A Machine Learning Factor-Based Interpretation for the Bond Risk Premia in the U.S.

Caio Vigo-Pereira

First version: March 4, 2020

This version: January 31, 2021

Click here for the latest version

Abstract

In this paper, we study the time variation of the risk premia in U.S. Treasuries bonds. We propose a novel approach for deriving a single spanning state factor consistent with a dynamic term-structure with unspanned risks theoretically motivated model. Using deep neural networks to uncover relationships in the full set of information from the yield curve, we derive a single state variable factor that provides a better approximation to the spanned space of all the information from the term-structure. We also introduce a way to obtain unspanned risks from the yield curve that is used to complete our state space. We show that this parsimonious number of state variables have predictive power for excess returns of bonds over 1-month holding period. Additionally, we provide an intuitive interpretation of derived factors and show what information from macroeconomic variables and sentiment-based measures they can capture.

JEL classification: G12, E43, E44, E47.

Keywords: Bond Premia. Deep Learning. Machine Learning. Bond Returns. Yield Curve. Unspanned Risk Factors.

1 Introduction

In recent years, many studies had shed light on a critical assumption in macro-finance models, the expectations hypothesis. As more evidence is gathered, there is a growing consensus in the literature to refute it, implying that excess returns of Treasuries bonds in some extent should be forecastable. Equally important is the spanning hypothesis, that can be summarized in the idea that the yield curve incorporates all the information useful for forecasting interest rates, and consequently, bonds returns. However, to what extent the spanning hypothesis holds true is still open in the literature.

An important question that could assist to elucidate the whole bond premia problem is related with the factor structure of expected returns. Is there a factor representation? If so, what is its structure? In this article, we study the time variation of the risk premia in U.S. Treasuries bonds. We provide a new approach for the factor structure of the expected returns of bonds. Recently, Cochrane (2015) argued that it is possible that there is a dominant single factor structure for bond returns, in such a way that risk premiums rise and fall together. A central question, in his words, is: *what is the linear combination of forecasting variables that captures common movement in expected returns across assets*?

In Cochrane and Piazzesi (2005), the authors took this path. Ludvigson and Ng (2009) derived a single factor as well, however not consistent with the spanning hypothesis. Recent papers (see, e.g., Cieslak and Povala (2015); Lee (2018)) obtained other factors as well, some of them not necessarily aligned with the spanning hypothesis. Nonetheless, Bauer and Hamilton (2018) argued that evidence against the spanning hypothesis for several recent studies should be weaker when more robust tests are used.

In this paper, we take a different route. We argue that this search for deriving, building and estimating factors that represent state variables in macro-finance models may be limited. We claim that the process done by financial economists of manually discovering and hand picking this list of factors may be leaving out unseen relationships between the state variables in their derivation.

We propose a novel approach for deriving a single state factor consistent with a dynamic term-structure with unspanned risks theoretically motivated model. To do so, we make use of one of the most powerful approaches in machine learning, namely a deep neural network to uncover relationships in the full set of information from the yield curve. We derive a single state variable factor that should provide a better approximation to the spanned space of all the information from the term-structure.

In our methodology, we introduce a way to obtain unspanned risks from the yield curve that is used to complete our state space. This unspanned factor can fill the gap left by

the spanning factor. The whole structure can be explained by a dynamic term-structure with unspanned risks, be macroeconomic, sentiment, or any other economy risk (since our methodology makes no differentiation or segregation among them) as an extension from the model proposed by Joslin et al. (2014). We show that a small numbers of state variables (in our framework only two), have predictive power for excess returns of bonds over 1-month holding period. Additionally, we provide an intuitive interpretation of derived factors, and show what information from macroeconomic variables and sentiment-based measures they can capture.

In our empirical analysis, we use the daily treasury parameters from Gürkaynak et al. (2007) to build the monthly excess returns. As we are interested in the short part of term structure, we deal with yields ranging up to five years ahead. Thus, we work with 60 observations at each month to better capture the information from the term structure. Our data spans from 1962 to 2017, and we use the data from 1962-1992 to initialize the process of obtaining our recursively updated latent factors in order to build a single spanning and unspanning factor for the period from 1993-2017.

The recursive process is done through a pre-designed architecture of a deep neural network that is fed solely with the high-dimensional set of monthly with different maturities. This architecture generates one latent factor for the two, three, four, and five years excess returns. Then, these recursively updated parameters are combined in a single spanning factor. The unspanned factor also has recursively updated parameters that are obtained through an orthogonalization process at each month, and then linearly combined into one single factor.

We evaluate our factors for the period of 1993-2017 and compare their predictability with the main factors in the literature, such as Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), and Fama and Bliss (1987). We also perform an out-of-sample experiment to assess the statistical evidence in bond return predictability for the period 1997-2017.

We contribute to the literature in a number of different ways. First, to the best of our knowledge, we are the first to introduce a deep neural network-based structure to generate a single spanning factor, as well as an unspanned factor, with recursively updated latent factors. Thus, we are able to introduce nonlinearities when modeling the bond risk premia in our first step of the recursive process, while still making use of a linear combination of these latent factors in the second step, to provide a novel interpretation for the bond risk premia, as proposed by Cochrane's question. This is consistent with recent findings from some works in empirical finance (see, e.g., Gu et al. (2018); Bianchi et al. (2019)) that document the importance of allowing for nonlinearities. It is known that with neural networks we can introduce flexibility and complex nonlinear relationships from the inputs while approximating arbitrarily well a rich set of smooth functions.

Second, motivated by Bauer and Hamilton (2018) that document that the use of over-lapping 12-month returns is prone to a number of problematic features, as it introduces substantial serial correlation in the predictive errors, we deviate from previous works that made use mainly of 12-month holding period, as we handle this issue with the use of non-overlapping returns. Furthermore, we are interested to obtain these factors to the short part of the term structure, and to do so we take into consideration the whole term structure at the higher frequency of 1-month holding period with maturities ranging up to 60 months ahead, allowing us to avoid to handpick only a subset of yields.

Third, we take a broader interpretation of the unspanning factor. Thus, our derived unspanned factor is more flexible, as it can be linked with other sources of risks, not limiting only to macroeconomics variables, but also with sentiment-based variables. And fourth, our approach avoids hand-picking the variables from the yield curve, as through our deep neural network we are able to recursively learn the best-approximating[1] function that condenses the yield curve into a single latent factor.

1.1 Related Literature

Our paper is related to several strands of the literature. This paper is related with the so known "spanning puzzle" which pertains a possible conflict between the theoretical spanning condition, in which the yield curve captures all the information for forecasting future yields and returns, and the use of unspanned macro information for these problems. Ludvigson and Ng (2009), Cooper and Priestley (2009), and Cieslak and Povala (2015) provide evidence that macroeconomic variables have predictive power for excess bond. On the other hand, Ghysels et al. (2018) show that the use of real time data substantially reduces the predictive power of macro variables for future bond. Bauer and Hamilton (2018) show that non spanning predictors proposed in the literature is weaker than expected, while Bauer and Rudebusch (2017) argue that the evidence from unspanned regressions cannot provide statistical basis for preferring either unspanned or spanned models.

Our work is also linked to a literature at the intersection of bond premia and sequential learning. Gargano et al. (2019) and Dubiel-Teleszynski et al. (2019) make use of a Bayesian learning approach in the context of bond risk premia. The former accounts for time-varying parameters, stochastic volatility, and parameter estimation error; while the latter implement the learning framework under a dynamic term structure model.

This paper is also related with the recent surge in the use of machine-learning and its growing impact in the field of economics, as reviewed in Mullainathan and Spiess (2017);

---

[1]For a given loss function.

Athey (2018); Athey and Imbens (2019); Varian (2014). This paper is also motivated by recent advances in the statistical properties of machine learning techniques, in especial the theoretical properties of inference using deep neural networks . Farrell et al. (2021), provide nonasymptotic bounds and convergence rates for nonparametric estimation using deep neural networks, and establishes valid inference on finite-dimensional parameters following first-step estimation using deep learning.

The use of recent machine-learning techniques in empirical finance has gotten special interest in the past years. Many papers have been interested in the dimensionality reduction, especially through penalized regression framework, such as LASSO, Ridge and Elastic-Net (see, e.g., Kozak et al. (2019); Freyberger et al. (2017)). Deep learning and some variations of deep neural networks, such as autoenconders, were used in some recent papers (Gu et al., 2018, 2020; Chen et al., 2019; Feng et al., 2018a,b; Heaton et al., 2017, 2016).

Specifically in macro-finance, Bianchi et al. (2019) and Huang and Shi (2019) make use of machine-learning techniques to model or evaluate bond return predictability. Huang and Shi (2019) use Supervised Adaptive Group LASSO to capture macroeconomic risks, and construct a single unspanned factor from a panel of 131 macro variables. Bianchi et al. (2019) seek to compare and evaluate several machine learning algorithms for the sole purpose of prediction of the US Treasury bonds excess returns. Their analysis ranges from penalized linear regressions, partial least squares, regression trees, random forests, and finally neural networks. The authors find evidence that non-linear methods can provide favorable out-of-sample prediction of bond excess returns. Our work builds on some of the insights from Bianchi et al. (2019) with regard of the use of deep neural networks to understanding bond premia. Nonetheless, our approach detours from their work in numerous ways, being the most important our goal to build a spanning and an unspanned factor from a deep neural network.

The structure of this paper is as follows. Next section introduces the general framework, contextualize the expectations and spanning hypothesis, and explain the deep-learning structure that we propose for bond premia. This section also provides an illustrative term-structure model. Section 3 explains our data, how we reconstruct the log yield of zero-coupons, and elucidate our empirical strategy. Section 4 presents the results. Finally, section 5 concludes. Additional results, tables and figures are presented in Appendix A.1.

2 Framework

2.1 Notation

Following the standard notation in the literature, let $p_t^{(n)}$ denote the natural logarithm of the price for a bond with n-period maturity at time t, and y represent its yield, such that:

$$y_t^{(n)} \equiv -\frac{1}{n}p_t^{(n)} \tag{1}$$

The holding period returns of a n-period maturity bond from time t to $t+\Delta$ is given by:

$$r_{t+\Delta}^{(n)} \equiv p_{t+\Delta}^{(n-\Delta)} - p_t^{(n)} \tag{2}$$

If integers of Δ represent years, then:

$$\begin{aligned} r_{t+h/12}^{(n)} &\equiv p_{t+h/12}^{(n-h/12)} - p_t^{(n)} \\ &= ny_t^{(n)} - (n-h/12)y_{t+h/12}^{(n-h/12)} \end{aligned} \tag{3}$$

where h is the frequency of the returns, measured in months. Thus, we can define the excess returns as

$$\begin{aligned} rx_{t+h/12}^{(n)} &\equiv \underbrace{p_{t+h/12}^{(n-h/12)} - p_t^{(n)}}_{\text{holding period return } r_{t+h/12}^{(n)}} - (h/12)y_t^{(h/12)} \\ &= ny_t^{(n)} - (n-h/12)y_{t+h/12}^{(n-h/12)} - (h/12)y_t^{(h/12)} \end{aligned} \tag{4}$$

Finally, we can define the forward rates at time t for loans between time $t + n - h/12$ and $t + n$ as

$$\begin{aligned} f_t^{(n)} &\equiv p_t^{(n-h/12)} - p_t^{(n)} \\ &= ny_t^{(n)} - (n-h/12)y_t^{(n-h/12)} \end{aligned} \tag{5}$$

2.2 Expectation Hypothesis and the Spanning Hypothesis

In its most common form, the *expectation hypothesis* states that yields of long maturity bonds should be the average of the future expected yield of short maturity bonds. Hence, it is equivalent with the statement that excess returns should not be predictable. Setting

$h = 1$ to express monthly frequency, the expectations hypothesis can be summarized as[2]

$$y_t^{(n)} \equiv \underbrace{\frac{1}{n}\mathbb{E}_t\left(y_t^{(1/12)} + y_{t+1/12}^{(1/12)} + \ldots + y_{t+n-1/12}^{(1/12)}\right)}_{\text{expectations component}} + \quad \text{yield risk premium} \quad . \tag{8}$$

In short, we can summarize the risk premium simply as the difference between a long rate and the expected average of future short rates. Knowing that we can express the *yield risk premium* as $\frac{1}{n}\mathbb{E}\left(rx_{t+1/12}^{(n)} + rx_{t+2/12}^{(n-1/12)} + rx_{t+3/12}^{(n-2/12)} + \ldots + +rx_{t+n-1/12}^{(2/12)}\right)$, then we can write

$$y_t^{(n)} \equiv \frac{1}{n}\mathbb{E}_t\left(y_t^{(1/12)} + y_{t+1/12}^{(1/12)} + \ldots + y_{t+n-1/12}^{(1/12)}\right) +$$

$$\tag{9}$$

$$\frac{1}{n}\mathbb{E}_t\left(rx_{t+1/12}^{(n)} + rx_{t+2/12}^{(n-1/12)} + rx_{t+3/12}^{(n-2/12)} + \ldots + +rx_{t+n-1/12}^{(2/12)}\right) \quad .$$

As in Duffee (2013), assuming that the agents' information set at time t can be summarized by a k-dimensional state vector $\mathbf{Z}_t$, from identity 9 we obtain

$$y_t^{(n)} = \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1}\mathbb{E}\left[y_{t+j\cdot h/12}^{(h/12)}|\mathbf{Z}_t\right]\right) + \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1}\left[rx_{t+h/12(j+1)}^{(n-j\cdot h/12)}|\mathbf{Z}_t\right]\right) \quad . \tag{10}$$

In equation (10), $\mathbf{Z}_t$ should contain all the information used by investors to forecast at time t the excess-returns for all future periods. If we stack all yields at time t in the vector $\boldsymbol{y}_t^{(n)}$, as

$$\boldsymbol{y}_t^{(n)} = f(\mathbf{Z}_t; N) \tag{11}$$

we can see that the yields must be a function only of the state vector $\boldsymbol{y}_t^{(n)}$ and the vector of maturities N. The essential assumption is the existence of an inverse function $f(\cdot)^{-1}$ that allow us to write $\mathbf{Z}_t = f\left(\boldsymbol{y}_t^{(n)}; N\cdot\right)^{-1}$. This holds true, as long as exists a correspondence in

---

[2]An accounting identity makes the link between the yield of bond to the sum of one-periods (h-periods) with the its excess returns for a n-period maturity bond as:

$$y_t^{(n)} \equiv \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1}y_{t+j\cdot h/12}^{(h/12)}\right) + \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1}rx_{t+h/12(j+1)}^{(n-j\cdot h/12)}\right) \tag{6}$$

where j are multiple of h-periods of the defined frequency. For annual frequency, i.e., $h = 12$ we have:

$$y_t^{(n)} \equiv \frac{1}{n}\left(\sum_{j=0}^{n-1}y_{t+j}^{(1)}\right) + \frac{1}{n}\left(\sum_{j=0}^{n-1}rx_{t+j+1}^{(n-j)}\right) \tag{7}$$

such a way that each $z_t \in \mathbf{Z}_t$ has its own effect on the yield curve $\boldsymbol{y}_t^{(n)}$. Thus, for a function $g(\cdot)$ we can write $\mathbb{E}_t\left(y_t^{(n)}\right) = g\left(\boldsymbol{y}_t^{(n)}; N\right)^3$.

As Duffee (2013) emphasizes, equation (9) determines that the expected returns depend on at most k state variables, and inverting equation (10) tells us that with the entire yield curve, we can disentangle shocks of the expected excess returns from shocks to expected future yields. What boils down to estimating the function $g(\cdot)$. The key takeaway is that the whole term-structure at time t contains all the information to predict $\mathbf{Z}_t$, and consequently the future yield curves.

However, the literature has gathered evidence against the expectations hypothesis. Influential studies from Fama and Bliss (1987), Campbell and Shiller (1991) and Cochrane and Piazzesi (2005) show some forecastability for excess returns. Among the most important approaches to test the predictability of the bonds' excess returns we have Fama and Bliss (1987), Cochrane and Piazzesi (2005), and Ludvigson and Ng (2009). Below we succinctly describe each one of them, as they will be used as our benchmarks.

