) that was concatenated to the earlier datasets.

components (there is a cubed factor in the formula). A regression of their predictor on our factors revealed a high R^2 meaning that the PCA factors also span LN. Desirous to understand which component predicts the term structure, we preferred to work with the individual 8 factors.

Next, additional financial data was sourced. Interest rates are determined by monetary policy, and as we have seen before the 2008 crash, stock market behavior affects central banker's decisions. A first set of potential predictors is the Fama-French (FF) factors downloaded from Kenneth French's website.[8] In addition, we downloaded from his site the momentum factor. We also downloaded the 1 month T-Bill (our risk-free rate) from his website.[9]

A large database of financial variables has been collected by Goyal et al. (2024).[10] The construction of their variables has been essentially geared towards the predictability of stocks. The advantage is that most recent factors found in the literature are in their database.

Duffee (2013) mentions that there are directly-observable and non-observable factors that may affect rates by changing agent's preferences. This has motivated researchers to proxy some of those non-observable factors. In the front line, there are measures of uncertainty and variables which may translate the result of economic agent's analysis.

Recent research on term-structure prediction has also emphasized the role played by macro and financial uncertainty. There are different ways to obtain measures of uncertainty. We first obtained the real-time data by Kilian (2009). Unfortunately, this data starts relatively late. Desirous to use data for as many time periods as possible, we next followed Jurado et al. (2015) and Ludvigson et al. (2021) and obtained their various indices of uncertainty (real, financial and macro-uncertainty).[11] We obtained directly the US Purchasing Manager's Index (PMI) from Refinitv/LSEG on 30 Sep 2025. Cieslak and Povala (2015) uncover an excellent predictor of the term structure based on the trend inflation. This pre-

[8]We downloaded this data on 27 Sept 2025 from the URL `https://mba.tuck.dartmouth.edu/pages/faculty/ ken.french/data_library.html`.

[9]The 1 month T-Bill rate needs to be multiplied by 12 to get an annualized rate.

[10]We downloaded this data on 27 Sept 2025 from the URL `https://sites.google.com/view/agoyal145`.

[11]We downloaded their data on 27 Sept 2025 `https://www.sydneyludvigson.com/macro-and-financial-uncertainty-indexes`.

dictor is named CPITR, and we remind briefly in the appendix how this predictor can be constructed. The full list of variables can be found in the Appendix.

4.3 Final variable choice

At this stage we have a very large number of predictors. These are predictors based on yields, or combinations of forward rates such as the CP measure of Cochrane and Piazzesi (2005), the macro-economic data from McCracken-Ng, or the Goyal-Welch financial variables in addition to some other popular predictors. The question then is how to introduce this large number of predictors into the NN. Should one use factor analysis of the data or regroup the variables according to their logic? At the same time, we know from Boivin and Ng (2006) that more variables are not necessary better in determining factors. For this reason, our choice of variables was guided by several considerations. First, we attempted to summarize the information contained in the non-yield predictors using PCA and machine-learning techniques. However, these analyses indicated that such an approach would introduce very noisy inputs into the neural network. We therefore opted to reduce the set of predictors by performing an exhaustive investigation using correlation and basic OLS analysis to exhibit only the strongest predictors. This procedure yielded the following variables: MOM, FINUNC, PMI, CPITR, and YC2 that had the strongest predictive power. This way of proceeding is in line with the observation by Bianchi et al. (2021) who compare the predictive performance after regrouping macro variables by ML techniques with a regrouping of variables based on human intelligence, finding in the end that the latter provides better estimators.

Second, we were guided by the recent findings of Andreasen et al. (2021), who use PMI, a real-time predictor, to condition the estimates of their regression. Our iterative process of variable selection and model estimation confirmed the fundamental relevance of PMI also in our setting. The relevance of the business cycle for model performance has also been emphasized by Gargano et al. (2019) and Sarno et al. (2016), among others. Motivated by this insight, we stratify the sample into three business-cycle regimes using the empirical distribution of PMI: a low-activity regime (PMI below its 20th-percentile value of 48.7), a high-activity regime (PMI above its 80th-percentile value of 58.1), and a medium regime comprising all observations between these two thresholds. In Figure 3 we plot the temporal evolution of the PMI index as well as official NBER recession periods: those are the horizontal dashes at the bottom of the figure. With the yellow/grey shading, we represent those periods where our PMI indicates that there are recessions or booming periods. It should be noticed that depressed periods of the economy measured by the PMI and the NBER coincide.

In Figure 4, we display the 2, 5, and 10 year excess holding return xr. Comparison of this figure with Figure 3 reveals that during recession period xr tends indeed to be higher than during other periods.

Descriptive statistics for the retained variables are reported in Table 3. We notice that the HPR increases with the time horizon. So does the variability of the HPR. In the short run, skewness and average kurtosis are very high characteristic for sudden market interventions by the Central Bank. We notice that the rates are significantly different according to the PMI level. Also booming economy with a high PMI is going hand in hand with low excess returns. During low PMI periods, the series is not very persistent.

When we consider the forward rates, we also observe that for longer maturities of the forward contract, the rate is higher. Volatility is lower for longer rates. Skewness is positive and kurtosis is very low. This data is highly persistent. This pattern repeats itself during the various PMI sub-periods. For the additional variables, we notice that PMI is on average above 50 translating on average optimistic purchasing managers. During the low PMI period, standard deviation is larger with a negative skewness corresponding to markets where managers rapidly lose confidence. The trend inflation is around 3.68% with less inflation when the PMI is high and more inflation when the PMI is low. $YC2$, the second factor from a PCA on the yields, which is most correlated with the slope of the yield curves, is also very persistent. By construction, this factor has an overall mean of 0. During low PMI periods, the slope becomes steeper. Volatility behaves similar during the periods.

5 Model Discussion

5.1 Estimation

Tables 4, 6, and 8 report the out-of-sample (OOS) performance measured by R^2_{OOS} of various forecasting models for excess treasury HPR from January 1990 onward. The results are presented for the full OOS period, as well as for sub-OOS periods based on the PMI level.

Among linear models, the regularized models (Lasso, Ridge, and Elastic Net) present better OOS performance than OLS, which is consistent with the drawback of OLS when predicting excess treasury HPR with multicollinearity. Elastic Net is consistently the best model among linear models. Moreover, a single MLP model $\{NN^{[i]}\}_{i=3,5,7}$ performs worse than the linear model overall, but its performance improves during periods of low PMI. This suggests that nonlinear models provide additional value during specific periods and brings intuition for the next linear-plus-nonlinear model. Moving on to the hybrid ENet-MLP model ($\{LN^{[i]}\}_{i=3,5,7}$), even if introducing a nonlinear residual correction using shallow MLPs on top of the Elastic net model leads to mixed results across all maturities and PMI periods, this hybrid model is still a robust prediction model. First, the mean R^2_{OOS} of ENet-MLP hybrid models for the entire OOS samples consistently outperforms the linear or nonlinear only models for various groups of predictors. Second, the ENet-MLP hybrid models show remarkably strong performance in predicting shorter-term excess treasury HPR, especially during normal- and low-PMI periods. Those findings are consistent with the findings in Andersen et al. (2024) and the results in Table 1 of this paper, as the Ramsay RESET test results indicate that the linear Elastic Net model shows stronger misspecification for the same maturities and PMI periods. Our ENet-MLP hybrid model performs as expected to effectively capture the nonlinearity in residuals which are omitted by the linear model.

When using only forward rates as predictors, the nonlinear improvements are modest. Moreover, when including the additional predictors TB, MOM, YC2, CPITR, and PMI, both the linear and nonlinear models are improved, and nonlinear models achieve larger improvements in predictive power. This suggests that additional predictors show predictive power beyond the yield curve and deliver more nonlinear features. It's worth noting that after including additional predictors, nonlinear NN models show stronger predictive power compared to linear models, especially for the performance of $NN_{ss}^{[5]}$ model as shown in Table 8.

Table 4 and 6 also show that using PMI as a binary switching criterion is effective. Both the soft-switching models, $\{NN_{ss}^{[i]}\}_{i=3,5,7}$, and the hard-switching models, $\{LN_{hs}^{[i]}\}_{i=3,5,7}$, outperform their respective non-switching counterparts. Specifically, the soft-switching neural networks perform better than the standard neural network models $\{NN^{[i]}\}_{i=3,5,7}$, and the hard-switching linear models outperform the standard linear models $\{LN^{[i]}\}_{i=3,5,7}$. In addition, the PMI-restricted-validation ENet–MLP hybrid models $\{LN^{[i],s}\}_{i=3,5,7}$ also outperform the non-restricted linear ENet–MLP hybrid models $\{LN^{[i]}\}_{i=3,5,7}$.

In Table 8, we evaluate the performance of various models when PMI is treated as a predictor. The results show that including PMI directly as a predictor generates stronger predictive power than when using PMI as a switching criterion. This finding shows that PMI itself as a leading indicator of the economy may convey more information than just identifying the economic state, and a nonlinear model can capture such features. The gradual changes in the interaction between PMI and other predictors can be better captured in the model than using the switching rules. Among all the models and groups of predictors, using forward rates and all 5 additional predictors to predict excess Treasury HPR with a simple $LN^{[3]}$ model achieves the best prediction performance. This model is denoted as the *best-performing model* in the following discussions.

Table 5, 7 and 8 show the results of testing whether the predictive model fails to outperform the EH benchmark forecast according to Clark and West (2007). Across all the tables, the p-value results indicate that our ENet-MLP hybrid models can not only produce numerically positive R_{OOS}^2, but also more importantly, are reliably enough to reject the null hypothesis of no predictability.

5.2 Economic Values

Apart from evaluating models by it's predictive power, we also test if the improvement in forecasts can increase an investor's economic gains. We report the economic performance with Certainty Equivalent Returns (CER) in Table 9 and Sharp Ratio (SR) in Table 10 under both power utility and mean-variance utility. We report the performance of OLS, and the best model for each model category and predictor group.

In terms of CER, the linear models (OLS and ENet) consistently show negative gains for short-term Treasury bonds, and positive and statistically significant gains for medium- and long-term ones. For short-term Treasury bonds, nonlinear models bring positive and statistically significant improvements in economic value. For longer maturities, the gains from nonlinear modeling become even stronger, both in magnitude and statistical significance. The ENet-MLP hybrid model bring stronger economic gains especially for 4-year Treasury bond ($xr_{t+h}^{(4)}$). Remarkably, including PMI as a predictor rather than a switching criterion bring higher and more significant economic gains under both power utility and mean-variance utility.

Regarding SR, compared with EH benchmark, linear models show higher SR for short-term Treasury bonds, but fail to outperform EH benchmark for medium- and long-term ones regardless of the selection of predictors. However, when nonlinear models and additional predictors are introduced, the nonlinear model alone already brings SR improvements, especially for medium- and long-term Treasury bonds. When including additional predictors, the SRs of nonlinear and hybrid models increase substantially under both power utility and mean-variance utility, while the improvements are higher if we use PMI as a switch criterion.

We conclude that CER and SR show that incorporating nonlinearities and additional predictors brings meaningful improvements in investor welfare. Particularly, longer-term Treasury bonds, which are more sensitive to the economic cycles, can enhance risk-adjusted investor's economic values.

5.3 Cross-elasticity

We want to first introduce the cross-elasticity. Let some function Φ depend on various inputs, say $x_1, \cdots, x_K$. The cross-elasticity between variables x_i and x_j is defined as

$$\eta_{i,j} := \frac{\partial \log(\Phi)}{\partial \log(x_i) \partial \log(x_j)} = \frac{\partial \Phi}{\partial x_i \partial x_j} \frac{x_i x_j}{\Phi}.$$

This cross-elasticity being a log variation of the output as the log of some input variable changes has an interpretation as a percentage.

Since we know that PMI plays a privileged role as a predictor of the future state of the economy, we will take PMI as one of the variables that we vary. Figure 5 and 6 displays the cross elasticities demonstrating show how the marginal effect of a predictor is affected by changes of the PMI. The three columns of figures correspond to the variations of the 1, 5 and 10 year excess return. Observation of those figures reveals striking facts. First, an horizontal scan reveals that the excess returns for different horizons are affected very differently by a given factor. A vertical scan reveals that the other input variables do have a very different impact. To our knowledge, there does not exist another econometric framework that would reveal such nonlinear patterns in a natural manner.

In Figure 5, the first row of figures concerns the impact of inflation on bond returns. It is known that inflation expectations is a key input to predict future returns. In a booming economy, a low rate of inflation is a sign that there will not be any changes in the interest rates an this has depressing effect on the short term interest rates. When we look at the long horizon interest rates, we notice that in a booming economy when the inflation expectation exceeds the Central Bank long term goal of having an inflation of about 2%, this generates rising excess returns.

Next, we observe that for a booming economy or a recession, a steep yield curve leads to expectations of an increase in the short term and the long term returns but to a decrease for the 5 year horizon. These patterns appear to capture the well established feature that when interest rate are high, the term structure tends to be flat but when interest rates are low, the term structure is steep and more convex than otherwise.

The next row is dedicated to financial uncertainty. This is a variable which may affect investors preferences and therefore indirectly required returns. We observe strikingly contrasting patterns between the 1, 5, and 10 year horizons. In well functioning economies, an increase in financial uncertainty may have a depressing impact on the short run rate. For the medium term, an increase in uncertainty appears to be related with falling returns.

The last row of this figure corresponds to the interaction between momentum in financial markets and PMI. A low level of PMI in a booming economy leads to increasing short-term excess bond returns. Similarly for the 5 year horizon. The 10 year horizon has an interesting feature. For a high momentum factor in a booming economy the excess returns are falling.

5.4 Temporal evolution of the predictive capabilities

Another important issue is how does the relevance of predictors evolve over time. To investigate this evolution, we evaluate the mean absolute Shapley values (SHAP) from the best-preforming model ($LN^{[3]}$), which uses forward rates and five additional predictors to forecast excess Treasury holding-period returns. The SHAP value measures how much an output changes with respect to some benchmark evaluation which is chosen to be the mean value of the inputs.

Figure 7 shows the mean absolute SHAP values for each predictor in each estimation window. The 5 additional predictors continuously show stronger influence on excess Treasury HPR forecasts than the forward rates. trend inflation (CPITR) dominates persistently, which is consistent with the model's better economic performance in predicting long-term excess Treasury HPR, as the inflation is an important driver of long-term interest rates. The slope of the yield curve (YC2) and PMI index also maintain substantial and stable influence, which reflect the expectation on inflation as well. The financial uncertainty (FINUNC) contributes little before the Dot-com Bubble crisis. Subsequently, its importance rises since 2000 and further increased after the global financial crisis.