Fama and Bliss (1987) builds forward rates spreads and use these variables as covariates. The forward rate spread between of a n-year maturity bond is defined as $fs_t^{(n,h)} \equiv f_t^{(n)} - y_t^{(h/12)}(h/12)$. The predictive regression in the Fama-Bliss approach is given by

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \epsilon_{t+h/12} \quad . \tag{12}$$

Cochrane and Piazzesi (2005) derive a single factor to use as predictor. The authors argue that their factor (CP_t^h), which has a peculiar tent-shape across maturities and is built from a linear combination of forward rates has a higher predictability of excess returns on one- to five-year maturity bonds. First, they estimate a vector $\boldsymbol{\gamma}$ by regressing the average of excess returns across maturities $n = 1, 2, 3, 4$ on all forward rates as

$$
\begin{aligned}
\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} &= \gamma_0 + \gamma_1 f_t^{(1)} + \gamma_2 f_t^{(2)} + \gamma_3 f_t^{(3)} + \gamma_4 f_t^{(4)} + \gamma_5 f_t^{(5)} + \bar{\epsilon}_{t+h/12} \\
\overline{rx}_{t+h/12} &= \underbrace{\boldsymbol{\gamma}^\top \mathbf{f}_t}_{CP_t^h} + \bar{\epsilon}_{t+h/12}
\end{aligned}
\tag{13}
$$

where $\boldsymbol{f}$ and $\boldsymbol{\gamma}$ are 6×1 vectors given by $\boldsymbol{f} \equiv \begin{bmatrix} 1 & f_t^{(1)} & f_t^{(2)} & f_t^{(3)} & f_t^{(4)} & f_t^{(5)} \end{bmatrix}^\top$, and $\boldsymbol{\gamma} \equiv \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 \end{bmatrix}^\top$. Denoting the estimated Cochrane-Piazzesi factor as $\widehat{CP}_t^h = \widehat{\boldsymbol{\gamma}}^\top \boldsymbol{f}_t$, the predictive regression in this approach is given by

---

³Which implies that $\mathbb{E}_t\left(rx_{t+h/12(j+1)}^{(n-j\cdot h/12)}\right) = g\left(\boldsymbol{y}_t^{(n)}; N\right)$ also holds.

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{CP}_t^h + \epsilon_{t+h/12} \quad . \tag{14}$$

Another important concept derived from the majority of macro-finance models is the *spanning hypothesis*. It says that all relevant information to forecast yields and excess returns can be found on the term-structure. Hence, under the *spanning hypothesis*, the yields curve fully spans all necessary information, and thus, no other variable or information already present in the term-structure should be necessary. As Bauer and Hamilton (2018) stress, the *spanning hypothesis* does not rules out the importance of macro variables (current or future). It only says that the yield curve completely reflects and spans this information.

Ludvigson and Ng (2009) show evidence against the spanning hypothesis. Using a large panel of macro variables, the authors build a single linear combination out of the first i estimated principal components $(\hat{g}_{i,t})$[4]. The authors start estimating a vector $\boldsymbol{\lambda}$ by regressing the average of excess returns across maturities $n = 1, 2, 3, 4$ on a subset of the first 8 principal components as

$$
\begin{aligned}
\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} &= \lambda_0 + \lambda_1\hat{g}_{1,t} + \lambda_2\hat{g}_{1,t}^3 + \lambda_3\hat{g}_{3,t} + \lambda_4\hat{g}_{4,t} + \lambda_5\hat{g}_{8,t} + \bar{\epsilon}_{t+h/12} \\
\overline{rx}_{t+h/12} &= \underbrace{\boldsymbol{\lambda}^\top \widehat{\boldsymbol{G}}_t}_{LN_t^h} + \bar{\epsilon}_{t+h/12}
\end{aligned}
\tag{16}
$$

where $\widehat{\boldsymbol{G}}_t$ and $\boldsymbol{\lambda}$ are 5×1 vectors given by $\widehat{\boldsymbol{G}}_t \equiv \begin{bmatrix} \hat{g}_{1,t} & \hat{g}_{1,t}^3 & \hat{g}_{3,t} & \hat{g}_{5,t} & \hat{g}_{8,t} \end{bmatrix}^\top$, and $\boldsymbol{\lambda} \equiv \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 \end{bmatrix}^\top$. Denoting the estimated Ludvigson-Ng factor as $\widehat{LN}_t^h = \widehat{\boldsymbol{\lambda}}^\top \widehat{\boldsymbol{G}}_t$, the predictive regression in this approach is given by

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{LN}_t^h + \epsilon_{t+h/12} \quad . \tag{17}$$

2.3 A Deep-Learning Structure for Bond Premia

The three main approaches presented in the last section seek to provide an explanation for the bond premia. We can summarize these approaches with the following predictive regression

$$rx_{t+h/12}^{(n)} = \boldsymbol{\beta}^\top \boldsymbol{Z}_t + \epsilon_{t+h/12} \tag{18}$$

---

[4]The authors consider a $T \times M$ panel of macroeconomic variables and assume that each macro variable $\{z_{j,t}^{macro}\}$ has a factor structure taking the form

$$z_{j,t}^{macro} = \boldsymbol{\nu}_t^\top \boldsymbol{g}_t + e_{j,t} \tag{15}$$

where $\boldsymbol{g}_t$ is an $s \times 1$ vector of latent common factors obtained through principal components analysis, and $\boldsymbol{\nu}_t$ is an $s \times 1$ vector of latent factor loadings. The essential point here is that $s \ll M$.

where $\boldsymbol{Z}_t = \left\{ \boldsymbol{Z}_t^{\mathrm{y}}, \boldsymbol{Z}_t^{\mathrm{y}^{\complement}} \right\}$ is a set of state variables that could potentially forecast the excess returns, and thus, provide evidence against the expectations hypothesis. If they rely on the spanning hypothesis $\boldsymbol{Z}_t = \{ \boldsymbol{Z}_t^{\mathrm{y}} \}$ and no macroeconomic variables are used to define the state space. Evidence against the spanning hypothesis is showed when $\boldsymbol{Z}_t^{\mathrm{y}^{\complement}} \neq \emptyset$.

In this paper we argue that this search for deriving, building and estimating factors that represent state variables in macro-finance models may be limited. We claim that the process done by financial economists of manually discovering and hand picking this list factors may be leaving unseen relationships between the state variables out in their derivation.

Hence, to assist in this process, we make use of one of the most powerful approaches in machine learning, namely a deep neural network. We aim to uncover relationships and derive a new single factor that could improve our understanding of the bond risk premia. We make use of deep feedforward network or multilayer perceptron (MLP)[5] and derive a single factor that has predictability in an our analysis.

Deep neural networks attempt to replicate the brain architecture in a computer, in a such a way that we must have many levels of processing information. As Murphy (2012) points out, it is believed that each level of learning features or representations at increasing levels of abstraction.

A deep feedforward network defines a mapping such as $rx_{t+h/12}^{(n)} = g\left(\boldsymbol{Z}_t, \boldsymbol{\theta}_t\right)$ to *learn* the parameter $\boldsymbol{\theta}_t$ that provides the best function approximation. In its most common structure, MLP can be represented in a direct acyclic graph with a chain of functions $g\left(\boldsymbol{Z}_t\right) = g^{(L)}\left(\ldots\left(g^{(2)}\left(g^{(1)}\left(\boldsymbol{Z}_t\right)\right)\right)\right)$. The name *network* comes from this chain and its inter-connectedness architecture, and *feedforward* because the information flows in one direction from $\boldsymbol{Z}_t$ through these functions, to finally obtain an output $rx_{t+h/12}^{(n)}$. The number of these functions L defines the *depth* of the network, motivating the use of the name "deep learning" to refer to this structure. We say that $g^{(1)}$ is the first layer, while the last one $g^{(L)}\left(\cdot\right)$ is the output layer.

As Goodfellow et al. (2016) discuss, deep feedforward network can capture the information between any two inputs, which is a limitation that linear models such as logistic and linear regressions face. This is done as an extension from a linear model, in such a way that we apply a nonlinear function $\phi(\cdot)$ in $\boldsymbol{Z}_t$, transforming our independent variable. Thus, the model can be represented as $rx_{t+h/12}^{(n)} = g\left(\boldsymbol{Z}_t, \boldsymbol{\theta}_t, \boldsymbol{w}\right) = \phi\left(\boldsymbol{Z}_t, \boldsymbol{\theta}_t\right)^{\top} \boldsymbol{w}$.

In this approach ϕ defines a hidden layer, $\boldsymbol{\theta}_t$ the parameters used to learn ϕ, and $\boldsymbol{w}$ are parameters mapping from $\phi\left(\boldsymbol{\theta}_t\right)$ to the output. An optimizing algorithm is responsible to find $\boldsymbol{\theta}_t$ that gives the best representation. The nonlinear function is called activation function, which is controlled by learned parameters. Hence, we define $g_l\left(\cdot\right) = g\left(\boldsymbol{W}_l^{\top} \boldsymbol{Z}_t + c\right)$, where

---

[5]They are also known as feedforward neural network

$\boldsymbol{W}$ is a set of weights and c the biases.

One advantage of deep neural networks is based on the universal approximation theorem (Hornik et al., 1989; Cybenko, 1989) that states that feedforward network with a linear output layer and at least one hidden layer with any activation function can approximate any function[6] from one finite-dimensional space to another with any desired nonzero amount of error. In short, this theorem says that a simple neural network can represent a wide variety of functions. However it does not guarantee the training algorithm will be able to learn the function. One implication from the universal approximation theorem is that there exists a network large enough to achieve any degree of accuracy.

In our framework, the single factor capturing the information from the yield curve is built in the following way. First, at each t we use the cross-section of the information on the term-structure to feed MLPs to obtain as output a factor derived from a learning network. We denote this deep neural network factor as $\mathfrak{f}_{DNN}$.

Aligned with the results from Bauer and Hamilton (2018) who gathered evidence that rejections of the spanning hypothesis by some recent papers is significantly weaker when more robust methods are used to deal especially with overlapping data, we use as input in our networks only $\boldsymbol{Z}_t^{\mathrm{y}}$, which is formed by the full set of information from the yield curve. We argue that given the superiority of deep feedforward networks to uncover relationships between the information found in $\boldsymbol{Z}_t^{\mathrm{y}}$, especially its capacity to nonlinear and more complex associations in the data, there is a potential gain of extracting more information out of the yield curve.

Figure 1 shows the deep feed forward architecture to obtain the DNN factor $\mathfrak{f}_{DNN}$. The depth, the width, the activation function of the deep neural network, as well as the loss function used for training at each t are variations discussed in section 3.

Notice that there are 4 separate groups of networks. Each one of them seek to find the function that provides the approximation g, such that the mapping is given by $g^{(n)} : \boldsymbol{Z}_t^{\mathrm{y}} \mapsto rx_{t+h/12}^{(n)}$, where $n \in \{2,3,4,5\}$, i.e., it is the mapping from the entire yield curve information to the excess returns in the next period $t+h/12$ for maturities ranging from 2 to 5 years.

Each group of network will deliver a factor associated with a maturity n at each t. After obtaining $\mathfrak{f}_{t,DNN}^{(n)}$, we estimate the single factor that summarizes the all the term-structure information to explain the excess returns. The idea is to describe the expected excess returns of all maturities with a unique factor, as proposed initially by Cochrane and Piazzesi (2005), and extended by others (Ludvigson and Ng, 2009; Cieslak and Povala, 2015). First, we regress the average of the excess returns of maturities $2, 3, 4$ and 5 years on all four factors

---

[6]Precisely, any Borel measurable function, i.e., any continuous function on a closed and bounded subset of $\mathbb{R}^n$.

Figure 1: Deep Neural Network
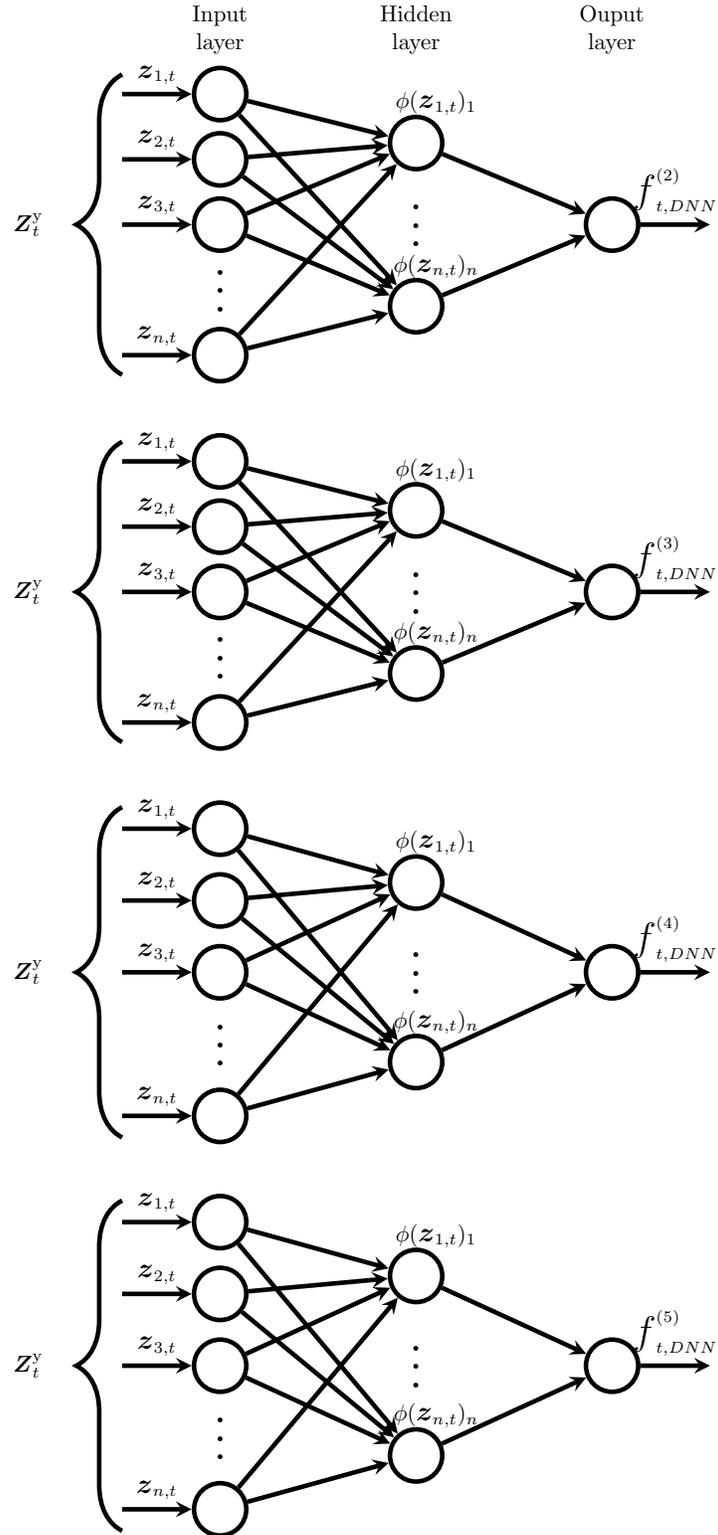
Figure 1 shows the general structure of the deep feed forward designed to obtain the DNN factor $\mathfrak{f}_{DNN}$. There are four groups of networks, each group for $n \in \{1, 2, 3, 4\}$. The inputs layer receives data from $\boldsymbol{Z}_t^y = \{\boldsymbol{z}_{1,t}, \boldsymbol{z}_{2,t}, \ldots, \boldsymbol{z}_{n,t}\}$. Each group of network n outputs a factor (DNN factor), which we denote by $\mathfrak{f}_{t,DNN}^{(n),h}$.

derived from our deep neural network, as below:

$$
\begin{aligned}
\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} &= \tau_0 + \tau_1 \mathfrak{f}_{t,DNN}^{(2),h} + \tau_2 \mathfrak{f}_{t,DNN}^{(3),h} + \tau_3 \mathfrak{f}_{t,DNN}^{(4),h} + \tau_4 \mathfrak{f}_{t,DNN}^{(5),h} + \bar{\epsilon}_{t+h/12} \\
&= \boldsymbol{\tau}^{\top} \widehat{\boldsymbol{\mathfrak{F}}}_t^h + \bar{\epsilon}_{t+h/12}
\end{aligned}
\tag{19}
$$

where $\widehat{\boldsymbol{\mathfrak{F}}}_t$ and $\boldsymbol{\tau}$ are 5×1 vectors given by $\widehat{\boldsymbol{\mathfrak{F}}}_t \equiv \begin{bmatrix} 1 & \mathfrak{f}_{t,DNN}^{(2),h} & \mathfrak{f}_{t,DNN}^{(3),h} & \mathfrak{f}_{t,DNN}^{(4),h} & \mathfrak{f}_{t,DNN}^{(5),h} \end{bmatrix}^{\top}$, and $\boldsymbol{\tau} \equiv \begin{bmatrix} \tau_0 & \tau_1 & \tau_2 & \tau_3 & \tau_4 \end{bmatrix}^{\top}$. The predictive regression in this approach is given by

$$
rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \left(\boldsymbol{\tau}^{\top} \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h + \epsilon_{t+h/12} \qquad n = 2,3,4,5 \quad .
\tag{20}
$$

Equation 20 tells us that a single factor $\left(\boldsymbol{\tau}^{\top} \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h$ defines the state variable driving the excess returns. Thus, starting from the spanning hypothesis, we feed a MLP with the entire information from the yield curve to approximate a function, and then derive a single linear combination of factors to explain the time-varying expected returns across maturities.

From the deep neural network we also would like to estimate a factor that represent the information not spanned by the term-structure. To do so, we design a new approach in which at each t and each group $n \in \{2,3,4,5\}$ of network for each maturity, we orthogonalize the excess returns by the deep neural network factor $\mathfrak{f}_{t,DNN}^{(n)}$, and denote it by $\boldsymbol{\xi}_t^{(n),h}$ as

$$
\xi_{t+h/12}^{(n),h} = rx_{t+h/12}^{(n)} - \beta_0 - \beta_1 \mathfrak{f}_{t,DNN}^{(n),h} \quad .
\tag{21}
$$

From equation (21), the factor $\xi_{t+h/12}^{(n),h}$ that lies in an orthogonal vector to the space spanned by $\mathfrak{f}_{t,DNN}^{(n)}$, can be seen as all the information not spanned by the term-structure captured by $\mathfrak{f}_{t,DNN}^{(n)}$ that affects the excess returns.

2.4   An Illustrative Term-Structure Model

In this section we make the link of our methodology with the main dynamic term-structure frameworks in the macro-finance literature. We follow Duffee (2013) and assume that interest rate dynamics are linear and homoskedastic with Gaussian shocks. The no-arbitrage assumption rely on the fundamental asset pricing equation:

$$
P_t^{(n)} = \mathbb{E}_t \left(\mathcal{M}_{t+1} P_{t+1}^{(n-1)} \right)
\tag{22}
$$

where $P_t^{(n)}$ is the price of a bond and $\mathcal{M}_{t+h/12}$ is the stochastic discount factor (SDF).

The economic agents value nominal bonds using the following SDF:

$$\mathcal{M}_{t+h/12} = \exp^{-r_t \frac{1}{2} \Lambda_t^\top \Lambda_t - \Lambda_t^\top \epsilon_{t+h/12}} \tag{23}$$

where Λ_t are the market prices of the risks, i.e., the amount of compensation required by investors to face the unit normal shock $\epsilon_{t+h/12}$. The yield on a one-period bond $r_t \equiv y^{(1)}$ is a function of $\boldsymbol{Z_t}$, as

$$r_t = \rho_0 + \rho_1 \boldsymbol{Z_t} \quad . \tag{24}$$

As we defined $\boldsymbol{Z}_t = \left\{ \boldsymbol{Z}_t^{\text{y}}, \boldsymbol{Z}_t^{\text{y}^\complement} \right\}$, we write the dynamics of $\boldsymbol{Z}_t$ that capture all the risks of the economy following a Gaussian VAR process given by:

$$\begin{bmatrix} \boldsymbol{Z}_t^{\text{y}} \\ \boldsymbol{Z}_t^{\text{y}^\complement} \end{bmatrix} = \boldsymbol{\mu} + \boldsymbol{\Phi} \begin{bmatrix} \boldsymbol{Z}_{t-1}^{\text{y}} \\ \boldsymbol{Z}_{t-1}^{\text{y}^\complement} \end{bmatrix} + \boldsymbol{\Sigma} \epsilon_t$$

$$\tag{25}$$

$$\boldsymbol{Z}_t = \boldsymbol{\mu} + \boldsymbol{\Phi} \boldsymbol{Z}_{t-1} + \boldsymbol{\Sigma} \epsilon_t \qquad \epsilon_t \sim N(0, \boldsymbol{I})$$

where $\boldsymbol{\mu}$ is a a $k \times 1$ vector, and $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ are $k \times k$ matrices, being k the number of state variables. In a similar fashion to Joslin et al. (2014), who developed an arbitrage-free dynamic term-structure model with unspanned macro risks, we can write:

$$\boldsymbol{Z}_t^{\text{y}^\complement} = \gamma_0 + \gamma_1 \boldsymbol{Z}_t^{\text{y}} + \boldsymbol{M}_{\boldsymbol{Z}_t^{\text{y}}} \boldsymbol{Z}_t^{\text{y}^\complement} \tag{26}$$

where $\boldsymbol{M}_{\boldsymbol{Z}_t^{\text{y}}} \boldsymbol{Z}_t^{\text{y}^\complement}$ is the annihilator matrix of the space spanned by $\boldsymbol{Z}_t^{\text{y}}$, i.e., $\boldsymbol{M}_{\boldsymbol{Z}_t^{\text{y}}} \boldsymbol{Z}_t^{\text{y}^\complement} \equiv \boldsymbol{Z}_t^{\text{y}^\complement} - \text{Proj} \left[\boldsymbol{Z}_t^{\text{y}^\complement} | \boldsymbol{Z}_t^{\text{y}} \right]$. Previous models have assumed that the $\boldsymbol{Z}_t^{\text{y}^\complement}$ was spanned by $\boldsymbol{Z}_t^{\text{y}}$, thus imposing the restriction of $\boldsymbol{Z}_t^{\text{y}^\complement} = \text{Proj} \left[\boldsymbol{Z}_t^{\text{y}^\complement} | \boldsymbol{Z}_t^{\text{y}} \right]$ in equation (26).

Aligned with Joslin et al. (2014), our methodology is also based on (i) a small number of risk factors, and (ii) the unspanned components of $\boldsymbol{Z}_t^{\text{y}^\complement}$ may contain predictive power for excess returns. However, we distinguish from Joslin et al. (2014) who provided the exact macroeconomic variables that are unspanned by the term-structure. Our unspanned factor, on the other hand, should be able to represent any other risk, be it macroeconomic or sentiment-based in the economy. In this sense, we say that our framework is more general. Additionally, to provide an intuitive interpretation, we analyze how $\boldsymbol{Z}_t^{\text{y}^\complement}$ is correlated with macroeconomic variables and sentiment-based measures.

From the above illustrative term-structure model, we make a set of propositions that makes the link between our deep-learning framework and a dynamic term structure model.

Proposition 1. *The state vector $\boldsymbol{Z}_t$ that encompasses all risks in the economy can be partitioned as $\boldsymbol{Z}_t = \left\{ \boldsymbol{Z}_t^y, \boldsymbol{Z}_t^{y^\complement} \right\}$, in such a way that $\boldsymbol{Z}_t^y$ contains information solely from the yield curve, and $\boldsymbol{Z}_t^{y^\complement}$ any other information not found in the term structure.*

Proposition 2. *Under the canonical arbitrage-free Gaussian term structure model as in Joslin et al. (2014), $\boldsymbol{Z}_t^y$ is given by the derived factor $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h$ from equation (19), and $\boldsymbol{Z}_t^{y^\complement}$ by a linear function $f\left(\cdot\right)$ of $\boldsymbol{\xi}_{t+h/12}^h$.*

Proposition 3. *As in the dynamic term structure model of Joslin et al. (2014), $f(\boldsymbol{\xi}_{t+h/12}^h)$ complete and fill the unspanned factor in the state space, in a such a way that $\left[\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h, f(\boldsymbol{\xi}_{t+h/12}^h) \right]$ and $\boldsymbol{Z}_t$ represent linear rotations of the same economy-wide risks underlying all tradable assets available to agents in the economy.*

3   Data & Strategy

3.1   Data

As emphasized by Bauer and Hamilton (2018), predictive regressions estimated using overlapping observations, approach commonly used by several previous studies, where monthly data is used and the annual excess bond return is the dependent variable, introduces serial correlation in the prediction errors, what results in inaccurate standard errors.

As done in Gargano et al. (2019), to overcome the issues generated by overlapping observations, we reconstruct the yield curve at the daily frequency, using the parameters estimated by Gürkaynak et al. (2007) and made available at the Federal Reserve Discussion Series website[7]. Thus, we reconstruct the log yield of a zero-coupon with n-period maturity at time t as

$$
\begin{aligned}
y_t^{(n)} = \beta_{0,t} \quad &+\beta_{1,t} \left(\frac{1 - \exp\left(-n/\tau_1\right)}{n/\tau_1} \right) \\
&+\beta_{2,t} \left(\frac{1 - \exp\left(-n/\tau_1\right)}{n/\tau_1} - \exp\left(-n/\tau_1\right) \right) \\
&+\beta_{3,t} \left(\frac{1 - \exp\left(-n/\tau_2\right)}{n/\tau_2} - \exp\left(-n/\tau_2\right) \right)
\end{aligned}
\tag{27}
$$

where the daily estimated parameters β_0, β_1, β_2, β_3, τ_1 and τ_2 are from Gürkaynak et al. (2007). The full period of analysis ranges from 1962:01 to 2017:12. We use these estimated parameters from the last day of each month to construct a monthly derived zero-coupon bonds log yields with maturities up to 60 months from each t. Figure 2 plots the log yields for all maturities. Figure 3 shows the 1-month excess returns for maturities with $n = 2, 3, 4$

---

[7]https://www.federalreserve.gov/econres/feds/2006.htm

and 5 years. In Appendix A.1, figure 14 plots for the same set of maturities the 12-month excess returns.

Figure 2: Derived zero-coupon bonds log yields for maturities (n) up to 60 months
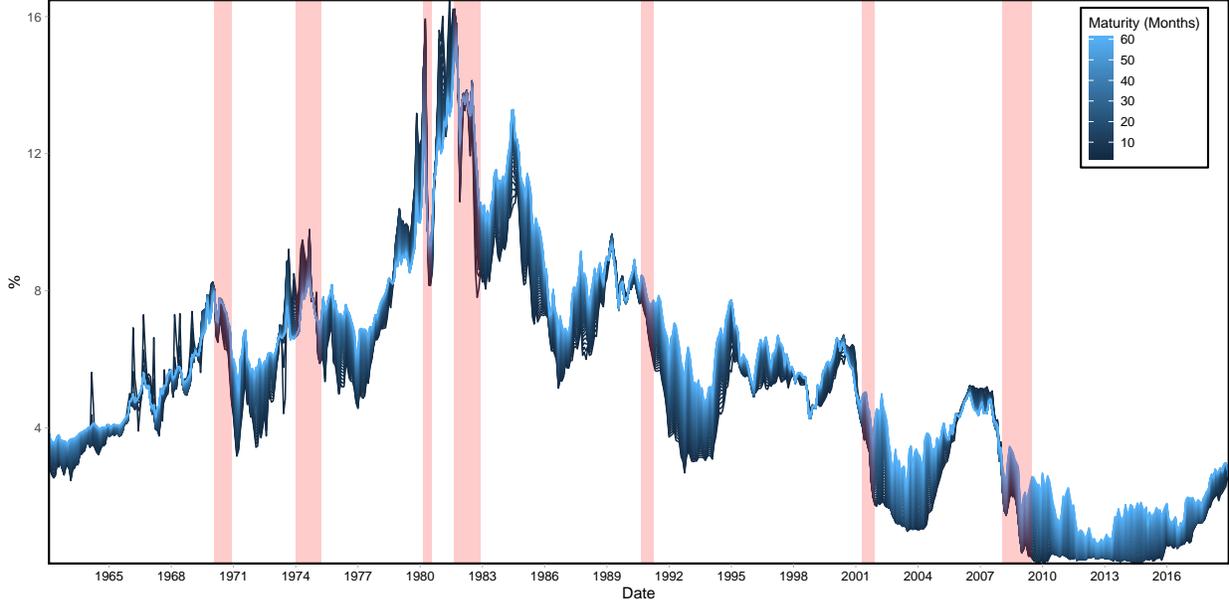
Figure 2 shows the log yields for all maturities we consider: $y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}$. At each month t, there are 60 yields represented by variation of color. The log yields of the zero-coupons bonds are reconstructed with equation (27), using the last day of each month estimated parameters from Gürkaynak et al. (2007) data. The entire sample ranges from 1962:01 to 2018:12.

Some papers have instead used the data from the Fama–Bliss Center for Research in Security Prices (CRSP) to build the series of log bond yields. Based on Fama and Bliss (1987), this approach constructs the yields sequentially from a set of estimated daily forward rates. As Gargano et al. (2019) point out, the differences between Fama and Bliss (1987) and Gürkaynak et al. (2007) are minimal. The correlation between both methods[8] when comparing yields and excess returns are both above 0.99 for all four maturities we use.

3.2 Empirical Strategy

In our first analysis we establish the period of evaluation from 1993:01 to 2017:12. We feed our deep neural network with three different sets of information from the term-structure: (i) set of forward rates from 2 to 60 months from t, i.e., $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)} \right\}$, (ii) set of zero-coupon yields with maturities ranging from 1 to 60 months from t, i.e., $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$, and finally (iii) a combination of the previous two groups,

---

[8]For a similar period: 1962:01 to 2015:12.

Figure 3: 1-Month Bonds Excess Returns
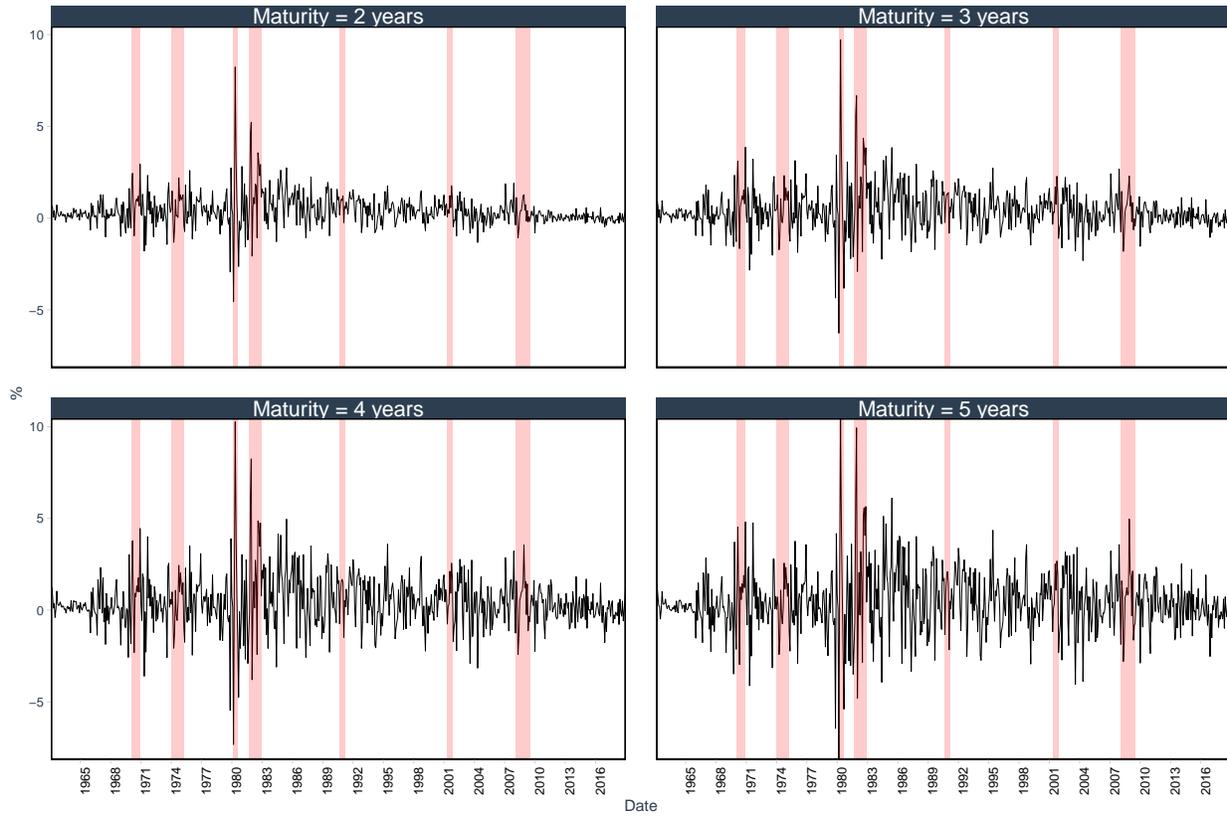
Figure 3 shows the 1-month excess returns for maturities with $n = 2, 3, 4$ and 5 years. The excess returns are calculeted as in equation (4), i.e., $rx_{t+1/12}^{(n)} = ny_t^{(n)} - (n+1/12)y_{t+1/12}^{(n-h/12)} - y_t^n$. Each panel represents one of the four maturities. The y-axis shows values in percentage (%). NBER-classified recessions are shaded in light red.

i.e., $\boldsymbol{Z}_t^{\text{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \dots, f_t^{(60/12)}, y_t^{(1/12)}, y_t^{(2/12)}, \dots, y_t^{(60/12)} \right\}$.

Goodfellow et al. (2016) discuss that the design of the hidden units is an extremely active area of research. This leads to many potential options for the nonlinear function in the hidden layers. As the authors mention, the rectified linear activation function (ReLU) is the default and recommended for use with the majority of feedforward neural networks. In all our hidden layers, for all groups and architectures, we make use of this activation function defined as

$$ReLU(x) = \begin{cases} 0 & , \text{if } x \leq 0 \\ x & , \text{otherwise} \end{cases} . \tag{28}$$

As Goodfellow et al. (2016) mention, applying this function to the output of a linear transformation yields a nonlinear transformation. Notice that, since ReLU units are nearly linear, they have the advantage of also retaining many of the properties from linear models, such as (i) efficiency to optimize with gradient-based methods, and (ii) ability to preserve the properties that make linear models generalize well.

All our neural networks share the same architecture as show in figure 1. To make use of the flexibility that MLP allows us, we designed three variation for the whole network. Bianchi et al. (2019) also developed several designs in their study, and we use some of their intuitions to design our deep neural networks architectures. The first (**DNN 1**) and second model (**DNN 2**) are feedfoward neural networks with 2 hidden layers ($L = 2$), with 16 and 4 nodes respectively, and finally an output layer for each group of maturity $n \in \{1, 2, 3, 4\}$. What differentiates **DNN 1** from **DNN 2** is the regularization function, where we use a ℓ_1-norm for **DNN 1** and a ℓ_1- and ℓ_2-norm for **DNN 2**. On the other hand, **DNN 3** has 4 hidden layers ($L = 4$), with 64, 32, 16 and 4 nodes. For **DNN 3** we use a ℓ_1- and ℓ_2-norm regularization function.