Forward rates also display the evolution over time. Short-term forward rates retain more overall relevance, but their importance gradually declines after 2008. This goes hand in hand with a switch in Central Bank policy where ad-hoc bond repurchases took over from markets setting freely rates. By contrast, longer-term forwards rates become increasingly relevant since 2008. Since the longer-term returns tend to be more oriented towards the state of the real economy, long-term forward rates reflecting markets expectations about future short term rates may impact the yield curve more.

When we consider the additional predictors, we notice that the momentum in financial markets may affect interest rates more in recent years. This observation goes hand in hand with the fact that also financial uncertainty matters more. When we look at the inflation component, we notice that this component plays a very strong role in the earlier times. The impact of this variables is the strongest among all variables. Since 2013 this variable has however been decreasing more and more. Next, we turn to the role of PMI, which plays steadily a role. To conclude, we examine the slope of the yield curve. We notice that this variable also plays systematically an important role. It has not varied over time but is always at a relatively high level.

6 Conclusion

The objective of this paper is to review the forecasting of monthly Treasury excess returns using an expanding set of predictors, ranging from forward rates to macroeconomic factors, uncertainty measures, and the PMI as a real-time indicator of economic activity. The predictions are performed with a large battery of models from the machine learning literature, ranging from regularized regression techniques to neural network estimators. After an exploratory analysis, we develop a hybrid estimation model that combines Elastic Net forecasts with a nonlinear correction of the residuals. The combination of both parts 'boosts' the forecasts from the backbone elastic-net model.

We exhibit strong differences in terms of predictability across normal, booming, and recessionary periods. We further demonstrate that an explicit switching model is not necessary to predict future excess returns. It is enough to consider our hybrid linear-nonlinear framework. Switching simply generates another type of nonlinearity. Inspection, via SHAP values, of what goes on in the network reveals very interesting patterns where in certain states (determined by the PMI), certain variables suddenly play a local role. Successful economic modeling of the yield curves should be able to capture those nonlinear patterns.

We show that the forecasts are economically valuable. Both certainty-equivalent returns and Sharpe ratios improve when hybrid models are used, with especially large gains for medium- and long-maturity Treasuries. We also check robustness of our findings by considering alternative datasets.

References

Andersen, D. B., Eriksen, J. N., Kjær, M. M., and Thyrsgaard, M. (2024). Predicting bond return predictability. *Management Science*, 70(2):931–951.

Andreasen, M. M., Engsted, T., Møller, S. V., and Sander, M. (2021). The yield spread and bond return predictability in expansions and recessions. *The Review of Financial Studies*, 34(6):2773–2812.

Bali, T. G., Goyal, A., Huang, D., Jiang, F., and Wen, Q. (2020). Predicting corporate bond returns: Merton meets machine learning. *Georgetown McDonough School of Business Research Paper*, (3686164):20–110.

Bauer, M. D. and Hamilton, J. D. (2018). Robust bond risk premia. *The Review of Financial Studies*, 31(2):399–448.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *The Journal of Finance*, 61(5):2187–2217.

Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample:

Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Viceira, L. M. (1999). Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics*, 114(2):433–495.

Cieslak, A. and Povala, P. (2015). Expected returns in treasury bonds. *The Review of Financial Studies*, 28(10):2859–2901.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Cochrane, J. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95:138–160.

Conley, T. G., Hansen, L. P., Luttmer, E. G., and Scheinkman, J. A. (1997). Short-term interest rates as subordinated diffusions. *The Review of Financial Studies*, 10(3):525–577.

Cooper, I. and Priestley, R. (2009). Time-varying risk premiums and the output gap. *The Review of Financial Studies*, 22(7):2801–2833.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics*, 131(1-2):309–338.

Duffee, G. R. (2013). Bond pricing and the macroeconomy. *Handbook of the Economics of Finance*, 2:907–967.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 77(4):680–692.

Filipović, D., Pelger, M., and Ye, Y. (2022a). Shrinking the term structure. *Swiss Finance Institute Research Paper Series*, (22-61).

Filipović, D., Pelger, M., and Ye, Y. (2022b). Stripping the discount curve – a robust machine learning approach. *Swiss Finance Institute Research Paper Series*, (22-24).

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Foster, D. P. and Nelson, D. B. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica*, 64(1):139–174.

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2019). Bond return predictability:

Economic value and links to the macroeconomy. *Management Science*, 65(2):508–540.

Ghysels, E., Horan, C., and Moench, E. (2018). Forecasting through the rearview mirror: Data revisions and bond return predictability. *The Review of Financial Studies*, 31(2):678–714.

Goyal, A., Welch, I., and Zafirov, A. (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 37(11):3490–3557.

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

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54:2291–2304.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, 2nd ed. edition.

Heaton, J. B., Polson, N. G., and Witte, J. H. (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Ioannidis, C. and Ka, K. (2021). Economic policy uncertainty and bond risk premia. *Journal of Money, Credit and Banking*, 53(6):1479–1522.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69.

Kynigakis, I. and Panopoulou, E. (2022). Does model complexity add value to asset allocation? evidence from machine learning forecasting models. *Journal of Applied Econometrics*, 37(3):603–639.

Litterman, R. (1991). Common factors affecting bond returns. *Journal of fixed income*, pages 54–61.

Liu, Y. and Wu, J. C. (2021). Reconstructing the yield curve. *Journal of Financial Economics*, 142(3):1395–1425.

Ludvigson, S. C., Ma, S., and Ng, S. (2021). Uncertainty and business cycles: exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, 13(4):369–410.

Ludvigson, S. C. and Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222.

Ludvigson, S. C. and Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, 22(12):5027–5067.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

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

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 31(2):350–371.

Sarno, L., Schneider, P., and Wagner, C. (2016). The economic value of predicting bond risk premia. *Journal of Empirical Finance*, 37:247–267.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3):788–829.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. *National Bureau of Economic Research Cambridge, Mass., USA*.

Tibshirani, R. (1997). The LASSO method for variable selection in the Cox model. *Statistics in Medicine*, 16(4):385–395.

Wan, R., Fulop, A., and Li, J. (2022). Real-time Bayesian learning and bond return predictability. *Journal of Econometrics*, 230(1):114–130.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320.

Appendix

A Data description

A.i Excess returns

Most of the earlier literature used annual holding periods of bonds in which case the formulas are simple. Gargano et al. (2019) are the first to introduce monthly holding period returns followed by many others. In this section we briefly review how to obtain the formula of monthly holding periods, in particular, it is important to determine what is annualized and what is on a time period basis.

It is good practice to express all time periods in terms of years. E.g. 1 month is 1/12. Denote by n the maturity (in years) of a bond, and h the holding period (also in years). Denote by $r^{(n)}_{t,t+h}$ the annualized continuous compounded holding period return (HPR) for an investor buying an n-period zero coupon bond at time t for a price of $P^{(n)}_t$ and selling a bond at a later date $t+h$ for a price $P^{(n)}_{t+h}$.[12] This HPR satisfies by definition $P^{(n)}_{t+h} = P^{(n)}_t\, e^{h \cdot r^{(n)}_{t,t+h}}$. From this we get

$$r^{(n)}_{t,t+h} = \frac{1}{h} \ln P^{(n)}_{t+h} / P^{(n)}_t. \tag{A.1}$$

The prices of pure discount bonds are related to the term structure of interest with the relation $P^{(n)}_t = e^{-n \cdot y^{(n)}_t}$ and so $\ln P^{(n)}_t = -n\, y^{(n)}_t$. Buying a bond at time t maturing n periods after, means holding a bond at time $t+h$ with a residual maturity $n-h$. Clearly,

$$\ln P^{(n-h)}_{t+h} = -(n-h)\, y^{(n-h)}_{t+h}.$$

Thus, equation (A.1) defining the annualized continuous compounded HPR on a bond ma-

[12] When there is a coupon, things are more complicated.

turing at time $t+m$ can be expressed as

$$r_{t,t+h}^{(n)} = \frac{1}{h}\left[-(n-h)\, y_{t+h}^{(n-h)} + n y_t^{(n)}\right]. \tag{A.2}$$

To get excess returns, let's introduce a benchmark risk-free asset. If the holding period is h, it is natural to choose at time t the rate $y_t^{(h)}$ which is known. The value of such an investment has as dynamic $B_{t+h} = B_t \exp(h \cdot y_t^{(h)})$. By expressing a discount bond in terms of this risk-free benchmark we obtain the annualized continuous compounded excess holding period return as $xr_{t,t+h}^{(n)} \equiv r_{t,t+h}^{(n)} - y_t^{(h)}$. Some trivial algebra gives

$$xr_{t,t+h}^{(n)} = \frac{n-h}{h}\left[y_t^{(n)} - y_{t+h}^{(n-h)}\right] + y_t^{(n)} - y_t^{(h)}. \tag{A.3}$$

When we perform a dimensional analysis, we notice that this expression has the dimension of a return.

A similar formulation is presented in Gargano et al. (2019) (equations 2 and 3). Comparison between both formulas reveals that their returns are defined over the holding period whereas we choose to work with annualized variables. With our convention, to find a return over the holding period instead of a year, it is necessary to multiply by h.

For an investor who holds an n-year bond for one year ($h=1$), the formula becomes

$$xr_{t,t+h} \equiv r_{t,t+h} - y_t^{(1)} = (n-1)\left[y_t^{(n)} - y_{t+h}^{(n-1)}\right] + y_t^{(n)} - y_t^{(1)}.$$

This is the expression one finds in the earlier literature dealing with annual returns. See Campbell and Shiller (1991), Cochrane and Piazzesi (2005), Sarno et al. (2016), and many others.

A.ii Forward rates

Forward rates are a key predictor that attracted lots of attention. See Fama and Bliss (1987), Cochrane and Piazzesi (2005), Ludvigson and Ng (2007). We define the annualized continuously compounded forward rate, $f_t^{(h,m)}$, as the rate agreed at time t, for a loan between $t + h$ and $t + m$. This rate is determined by the no-arbitrage relationship between zero-coupon bonds.

$$e^{h \cdot y_t^{(h)}} e^{(m-h) \cdot f_t^{(h,m)}} = e^{m \cdot y_t^{(n)}}$$
$$\Rightarrow f_t^{(h,m)} = \frac{1}{m-h} \left[m \cdot y_t^{(n)} - h \cdot y_t^{(h)} \right]$$

It is possible to express this formula in terms of the time between $t + h$ and $t + m$ by introducing $k = m - h$. This allows us to write

$$f_t^{(m-k,m)} = \frac{1}{k} \left[m \cdot y_t^{(n)} - (m-k) \cdot y_t^{(m-k)} \right], \tag{A.4}$$

which corresponds to Gargano et al. (2019) formula (4). This expression reveals that their forward rate is for the time the loan lasts, and not as here annualized.[13]

A.iii Benchmark models

Cieslak-Povala

They introduce a component which is orthogonal to trend inflation in predicting yields show that this component has strong predictive power for excess returns. Denote by CPI the consumer price index. The inflation rate is $\pi_t = \ln \text{CPI}_t / \text{CPI}_{t-1}$. Trend inflation is defined as the exponentially weighted inflation: $\tau_t^{\text{CPI}} = (1-\nu) \sum_{i=0}^{t-1} v^i \pi_{t-1-i}$. To empirically construct this measure one uses for π_t the year-over-year inflation with monthly sampling. The parameter ν weighs the different observations and is set to 0.987. The sum above gets truncated at 120. The expression above can be rewritten as $\tau_t^{\text{CPI}} = \tau_{t-1}^{\text{CPI}} + (1-\nu)(\pi_t - \tau_{t-1}^{\text{CPI}})$ showing that trend inflation is a mixture of previous trend inflation updated by the difference between the prediction error of previous trend inflation of the current inflation.

They then consider the regression $y_t^{(n)} = a_n + b_n^{\tau} \tau_t^{\text{CPI}} + \varepsilon_t$, for bond maturities ranging from 2 to 20 and define the prediction error as $c_t^{(n)} = y_t^{(n)} - \hat{a}_n - \hat{b}_n^{\tau} \tau_t^{\text{CPI}}$ which they call the maturity-specific cycle. Inspired by the work by Cochrane and Piazzesi (2005) they construct

[13] Their rate is $k \cdot f_t^{(m-k,m)}$ with our notations.

an univariate time series as common predictor of various future excess returns. To do so, they define the average excess returns $\overline{xr}_{t+h} \equiv \frac{1}{19}\sum_{n=2}^{20} xr_{t+h}^{(n)}$. Similarly, the construct for all maturities the average prediction error $\bar{c}_t \equiv \frac{1}{19}\sum_{n=2}^{20} c_t^{(n)}$. Eventually, they run regressions such as $\overline{xr}_{t+h}^{(n)} = d_0 + d_1\bar{c}_t + d_2 c_t^{(1)} + \varepsilon_{t+h}$ and consider as predictor of future excess returns $\widehat{cf}_{t+h|t} = \hat{d}_0 + \hat{d}_1\bar{c}_t + \hat{d}_2 c_t^{(1)}$. We verified that the predictor $\widehat{cf}_{t+h|t}$ is an excellent predictor of the future term structure.

Careful inspection of their paper reveals that the predictive R^2_{OOS} of using the $\widehat{cf}_{t+h|t}$ is just as good as when performing a regression using the short-term T-Bill and trend inflation which we denote by CPITR. For this reason, we decided to introduce in the NN both the short-term T-Bill rate and trend inflation. The NN can then decide how to combine the T-Bill and trend inflation to possibly generate an even better nonlinear estimate.

To construct the CPITR trend inflation, we downloaded the “Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL)” from FRED on 4 Oct 2025. The index is 1982-1984=100. This series is seasonally adjusted.