The process of obtaining $\widehat{\widetilde{\mathfrak{S}}}_t^h$ from equation (19) at each t can be summarized in the following way. First, for each set of $\boldsymbol{Z}_t^{\text{y}}$ in consideration, we feed each one of the three DNNs architectures with the entire past information of each variable in $\boldsymbol{Z}_t^{\text{y}}$. We use the 10% most recent data in each $\boldsymbol{z}_t^{\text{y}} \in \boldsymbol{Z}_t^{\text{y}}$ for validation. After the set of weights are chosen, with the final set of weights and the final approximated function, we use $\boldsymbol{Z}_{t-1}^{\text{y}}$ to predict $rx_t^{(n)}$ in each group of maturity $n \in \{2, 3, 4, 5\}$. Thus, we form the 4×1 vector of $\widehat{\widetilde{\mathfrak{S}}}_t^h$ as the factor at t generated by each DNN_i. As a final step we run univariate regressions to obtain $\boldsymbol{\xi}_t^h$, as shown in equation 21. Similarly, we build the 4×1 vector using the observation t residuals as $\widehat{\boldsymbol{\xi}}_t^h = \begin{bmatrix} \widehat{\xi}_t^{(2),h} & \widehat{\xi}_t^{(3),h} & \widehat{\xi}_t^{(4),h} & \widehat{\xi}_t^{(5),h} \end{bmatrix}$. The whole process is summarized in the pseudocode given in algorithm 1.

Algorithm 1: Recursively generated factors with updated parameters

Initialization:

Start with a set of information from the term structure collected in $\boldsymbol{Z}^{\mathrm{y}}$. Partitionate your sample $\{t_0, \ldots, t_{split}, \tau, \tau+1, \ldots, T\}$ between the data to be used to initialize the process $\{t_0, \ldots, t_{split}\}$, and to obtain the recursively generated factors $\{\tau, \tau+1, \ldots, T\}$;

Define the deep neural network architecture to be used (number of hidden layers L, and number of nodes in each layer $l \in L$);

For a pre-defined deep neural network architecture DNN_i, set the activation function $\phi(\cdot)$ in each node at layer $l \in L$ as the ReLu defined in equation (28);

for $n \in \{2, 3, 4, 5\}$ **do**

    for $t \in \{\tau, \tau+1, \ldots, T\}$ **do**

        Feed DNN_i with lagged data $\boldsymbol{Z}_{t-1}^{\mathrm{y}} = \left\{\boldsymbol{z}_{t_0}^{\mathrm{y}}, \boldsymbol{z}_{t_0+1}^{\mathrm{y}}, \ldots, \boldsymbol{z}_{t-1}^{\mathrm{y}}\right\}$ to learn/aproximate with output $rx_t^{(n)}$, and use the last 10% of the data for validation;

        Obtain the learned parameters;

$$\widehat{\mathfrak{f}}_{t,DNN}^{(n),h} \leftarrow g\left(\boldsymbol{Z}_{t-1}^{\mathrm{y}}, \boldsymbol{\theta}_{t-1}\right)$$

        Obtain the t-*th* element that lies in the orthogonal vector from the space generated by the $\mathfrak{f}_{t-1,DNN}^{(n),h}$ through:

$$\widehat{\xi}_t^{(n),h} \leftarrow rx_t^{(n)} - \widehat{\beta}_0 - \widehat{\beta}_1 \mathfrak{f}_{t-1,DNN_i}^{(n),h}$$

    end

end

Result:

$$\widehat{\boldsymbol{\mathfrak{F}}}_{t,DNN_i} \equiv \begin{bmatrix} \widehat{\mathfrak{f}}_{t,DNN_i}^{(2),h} \\ \widehat{\mathfrak{f}}_{t,DNN_i}^{(3),h} \\ \widehat{\mathfrak{f}}_{t,DNN_i}^{(4),h} \\ \widehat{\mathfrak{f}}_{t,DNN_i}^{(5),h} \end{bmatrix} = \begin{bmatrix} \widehat{\mathfrak{f}}_{\tau,DNN_i}^{(2),h} & \widehat{\mathfrak{f}}_{\tau,DNN_i}^{(3),h} & \widehat{\mathfrak{f}}_{\tau,DNN_i}^{(4),h} & \widehat{\mathfrak{f}}_{\tau,DNN_i}^{(5),h} \\ \widehat{\mathfrak{f}}_{\tau+1,DNN_i}^{(2),h} & \widehat{\mathfrak{f}}_{\tau+1,DNN_i}^{(3),h} & \widehat{\mathfrak{f}}_{\tau+1,DNN_i}^{(4),h} & \widehat{\mathfrak{f}}_{\tau+1,DNN_i}^{(5),h} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\mathfrak{f}}_{T,DNN_i}^{(2),h} & \widehat{\mathfrak{f}}_{T,DNN_i}^{(3),h} & \widehat{\mathfrak{f}}_{T,DNN_i}^{(4),h} & \widehat{\mathfrak{f}}_{T,DNN_i}^{(5),h} \end{bmatrix}$$

And,

$$\widehat{\boldsymbol{\xi}}_t^h \equiv \begin{bmatrix} \widehat{\xi}_{\tau,DNN_i}^{(2),h} & \widehat{\xi}_{\tau,DNN_i}^{(3),h} & \widehat{\xi}_{\tau,DNN_i}^{(4),h} & \widehat{\xi}_{\tau,DNN_i}^{(5),h} \\ \widehat{\xi}_{\tau+1,DNN_i}^{(2),h} & \widehat{\xi}_{\tau+1,DNN_i}^{(3),h} & \widehat{\xi}_{\tau+1,DNN_i}^{(4),h} & \widehat{\xi}_{\tau+1,DNN_i}^{(5),h} \\ \vdots & \vdots & vdots & \vdots \\ \widehat{\xi}_{T,DNN_i}^{(2),h} & \widehat{\xi}_{T,DNN_i}^{(3),h} & \widehat{\xi}_{T,DNN_i}^{(4),h} & \widehat{\xi}_{T,DNN_i}^{(5),h} \end{bmatrix}$$

Notes: For our analysis the derived factors are calculated for the period τ=1993:01 to T=2017:12, while the data ranging from t_0 =1962:01 to t_{split} =1992:12 is used as a burn-in data to initiate the recursive process of obtaining the the derived factors $\widehat{\mathfrak{f}}_{t,DNN}^{(n),h}$ and $\widehat{\xi}_t^{(n),h}$.

18

At the end of our period of analysis, we use the entire series of $\widehat{\mathfrak{F}}_t^h$ to obtain our single factor $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$ that spans the yield curve information as in equation (19). To complete the factor space of our dynamic term-structure model, we define the unspanned factor as a function of $\boldsymbol{\xi}_t^h$ to build a single factor as well for $\boldsymbol{Z}_t^{y^{\complement}}$. We investigate two alternatives for $f(\boldsymbol{\xi}_t^h)$. In the first one, $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^h$ is the unique factor obtained as the projection of $\overline{rx}_{t+h/12}$ in $\widehat{\boldsymbol{\xi}}_t$. The second alternative is a similar projection, however for each maturity $n \in \{2, 3, 4, 5\}$ we regress $rx_{t+h/12}^{(n)}$ on $\widehat{\boldsymbol{\xi}}_t^{(-n),h} \equiv \widehat{\boldsymbol{\xi}}_t^h \setminus \widehat{\xi}_t^{(n),h}$, i.e., on the set $\boldsymbol{\xi}_t^h$ excluding its own $\widehat{\xi}_t^{(n)}$. We denote the factors generated by this second approach as $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$.

Consistent with our adapted dynamic term-structure model, the orthogonal vector from $\text{Proj}\left[f(\boldsymbol{\xi}_t^{(n)})|\boldsymbol{Z}_t^y\right]$ has predictive power for excess returns. Thus, we use the projection error $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^h$ for alternative 1 and $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$ for alternative 2 in our predictive analysis in the following section.

The intuition that motivates our construction of $\boldsymbol{Z}_t^{y^{\complement}}$ lies in the fact that at each t, $\widehat{\xi}_t^{(n)}$ is orthogonal to $\mathfrak{f}_{t,DNN}^{(n),h}$, allowing the interpretation that, for each maturity group n in our DNN, anything not captured by the neural network process of approximating $g(\cdot)$ from the yield curve information $\boldsymbol{Z}_t^y$, are unspanned and should be in an orthogonal space. Hence, the unspanned information in $\widehat{\boldsymbol{\xi}}_t^h$ could be capturing macroeconomic information or sentiment measures not spanned by the term-structure information that affects the bonds' excess returns. Alternative 1 builds an unique factor for $\boldsymbol{Z}_t^{y^{\complement}}$ in a such a way that a single linear combination of orthogonal variables is the state variable that completes the state space for time-varying expected returns on all maturities. On the other hand, alternative 2 tells us that a single linear combination of three orthogonal variables from the remaining maturities complete the state space for time-varying expected returns for maturity n.

4 Empirical Results

For the period 1993:01 to 2017:12, we generated $\mathfrak{f}_{t,DNN}^{(n),h}$ in a recursive way. Figure 4 shows the derived DNN factor for each scenario under consideration. Each column uses a different set of information from the term-structure to derive the factor $\mathfrak{f}_{t,DNN}^{(n),h}$. Column (1) shows the the derived DNN factors when we feed the MLP with all the forward rates, i.e., $\boldsymbol{Z}_t^y = \left\{f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)}\right\}$, column (2) when the set of yields is used, i.e., $\boldsymbol{Z}_t^y = \left\{y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}\right\}$ and column (3) when both previous sets are combined, i.e., $\boldsymbol{Z}_t^y = \left\{f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60)}, y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}\right\}$. Each row represents one of the four groups of maturities. Finally, different colors represent the three variations of DNN

as explained in section 3.2. A quick inspection in figure 4 shows how the different structures of neural networks result in different factors. Clearly, **DNN 3** distinguishes from the other two. We also see that **DNN 1** and **DNN 2** have an evident mean reverting tendency.

In order to better investigate how the factors $\mathfrak{f}_{t,DNN}^{(n),h}$ behave, we plot in figure 5 only for the **DNN 2** factors generated by the set of yields in terms of maturity for the period of analysis (1993:01 - 2017:12). Some patterns become evident when we inspect this figure. First, on average the set $\left\{\mathfrak{f}_{t,DNN}^{(2),h}, \mathfrak{f}_{t,DNN}^{(3),h}, \mathfrak{f}_{t,DNN}^{(4),h}, \mathfrak{f}_{t,DNN}^{(5),h}\right\}$ throughout the period of analysis, we notice that it behaves as an increasing function of the maturity (n). In the first months we see that the DNN factors behave quite erratically, what could be interpreted as the neural network changing the weights in its functions more intensively to try to improve the learning process. Another clear pattern inferred from figure 5 is that the curve generated at each t apparently moves in synchrony across maturities. This is more evident when we take in consideration the two recessions (2001:04 - 2001:11 and 2008:01 - 2009:06) in the period of analysis. We see that the curve of generated factors move down for all maturities following a recession and for some time after it the values of $\mathfrak{f}_{t,DNN}^{(n),h}$ are low. As the recession fades, the curve $\mathfrak{f}_{t,DNN}^{(n),h}$ slowly start to move up as well.

Following our methodology, we use equation (19) to obtain our single factor $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_t\right)_t^h$ as a linear combination of the derived factors $\mathfrak{f}_{t,DNN}^{(n),h}$. Figure 6 plots $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_t\right)_t^h$ for each DNN architecture and the three different sets of $\boldsymbol{Z}_t^{\mathcal{Y}}$, our unique factor from the state space in the dynamic term-structure model that captures all the information from the yield curve. Notice that, based on which DNN structure we use, the factor $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_t\right)_t^h$ behaves quite differently. In the first years of analysis, the single factor seems to be correlated. However, consistent with our comments from figure 4, as the training process of the neural network advances, the three DNNs produce distincts $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_t\right)_t^h$, being more evident the contrast of the factor produced by **DNN 3**, since the its structure is the most different one.

4.1 Predictive Regressions

In table 1 we have the predictive regressions for the period from 1993:01 to 2017:12 using our derived state variables: $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_t\right)_t^h$ and $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_t^h$ (alternative 1) or $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$ (alternative 2). We split the regressions in 4 panels, one for each maturity. We evaluate three different predictive regression models. The first one is shown in equation (20), where we only use $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_t\right)_t^h$ as the state variable. In our model, to complete the state space, we use the orthogonal vector from the projection of $f(\boldsymbol{\xi}_{t+h/12}^{(n)})$ on $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_t\right)_t^h$. The two alternatives for the single factor that captures the unspanned information from the yield curve are the

Figure 4: DNN Factor $\mathfrak{f}_{t,DNN}^{(n),h}$ by MLP Architecture and Choice of $\boldsymbol{Z}_t^{\text{y}}$
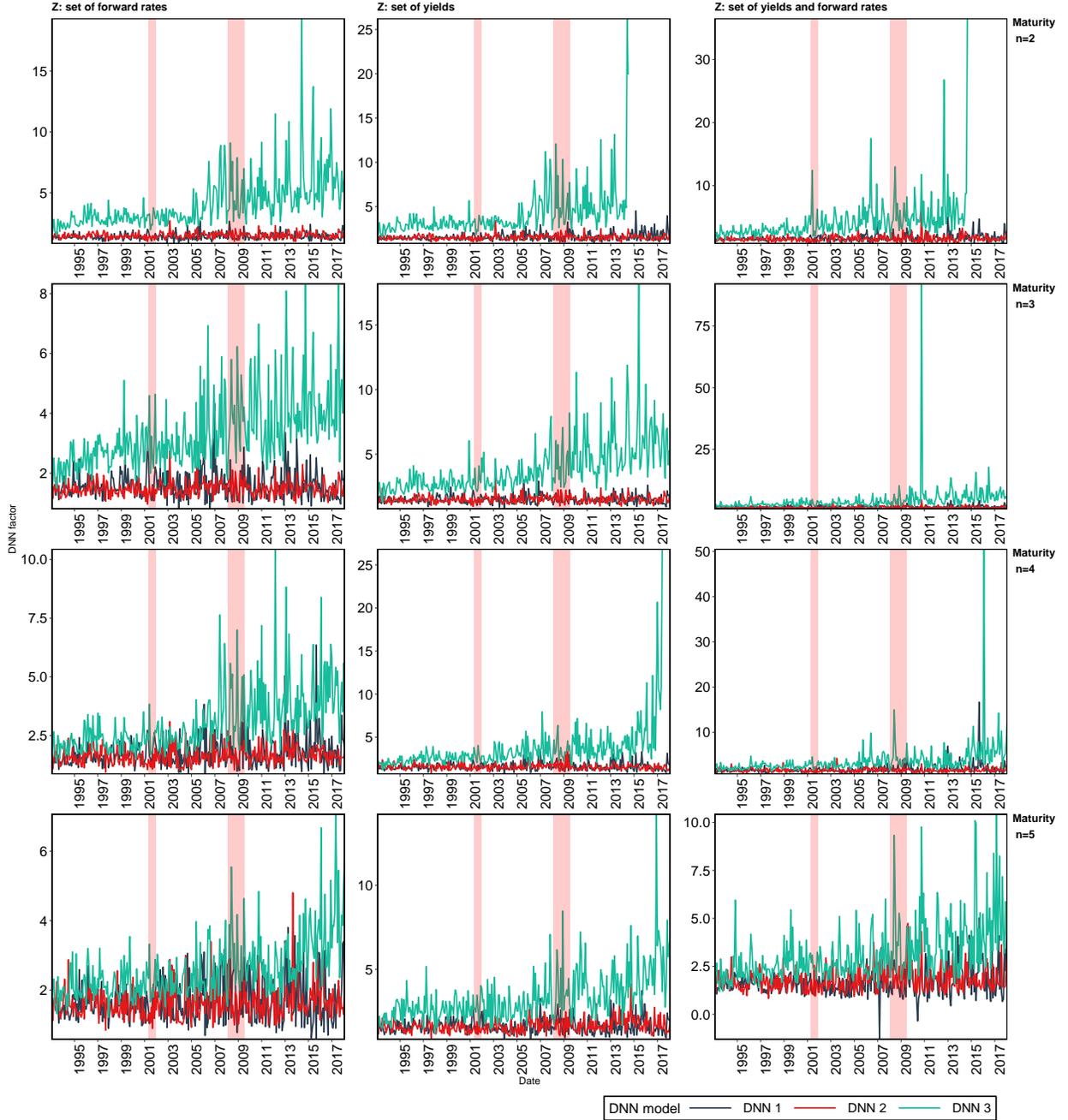
Figure 4 shows the derived $\mathfrak{f}_{t,DNN}^{(n),h}$ for each scenario under consideration. Each column uses a different set of information from the term-structure to derive the factor $\mathfrak{f}_{t,DNN}^{(n),h}$. Column (1) shows the the derived DNN factors for $\boldsymbol{Z}_t^{\text{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60)} \right\}$, column (2) for $\boldsymbol{Z}_t^{\text{y}} = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$ and column (3) for $\boldsymbol{Z}_t^{\text{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)}, y_t^{(1)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$. Each row represents one of the four groups of maturities. Finally, different colors represent the three variations of DNN considered, as explained in section 3.2. The derived factors are calculated for the period 1993:01 to 2017:12, where we use the data from 1962:01 to 1992:12 as a burn-in data to initiate the recursive process of obtaining the the derived factors $\mathfrak{f}_{t,DNN}^{(n),h}$.