B Summary of retained variables

Predictor	Description	Source
TB	1-month Treasury bill rate	KF
CP	Linear combination of forward rates	GSW
OUTINC	Output and Income	FRED; GWZ
LABOR	Labor Market	FRED; GWZ
CONSMP	Consumption and Orders	FRED; GWZ
ORDINV	Orders and Inventories	FRED; GWZ
MONCRD	Money and Credit	FRED; GWZ
INTEXR	Interest rate and Exchange Rates	FRED; GWZ
PRICE	Prices	FRED; GWZ
STOCK	Stock Market	FRED; GWZ
PC1-8	First 8 principal components of the macroeconomic variables	FRED
SMB	Size factor (small minus big)	KF
HML	Value factor (high minus low)	KF
MOM	Momentum factor	KF
KILUNC	Economic uncertainty index by Kilian	Dallas Fed
FINUNC1	Financial uncertainty index by JLN	SL
MACUNC1	Macroeconomic uncertainty index by LMN	SL
RLUNC	Real activity uncertainty index by JLN	SL
OGAP	Output gap	GWZ
SNTM	Economic sentiment indicator	GWZ
PMI	U.S. Purchasing Managers' Index	Refinitiv
CPITR	Trend inflation by Cieslak-Povala	FRED
YC1	1st principal component of the U.S. Treasury yield curve	GSW
YC2	2nd principal component of the U.S. Treasury yield curve	GSW
YC3	3rd principal component of the U.S. Treasury yield curve	GSW

Table B1: This table reports descriptions of all the predictor variables used in the analysis. Data sources are as follows: KF denotes the Kenneth French Data Library; GWZ refers to the Goyal–Welch–Zafirov dataset; JLN corresponds to the Jurado–Ludvigson–Ng macro-uncertainty measures; LMN introduce the financial uncertainty index. Dallas Fed indicates series obtained from the Federal Reserve Bank of Dallas; SL refers to data from Sydney C. Ludvigson's website; and GSW represents the Treasury yield curve dataset. OGAP follows from Cooper and Priestley (2009).

Tables

	$xr^{(1)}$	$xr^{(2)}$	$xr^{(3)}$	$xr^{(4)}$	$xr^{(5)}$	$xr^{(7)}$	$xr^{(10)}$
All periods	22.22***	7.35***	4.02***	4.20***	5.93***	11.73***	16.29***
High PMI period	12.57***	3.53**	1.77	1.52	1.61	1.87	2.31*
Normal PMI period	14.04***	3.42**	1.10	1.46	2.44*	5.63***	9.72***
Low PMI period	4.69***	2.41*	2.41*	3.34**	4.61***	7.43***	10.95***

Table 1: This table reports the Ramsey RESET test statistics for Elastic Net residuals from predicting Treasury bond excess returns using forwards rates across maturities and PMI periods. Significance levels are denoted by *, **, and *** for 10%, 5%, and 1%, respectively.

	YC1	YC2	YC3
Level	1.0000	0.0075	0.0002
Slope	0.2099	0.9605	0.1561
Curvature	-0.5434	-0.2911	0.7167

Table 2: Correlations between the three yield curve moments (Level, Slope, Curvature) and the first three principal components of the yield curve (YC1, YC2, YC3).

	All Periods					High PMI Period					Normal PMI Period					Low PMI Period				
	mean	std	skew	kurt	AC(1)	mean	std	skew	kurt	AC(1)	mean	std	skew	kurt	AC(1)	mean	std	skew	kurt	AC(1)
Panel A: Excess Returns																				
$xr^{(1)}$	0.71	4.82	1.20	19.06	0.11	-0.19	3.44	-0.25	3.30	-0.07	0.47	3.87	-1.47	11.73	0.14	2.29	7.44	1.73	11.20	0.18
$xr^{(2)}$	1.03	9.34	0.59	14.16	0.13	-0.76	6.39	-0.31	2.59	-0.02	0.76	8.06	-1.55	11.23	0.18	3.62	13.73	1.41	8.76	0.17
$xr^{(3)}$	1.28	13.20	0.25	8.80	0.13	-1.25	8.90	-0.35	2.55	-0.01	1.09	11.90	-1.20	7.24	0.16	4.41	18.69	0.99	5.56	0.17
$xr^{(4)}$	1.48	16.73	0.08	5.48	0.12	-1.67	11.20	-0.42	2.63	-0.01	1.40	15.53	-0.92	4.54	0.13	4.94	23.00	0.66	3.37	0.15
$xr^{(5)}$	1.65	20.12	0.02	3.90	0.10	-2.06	13.44	-0.47	2.70	-0.02	1.67	18.98	-0.73	3.15	0.12	5.35	27.15	0.51	2.30	0.13
$xr^{(7)}$	1.88	26.78	0.08	2.94	0.08	-2.84	18.10	-0.57	2.84	-0.04	2.13	25.45	-0.53	2.09	0.10	5.97	35.81	0.52	1.73	0.08
$xr^{(10)}$	2.03	36.97	0.19	2.72	0.04	-4.13	25.93	-0.68	3.21	-0.08	2.68	34.83	-0.33	1.74	0.08	6.47	49.72	0.58	1.62	0.03
Panel B: Forward Rates																				
$f^{(1)}$	5.03	3.28	0.50	0.10	0.99	5.01	3.11	0.38	-0.37	0.97	4.60	3.10	0.38	-0.16	0.98	6.26	3.66	0.57	-0.01	0.96
$f^{(2)}$	5.36	3.16	0.50	-0.01	0.99	5.37	3.04	0.50	-0.09	0.98	4.97	2.96	0.36	-0.22	0.99	6.48	3.55	0.50	-0.32	0.98
$f^{(3)}$	5.62	3.02	0.54	-0.00	0.99	5.60	2.96	0.62	0.10	0.98	5.26	2.81	0.41	-0.17	0.99	6.67	3.42	0.47	-0.45	0.98
$f^{(4)}$	5.83	2.91	0.58	0.03	0.99	5.75	2.90	0.69	0.20	0.98	5.50	2.69	0.45	-0.10	0.99	6.84	3.30	0.46	-0.47	0.98
$f^{(5)}$	6.00	2.82	0.59	0.06	0.99	5.86	2.84	0.71	0.22	0.98	5.70	2.59	0.46	-0.06	0.99	6.99	3.20	0.44	-0.44	0.97
$f^{(6)}$	6.14	2.76	0.59	0.08	0.99	5.94	2.81	0.72	0.20	0.98	5.87	2.52	0.46	-0.04	0.99	7.12	3.13	0.44	-0.39	0.97
$f^{(7)}$	6.25	2.71	0.58	0.09	0.99	6.00	2.78	0.71	0.16	0.98	6.00	2.47	0.45	-0.02	0.99	7.22	3.08	0.43	-0.35	0.97
$f^{(8)}$	6.34	2.68	0.57	0.10	0.99	6.04	2.76	0.70	0.11	0.98	6.10	2.43	0.44	0.00	0.99	7.31	3.05	0.43	-0.32	0.97
$f^{(9)}$	6.40	2.66	0.57	0.12	0.99	6.07	2.75	0.69	0.07	0.98	6.18	2.40	0.43	0.02	0.99	7.38	3.03	0.44	-0.32	0.97
f^{10}	6.45	2.65	0.58	0.13	0.99	6.09	2.74	0.68	0.04	0.98	6.24	2.39	0.43	0.04	0.99	7.42	3.02	0.46	-0.32	0.97
Panel C: Additional Predictors																				
MOM	0.61	4.15	-1.32	9.92	0.03	0.80	3.40	0.09	2.22	0.04	0.74	3.58	0.03	3.16	-0.00	0.05	5.97	-2.04	8.47	0.02
FINUNC	0.91	0.17	0.59	0.24	0.98	0.88	0.16	0.71	0.49	0.90	0.89	0.15	0.46	-0.41	0.94	1.01	0.18	0.46	0.53	0.88
PMI	53.07	6.29	-0.54	1.10	0.94	60.94	2.87	1.68	2.83	0.77	53.43	2.56	-0.10	-1.10	0.78	44.05	4.51	-1.25	0.88	0.84
CPITR	3.68	2.09	1.12	0.31	1.00	3.44	2.21	0.90	-0.53	0.99	3.46	1.86	1.27	0.81	1.00	4.55	2.35	0.93	-0.44	0.99
YC2	-0.00	1.41	-0.10	-0.63	0.95	0.06	1.34	-0.26	-0.89	0.85	-0.08	1.36	-0.18	-0.63	0.92	0.16	1.59	0.08	-0.69	0.86

Table 3: This table reports descriptive statistics for excess returns and all predictors used in this paper. Panel A reports summary statistics for excess returns ($xr^{(n)}$ for maturities $n \in \{1, 2, 3, 4, 5, 7, 10\}$ in years). Panel B reports statistics for forward rates (f_t^n), where the n indicates the forward-rate horizon. Panel C reports statistics for the additional predictors that complement the forward rates. For each variable, the table presents the mean, standard deviation (std), skewness (skew), kurtosis (kurt), and first-order autocorrelation (AC(1)) across all periods, as well as separately for high, normal, and low PMI periods.

	OLS	$L1$	$L2$	$ENet$	$NN^{[3]}$	$NN^{[5]}$	$NN^{[7]}$	$NN_{ss}^{[3]}$	$NN_{ss}^{[5]}$	$NN_{ss}^{[7]}$	$LN^{[3]}$	$LN^{[5]}$	$LN^{[7]}$	$LN^{[3],s}$	$LN^{[5],s}$	$LN^{[7],s}$	$LN_{hs}^{[3]}$	$LN_{hs}^{[5]}$	$LN_{hs}^{[7]}$	$LN_{hs}^{[3],s}$	$LN_{hs}^{[5],s}$	$LN_{hs}^{[7],s}$
Panel A: High PMI Period																						
$xr_{t+h}^{(1)}$	-49.01	-5.22	2.24	0.95	-16.50	-28.38	0.47	-17.62	-2.56	-2.26	2.15	-0.85	-0.71	2.56	-1.32	-0.57	1.57	-0.23	-0.99	1.36	-1.19	-1.10
$xr_{t+h}^{(2)}$	-16.55	2.71	5.99	5.28	-12.68	-6.66	5.16	-11.44	-0.83	1.71	0.96	1.11	4.24	-0.56	1.35	4.02	2.06	3.59	4.07	1.01	2.95	3.43
$xr_{t+h}^{(3)}$	-4.38	5.55	7.06	6.73	2.19	-3.31	7.58	4.80	3.19	4.86	3.72	3.38	4.16	5.03	3.43	3.61	4.15	5.29	5.67	4.67	4.66	5.31
$xr_{t+h}^{(4)}$	2.31	7.38	7.97	7.89	3.47	3.88	7.45	3.23	5.52	2.58	5.30	5.45	5.83	5.14	4.48	5.47	6.41	6.76	6.76	6.39	6.42	6.70
$xr_{t+h}^{(5)}$	6.52	9.00	9.11	9.15	1.24	3.08	7.85	0.31	7.53	4.77	6.95	7.83	7.05	8.02	7.99	7.60	7.42	8.76	7.84	7.89	8.46	7.98
$xr_{t+h}^{(7)}$	11.32	11.49	11.22	11.35	4.29	5.41	10.26	6.89	6.26	5.03	10.70	8.93	11.50	10.60	8.90	11.84	10.41	10.13	10.97	10.29	10.12	11.15
$xr_{t+h}^{(10)}$	12.25	11.65	11.52	11.50	8.82	7.77	12.37	6.35	9.36	8.55	10.81	11.19	11.25	11.06	11.30	11.21	11.04	11.55	11.58	11.44	11.50	11.52
Mean	-5.36	6.08	7.87	7.55	-1.31	-2.60	7.31	-1.07	4.07	3.61	5.80	5.29	6.19	5.98	5.16	6.17	6.15	6.55	6.56	6.15	6.13	6.43
Panel B: Normal PMI Period																						
$xr_{t+h}^{(1)}$	-23.30	2.36	3.20	3.70	4.22	2.27	5.65	6.69	7.40	7.18	3.70	4.84	4.57	4.50	4.57	4.86	3.94	4.87	4.67	4.57	4.69	4.84
$xr_{t+h}^{(2)}$	-13.28	1.64	1.57	1.89	1.57	2.32	2.44	4.41	3.85	3.36	5.59	3.39	4.13	5.21	3.90	4.59	5.52	3.59	4.13	5.66	4.16	4.73
$xr_{t+h}^{(3)}$	-7.11	2.81	2.53	2.79	1.38	2.35	-0.19	1.91	2.44	2.00	5.73	5.18	4.11	5.52	5.37	4.44	5.74	5.27	4.30	5.65	5.64	4.65
$xr_{t+h}^{(4)}$	-3.06	4.32	4.03	4.27	2.77	2.16	0.77	1.83	2.66	2.95	6.99	6.26	5.49	7.06	6.11	5.63	6.94	6.28	5.61	7.00	6.29	5.83
$xr_{t+h}^{(5)}$	-0.61	5.38	5.21	5.41	2.19	2.41	0.78	2.82	0.92	0.97	7.63	6.06	7.66	8.01	5.37	7.43	7.56	6.12	7.68	8.03	5.58	7.53
$xr_{t+h}^{(7)}$	1.40	5.94	6.11	6.18	2.02	1.50	0.62	1.29	2.76	1.93	7.20	8.00	6.94	7.03	7.59	6.93	7.20	7.97	6.95	7.14	7.69	7.02
$xr_{t+h}^{(10)}$	0.40	3.80	4.28	4.22	0.52	1.59	1.71	2.49	2.53	1.39	4.36	4.71	4.80	4.14	4.39	4.76	4.38	4.65	4.84	4.20	4.40	4.83
Mean	-6.51	3.75	3.85	4.07	2.10	2.09	1.68	3.06	3.22	2.83	5.89	5.49	5.39	5.92	5.33	5.52	5.90	5.53	5.45	6.04	5.49	5.63
Panel C: Low PMI Period																						
$xr_{t+h}^{(1)}$	-4.88	3.03	4.19	4.49	10.50	8.25	5.96	11.09	7.30	6.58	7.10	6.76	6.83	7.54	7.54	6.92	7.10	6.76	6.83	7.54	7.54	6.92
$xr_{t+h}^{(2)}$	0.76	2.70	3.45	3.47	9.77	6.21	3.98	10.65	6.72	5.80	8.72	6.21	7.10	10.26	8.04	7.92	8.72	6.21	7.10	10.26	8.04	7.92
$xr_{t+h}^{(3)}$	5.22	4.64	5.16	5.09	5.83	7.41	3.02	7.28	7.68	7.03	7.99	7.54	9.36	7.70	8.82	9.51	7.99	7.54	9.36	7.70	8.82	9.51
$xr_{t+h}^{(4)}$	8.03	6.24	6.70	6.61	8.84	6.79	6.08	8.61	8.01	8.80	8.20	8.77	8.59	8.24	9.11	9.47	8.20	8.77	8.59	8.24	9.11	9.47
$xr_{t+h}^{(5)}$	9.63	7.41	7.85	7.73	8.89	8.69	7.41	9.26	7.24	8.05	9.58	8.22	9.72	9.17	8.72	10.08	9.58	8.22	9.72	9.17	8.72	10.08
$xr_{t+h}^{(7)}$	10.66	8.70	8.98	8.89	8.50	8.01	6.49	8.01	8.12	8.42	9.24	9.29	9.02	9.14	9.56	9.17	9.24	9.29	9.02	9.14	9.56	9.17
$xr_{t+h}^{(10)}$	9.65	8.85	8.68	8.69	7.02	7.46	6.92	7.88	7.61	7.35	9.02	8.40	9.19	8.98	8.14	9.27	9.02	8.40	9.19	8.98	8.14	9.27
Mean	5.58	5.94	6.43	6.42	8.48	7.55	5.69	8.97	7.52	7.43	8.55	7.88	8.54	8.72	8.56	8.91	8.55	7.88	8.54	8.72	8.56	8.91
Panel D: All Periods																						
$xr_{t+h}^{(1)}$	-18.86	2.09	3.42	3.75	5.09	2.54	5.42	6.76	6.78	6.42	4.68	5.11	4.96	5.34	5.16	5.18	4.80	5.16	5.01	5.32	5.24	5.14
$xr_{t+h}^{(2)}$	-10.30	1.08	1.53	1.68	2.16	2.01	2.22	4.29	3.51	3.09	5.28	3.13	4.10	5.38	4.00	4.61	5.31	3.42	4.09	5.77	4.27	4.66
$xr_{t+h}^{(3)}$	-4.73	2.25	2.35	2.47	1.50	2.18	0.06	2.42	2.75	2.41	4.96	4.48	4.32	4.84	4.96	4.51	5.00	4.67	4.54	4.90	5.22	4.77
$xr_{t+h}^{(4)}$	-1.00	3.69	3.69	3.81	3.15	2.22	1.38	2.46	2.99	3.19	5.84	5.53	5.00	5.88	5.45	5.30	5.88	5.63	5.14	5.93	5.71	5.52
$xr_{t+h}^{(5)}$	1.30	4.77	4.80	4.90	2.56	2.77	1.72	2.99	1.71	1.79	6.69	5.38	6.74	6.89	5.07	6.73	6.67	5.48	6.81	6.90	5.24	6.82
$xr_{t+h}^{(7)}$	3.31	5.64	5.82	5.85	2.54	2.15	1.51	2.09	3.02	2.52	6.53	6.91	6.33	6.38	6.73	6.38	6.51	6.98	6.30	6.43	6.88	6.39
$xr_{t+h}^{(10)}$	2.60	4.39	4.62	4.59	1.58	2.29	2.56	2.88	3.03	2.20	4.73	4.77	5.08	4.60	4.50	5.08	4.75	4.76	5.12	4.66	4.52	5.14
Mean	-3.96	3.42	3.75	3.86	2.65	2.31	2.12	3.41	3.40	3.09	5.53	5.04	5.22	5.61	5.13	5.40	5.56	5.16	5.29	5.70	5.30	5.49

Table 4: This table reports R^2_{OOS} (in percentage) for one-month excess U.S. Treasury bond HPR forecasts using forward rates $\{f_t^{(j)}\}_{j=1}^{10}$ as predictors. Out-of-sample evaluation begins in January 1990 and is presented for the full period, and separately for high, normal, and low PMI periods. Forecasting models include: (i) linear models (OLS, Lasso (L1), Ridge (L2), and Elastic Net (ENet)); (ii) MLP models ($NN^{[3,5,7]}$); (iii) soft-switch on MLPs ($NN_{ss}^{[3,5,7]}$); (iv) ENet–MLP hybrids ($LN^{[3,5,7]}$); (v) PMI-restricted-validation ENet–MLP hybrids ($LN^{[3,5,7],s)}$); (vi) hard-switch on ENet–MLP hybrids ($LN_{hs}^{[3,5,7]}$ and $LN_{hs}^{[3,5,7],s}$.)

	OLS	$L1$	$L2$	$ENet$	$NN^{[3]}$	$NN^{[5]}$	$NN^{[7]}$	$NN_{ss}^{[3]}$	$NN_{ss}^{[5]}$	$NN_{ss}^{[7]}$	$LN^{[3]}$	$LN^{[5]}$	$LN^{[7]}$	$LN^{[3],s}$	$LN^{[5],s}$	$LN^{[7],s}$	$LN_{hs}^{[3]}$	$LN_{hs}^{[5]}$	$LN_{hs}^{[7]}$	$LN_{hs}^{[3],s}$	$LN_{hs}^{[5],s}$	$LN_{hs}^{[7],s}$
Panel A: High PMI Period																						
$xr^{(1)}_{t+h}$	0.025	0.021	0.014	0.016	0.097	0.118	0.059	0.095	0.074	0.070	0.017	0.018	0.016	0.013	0.016	0.014	0.016	0.016	0.016	0.015	0.016	0.015
$xr^{(2)}_{t+h}$	0.017	0.026	0.027	0.028	0.103	0.097	0.070	0.101	0.080	0.076	0.034	0.034	0.028	0.038	0.035	0.030	0.028	0.031	0.028	0.031	0.032	0.029
$xr^{(3)}_{t+h}$	0.014	0.025	0.030	0.030	0.086	0.107	0.064	0.074	0.068	0.071	0.032	0.041	0.038	0.033	0.044	0.040	0.031	0.030	0.031	0.032	0.033	0.032
$xr^{(4)}_{t+h}$	0.012	0.020	0.025	0.025	0.076	0.089	0.063	0.080	0.063	0.082	0.032	0.033	0.034	0.032	0.037	0.036	0.024	0.024	0.025	0.025	0.026	0.026
$xr^{(5)}_{t+h}$	0.009	0.013	0.017	0.017	0.089	0.085	0.055	0.094	0.051	0.079	0.020	0.025	0.024	0.022	0.020	0.023	0.015	0.017	0.016	0.018	0.019	0.017
$xr^{(7)}_{t+h}$	0.005	0.006	0.007	0.007	0.079	0.080	0.033	0.064	0.053	0.074	0.010	0.015	0.007	0.010	0.015	0.008	0.008	0.008	0.007	0.009	0.008	0.008
$xr^{(10)}_{t+h}$	0.003	0.002	0.003	0.003	0.050	0.066	0.015	0.078	0.040	0.055	0.004	0.005	0.005	0.004	0.005	0.006	0.003	0.004	0.003	0.003	0.004	0.003
Mean	0.012	0.016	0.018	0.018	0.083	0.092	0.051	0.084	0.061	0.072	0.021	0.024	0.022	0.022	0.025	0.022	0.018	0.019	0.018	0.019	0.020	0.019
Panel B: Normal PMI Period																						
$xr^{(1)}_{t+h}$	0.000	0.000	0.000	0.000	0.005	0.027	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(2)}_{t+h}$	0.001	0.000	0.001	0.001	0.050	0.017	0.007	0.005	0.006	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(3)}_{t+h}$	0.001	0.000	0.002	0.001	0.019	0.016	0.028	0.010	0.013	0.013	0.001	0.000	0.001	0.000	0.000	0.001	0.001	0.000	0.001	0.000	0.000	0.001
$xr^{(4)}_{t+h}$	0.002	0.000	0.001	0.001	0.009	0.015	0.022	0.017	0.012	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(5)}_{t+h}$	0.002	0.000	0.001	0.001	0.011	0.012	0.038	0.012	0.023	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(7)}_{t+h}$	0.004	0.000	0.001	0.001	0.016	0.028	0.041	0.031	0.011	0.022	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
$xr^{(10)}_{t+h}$	0.016	0.005	0.006	0.006	0.074	0.039	0.023	0.016	0.015	0.032	0.006	0.004	0.004	0.006	0.003	0.003	0.006	0.005	0.004	0.006	0.003	0.004
Mean	0.004	0.001	0.002	0.002	0.026	0.022	0.023	0.013	0.011	0.016	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.000	0.001
Panel C: Low PMI Period																						
$xr^{(1)}_{t+h}$	0.005	0.003	0.005	0.004	0.005	0.008	0.005	0.004	0.005	0.006	0.002	0.003	0.002	0.002	0.003	0.002	0.002	0.003	0.002	0.002	0.003	0.002
$xr^{(2)}_{t+h}$	0.008	0.008	0.010	0.009	0.028	0.027	0.014	0.020	0.019	0.017	0.008	0.010	0.008	0.009	0.008	0.009	0.008	0.010	0.008	0.009	0.008	0.009
$xr^{(3)}_{t+h}$	0.010	0.015	0.015	0.015	0.039	0.047	0.030	0.034	0.031	0.034	0.015	0.017	0.015	0.020	0.016	0.017	0.015	0.017	0.015	0.020	0.016	0.017
$xr^{(4)}_{t+h}$	0.012	0.021	0.020	0.021	0.055	0.049	0.038	0.040	0.051	0.047	0.023	0.023	0.026	0.023	0.024	0.025	0.023	0.023	0.026	0.023	0.024	0.025
$xr^{(5)}_{t+h}$	0.013	0.027	0.024	0.025	0.062	0.059	0.041	0.063	0.048	0.051	0.027	0.029	0.027	0.031	0.028	0.028	0.027	0.029	0.027	0.031	0.028	0.028
$xr^{(7)}_{t+h}$	0.017	0.034	0.031	0.032	0.059	0.060	0.049	0.061	0.060	0.059	0.035	0.038	0.033	0.037	0.037	0.037	0.035	0.038	0.033	0.037	0.037	0.037
$xr^{(10)}_{t+h}$	0.025	0.038	0.040	0.040	0.061	0.071	0.055	0.063	0.064	0.062	0.041	0.044	0.043	0.041	0.046	0.044	0.041	0.044	0.043	0.041	0.046	0.044
Mean	0.013	0.021	0.021	0.021	0.044	0.046	0.033	0.041	0.040	0.039	0.022	0.023	0.022	0.023	0.023	0.023	0.022	0.023	0.022	0.023	0.023	0.023
Panel D: All Periods																						
$xr^{(1)}_{t+h}$	0.000	0.000	0.000	0.000	0.002	0.026	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(2)}_{t+h}$	0.001	0.000	0.001	0.001	0.056	0.034	0.006	0.011	0.007	0.005	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$xr^{(3)}_{t+h}$	0.001	0.000	0.001	0.001	0.027	0.039	0.026	0.011	0.012	0.012	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.000	0.001
$xr^{(4)}_{t+h}$	0.001	0.000	0.001	0.000	0.022	0.020	0.019	0.018	0.015	0.018	0.000	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001
$xr^{(5)}_{t+h}$	0.000	0.000	0.000	0.000	0.038	0.025	0.024	0.034	0.019	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(7)}_{t+h}$	0.000	0.000	0.000	0.000	0.035	0.035	0.025	0.030	0.020	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(10)}_{t+h}$	0.001	0.001	0.001	0.001	0.050	0.045	0.015	0.030	0.019	0.033	0.002	0.001	0.002	0.002	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.001
Mean	0.001	0.000	0.001	0.000	0.033	0.032	0.016	0.019	0.013	0.019	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000

Table 5: This table reports p-value for the null hypothesis $R^2_{OOS} < 0$ for one-month U.S. Treasury bond excess returns forecasts using forward rates $\{f_t^{(j)}\}_{j=1}^{10}$ as predictors. Out-of-sample evaluation begins in January 1990 and is presented for the full period, and separately for high, normal, and low PMI periods. The models are identical to those in Table 4.