Figure 5: Derived Factors $\mathfrak{f}_{t,DNN}^{(n),h}$ for **DNN 2** Generated Using the Set of Yields
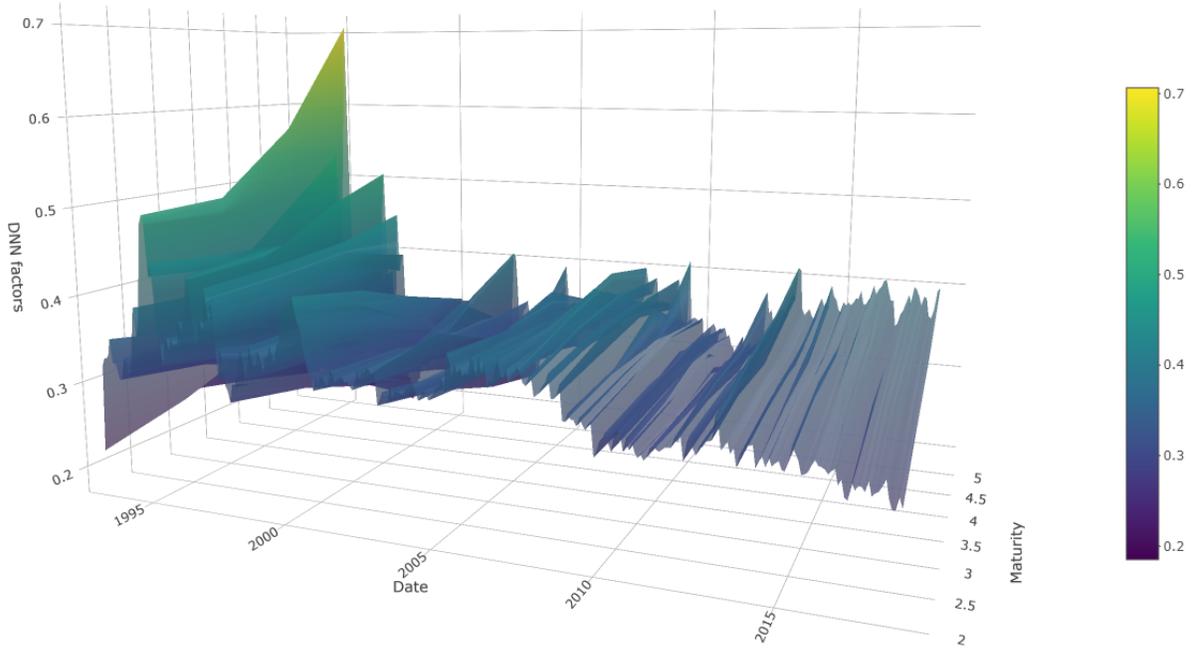
Figure 5 shows a 3D representation of $\mathfrak{f}_{t,DNN}^{(n),h}$ generated by the MLP architecture **DNN 2** using the set of yields $\boldsymbol{Z}_t^Y = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$ in terms of maturity for the period of analysis. **DNN 2** is a feedfoward neural networks architecture with 2 hidden layers ($L = 2$), with 16 and 4 nodes respectively, and an output layer for each group of maturity $n \in \{1, 2, 3, 4\}$. Period of analysis ranges from 1993:01 to 2017:12.

Figure 6: Single Factor $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ Series by DNN Architecture and Choice of $\boldsymbol{Z}_t^{\mathrm{y}}$
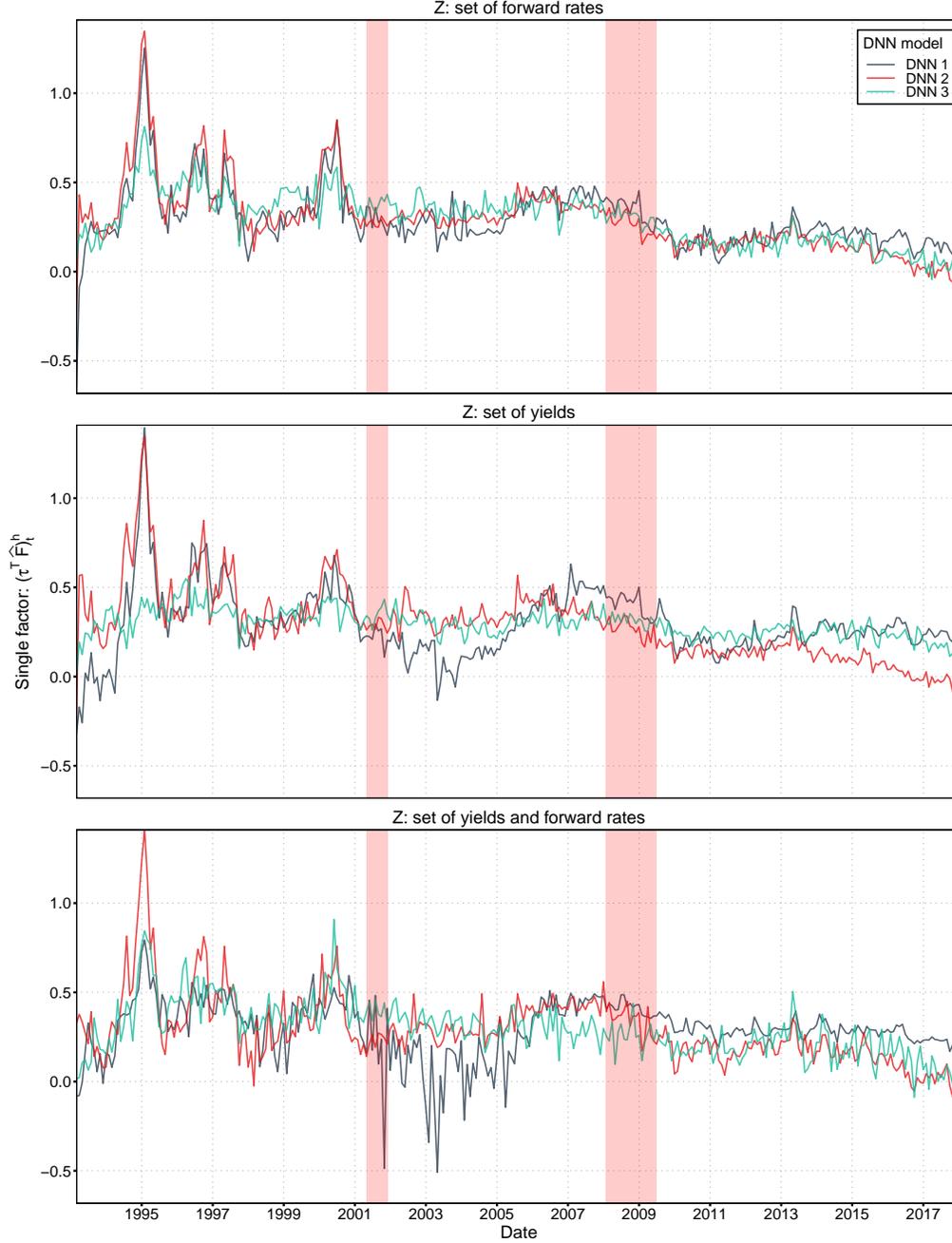
Figure 6 shows $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ for each DNN architecture and the three different sets of $\boldsymbol{Z}_t^{\mathrm{y}}$. The first panel plots $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ when $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60)}\right\}$ is used to obtain $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ from the DNN derived factors $\mathfrak{f}_{t,DNN}^{(n),h}$. The panel in the center plots the single factor when $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}\right\}$ is used. Finally, the third panel plots $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ when $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)}, y_t^{(1)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}\right\}$ is used. Different colors represent the three variations of DNN considered, as explained in section 3.2. The derived factors are calculated for the period 1993:01 to 2017:12, where we use the data from 1962:01 to 1992:12 as a burn-in data to initiate the recursive process.

23

following two regression models. In each panel, we show the these three regressions depending on wich DNN architecture was used to build the single state factor $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_{t}\right)_{t}^{h}$.

From table 1 we see that for 1-month holding period, with no overlapping returns to affect the robustness of our tests, our state variable $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_{t}\right)_{t}^{h}$, when used as the only predictor, is always statistically significant for **DNN 1** and **DNN 2**. For **DNN 3**, the single factor loses statistically significance when the maturity increases. More importantly, the adjusted R^2 ranges for maturity of 2 years, for maturity of 2 years, for maturity of 2 years, and for maturity of 5 years. When we add the second state variable that captures the unspanned factors, we keep seeing statistically significance for the same cases, and the adjusted R^2 raises quite substantially, either for alternative 1 $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_{t}^{h}$, or alternative 2 $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_{t}^{(-n),h}$.

As we discussed above, for each DNN architecture and each set of Z used, we obtained a state factor $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_{t}\right)_{t}^{h}$ for the time varying of the expected returns across all maturities. Out of the 9 different specifications for this single factor, we will focus only on the one formed using the $\mathfrak{f}_{t,DNN}^{(n),h}$ from the **DNN 2** fed with the entire set of yields $\boldsymbol{Z}_{t}^{\mathrm{y}} = \left\{y_{t}^{(1/12)}, y_{t}^{(2/12)}, \ldots, y_{t}^{(60/12)}\right\}$. We do so motivated by two reasons. First, because as shown in Gu et al. (2018), higher complexity with a much "deeper" network it is not necessary associated with better out-of-sample results. And second, because this pair of choices result in smaller MSE in our period of analysis.

Table 2 presents the correlation between our state variables, i.e., $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_{t}\right)_{t}^{h}$, $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_{t}^{h}$ and $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_{t}^{(-n),h}$, as well as with the Cochrane-Piazzesi and Ludvingson-Ng factors calculated as explained in section 2.2. By definition the correlation between our factor that summarizes the information from the term-structure and the alternatives for the one(s) that complete the state space is 0, which we can see in table 2. Now the correlation between our factors for the unspanned information from the yield curve are always high, ranging from .84 to .99. We see that the correlation between $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_{t}^{h}$ and the factor for low maturities, especially $n = 2$, is high (.99). For the remaining one, we notice that the correlation decays.

Table 1: Predictive Regressions Using $\left(\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}_t\right)_t^h$, $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)_t^h$ and $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$ as State Variables

Panel A:

$rx^{(2)}_{t+h/12}$

	DNN 1			DNN 2			DNN 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$(\boldsymbol{\tau}^\top\widehat{\mathfrak{S}})_t^h$	0.810***	0.810***	0.810***	0.811***	0.811***	0.811***	1.419***	1.419***	1.419***
	(0.160)	(0.149)	(0.147)	(0.131)	(0.119)	(0.119)	(0.414)	(0.377)	(0.356)
$M_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-2),h}$		0.760***			0.779***			0.875***	
		(0.204)			(0.180)			(0.211)	
$M_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$			0.591***			0.525***			0.679***
			(0.139)			(0.126)			(0.138)
Constant	−0.010	−0.010	−0.010	−0.010	−0.010	−0.010	−0.189*	−0.189*	−0.189**
	(0.054)	(0.050)	(0.049)	(0.039)	(0.035)	(0.035)	(0.110)	(0.101)	(0.094)
Observations	300	300	300	300	300	300	300	300	300
Adjusted R^2	0.100	0.148	0.159	0.119	0.178	0.175	0.046	0.105	0.124

Panel B:

$rx^{(3)}_{t+h/12}$

	DNN 1			DNN 2			DNN 3		
$(\boldsymbol{\tau}^\top\widehat{\mathfrak{S}})_t^h$	0.959***	0.959***	0.959***	0.943***	0.943***	0.943***	1.175*	1.175**	1.175**
	(0.248)	(0.234)	(0.233)	(0.199)	(0.188)	(0.184)	(0.630)	(0.566)	(0.559)
$M_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-3),h}$		0.799***			0.789***			0.984***	
		(0.234)			(0.219)			(0.236)	
$M_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^h$			0.765***			0.757***			0.929***
			(0.225)			(0.205)			(0.224)
Constant	−0.008	−0.008	−0.008	−0.003	−0.003	−0.003	−0.072	−0.072	−0.072
	(0.087)	(0.082)	(0.082)	(0.063)	(0.060)	(0.059)	(0.169)	(0.153)	(0.150)
Observations	300	300	300	300	300	300	300	300	300
Adjusted R^2	0.055	0.092	0.093	0.063	0.100	0.109	0.010	0.067	0.067

(Continued)

Table 1: Predictive Regressions Using $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{S}}_t\right)_t^h$, $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^h$ and $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$ as State Variables *(Continued)*

Panel C: $rx_{t+h/12}^{(4)}$

$(\boldsymbol{\tau}^\top \widehat{\mathfrak{S}}_t)_t^h$	1.073***	1.073***	1.073***	1.065***	1.065***	1.065***	0.864	0.864	0.864
	(0.334)	(0.320)	(0.317)	(0.264)	(0.253)	(0.248)	(0.835)	(0.759)	(0.755)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-4),h}$		0.795***				0.807***			1.038***
		(0.291)				(0.288)			(0.289)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$			0.902***			0.945***			1.144***
			(0.312)			(0.284)			(0.313)
Constant	0.002	0.002	0.002	0.004	0.004	0.004	0.063	0.063	0.063
	(0.120)	(0.116)	(0.115)	(0.088)	(0.086)	(0.085)	(0.228)	(0.209)	(0.207)
Observations	300	300	300	300	300	300	300	300	300
Adjusted R^2	0.036	0.060	0.063	0.042	0.069	0.080	0.001	0.046	0.046

Panel D: $rx_{t+h/12}^{(5)}$

$(\boldsymbol{\tau}^\top \widehat{\mathfrak{S}}_t)_t^h$	1.158***	1.158***	1.158***	1.181***	1.181***	1.181***	0.542	0.542	0.542
	(0.415)	(0.395)	(0.398)	(0.325)	(0.312)	(0.309)	(1.025)	(0.949)	(0.939)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-5),h}$		0.854**				0.848***			1.069***
		(0.336)				(0.318)			(0.339)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$			1.000**			1.081***			1.322***
			(0.398)			(0.363)			(0.404)
Constant	0.017	0.017	0.017	0.010	0.010	0.010	0.198	0.198	0.198
	(0.152)	(0.146)	(0.147)	(0.114)	(0.111)	(0.111)	(0.284)	(0.267)	(0.263)
Observations	300	300	300	300	300	300	300	300	300
Adjusted R^2	0.025	0.049	0.046	0.032	0.060	0.062	-0.002	0.033	0.036

Note: *p<0.1; **p<0.05; ***p<0.01

Table 1 reports the predictive regressions using $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{S}}_t\right)_t^h$, $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^h$ and $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$ as state variables for 1-month holding period ($h = 1$). Panel A reports the predictive regressions for maturity $n = 2$ years. Panel B reports the predictive regressions for maturity $n = 3$ years. Panel C reports the predictive regressions for maturity $n = 4$ years. Panel D reports the predictive regressions for maturity $n = 5$ years. The state factor $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{S}}_t\right)_t^h$ reported in this table used only the set of yields $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$ to feed the MLP. We use Newey-West robust standard errors. Sample ranges from $1993:01$ to $2017:12$.

	$(\tau^\top \widehat{\mathfrak{F}})_t^h$	$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^h$	$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^{(-2),h}$	$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^{(-3),h}$	$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^{(-4),h}$	$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^{(-5),h}$	$\widehat{CP}_t^h$	$\widehat{LN}_t^h$
$(\tau^\top \widehat{\mathfrak{F}})_t^h$	1	0	0	0	0	0	0.556	-0.059
$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^h$	0	1	0.995	0.912	0.904	0.919	0.129	0.171
$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^{(-2),h}$	0	0.995	1	0.938	0.900	0.888	0.135	0.174
$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^{(-3),h}$	0	0.912	0.938	1	0.947	0.849	0.170	0.203
$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^{(-4),h}$	0	0.904	0.900	0.947	1	0.959	0.173	0.204
$M_{\tau^\top \widehat{\mathfrak{s}}}(\kappa^\top \widehat{\xi})_t^{(-5),h}$	0	0.919	0.888	0.849	0.959	1	0.146	0.178
$\widehat{CP}_t^h$	0.556	0.129	0.135	0.170	0.173	0.146	1	-0.007
$\widehat{LN}_t^h$	-0.059	0.171	0.174	0.203	0.204	0.178	-0.007	1

Table 2 reports the correlation between our single factor $\left(\tau^\top \widehat{\mathfrak{F}}_t\right)_t^h$, with the factors that complete the state space in our dynamic term-structure model. The first alternative is $\left(\kappa^\top \widehat{\xi}\right)_t^h$, which is the unique factor obtained as the projection of $\overline{rx}_{t+h/12}$ in $\widehat{\xi}_{t+h/12}$. The second alternative is a similar projection, however for each maturity $n \in \{2,3,4,5\}$ we regress $rx_{t+h/12}^{(n)}$ on $\widehat{\xi}_{t+h/12}^{(-n),h} \equiv \widehat{\xi}_{t+h/12}^h \setminus \widehat{\xi}_{t+h/12}^{(n),h}$. We use orthogonal vector from the projection of each one of them on $\left(\tau^\top \widehat{\mathfrak{F}}_t\right)_t^h$ to complete our state space. The table also reports the correlation with the Cochrane-Piazzesi and Ludvingson-Ng factors calculated as explained in section 2.2. The period of analysis ranges from $1993:01$ to $2017:12$.

4.2 Comparison with Other Factors from the Literature

In this section, we are interested in evaluating how our derived and theoretically motivated factors compare with the other factors and frameworks that were proposed in the literature to explain the time-varying expected excess returns. Figure 7 shows in two separated panels our single factor that spans the information from the term-structure, $\left(\tau^\top \widehat{\mathfrak{F}}_t\right)_t^h$, as well as the factor with the spanned risks from alternative 1, $\left(\kappa^\top \widehat{\xi}\right)_t^h$, along with the Cochrane-Piazzesi and Ludvingson-Ng factors. Aligned with the correlation in table 2, we see that our factor has some positive correlation (.56) with the Cochrane-Piazzesi factor. However, this correlation is not strong enough to claim that both are capturing the same information. We must say that this should be an expected result, given that both factors capture information from the term-structure.

On the other hand, $\left(\tau^\top \widehat{\mathfrak{F}}_t\right)_t^h$ seems to be uncorrelated $(-.06)$ with the Ludvingson-Ng factor. Now, the time series in figure 7 with the correlation shows us an interesting result. Consistent with our framework, the unspanned risks from the term-structure should be captured by our orthogonal factor $\left(\kappa^\top \widehat{\xi}\right)_t^h$, or $\left(\kappa^\top \widehat{\xi}\right)_t^{(-n),h}$. Given that Ludvingson-Ng factor is solely based in macroeconomic variables information, we see that the correlation of $\widehat{LN}_t^h$ with our unspanned risks factors ranges from .17 to .20. This could be understood as the risk factors not spanned by the yield-curve, that are captured by our orthogonal state

Figure 7: Time Series of our Derived Factors $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)^h_t$ and $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)^h_t$, along with $\widehat{CP}^h_t$ and $\widehat{LN}^h_t$
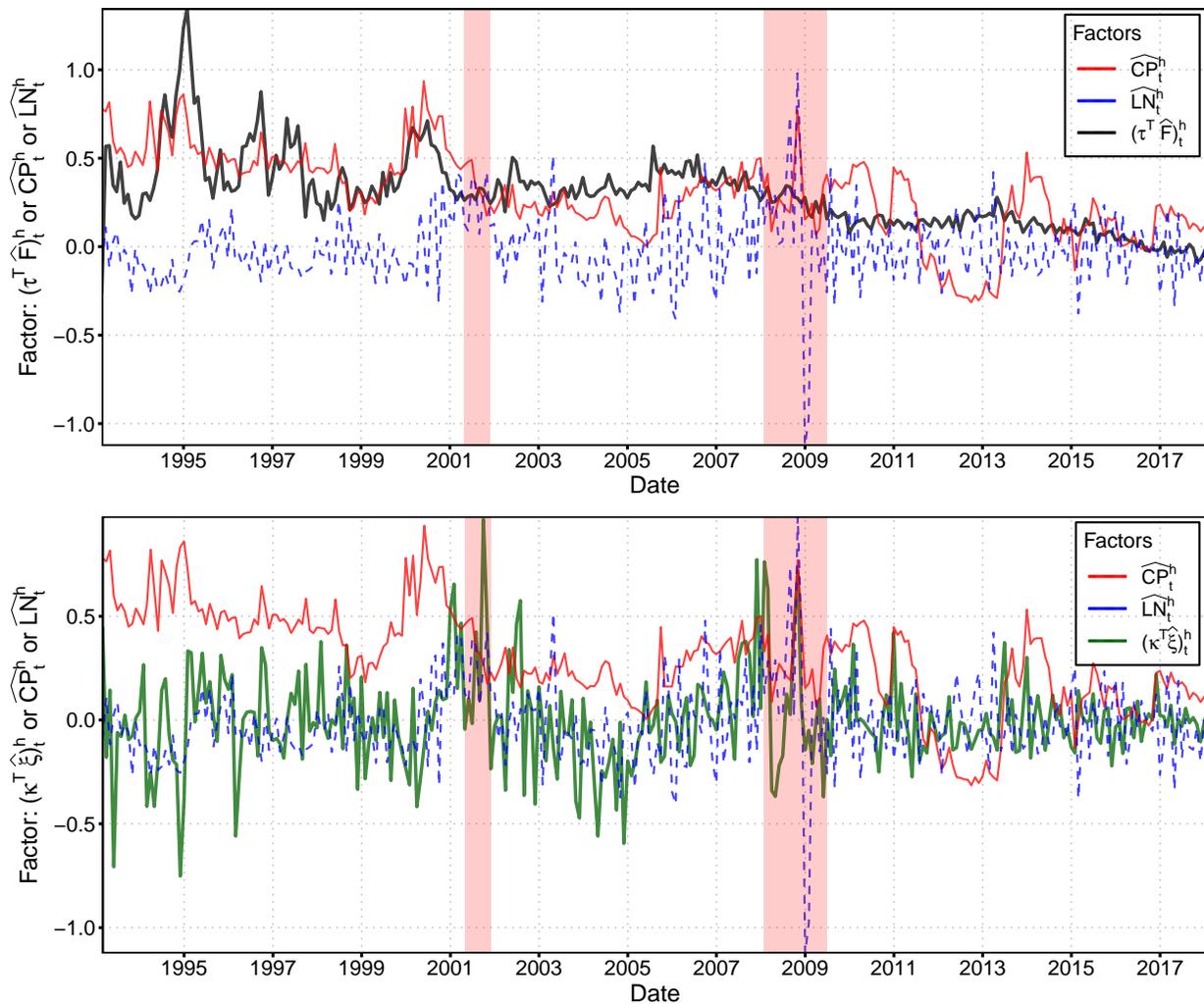
Figure 7 plots in two separated panels our single factor that spans the information from the term-structure, as well as the factor with the spanned risks (alternative 1). The graph in the top plots $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)^h_t$ along with the Cochrane-Piazzesi and Ludvingson-Ng factors. The bottom graph plots $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)^h_t$ along with the same factors. The period of analysis ranges from $1993:01$ to $2017:12$.

Table 3: Predictive Regressions with $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)^h_t$ and $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)^{(-n),h}_t$, along with the Cochrane-Piazzesi and Ludvingson-Ng factors, and Fama-Bliss Regressions with Forward Spreads

Panel A: $rx^{(2)}_{t+h/12}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}})^h_t$	0.847***	0.842***	0.853***	0.824***	0.525***	0.582***	0.582***	0.614***
	(0.124)	(0.115)	(0.128)	(0.117)	(0.154)	(0.140)	(0.145)	(0.135)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})^{(-2),h}_t$		0.658***		0.745***		0.704***		0.558***
		(0.172)		(0.182)		(0.182)		(0.185)
$\widehat{LN}^h_t$	0.617***	0.529***					0.559***	0.518***
	(0.127)	(0.120)					(0.110)	(0.110)
$fs^{(n,h)}_t$			−0.746	−0.225			−0.570	−0.172
			(0.476)	(0.438)			(0.437)	(0.429)
$\widehat{CP}^h_t$					0.454***	0.364***	0.465***	0.375***
					(0.126)	(0.112)	(0.112)	(0.109)
Constant	−0.013	−0.012	0.031	0.002	−0.060	−0.050	−0.031	−0.044
	(0.037)	(0.034)	(0.051)	(0.047)	(0.039)	(0.036)	(0.045)	(0.043)
Observations	300	300	300	300	300	300	300	300
Adjusted R^2	0.183	0.223	0.128	0.177	0.150	0.197	0.215	0.240

Panel B: $rx^{(3)}_{t+h/12}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}})^h_t$	0.996***	0.989***	0.940***	0.947***	0.559**	0.648***	0.626***	0.719***
	(0.190)	(0.184)	(0.199)	(0.188)	(0.245)	(0.234)	(0.238)	(0.237)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})^{(-3),h}_t$		0.620***		0.852***		0.692***		0.585**
		(0.209)		(0.228)		(0.226)		(0.237)
$\widehat{LN}^h_t$	0.921***	0.800***					0.900***	0.823***
	(0.209)	(0.201)					(0.194)	(0.191)
$fs^{(n,h)}_t$			−0.215	0.410			−0.053	0.394
			(0.554)	(0.532)			(0.525)	(0.542)
$\widehat{CP}^h_t$					0.608***	0.467**	0.583***	0.437**
					(0.205)	(0.195)	(0.188)	(0.198)
Constant	−0.007	−0.006	0.021	−0.049	−0.070	−0.054	−0.064	−0.098
	(0.060)	(0.059)	(0.091)	(0.087)	(0.063)	(0.061)	(0.082)	(0.082)
Observations	300	300	300	300	300	300	300	300
Adjusted R^2	0.120	0.141	0.060	0.099	0.084	0.111	0.136	0.151

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3 reports the predictive regressions using $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)^h_t$ and $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)^{(-n),h}_t$, along with $\widehat{CP}^h_t$, $\widehat{LN}^h_t$ and the Fama-Bliss regressions with forward spreads for 1-month holding period ($h = 1$). Panel A reports the predictive regressions for maturity $n = 2$ years. Panel B reports the predictive regressions for maturity $n = 3$ years. The state factor $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)^h_t$ reported in this table used only the set of yields $\boldsymbol{Z}^{\text{y}}_t = \left\{y^{(1/12)}_t, y^{(2/12)}_t, \ldots, y^{(60/12)}_t\right\}$ to feed the MLP. We use Newey-West robust standard errors. Sample ranges from $1993:01$ to $2017:12$.

Table 4: **(Continued)** Predictive Regressions with $\left(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ and $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$, along with the Cochrane-Piazzesi and Ludvingson-Ng factors, and Fama-Bliss Forward Spreads

Panel C: $rx^{(4)}_{t+h/12}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h$	1.135***	1.127***	1.082***	1.108***	0.547	0.651**	0.685**	0.790**
	(0.254)	(0.247)	(0.270)	(0.257)	(0.335)	(0.323)	(0.329)	(0.329)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-4),h}$		0.609**		0.872***		0.688**		0.555**
		(0.262)		(0.289)		(0.291)		(0.274)
$\widehat{LN}_t^h$	1.218***	1.079***					1.222***	1.118***
	(0.307)	(0.287)					(0.285)	(0.273)
$fs_t^{(n,h)}$			0.260	0.665			0.386	0.655
			(0.622)	(0.595)			(0.593)	(0.587)
$\widehat{CP}_t^h$					0.822***	0.657**	0.755***	0.606**
					(0.290)	(0.276)	(0.265)	(0.272)
Constant	−0.0003	0.0002	−0.038	−0.103	−0.085	−0.068	−0.144	−0.171
	(0.085)	(0.084)	(0.130)	(0.124)	(0.089)	(0.087)	(0.121)	(0.118)
Observations	300	300	300	300	300	300	300	300
Adjusted R²	0.095	0.108	0.039	0.070	0.063	0.081	0.112	0.122

Panel D: $rx^{(5)}_{t+h/12}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h$	1.268***	1.258***	1.247***	1.263***	0.511	0.626	0.736*	0.834**
	(0.315)	(0.305)	(0.334)	(0.318)	(0.422)	(0.401)	(0.409)	(0.400)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-5),h}$		0.673**		0.872***		0.738**		0.590**
		(0.281)		(0.312)		(0.315)		(0.279)
$\widehat{LN}_t^h$	1.501***	1.337***					1.518***	1.386***
	(0.421)	(0.381)					(0.387)	(0.360)
$fs_t^{(n,h)}$			0.633	0.789			0.739	0.848
			(0.698)	(0.656)			(0.658)	(0.632)
$\widehat{CP}_t^h$					1.064***	0.882**	0.967***	0.818**
					(0.380)	(0.352)	(0.343)	(0.337)
Constant	0.005	0.005	−0.116	−0.147	−0.106	−0.086	−0.248	−0.253*
	(0.111)	(0.109)	(0.166)	(0.158)	(0.117)	(0.115)	(0.158)	(0.152)
Observations	300	300	300	300	300	300	300	300
Adjusted R²	0.082	0.098	0.031	0.062	0.054	0.074	0.103	0.114

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4 reports the predictive regressions using $\left(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ and $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$, along with $\widehat{CP}_t^h$, $\widehat{LN}_t^h$ and the Fama-Bliss regressions with forward spreads for 1-month holding period ($h=1$). Panel C reports the predictive regressions for maturity $n=4$ years. Panel D reports the predictive regressions for maturity $n=5$ years. The state factor $\left(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ reported in this table used only the set of yields $\boldsymbol{Z}_t^y = \left\{y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}\right\}$ to feed the MLP. We use Newey-West robust standard errors. Sample ranges from $1993:01$ to $2017:12$.

variable and Ludvingson-Ng approach.

Next, we run predictive regressions using our factors with the main factors proposed in the literature. Tables 3 and 4 reports the predictive regressions using $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$, $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_t^h$ and $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$, along with $\widehat{CP}_t^h$, $\widehat{LN}_t^h$ and the Fama-Bliss regressions with forward spreads for 1-month holding period ($h = 1$). For each maturity (in each one of the four panels), there are 8 different regressions specifications. In pairs, we run predictive regressions first only with our state variable that spans the term-structure along with a proposed factor from the literature. Then, we add our state variable of the unspanned risks.

The results, consistent with table 1, shows that our state variables are still significant when adding either CP LN or forward spreads especially for maturities $n = 2$ and $n = 3$. Interestingly, the forward spreads loose statistical significance. In column (8) we see that our factors remain significant if we still add all factors, including $\widehat{CP}_t^h$, $\widehat{LN}_t^h$ and the forward spreads. As already mentioned, nonetheless, for higher maturities our factors loose some statistical significance.

4.3   Economic Interpretation

Some natural questions may arise at this stage. What are the economic interpretation of these factors derived from a deep neural network? How are they linked with the macroeconomic variables? What macroeconomic and possibly other risk measures do they capture? In order to answer these questions, we make use of the the FRED-MD dataset (McCracken and Ng, 2016), which is a large macroeconomic database and monthly updated by the FRED[9] that shares the predictive content of a widespread dataset known in the literature as Stock-Watson (Stock and Watson (1996)). It is a balanced panel consisting of 128 macroeconomic and financial variables. The variables are split in 8 groups: (1) output and income, (2) labor market, (3) housing, (4) consumption, orders, and inventories, (5) money and credit, (6) interest and exchange rates, (7) prices, and (8) stock market. In Appendix A.1, table 11 list all the variables, codes and their groups.