	OLS	$L1$	$L2$	$ENet$	$NN^{[3]}$	$NN^{[5]}$	$NN^{[7]}$	$NN^{[3]}_{ss}$	$NN^{[5]}_{ss}$	$NN^{[7]}_{ss}$	$LN^{[3]}$	$LN^{[5]}$	$LN^{[7]}$	$LN^{[3],s}$	$LN^{[5],s}$	$LN^{[7],s}$	$LN^{[3]}_{hs}$	$LN^{[5]}_{hs}$	$LN^{[7]}_{hs}$	$LN^{[3],s}_{hs}$	$LN^{[5],s}_{hs}$	$LN^{[7],s}_{hs}$
Panel A: High PMI Period																						
$xr^{(1)}_{t+h}$	-77.70	-4.92	4.96	5.73	-18.56	-9.01	-5.86	-8.16	-2.10	0.79	1.37	6.37	0.96	5.48	5.52	0.24	4.23	6.26	3.85	5.06	5.91	3.85
$xr^{(2)}_{t+h}$	-32.51	-0.56	6.55	6.02	1.40	1.15	3.43	-1.77	2.99	4.49	-1.09	2.40	5.10	0.09	2.45	5.95	3.58	4.63	5.58	3.96	4.50	5.87
$xr^{(3)}_{t+h}$	-16.45	1.83	7.12	6.48	2.33	5.87	2.97	-2.33	4.76	4.85	5.43	2.41	6.74	5.41	2.37	6.21	6.06	5.28	6.63	6.21	5.04	6.39
$xr^{(4)}_{t+h}$	-7.54	3.45	7.61	6.97	6.56	1.95	9.24	4.60	5.16	5.93	2.56	6.22	6.38	5.17	6.52	6.09	4.88	6.62	6.68	5.69	6.78	6.63
$xr^{(5)}_{t+h}$	-1.97	5.06	8.41	7.78	7.71	4.29	6.37	4.19	6.58	6.44	5.68	5.17	5.20	7.62	6.37	6.66	6.53	6.41	6.46	7.33	6.98	7.18
$xr^{(7)}_{t+h}$	3.84	8.19	10.50	9.86	8.72	10.71	8.73	8.58	10.71	9.01	8.95	8.79	8.96	9.71	9.36	9.41	9.07	9.17	9.10	9.48	9.52	9.41
$xr^{(10)}_{t+h}$	3.42	11.52	13.08	12.37	12.01	10.74	11.70	9.68	12.44	10.92	11.63	11.56	11.68	12.07	11.72	11.61	11.75	11.57	11.64	11.90	11.63	11.59
Mean	-18.42	3.51	8.32	7.89	2.88	3.67	5.22	2.11	5.79	6.06	4.93	6.13	6.43	6.51	6.33	6.59	6.59	7.13	7.14	7.09	7.20	7.27
Panel B: Normal PMI Period																						
$xr^{(1)}_{t+h}$	-37.57	5.96	3.10	4.35	5.69	5.23	9.35	7.98	7.89	7.54	5.96	2.79	6.57	6.59	2.97	7.63	5.80	2.94	6.51	6.55	3.12	7.47
$xr^{(2)}_{t+h}$	-21.68	5.25	3.97	4.39	3.83	4.06	6.89	4.27	6.00	5.56	7.96	5.70	4.52	8.24	6.22	4.05	7.95	5.70	4.51	8.15	6.14	4.13
$xr^{(3)}_{t+h}$	-13.57	4.92	3.88	4.06	-0.80	3.52	5.58	3.73	3.82	4.53	5.07	5.37	3.37	5.13	5.51	3.50	5.11	5.37	3.58	5.24	5.44	3.67
$xr^{(4)}_{t+h}$	-8.47	4.82	3.91	3.97	1.38	3.79	1.63	3.01	3.98	1.27	6.23	4.49	4.19	5.73	4.51	4.19	5.96	4.34	4.11	5.50	4.42	4.14
$xr^{(5)}_{t+h}$	-5.44	4.81	4.01	3.98	0.46	3.92	4.48	3.46	4.30	3.62	5.18	5.22	4.76	4.90	5.03	4.35	5.05	4.94	4.79	4.76	4.82	4.45
$xr^{(7)}_{t+h}$	-3.08	4.82	4.14	4.02	-0.11	-1.37	3.57	2.62	3.58	3.92	5.14	4.99	4.99	4.78	4.68	4.66	5.07	4.86	4.83	4.79	4.67	4.52
$xr^{(10)}_{t+h}$	-4.32	4.74	4.09	3.96	-0.68	1.79	1.98	2.87	3.63	3.52	4.61	4.94	5.10	4.34	5.14	5.05	4.54	4.78	4.93	4.31	4.92	4.84
Mean	-13.45	5.04	3.87	4.10	1.40	2.99	4.78	3.99	4.74	4.28	5.74	4.79	4.79	5.67	4.87	4.77	5.64	4.70	4.75	5.61	4.79	4.74
Panel C: Low PMI Period																						
$xr^{(1)}_{t+h}$	-2.25	1.91	5.96	5.64	7.41	7.49	9.12	10.75	9.75	6.39	7.97	4.68	8.00	9.18	5.16	8.42	7.97	4.68	8.00	9.18	5.16	8.42
$xr^{(2)}_{t+h}$	4.98	5.87	7.58	7.17	5.17	7.24	7.77	7.80	8.94	6.56	12.07	9.89	7.78	12.06	10.34	7.49	12.07	9.89	7.78	12.06	10.34	7.49
$xr^{(3)}_{t+h}$	9.55	7.63	8.88	8.52	2.66	7.37	9.85	9.24	9.27	8.76	9.61	9.91	8.46	9.41	9.96	8.92	9.61	9.91	8.46	9.41	9.96	8.92
$xr^{(4)}_{t+h}$	12.16	8.21	9.47	9.18	5.87	9.27	5.55	8.03	9.74	7.64	11.04	9.38	9.21	10.62	9.29	9.67	11.04	9.38	9.21	10.62	9.29	9.67
$xr^{(5)}_{t+h}$	13.51	8.26	9.65	9.41	6.64	9.33	9.79	9.99	9.82	10.17	10.53	10.06	10.79	10.17	9.91	10.49	10.53	10.06	10.79	10.17	9.91	10.49
$xr^{(7)}_{t+h}$	14.22	7.82	9.40	9.24	6.87	5.02	8.67	8.69	8.85	9.31	9.91	9.80	9.74	10.08	9.42	9.83	9.91	9.80	9.74	10.08	9.42	9.83
$xr^{(10)}_{t+h}$	13.40	7.30	8.89	8.83	5.63	8.64	7.55	8.52	9.41	9.19	9.29	9.24	9.17	9.69	9.06	9.57	9.29	9.24	9.17	9.69	9.06	9.57
Mean	9.37	6.72	8.55	8.28	5.75	7.76	8.33	9.00	9.40	8.29	10.06	8.99	9.02	10.17	9.02	9.20	10.06	8.99	9.02	10.17	9.02	9.20
Panel D: All Periods																						
$xr^{(1)}_{t+h}$	-28.55	4.00	4.06	4.78	4.90	5.16	8.39	7.95	7.88	6.75	6.29	3.54	6.67	7.32	3.75	7.42	6.33	3.62	6.78	7.28	3.87	7.51
$xr^{(2)}_{t+h}$	-15.51	4.10	4.34	4.44	3.17	3.94	6.05	4.04	5.74	4.88	7.65	5.82	4.64	7.90	6.29	4.32	7.96	5.97	4.66	8.11	6.38	4.36
$xr^{(3)}_{t+h}$	-8.59	4.21	4.31	4.27	-0.84	3.55	5.38	3.59	4.21	4.51	5.13	5.21	3.81	5.10	5.32	3.99	5.20	5.41	3.94	5.22	5.46	4.11
$xr^{(4)}_{t+h}$	-4.13	4.29	4.38	4.28	1.64	3.85	1.92	3.16	4.32	2.01	5.95	4.61	4.40	5.70	4.62	4.50	5.94	4.55	4.37	5.59	4.58	4.51
$xr^{(5)}_{t+h}$	-1.40	4.34	4.48	4.34	1.27	4.02	4.67	3.91	4.56	4.21	5.29	5.13	5.04	5.15	5.05	4.81	5.26	5.04	5.15	5.04	4.96	4.91
$xr^{(7)}_{t+h}$	0.91	4.40	4.62	4.46	1.04	-0.18	3.92	3.29	4.08	4.32	5.29	5.16	5.14	5.18	4.89	5.00	5.25	5.10	5.05	5.17	4.90	4.91
$xr^{(10)}_{t+h}$	0.21	4.42	4.67	4.53	0.64	3.05	2.85	3.55	4.50	4.26	5.01	5.20	5.27	5.02	5.28	5.36	4.98	5.10	5.16	4.99	5.13	5.23
Mean	-8.15	4.25	4.41	4.44	1.69	3.34	4.74	4.21	5.04	4.42	5.80	4.95	5.00	5.91	5.03	5.06	5.85	4.97	5.01	5.91	5.04	5.08

Table 6: This table reports R^2_{OOS} (in percentage) of one-month U.S. Treasury bond excess returns forecasts using forward rates $\{f^{(j)}_t\}^{10}_{j=1}$, TB, MOM, YC2, and CPITR as predictors. Out-of-sample evaluation begins in January 1990 and is presented for the full period, and separately for high, normal, and low PMI periods. The models are identical to those in Table 4.

	OLS	$L1$	$L2$	$ENet$	$NN^{[3]}$	$NN^{[5]}$	$NN^{[7]}$	$NN^{[3]}_{ss}$	$NN^{[5]}_{ss}$	$NN^{[7]}_{ss}$	$LN^{[3]}$	$LN^{[5]}$	$LN^{[7]}$	$LN^{[3],s}$	$LN^{[5],s}$	$LN^{[7],s}$	$LN^{[3]}_{hs}$	$LN^{[5]}_{hs}$	$LN^{[7]}_{hs}$	$LN^{[3],s}_{hs}$	$LN^{[5],s}_{hs}$	$LN^{[7],s}_{hs}$
Panel A: High PMI Period																						
$xr^{(1)}_{t+h}$	0.125	0.117	0.026	0.032	0.099	0.083	0.061	0.082	0.061	0.053	0.043	0.034	0.046	0.025	0.037	0.049	0.033	0.032	0.032	0.031	0.032	0.032
$xr^{(2)}_{t+h}$	0.030	0.068	0.036	0.043	0.076	0.078	0.064	0.081	0.054	0.054	0.057	0.050	0.041	0.051	0.050	0.039	0.044	0.043	0.040	0.044	0.043	0.039
$xr^{(3)}_{t+h}$	0.013	0.060	0.036	0.043	0.091	0.056	0.062	0.096	0.055	0.053	0.046	0.058	0.043	0.048	0.057	0.045	0.045	0.048	0.044	0.044	0.049	0.045
$xr^{(4)}_{t+h}$	0.007	0.054	0.031	0.040	0.067	0.082	0.038	0.072	0.051	0.050	0.053	0.041	0.039	0.047	0.038	0.037	0.050	0.040	0.039	0.045	0.038	0.036
$xr^{(5)}_{t+h}$	0.004	0.044	0.024	0.032	0.052	0.060	0.039	0.064	0.038	0.038	0.030	0.040	0.042	0.028	0.037	0.037	0.032	0.037	0.038	0.029	0.035	0.035
$xr^{(7)}_{t+h}$	0.001	0.025	0.010	0.017	0.051	0.021	0.030	0.043	0.016	0.016	0.016	0.023	0.016	0.017	0.021	0.014	0.017	0.021	0.017	0.017	0.020	0.016
$xr^{(10)}_{t+h}$	0.001	0.013	0.003	0.008	0.009	0.027	0.008	0.034	0.008	0.013	0.011	0.011	0.010	0.012	0.011	0.008	0.010	0.010	0.009	0.011	0.010	0.008
Mean	0.026	0.054	0.024	0.031	0.064	0.058	0.043	0.067	0.040	0.040	0.037	0.037	0.034	0.033	0.036	0.033	0.033	0.033	0.031	0.032	0.032	0.030
Panel B: Normal PMI Period																						
$xr^{(1)}_{t+h}$	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(2)}_{t+h}$	0.000	0.000	0.000	0.000	0.002	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(3)}_{t+h}$	0.000	0.000	0.000	0.000	0.037	0.001	0.000	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(4)}_{t+h}$	0.000	0.000	0.000	0.001	0.010	0.002	0.004	0.006	0.000	0.003	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000
$xr^{(5)}_{t+h}$	0.000	0.000	0.000	0.001	0.020	0.001	0.000	0.005	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001
$xr^{(7)}_{t+h}$	0.001	0.000	0.001	0.001	0.031	0.026	0.002	0.006	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000
$xr^{(10)}_{t+h}$	0.004	0.001	0.001	0.001	0.021	0.014	0.004	0.007	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001
Mean	0.001	0.000	0.000	0.001	0.017	0.007	0.001	0.004	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Panel C: Low PMI Period																						
$xr^{(1)}_{t+h}$	0.002	0.006	0.004	0.005	0.007	0.005	0.007	0.003	0.002	0.005	0.003	0.004	0.003	0.002	0.003	0.004	0.003	0.004	0.003	0.002	0.003	0.004
$xr^{(2)}_{t+h}$	0.003	0.008	0.005	0.007	0.030	0.009	0.010	0.015	0.008	0.006	0.005	0.006	0.005	0.005	0.006	0.005	0.005	0.006	0.005	0.005	0.006	0.005
$xr^{(3)}_{t+h}$	0.003	0.015	0.009	0.012	0.039	0.024	0.015	0.041	0.012	0.016	0.011	0.015	0.014	0.013	0.016	0.015	0.011	0.015	0.014	0.013	0.016	0.015
$xr^{(4)}_{t+h}$	0.004	0.022	0.012	0.017	0.048	0.028	0.019	0.043	0.017	0.016	0.018	0.018	0.015	0.016	0.019	0.015	0.018	0.018	0.015	0.016	0.019	0.015
$xr^{(5)}_{t+h}$	0.004	0.027	0.014	0.019	0.042	0.037	0.026	0.038	0.020	0.019	0.015	0.023	0.019	0.017	0.024	0.019	0.015	0.023	0.019	0.017	0.024	0.019
$xr^{(7)}_{t+h}$	0.005	0.032	0.016	0.022	0.042	0.041	0.033	0.043	0.027	0.025	0.021	0.024	0.021	0.018	0.025	0.021	0.021	0.024	0.021	0.018	0.025	0.021
$xr^{(10)}_{t+h}$	0.007	0.037	0.019	0.025	0.047	0.036	0.037	0.052	0.026	0.024	0.025	0.028	0.030	0.025	0.029	0.027	0.025	0.028	0.030	0.025	0.029	0.027
Mean	0.004	0.021	0.011	0.015	0.036	0.026	0.021	0.034	0.016	0.016	0.014	0.017	0.015	0.014	0.017	0.015	0.014	0.017	0.015	0.014	0.017	0.015
Panel D: All Periods																						
$xr^{(1)}_{t+h}$	0.000	0.000	0.000	0.000	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(2)}_{t+h}$	0.000	0.000	0.000	0.000	0.007	0.002	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(3)}_{t+h}$	0.000	0.001	0.000	0.000	0.044	0.002	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(4)}_{t+h}$	0.000	0.001	0.000	0.000	0.014	0.005	0.002	0.009	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$xr^{(5)}_{t+h}$	0.000	0.001	0.000	0.000	0.014	0.005	0.001	0.007	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.000
$xr^{(7)}_{t+h}$	0.000	0.001	0.000	0.000	0.020	0.013	0.002	0.007	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(10)}_{t+h}$	0.000	0.001	0.000	0.000	0.008	0.007	0.002	0.011	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000
Mean	0.000	0.001	0.000	0.000	0.016	0.005	0.001	0.008	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 7: This table reports p-value for the null hypothesis $R^2_{OOS} < 0$ for one-month U.S. Treasury bond excess returns forecasts using forward rates $\{f^{(j)}_t\}^{10}_{j=1}$, TB, MOM, YC2, and CPITR as predictors. Out-of-sample evaluation begins in January 1990 and is presented for the full period, and separately for high, normal, and low PMI periods. The models are identical to those in Table 4.