In a similar fashion to Ludvigson and Ng (2009), we find the marginal R^2 of our factors $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$, $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ and $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^h$. The marginal R^2 simply is the goodness-of-fit of the regression of each one the 128 variables from the FRED-MD on our state variables. Figure 8 reports the marginal R^2 as bar charts using colors to split the 8 groups. A quick inspection in this figure reveals that $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ has a high R^2 with many macroeconomic variables. However, this is not evenly distributed within and across groups. We can see that especially the groups (7) *prices* and (5) *money and credit* have R^2 above or

Figure 8: Marginal R^2 of the factors $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{S}}_t\right)_t^h$ and $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{s}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^h$
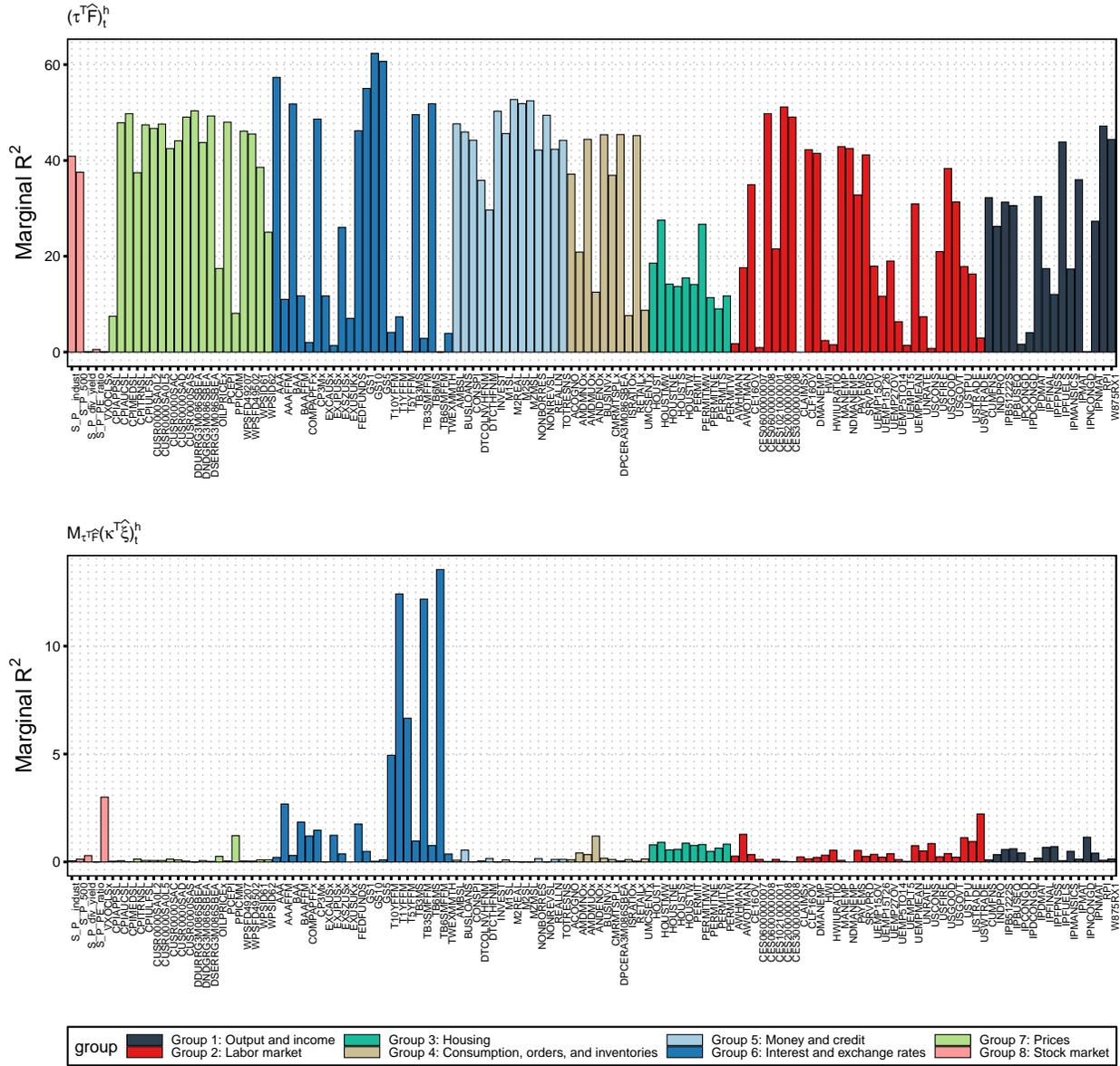
Figure 8 reports the marginal R^2 of the factor $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{S}}_t\right)_t^h$ in the top panel, and $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{s}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^h$ in the bottom panel. The marginal R^2 is obtained with the regression of each one the 128 variables from the FRED-MD on $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{S}}_t\right)_t^h$ or $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{s}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^h$. Sample ranges from $1993:01$ to $2017:12$.

Figure 9: Marginal R^2 of the factors $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$
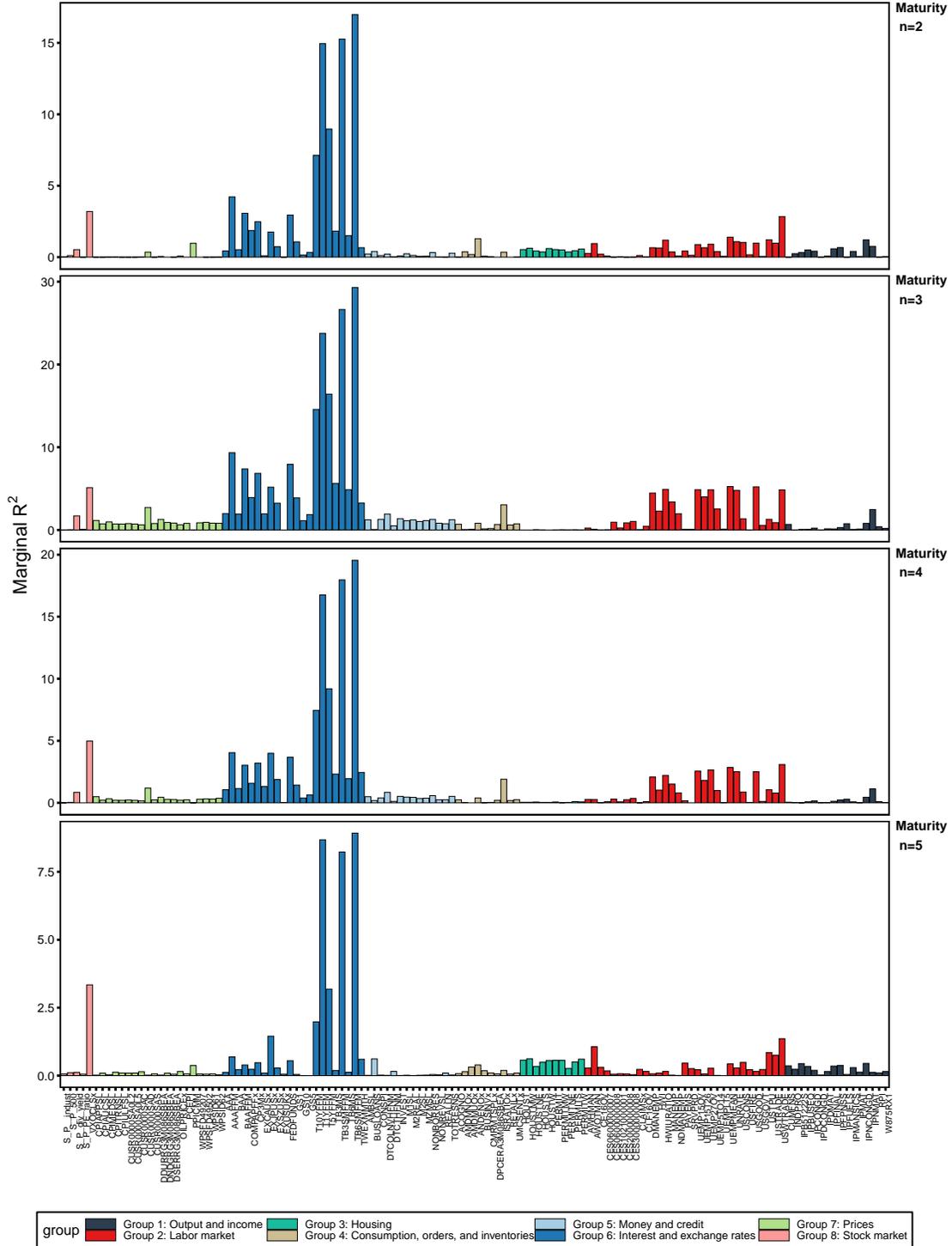
Figure 9 reports the marginal R^2 of the factors $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ by maturity. The marginal R^2 is obtained with the regression of each one the 128 variables from the FRED-MD on $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$. Sample ranges from $1993:01$ to $2017:12$.

Figure 10: Marginal R^2 Using Sentiment-Based Measures
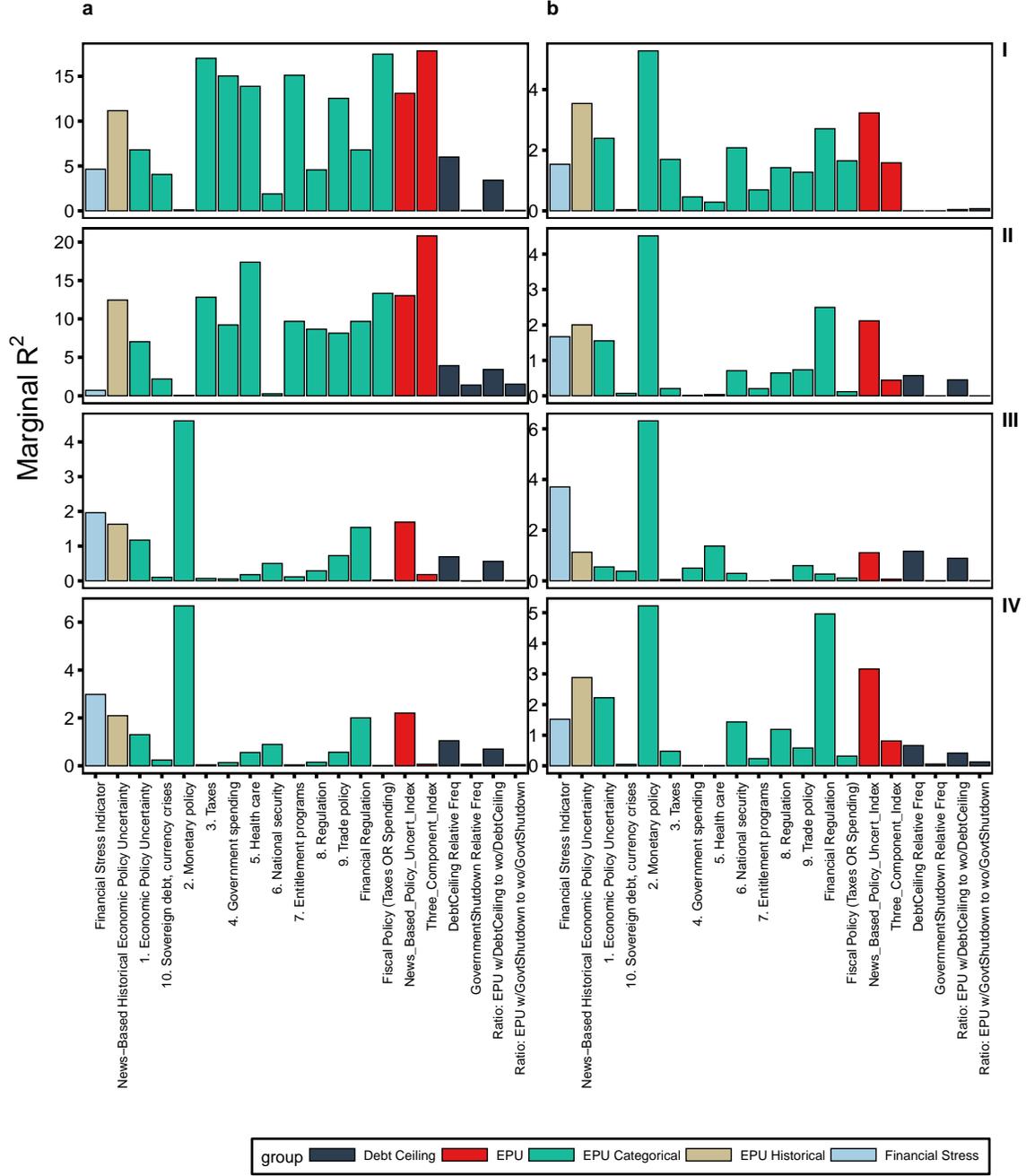
Figure 10 reports the marginal R^2 obtained from sentiment-based measures. It is obtained with the regression of each one these indexes on our state variables. For comparison, we also report for the Cochrane-Piazzesi and Ludvingson-Ng factors. Row (I), panel (a) shows the marginal R^2 for $\widehat{CP}_t^h$, and panel (b) plots for $\widehat{LN}_t^h$. Row (II) panel (a) plots for our spanning factor $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$, and panel (b) for the unspanned factor $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^h$. Rows (III) and (IV) plots for the other derived unspanned state variables: in panel (III-a) we have $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-2),h}$, panel (III-b) plots $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-3),h}$, panel (IV-a) shows the marginal R^2 for $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-4),h}$, and panel (IV-b) for $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-5),h}$. Sample ranges from $1993:01$ to $2017:12$.

around .40 for most of their variables. Even though the groups (6) *interest and exchange rates* and (2) *labor market* have some variables with high R^2, there are many others within the group that do not. Thus, apparently, the state variable spanning the yield curve loads more in monetary variables movements, what should be expected. Nonetheless, it also captures a wide range of macroeconomic variables.

An interesting pattern stands out when we calculate the marginal R^2 with our unspanned factor from the term-structure. The graph in the bottom of figure 8 shows the marginal R^2 for $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^h$, while figure 9 shows for each $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$ in different panels. It is clear that our state variable of the unspanned risks capture important information left out by the spanning factor. We see that especially some variables from the group (6) *interest and exchange rates* stand out. This pattern is consistent either for $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^h$, or the four factors $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$. Among these variables that have a high R^2 with our unspanned factor, there are relevant variables, such as *6-Month Treasury C Minus FEDFUNDS* (TB6SMFFM), *1-Year Treasury C Minus FEDFUNDS* (T1YFFM) and *3-Month Treasury C Minus FEDFUNDS* (TB3SMFFM).

Next, we evaluate if our factors capture any sentiment information. To do so, we make use of several indexes recently proposed in the literature that seek to estimate the state of the sentiment in the economy. The first one is the economic policy uncertainty measure (EPU) from Baker et al. (2016). The EPU is an index that proxies for movements in policy-related economic uncertainty for U.S., being based on newspaper coverage frequency. The authors also calculated a categorical EPU, which is derived using results from the Access World News database of over 2,000 US newspapers, in such a way that each one of the sub-indexes requires the economic uncertainty term, as well as a set of categorical policy terms[10].

In the sense of the EPU, we also use the financial stress indicator (FSI) for the U.S from Püttmann (2018). The essence of the FSI is being an indicator of negative financial sentiment. It is based on the reporting in five major US newspapers[11]. Püttmann (2018) shows that the FSI is a robust indicator, such that an increase in negative financial sentiment is followed by a fall in output, higher unemployment, lower stock market returns and rising corporate bond spreads.

Figure 10 plots in each panel the marginal R^2 obtained using these sentiment-based measures, where we use colors to split between each index. Row (I), panel (a) shows the

---

[10]As an example, the category *Monetary policy* has the following terms: Monetary policy - federal reserve, the fed, money supply, open market operations, quantitative easing, monetary policy, fed funds rate, overnight lending rate, Bernanke, Volcker, Greenspan, central bank, interest rates, fed chairman, fed chair, lender of last resort, discount window, European Central Bank, ECB, Bank of England, Bank of Japan, BOJ, Bank of China, Bundesbank, Bank of France, Bank of Italy

[11]Boston Globe, Chicago Tribune, Los Angeles Times, Wall Street Journal and Washington Post.

marginal R^2 for $\widehat{CP}_t^h$, and panel (b) plots for $\widehat{LN}_t^h$ for comparison. Row (II) panel (a) plots for our spanning factor $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$, and panel (b) for the unspanned factor $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^h$. Rows (III) and (IV) plots for the other derived unspanned state variables: in panel (III-a) we have $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-2),h}$, panel (III-b) plots $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-3),h}$, panel (IV-a) shows the marginal R^2 for $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-4),h}$, and panel (IV-b) for $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-5),h}$.

It is clear 10 from figure three facts: (i) our spanning factor and the Cochrane-Piazzesi factor have similar marginal R^2, (ii) our unspanned state factors and the Ludvingson-Ng also have similar marginal R^2, and most important (iii) our unspanned factors has their highest R^2 with the categorical EPU related to monetary policy. Therefore, there is some evidence that the spanned factor $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$, or even the Cochrane-Piazzesi factor, cannot capture some economy sentiment associated with possible changes in the monetary policy.

4.4   Out-of-Sample Forecasting Performance

In this section we are interested in to know how the predictive regressions using our DNN derived state variables behave in an out-of-sample (OoS) analysis. Following Campbell and Thompson (2007); Gargano et al. (2019), we compute the out-of-sample R^2 for all possible predictive regression models from tables 1 and 3. Additionally, we also consider univariate predictive regressions using only $\widehat{LN}_t^h$, or $fs_t^{(n,h)}$, or $\widehat{CP}_t^h$. We set the out-of-sample period to range from $1997:01$ to $2017:12$, where the data from $1993:01$ to $1996:12$ is used to initiate the analysis. To avoid any look-ahead bias, at each $\tau \in \tau_{OoS}$, where τ_{OoS} is the OoS subsample, we use all the previous information up to $\tau - 1$ to obtain the point forecast of $rx^{(n)}$ for the month τ. The out-of-sample R^2 is computed as

$$R_{OoS,i}^{2(n)} = 1 - \frac{\sum_{\tau \in \tau_{OoS}} \left(rx_{t+h/12|t}^{(n)} - \widehat{rx}_{t+h/12|t}^{(n)}\right)^2}{\sum_{\tau \in \tau_{OoS}} \left(rx_{t+h/12|t}^{(n)} - \overline{rx}_{t+h/12|t}^{(n)}\right)^2} \tag{29}$$

where $\widehat{rx}_{t+h/12|t}^{(n)}$ is the estimate of the conditional mean of the excess returns for the bond with maturity (n), and $\overline{rx}_{t+h/12|t}^{(n)}$ is the estimate of the conditional mean assuming that the excess returns are constant (as under the expectation hypothesis), implying that the βs from all predictive regressions are assumed to be zero for the same bond with maturity (n). Notice that evidence of time-varying return predictability is obtained when the out-of-sample R^2 is positive.