	R^2 OOS												
	OLS	*L1*	*L2*	*ENet*	$NN^{[3]}$	$NN^{[5]}$	$NN^{[7]}$	$NN^{[3]}_{ss}$	$NN^{[5]}_{ss}$	$NN^{[7]}_{ss}$	$LN^{[3]}$	$LN^{[5]}$	$LN^{[7]}$
Panel A: High PMI Period													
$xr^{(1)}_{t+h}$	-78.59	0.45	-1.05	0.82	1.66	11.12	-0.58	5.63	4.50	5.21	0.26	-7.12	-2.98
$xr^{(2)}_{t+h}$	-28.81	4.49	7.65	7.28	4.56	7.18	8.29	4.86	6.90	6.56	2.92	6.32	4.99
$xr^{(3)}_{t+h}$	-13.52	5.46	8.46	7.82	3.02	7.80	5.23	6.19	7.81	7.26	7.21	6.00	5.83
$xr^{(4)}_{t+h}$	-5.10	6.16	8.71	8.01	7.56	6.42	5.34	6.20	8.06	8.26	4.16	7.59	6.20
$xr^{(5)}_{t+h}$	0.12	7.13	9.22	8.59	6.40	6.49	8.90	5.09	8.11	9.44	5.39	8.04	7.83
$xr^{(7)}_{t+h}$	5.27	9.47	10.87	10.38	9.70	8.76	9.78	9.83	8.19	9.89	8.33	9.32	9.69
$xr^{(10)}_{t+h}$	3.87	12.22	13.08	12.72	10.27	11.43	10.22	11.93	12.62	12.17	12.20	12.13	12.53
Mean	-16.68	6.48	8.13	7.95	6.17	8.46	6.74	7.10	8.03	8.40	5.78	6.04	6.30
Panel B: Normal PMI Period													
$xr^{(1)}_{t+h}$	-28.64	7.04	5.27	5.89	1.72	3.48	7.09	7.47	8.96	7.74	8.28	4.42	6.08
$xr^{(2)}_{t+h}$	-17.42	5.59	4.88	5.32	2.48	1.77	2.58	5.49	7.39	4.83	6.39	7.03	4.59
$xr^{(3)}_{t+h}$	-11.24	4.93	4.34	4.77	0.90	2.44	2.27	4.27	5.11	4.34	5.68	5.92	6.01
$xr^{(4)}_{t+h}$	-7.06	4.71	4.19	4.58	0.91	0.78	0.84	3.42	5.02	2.80	6.20	4.64	6.02
$xr^{(5)}_{t+h}$	-4.53	4.65	4.20	4.56	1.24	2.72	2.90	3.95	4.73	2.64	5.54	5.24	4.49
$xr^{(7)}_{t+h}$	-2.67	4.65	4.25	4.57	-1.63	3.24	1.75	3.86	3.81	3.49	5.56	5.34	5.03
$xr^{(10)}_{t+h}$	-4.22	4.61	4.09	4.33	-2.50	0.36	-2.04	3.81	2.61	3.51	4.69	4.60	3.65
Mean	-10.83	5.17	4.46	4.86	0.45	2.11	2.20	4.61	5.38	4.19	6.05	5.31	5.12
Panel C: Low PMI Period													
$xr^{(1)}_{t+h}$	0.06	10.30	12.33	12.41	3.83	2.42	6.96	8.40	13.05	10.30	13.94	13.39	13.15
$xr^{(2)}_{t+h}$	8.80	9.75	11.05	10.97	6.52	6.13	4.95	8.16	10.97	9.74	11.86	12.49	12.25
$xr^{(3)}_{t+h}$	13.11	9.91	11.20	10.99	5.09	8.20	8.02	9.08	10.08	10.54	11.41	12.27	11.62
$xr^{(4)}_{t+h}$	15.42	9.85	11.28	11.17	5.78	7.16	6.81	10.05	10.42	10.53	13.00	11.69	11.90
$xr^{(5)}_{t+h}$	16.36	9.59	11.18	11.03	7.90	8.90	8.13	10.84	10.31	9.54	12.25	11.14	10.94
$xr^{(7)}_{t+h}$	16.04	8.85	10.59	10.47	6.17	8.91	8.29	9.47	10.13	10.52	11.34	11.14	10.79
$xr^{(10)}_{t+h}$	13.82	8.06	9.74	9.67	6.02	6.93	5.72	9.56	9.25	9.83	10.36	10.34	9.16
Mean	11.94	9.47	11.05	10.96	5.90	6.95	6.98	9.37	10.60	10.14	12.02	11.78	11.40
Panel D: All Periods													
$xr^{(1)}_{t+h}$	-22.22	7.69	7.18	7.68	2.35	3.47	6.56	7.59	9.97	8.36	9.58	6.65	7.83
$xr^{(2)}_{t+h}$	-11.39	5.85	6.05	6.26	2.98	2.62	2.82	5.38	7.54	5.54	6.85	7.70	6.02
$xr^{(3)}_{t+h}$	-5.83	5.14	5.37	5.53	1.03	3.27	2.92	4.54	5.49	5.09	6.16	6.48	6.35
$xr^{(4)}_{t+h}$	-2.10	4.87	5.15	5.31	1.39	1.63	1.46	4.08	5.37	3.99	6.55	5.47	6.30
$xr^{(5)}_{t+h}$	0.16	4.76	5.09	5.23	2.04	3.28	3.36	4.51	5.07	3.62	5.95	5.65	5.09
$xr^{(7)}_{t+h}$	1.81	4.68	5.07	5.20	-0.07	3.77	2.72	4.39	4.44	4.48	5.93	5.80	5.53
$xr^{(10)}_{t+h}$	0.43	4.63	4.95	5.04	-0.45	1.67	-0.27	4.62	3.85	4.55	5.45	5.36	4.43
Mean	-5.59	5.37	5.55	5.75	1.32	2.82	2.80	5.02	5.96	5.09	6.64	6.16	5.94

	p-value												
	OLS	*L1*	*L2*	*ENet*	$NN^{[3]}$	$NN^{[5]}$	$NN^{[7]}$	$NN^{[3]}_{ss}$	$NN^{[5]}_{ss}$	$NN^{[7]}_{ss}$	$LN^{[3]}$	$LN^{[5]}$	$LN^{[7]}$
Panel A: High PMI Period													
$xr^{(1)}_{t+h}$	0.205	0.088	0.026	0.029	0.091	0.040	0.077	0.046	0.044	0.040	0.039	0.067	0.041
$xr^{(2)}_{t+h}$	0.043	0.051	0.026	0.029	0.073	0.065	0.054	0.069	0.047	0.047	0.042	0.030	0.026
$xr^{(3)}_{t+h}$	0.010	0.046	0.025	0.029	0.088	0.052	0.069	0.051	0.045	0.044	0.030	0.035	0.032
$xr^{(4)}_{t+h}$	0.003	0.042	0.021	0.026	0.061	0.070	0.069	0.053	0.036	0.034	0.040	0.022	0.030
$xr^{(5)}_{t+h}$	0.001	0.034	0.015	0.020	0.067	0.055	0.041	0.058	0.036	0.027	0.031	0.020	0.016
$xr^{(7)}_{t+h}$	0.000	0.018	0.005	0.008	0.037	0.030	0.029	0.023	0.038	0.022	0.006	0.011	0.007
$xr^{(10)}_{t+h}$	0.000	0.009	0.002	0.003	0.024	0.015	0.021	0.021	0.008	0.007	0.005	0.002	0.002
Mean	0.037	0.041	0.017	0.021	0.063	0.047	0.051	0.046	0.036	0.032	0.028	0.027	0.022
Panel B: Normal PMI Period													
$xr^{(1)}_{t+h}$	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(2)}_{t+h}$	0.000	0.001	0.000	0.000	0.009	0.008	0.004	0.001	0.000	0.000	0.000	0.000	0.000
$xr^{(3)}_{t+h}$	0.000	0.001	0.000	0.000	0.021	0.007	0.009	0.001	0.000	0.000	0.000	0.000	0.000
$xr^{(4)}_{t+h}$	0.000	0.001	0.000	0.000	0.018	0.019	0.018	0.002	0.001	0.001	0.000	0.000	0.000
$xr^{(5)}_{t+h}$	0.001	0.001	0.000	0.000	0.018	0.006	0.004	0.002	0.000	0.002	0.000	0.000	0.000
$xr^{(7)}_{t+h}$	0.001	0.001	0.001	0.001	0.036	0.003	0.013	0.002	0.001	0.001	0.000	0.000	0.000
$xr^{(10)}_{t+h}$	0.005	0.001	0.001	0.001	0.074	0.035	0.050	0.002	0.002	0.001	0.000	0.000	0.001
Mean	0.001	0.001	0.000	0.000	0.026	0.011	0.014	0.001	0.001	0.001	0.000	0.000	0.000
Panel C: Low PMI Period													
$xr^{(1)}_{t+h}$	0.002	0.006	0.004	0.005	0.008	0.006	0.002	0.007	0.004	0.006	0.009	0.004	0.005
$xr^{(2)}_{t+h}$	0.004	0.017	0.009	0.011	0.018	0.012	0.016	0.021	0.010	0.011	0.012	0.010	0.008
$xr^{(3)}_{t+h}$	0.005	0.027	0.015	0.018	0.045	0.028	0.029	0.024	0.025	0.019	0.020	0.020	0.017
$xr^{(4)}_{t+h}$	0.006	0.033	0.017	0.021	0.043	0.043	0.043	0.041	0.029	0.024	0.022	0.016	0.023
$xr^{(5)}_{t+h}$	0.006	0.036	0.018	0.022	0.051	0.051	0.046	0.037	0.038	0.026	0.023	0.023	0.022
$xr^{(7)}_{t+h}$	0.007	0.039	0.020	0.023	0.047	0.054	0.041	0.040	0.034	0.025	0.020	0.024	0.021
$xr^{(10)}_{t+h}$	0.010	0.043	0.023	0.027	0.051	0.056	0.038	0.034	0.029	0.027	0.024	0.024	0.028
Mean	0.006	0.029	0.015	0.018	0.038	0.036	0.031	0.029	0.024	0.020	0.019	0.017	0.018
Panel D: All Periods													
$xr^{(1)}_{t+h}$	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$xr^{(2)}_{t+h}$	0.000	0.000	0.000	0.000	0.007	0.004	0.003	0.001	0.000	0.000	0.000	0.000	0.000
$xr^{(3)}_{t+h}$	0.000	0.001	0.000	0.000	0.027	0.005	0.009	0.001	0.000	0.000	0.000	0.000	0.000
$xr^{(4)}_{t+h}$	0.000	0.001	0.000	0.000	0.017	0.016	0.020	0.004	0.001	0.000	0.000	0.000	0.000
$xr^{(5)}_{t+h}$	0.000	0.001	0.000	0.000	0.020	0.010	0.005	0.003	0.001	0.001	0.000	0.000	0.000
$xr^{(7)}_{t+h}$	0.000	0.001	0.000	0.000	0.022	0.006	0.008	0.002	0.001	0.001	0.000	0.000	0.000
$xr^{(10)}_{t+h}$	0.000	0.001	0.000	0.000	0.037	0.018	0.021	0.002	0.001	0.001	0.000	0.000	0.000
Mean	0.000	0.001	0.000	0.000	0.019	0.008	0.009	0.002	0.001	0.000	0.000	0.000	0.000

Table 8: This table reports R^2_{OOS} (in percentage) and corresponding p-value for the null hypothesis $R^2_{OOS} < 0$, based on one-month U.S. Treasury bond excess returns forecasts using forward rates $\{f^{(j)}_t\}^{10}_{j=1}$, TB, MOM, YC2, CPITR, and PMI as predictors. Out-of-sample evaluation begins in January 1990 and is presented for the full period, and separately for high, normal, and low PMI periods. Forecasting models include: (i) linear models (OLS, Lasso (L1), Ridge (L2), and Elastic Net (ENet)); (ii) MLP models ($NN^{[3,5,7]}$); (iii) soft-switch on MLPs ($NN^{[3,5,7]}_{ss}$); (iv) ENet–MLP hybrids ($LN^{[3,5,7]}$.

	$xr^{(1)}_{t+h}$	$xr^{(2)}_{t+h}$	$xr^{(3)}_{t+h}$	$xr^{(4)}_{t+h}$	$xr^{(5)}_{t+h}$	$xr^{(7)}_{t+h}$	$xr^{(10)}_{t+h}$	Joint
Panel A: Power Utility								
Forward rates Only								
OLS	−0.16	0.22	0.32	0.51	0.72*	0.91**	0.63*	−2.57
ENet	−0.12	−0.05	0.09	0.39	0.63**	0.84**	0.65*	−7.19
$NN^{[3]}$	0.03**	0.31***	−0.06	0.19**	0.19***	0.10***	−0.03	0.50
$NN^{[3]}_{ss}$	0.05*	0.28***	−0.13	−0.03	0.21***	−0.13	0.17***	5.65**
$LN^{[3],s}_{hs}$	−0.09	0.39	0.36	0.58*	0.93***	0.96***	0.66*	1.32
Forward rates + 4 predictors								
OLS	−0.02	0.38	0.50	0.72	0.91*	1.12**	0.96**	−1.10
ENet	−0.34	−0.32	−0.07	0.06	0.14	0.22	0.28*	3.82**
$NN^{[7]}$	0.06	0.03	0.16	−0.14	0.17	0.15	0.08	7.45***
$NN^{[5]}_{ss}$	−0.00	−0.13	−0.17	0.08	0.22	0.20	0.29*	−2.01
$LN^{[3],s}_{hs}$	−0.10	0.32	−0.08	0.38**	0.28	0.31*	0.32**	4.14
Forward rates + 5 predictors								
OLS	−0.05	0.41	0.60	0.80*	1.06**	1.24**	0.99**	−2.97
ENet	−0.07	0.03	0.18	0.27	0.32	0.38*	0.41*	2.89
$NN^{[5]}$	−0.25	−0.03	0.01	−0.07	0.12*	0.22**	−0.02	0.54
$NN^{[5]}_{ss}$	0.03	0.15	0.20	0.29	0.27*	0.29**	0.21	0.64
$LN^{[3]}$	0.11	0.26	0.28	0.48**	0.44**	0.55**	0.48**	−0.60
Panel B: Mean Variance Utility								
Forward rates Only								
OLS	−0.17	0.22	0.31	0.49	0.69*	0.89**	0.62*	−8.64
ENet	−0.12	−0.06	0.08	0.38	0.61*	0.83**	0.65*	−9.27
NN3	0.03**	0.31***	−0.06	0.19**	0.19***	0.10***	−0.04	−4.52
$NN^{[3]}_{ss}$	0.05	0.28***	−0.14	−0.04	0.21***	−0.13	0.18***	3.16
$LN^{[3],s}_{hs}$	−0.09	0.38	0.37	0.57*	0.92***	0.96***	0.67*	0.40
Forward rates + 4 predictors								
OLS	−0.02	0.38	0.49	0.69	0.86*	1.08**	0.92**	−5.46
ENet	−0.36	−0.34	−0.08	0.05	0.13	0.21	0.25	0.90
NN7	0.06	0.01	0.15	−0.16	0.16	0.14	0.05	0.98
$NN^{[5]}_{ss}$	−0.01	−0.15	−0.18	0.07	0.20	0.18	0.27	−5.16
$LN^{[3],s}_{hs}$	−0.12	0.30	−0.09	0.37**	0.27	0.29*	0.30*	0.38
Forward rates + 5 predictors								
OLS	−0.05	0.41	0.59	0.79*	1.00**	1.21**	0.97**	−6.21
ENet	−0.09	0.01	0.16	0.26	0.30	0.36*	0.39*	−0.32
NN5	−0.28	−0.04	0.01	−0.08	0.11*	0.22*	−0.04	0.88
$NN^{[5]}_{ss}$	0.03	0.13	0.19	0.28	0.26*	0.29**	0.19	−1.19
$LN^{[3]}$	0.11	0.24	0.26	0.48**	0.43**	0.53**	0.46**	−1.25

Table 9: This table reports certainty-equivalent return (CER) gains, expressed in percentage terms, over the expectations hypothesis (EH) benchmark. Results are presented for models estimated using forward rates only as well as for models that augment forward rates with additional predictors. For each predictor set and each model category, we report results for the best-performing model. CER gains are computed under both power-utility and mean–variance investor preferences. Significance levels are denoted by *, **, and *** for 10%, 5%, and 1%, respectively.

	$xr^{(1)}_{t+h}$	$xr^{(2)}_{t+h}$	$xr^{(3)}_{t+h}$	$xr^{(4)}_{t+h}$	$xr^{(5)}_{t+h}$	$xr^{(7)}_{t+h}$	$xr^{(10)}_{t+h}$	Joint
Panel A: Power Utility								
Forward rates Only								
OLS	0.636	0.464	0.390	0.354	0.355	0.337	0.293	0.095
ENet	0.625	0.577	0.544	0.522	0.463	0.368	0.318	0.043
NN3	0.563	0.426	0.533	0.510	0.459	0.483	0.506	0.095
$NN^{[3]}_{ss}$	0.584	0.491	0.644	0.608	0.499	0.553	0.448	0.192
$LN^{[3],s}_{hs}$	0.618	0.518	0.488	0.520	0.439	0.355	0.316	0.132
Forward rates + 4 predictors								
OLS	0.665	0.485	0.400	0.371	0.369	0.359	0.331	0.103
ENet	0.700	0.711	0.747	0.743	0.708	0.627	0.551	0.157
NN7	0.608	0.636	0.682	0.841	0.676	0.600	0.677	0.223
$NN^{[5]}_{ss}$	0.693	0.696	0.705	0.675	0.697	0.678	0.491	0.125
$LN^{[3],s}_{hs}$	0.745	0.691	0.735	0.735	0.677	0.582	0.534	0.167
Forward rates + 5 predictors								
OLS	0.645	0.482	0.408	0.375	0.354	0.334	0.307	0.081
ENet	0.654	0.635	0.687	0.686	0.650	0.566	0.493	0.162
NN5	0.730	0.580	0.752	0.615	0.610	0.523	0.614	0.100
$NN^{[5]}_{ss}$	0.642	0.652	0.671	0.663	0.608	0.549	0.509	0.136
$LN^{[3]}$	0.642	0.666	0.629	0.627	0.628	0.549	0.479	0.110
EH	0.549	0.456	0.449	0.476	0.488	0.480	0.441	0.127
Panel B: Mean Variance Utility								
Forward rates Only								
OLS	0.637	0.457	0.378	0.338	0.332	0.320	0.291	0.050
ENet	0.628	0.579	0.519	0.500	0.448	0.364	0.330	0.019
NN3	0.568	0.439	0.594	0.597	0.557	0.682	0.857	0.063
$NN^{[3]}_{ss}$	0.592	0.510	0.735	0.730	0.611	0.799	0.730	0.217
$LN^{[3],s}_{hs}$	0.622	0.530	0.489	0.513	0.442	0.359	0.334	0.178
Forward rates + 4 predictors								
OLS	0.663	0.477	0.382	0.348	0.337	0.328	0.309	0.079
ENet	0.707	0.723	0.755	0.767	0.747	0.686	0.636	0.304
NN7	0.619	0.680	0.759	0.809	0.777	0.803	0.706	0.178
$NN^{[5]}_{ss}$	0.707	0.726	0.731	0.741	0.751	0.723	0.588	0.113
$LN^{[3],s}_{hs}$	0.754	0.720	0.761	0.824	0.710	0.648	0.666	0.171
Forward rates + 5 predictors								
OLS	0.645	0.470	0.391	0.350	0.327	0.310	0.289	0.083
ENet	0.660	0.645	0.689	0.692	0.664	0.592	0.537	0.156
NN5	0.740	0.606	0.854	0.733	0.796	0.704	0.912	0.175
$NN^{[5]}_{ss}$	0.650	0.673	0.709	0.725	0.699	0.681	0.589	0.163
$LN^{[3]}$	0.649	0.681	0.633	0.688	0.671	0.585	0.524	0.144
EH	0.552	0.471	0.486	0.551	0.601	0.678	0.749	0.165

Table 10: This table reports the Sharpe ratios (SR) under power-utility (Panel A) and mean–variance (Panel B) preferences. We report models estimated using forward rates only, as well as forward rates augmented with additional predictors. For each predictor set and model category, we report the SR of the best-performing model. The expectations hypothesis (EH) benchmark is shown at the bottom of each panel for comparison.

Figures

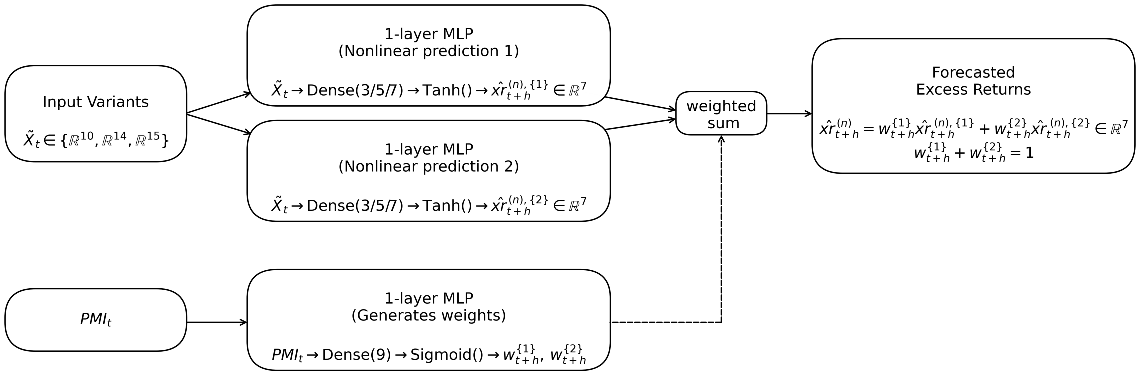


Figure 1: Architecture of the nonlinear soft-switch on MLPs forecasting model. The predictor vector $\tilde{X}_t^i$ belongs to one of three sets: $\mathbb{R}^{10}$, containing the ten forward rates $\{f_t^{(j)}\}_{j=1}^{10}$; $\mathbb{R}^{14}$, augmenting these forward rates with four additional predictors:MOM, FINUNC, CPITR, YC2; or $\mathbb{R}^{15}$, further including the PMI. The predictor vector $\tilde{X}_t$ is processed by two parallel one-layer MLPs, producing nonlinear excess-return forecasts $\hat{xr}_{t+h}^{(n),\{1\}}$ and $\hat{xr}_{t+h}^{(n),\{2\}}$ for maturities $n \in \{1, 2, 3, 4, 5, 7, 10\}$. A third MLP maps PMI_t into weights $w_{t+h}^{\{1\}}$ and $w_{t+h}^{\{2\}}$, constrained to satisfy $w_{t+h}^{\{1\}} + w_{t+h}^{\{2\}} = 1$. The final forecast $\hat{xr}_{t+h}^{n}$ is a weighted combination of the two nonlinear predictions.

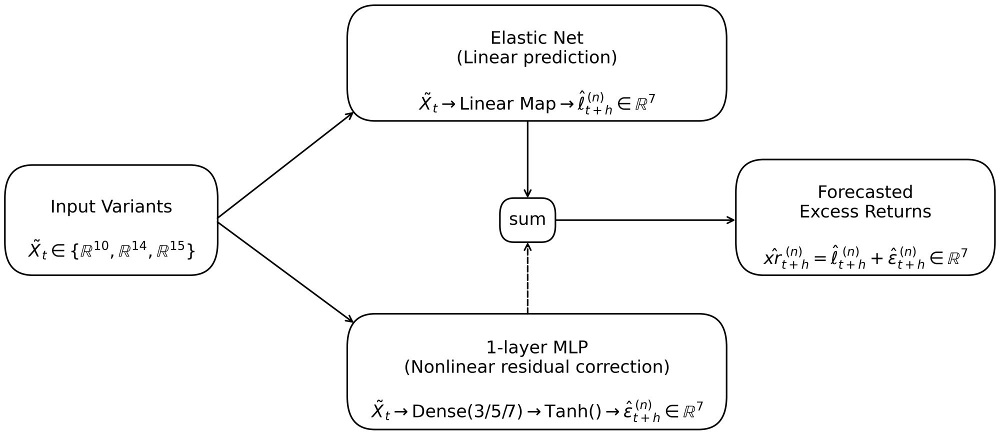


Figure 2: Architecture of the Elastic Net–MLP hybrid model for forecasting one-month excess bond returns. The predictor vector $\tilde{X}^i_t$ belongs to one of three sets: $\mathbb{R}^{10}$, containing the ten forward rates $\{f^{(j)}_t\}^{10}_{j=1}$; $\mathbb{R}^{14}$, augmenting these forward rates with four additional predictors:MOM, FINUNC, CPITR, YC2; or $\mathbb{R}^{15}$, further including the PMI. The Elastic Net produces a linear forecast $\hat{\ell}^{(n)}_{t+h}$, and the one-layer MLP estimates the nonlinear residual $\hat{\varepsilon}^{(n)}_{t+h}$. The final prediction $\{xr^{(n)}_{t+h}\}_{n\in\{1,2,3,4,5,7,10\}}$ is their sum. In the hard-switch models, the MLP component $\hat{\varepsilon}^{n}_{t+h}$ is included only when the PMI is below its 80th-percentile threshold.

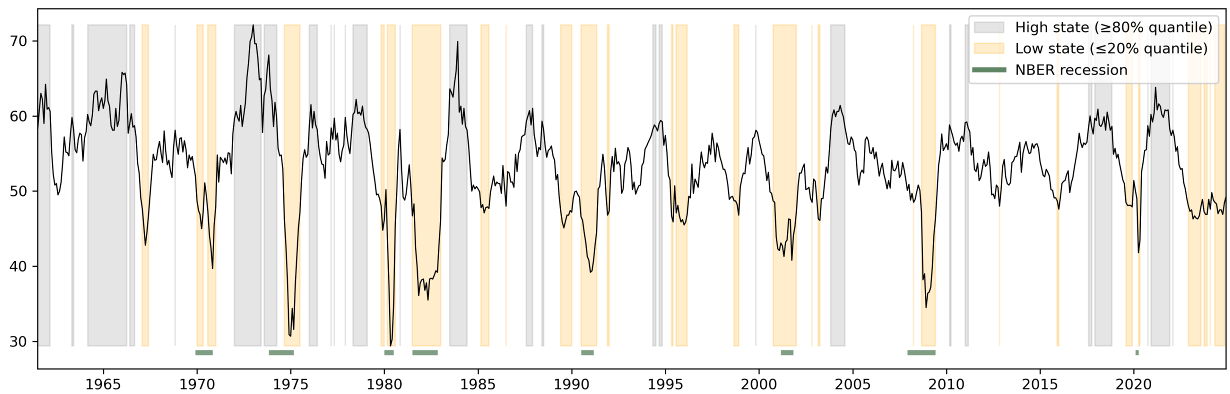
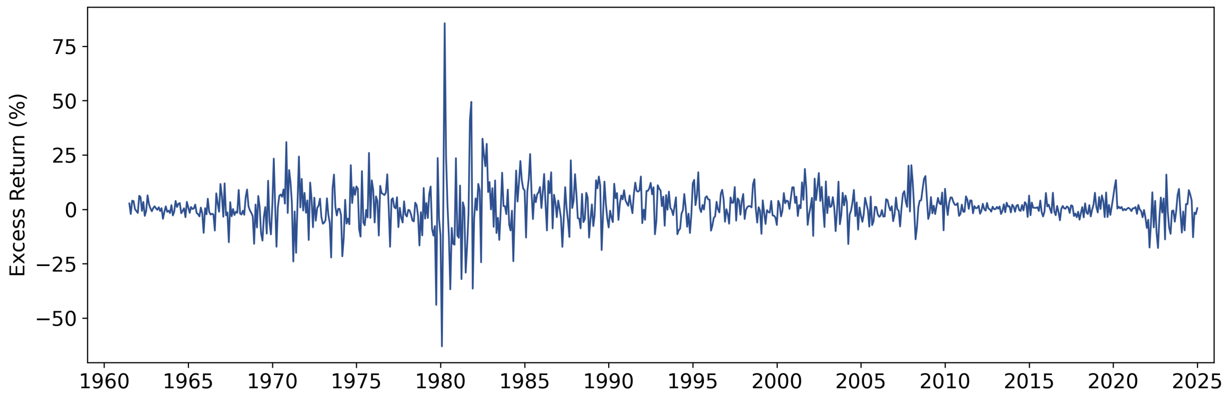
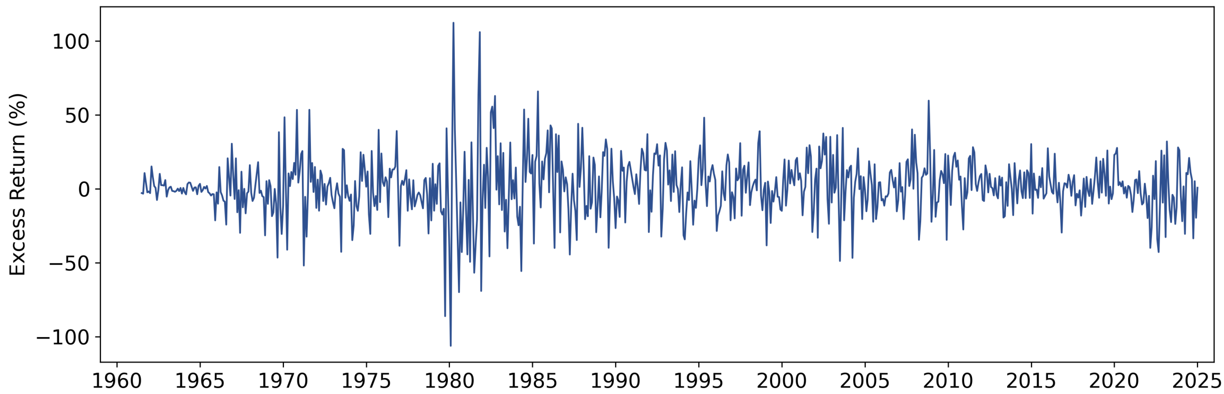


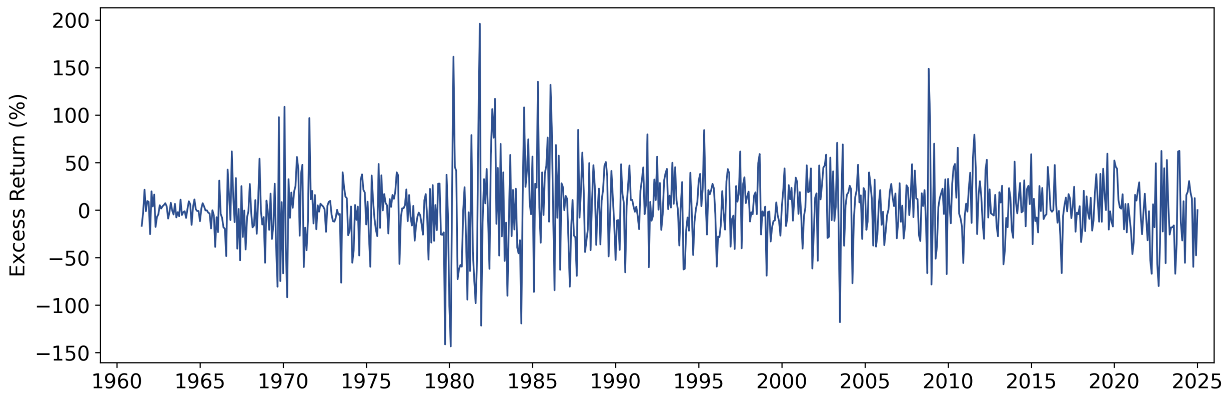
Figure 3: This figure shows the time series of the Purchasing Managers' Index (PMI). Shaded areas indicate periods of high and low activity, defined by the 80% and 20% quantiles of the time series. The sample covers from July 1961 to December 2024. The green bars at the bottom indicate NBER recession periods.



(a) 2-year excess return



(b) 5-year excess return



(c) 10-year excess return

Figure 4: This figure shows the time series of 1-month annualized U.S. Treasury bond excess returns, measured in percentage terms, for the 2-, 5-, and 10-year maturities. The sample covers August 1961 to January 2025. Excess returns are computed from the GSW term structure data and are measured relative to the 1-month T-bill rate.

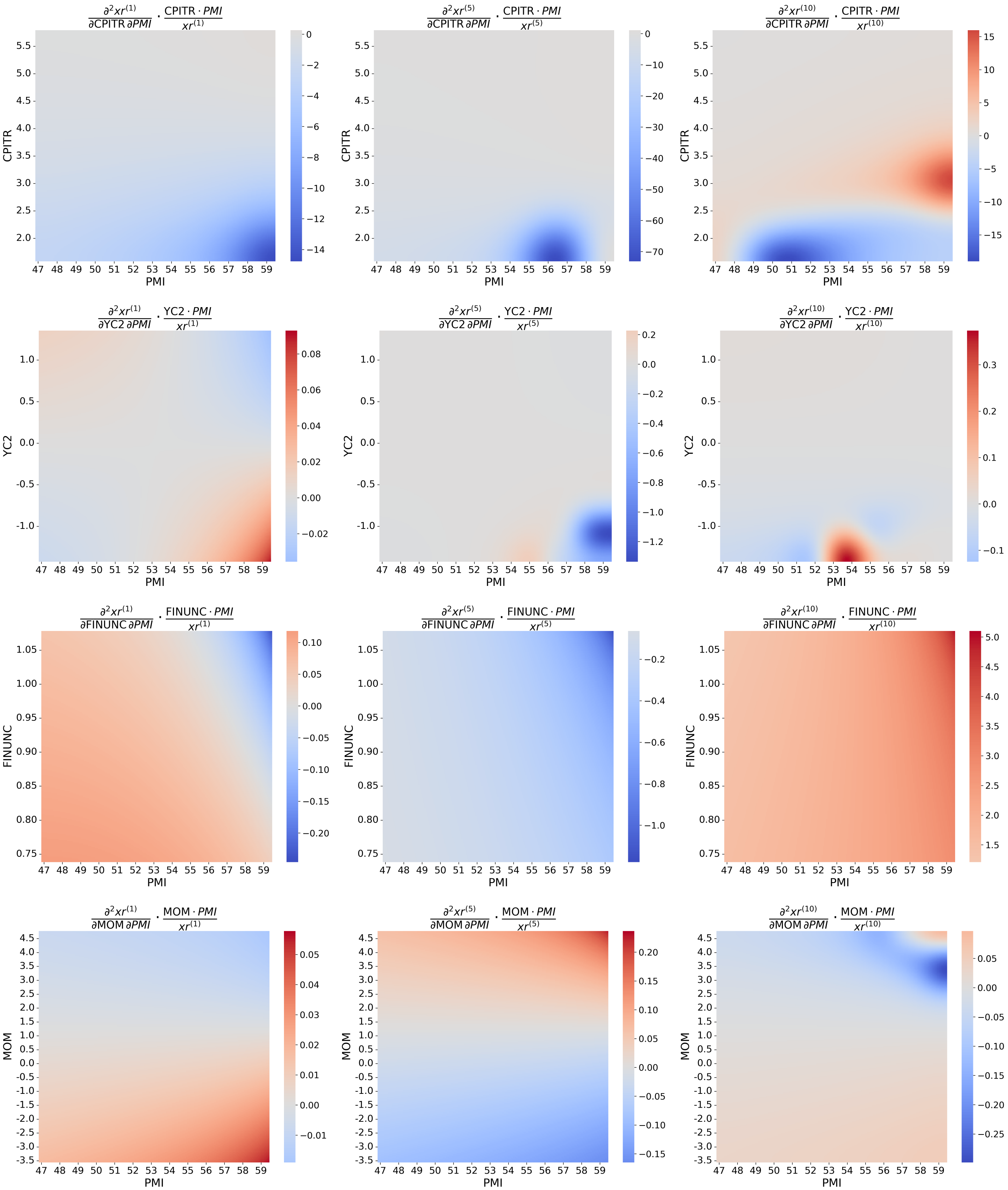


Figure 5: These heatmaps show the cross elasticity of the excess Treasury HPR $\{xr^{(n)}\}_{n=1,5,10}$ with respect to changes in PMI and one additional predictor (CPITR, YC2, FINUNC, and MOM). Heatmap values are in percentage.

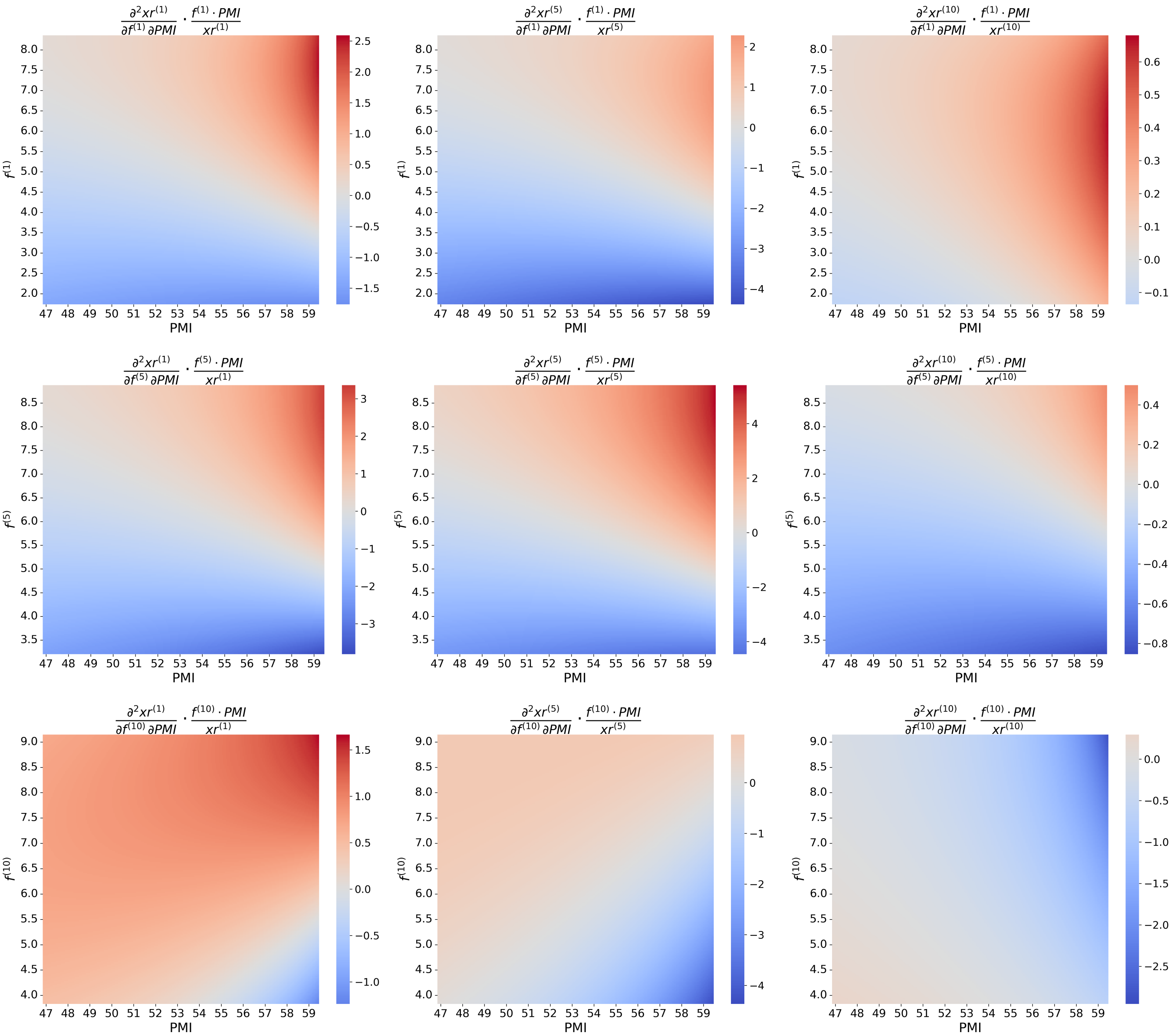


Figure 6: These heatmaps show the cross elasticity of the excess Treasury HPR $\{xr^{(n)}\}_{n=1,5,10}$ with respect to changes in PMI and one forward rate $\{f^{(j)}\}_{j=1,5,10}$. Heatmap values are in percentage.

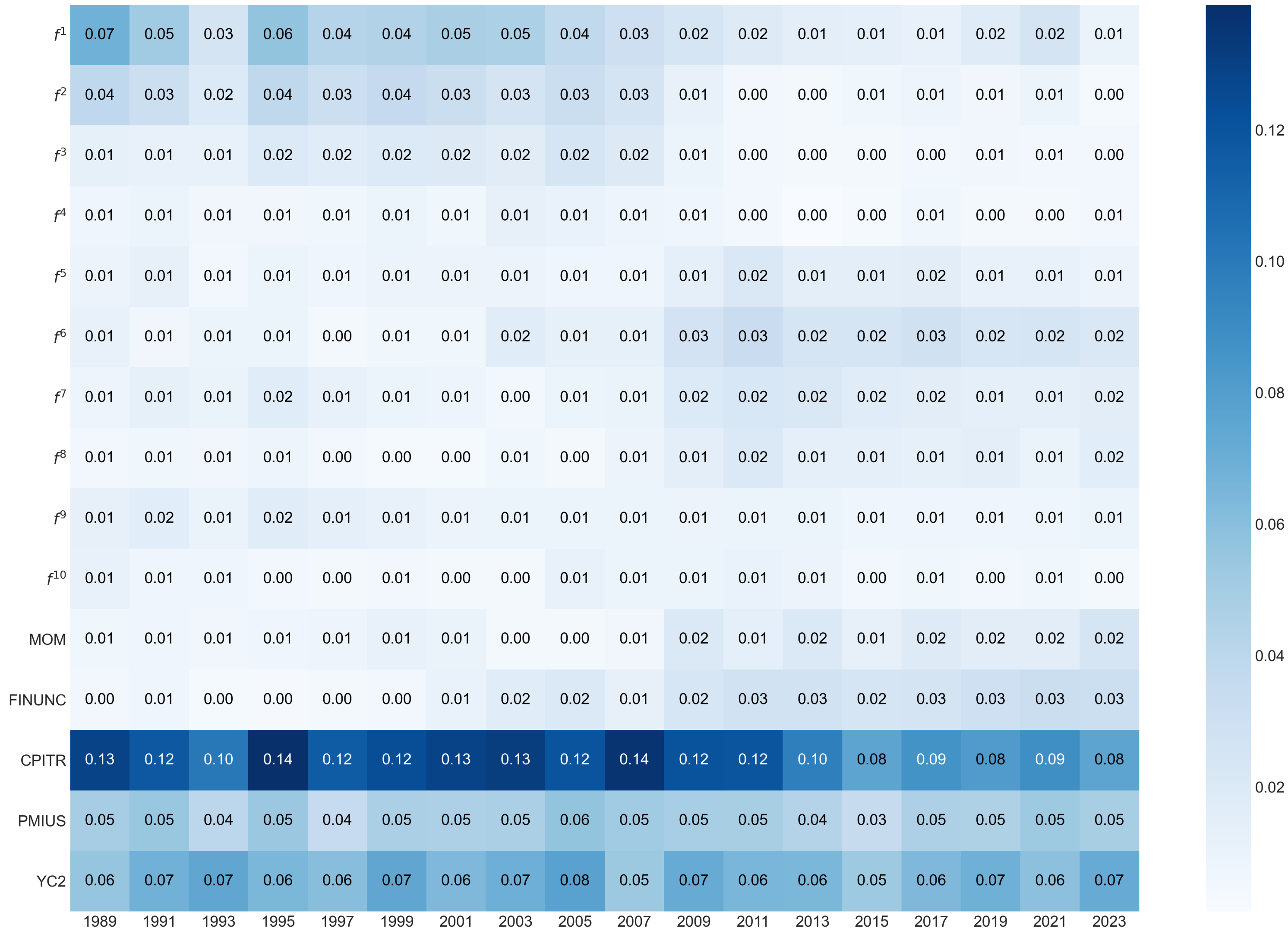


Figure 7: This figure reports a SHAP value heatmap showing the time-varying importance of the forward rates $\{f_t^{(j)}\}_{j=1}^{10}$ and additional predictors in the best-performing $LN^{[3]}$ forecasting model. Each cell shows the mean absolute SHAP contribution for a given input over the corresponding forecast window. The first set of models are estimated in December 1989. Subsequently, parameters are re-estimated every 2 years using an expanding window that incorporates all the available data up to each re-estimation date.