Table 5: Out-of-Sample R^2

Regression	Maturity $n=2$	Maturity $n=3$	Maturity $n=4$	Maturity $n=5$
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t)_t^h + \epsilon_{t+h/12}$	0.17	0.03	-0.02	-0.04
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{s}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-n),h} + \epsilon_{t+h/12}$	0.21	0.05	-0.01	-0.02
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{s}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h + \epsilon_{t+h/12}$	0.22	0.05	-0.01	-0.03
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2 \widehat{LN}_t^h + \epsilon_{t+h/12}$	0.21	0.04	-0.03	-0.05
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2 \boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{s}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-n),h} + \beta_3 \widehat{LN}_t^h + \epsilon_{t+h/12}$	0.23	0.04	-0.02	-0.05
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2 fs_t^{(n,h)} + \epsilon_{t+h/12}$	0.26	0.08	0.02	-0.00
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2 \boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{s}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-n),h} + \beta_3 fs_t^{(n,h)} + \epsilon_{t+h/12}$	0.27	0.08	0.02	-0.00
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2 \widehat{CP}_t^h + \epsilon_{t+h/12}$	0.20	0.01	-0.06	-0.09
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2 \boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{s}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-n),h} + \beta_3 \widehat{CP}_t^h + \epsilon_{t+h/12}$	0.22	0.01	-0.06	-0.08
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2 \widehat{LN}_t^h + \beta_3 fs_t^{(n,h)} + \beta_4 \widehat{CP}_t^h + \epsilon_{t+h/12}$	0.19	-0.03	-0.10	-0.13
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2 \boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{s}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-n),h} + \beta_3 \widehat{LN}_t^h + \beta_4 fs_t^{(n,h)} + \beta_5 \widehat{CP}_t^h + \epsilon_{t+h/12}$	0.19	-0.04	-0.11	-0.13
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{LN}_t^h + \epsilon_{t+h/12}$	0.12	-0.02	-0.06	-0.07
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \epsilon_{t+h/12}$	0.18	0.05	0.00	-0.01
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{CP}_t^h + \epsilon_{t+h/12}$	0.15	-0.02	-0.08	-0.10

Table 5 reports the OoS R^2 of our predictive regressions. The first three rows present the same regressions models from table 1, the last three rows report the univariate predictive regression using the other factors from the literature: $\widehat{LN}_t^h$, $fs_t^{(n,h)}$, and $\widehat{CP}_t^h$. Finally, the remaining rows are the same regressions models from table 3. The out-of-sample period ranges from $1997:01$ to $2017:12$, where the data from $1993:01$ to $1996:12$ is used to initiate the analysis.

Table 5 summarizes the R^2_{OoS} of our predictive regressions. The first three rows present the same regressions models from table 1, the last three rows report the univariate predictive regression using the other factors from the literature: the Cochrane-Piazzesi and Ludvingson-Ng factors, and Fama-Bliss regressions with forward spreads. Finally, the remaining rows are the same regressions models from table 3.

It is clear that for $n = 2$ and $n = 3$, we see evidence of time-varying return predictability. Also, we can see indication that the parsimonious regressions using either $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)^h_t$ or our unspanned factors, provide comparable better R^2_{OoS}, especially for low maturities. Notice that our factors provide higher R^2_{OoS} when compared to univariate predictive regressions using other factors from the literature. For longer maturities, especially $n = 5$, no regression model provided evidence of time-varying return predictability. However, the higher R^2_{OoS} are still those obtained using the DNN factors.

4.5   Relation with PCs

A natural question is how these factors relate with the first principal components from the term-structure. In table 6 we present the correlation between the first five principal components and our state variables, the spanning factor $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)^h_t$ and the unspannned factor $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_t$. We also computed the correlations with the other representation of the unspanned factors $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-n),h}_t$, with the Cochrane-Piazzesi and Ludvingson-Ng factors, and for the sake of completeness with the latent DNN factors $\mathfrak{f}^{(n),h}_{t,DNN}$ derived from the selected neural network.

The first five principal components is correlated out of the monthly set of yields $\boldsymbol{Z}^y_t = \left\{y^{(1/12)}_t, y^{(2/12)}_t, \ldots, y^{(60/12)}_t\right\}$ from 1993 to 2017. We see that the first principal component (level) is negatively correlated with the spanning factor (-0.7549) and almost zero correlation with the unspanned factor(s). The PC1 (level) has a similar correlation with the Cochrane-Piazzesi factor (-0.717), and therefore an irrelevant correlation with the Ludvingson-Ng factor (0.0492) as well. The following two principal components have a small correlation with the spanned factor $(0.139$ and -0.0498, respectively). However, we see a somewhat interesting positive correlation (0.428) between the curvature component and the unspanned factor. The takeaway from table 6 is that the spanning factor captures negatively the slope of the yield curve, and the unspanning factor has some positive linear association with the curvature of the term-structure.

Additionally, we project our spanning and unspanning factors in an increasing set of first principal components (up to PC5). The R-squared of these regressions are reported in table 7. The regressions are computed in such a way that the first column uses PC1 as

Table 6: Correlations with Principal Components of the Term-Structure

		PC1	PC2	PC3	PC4	PC5
		Level	Slope	Curvature		
Spanning Factor	$(\boldsymbol{\tau}^\top\widehat{\mathfrak{S}})_t^h$	-0.7549	0.1390	-0.0498	0.1685	-0.0731
Unspanning Factors	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$	-0.0212	-0.1854	0.4276	-0.0313	0.0907
	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-2),h}$	-0.0532	-0.2055	0.4353	-0.0635	0.0953
	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-3),h}$	-0.1426	-0.2531	0.4837	-0.1288	0.0909
	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-4),h}$	-0.0709	-0.1939	0.4791	-0.0424	0.0744
	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^{(-5),h}$	0.0204	-0.1285	0.4284	0.0397	0.0671
Other Factors	$\widehat{CP}_t^h$	-0.7170	0.2799	0.2859	0.3689	0.0413
	$\widehat{LN}_t^h$	0.0492	-0.1539	0.2745	-0.0328	-0.0240
Derived $\mathfrak{f}_{t,DNN}^{(n),h}$	$\mathfrak{f}_{t,DNN}^{(2),h}$	-0.7378	-0.0243	-0.0424	0.1556	-0.0399
	$\mathfrak{f}_{t,DNN}^{(3),h}$	-0.4654	-0.1406	-0.1674	0.1332	0.0350
	$\mathfrak{f}_{t,DNN}^{(4),h}$	-0.1252	-0.0483	-0.3031	0.1284	0.0389
	$\mathfrak{f}_{t,DNN}^{(5),h}$	0.1284	0.0498	-0.3747	0.0690	0.1224

Table 6 reports the correlation of the first five principal components, PC1 (level), PC2 (slope), PC3 (curvature), PC4, and PC5 computed from the the monthly set of yields $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$. The correlations are calculated against the spanning factor, the unspanned factor(s), the Cochrane-Piazzesi and Ludvingson-Ng factors, and for the sake of completeness with the latent DNN factors. The period of analysis ranges from 1993:01 to 2017:12.

the independent variable, the second column PC1 and PC2, and so on. We see that our single spanning factor is associated with a larger set of principal components. In column (III), where the R-squared of the regressions using level, slope and curvature components are used, we see an R-squared of 0.592 for the spanning factor, 0.218 for the unspanning factor $\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$. For the first three principal component, we see a similar pattern with the Cochrane-Piazzesi factor, with a little higher values; and it is expected given the nature of the derivation of the Ludvingson-Ng factor, the R-squared remain low for all cases.

4.6 A Robustness Check

Recent studies in the forecasting literature raised the issue that defining the sample split may be data-mined (Hansen and Timmermann, 2012; Kelly and Pruitt, 2013; Rossi and Inoue, 2012). As a robustness check, we seek to know if the results reported of the statistical significance of our state factors could be a sample-specific fact. To demonstrate the robustness of our estimates to alternative sample splits, we re-run the same regressions from table 1 restricting the series up to the last month of each year from 1994 up to the last year of analysis, 2017.

Figure 11
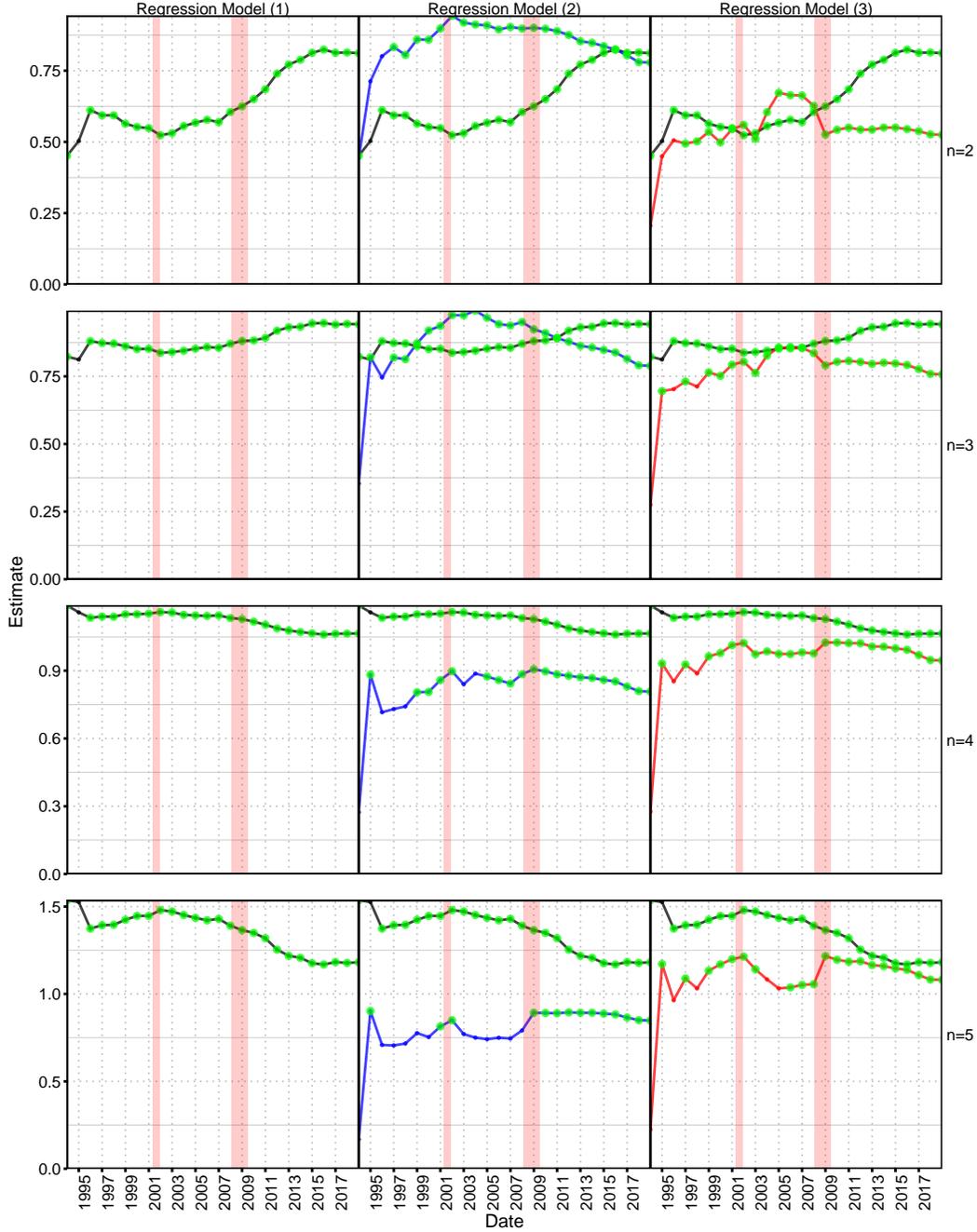
Figure 11 reports the estimates of each one of the three regressions models reported similiar to those summarized in table 1, where the x-axis defines the end of the sample. All samples start in 1993 : 01. Regression Model (1) is given by $rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}_t)_t^h + \epsilon_{t+h/12}$. Regression Model (2) is given by $rx_{t+h/12}^{(n)} = \beta_0 + \beta_1\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h} + \epsilon_{t+h/12}$. Regression Model (3) is given by $rx_{t+h/12}^{(n)} = \beta_0 + \beta_1\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^h + \epsilon_{t+h/12}$. The black line represents the estimates of $\left(\boldsymbol{\tau}^\top\widehat{\mathfrak{S}}_t\right)_t^h$. The blue line represents the estimates of $\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$. Finally, the red represents the estimates of $\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^h$. The figure is split in four panels, each panel representing one maturity. Statistically significant coefficients are presented as green points.

Figure 12: Regression Coefficients of $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ Over Time as a Function of Maturity (n)
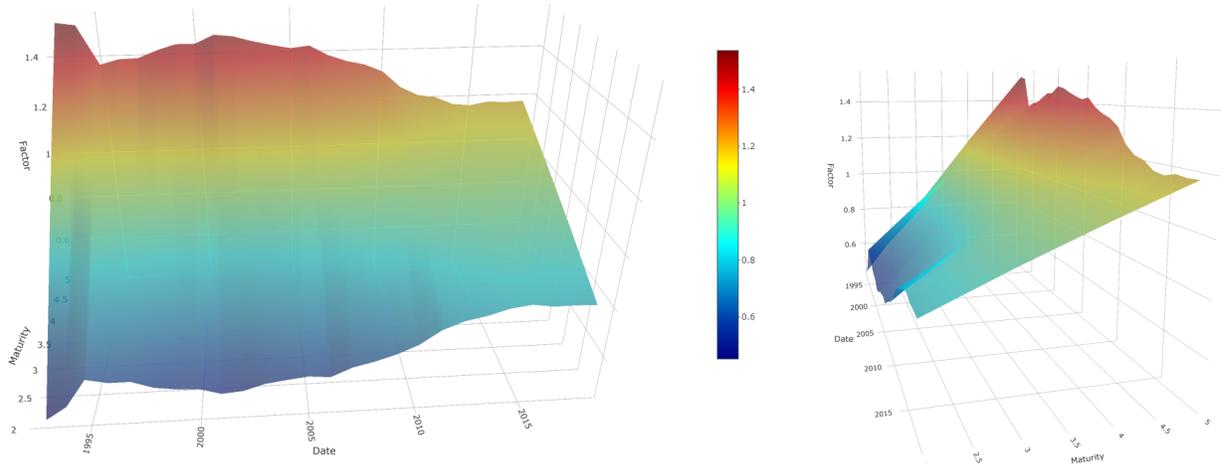
Figure 12 plots the behavior of our spanning factor $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ as a function of maturity (n) over the period of analysis (1993-2017).

Figure 13: Regression Coefficients of $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ Over Time as a Function of Maturity (n)
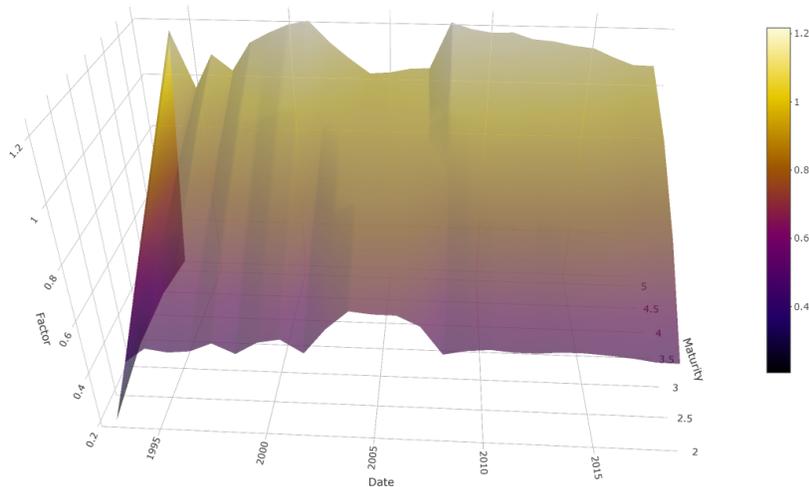
Figure 13 plots the behavior of our unspanned factor $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ as a function of maturity (n) over the period of analysis (1993-2017).

Table 7: R-Squared of Projections - Principal Components of the Term-Structure

		(I)	(II)	(III)	(IV)	(V)
		PC1	(I) + PC2	(II) + PC3	(III) + PC4	(IV) + PC5
Spanning Factor	$(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}})_{t}^{h}$	0.5699	0.5892	0.5917	0.6201	0.6255
Unspanning Factors	$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t}^{h}$	0.0005	0.0348	0.2177	0.2187	0.2269
	$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t}^{(-2),h}$	0.0028	0.0451	0.2346	0.2386	0.2477
	$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t}^{(-3),h}$	0.0203	0.0844	0.3184	0.3350	0.3432
	$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t}^{(-4),h}$	0.0050	0.0426	0.2722	0.2740	0.2795
	$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t}^{(-5),h}$	0.0004	0.0169	0.2005	0.2021	0.2066
Other Factors	$\widehat{CP}_{t}^{h}$	0.5140	0.5924	0.6741	0.8102	0.8119
	$\widehat{LN}_{t}^{h}$	0.0024	0.0261	0.1014	0.1025	0.1031
Derived $\mathfrak{f}_{t,DNN}^{(n),h}$	$\mathfrak{f}_{t,DNN}^{(2),h}$	0.5443	0.5449	0.5467	0.5709	0.5725
	$\mathfrak{f}_{t,DNN}^{(3),h}$	0.2166	0.2363	0.2643	0.2821	0.2833
	$\mathfrak{f}_{t,DNN}^{(4),h}$	0.0157	0.0180	0.1099	0.1264	0.1279
	$\mathfrak{f}_{t,DNN}^{(5),h}$	0.0165	0.0190	0.1593	0.1641	0.1791

Table 7 reports the R-squared of the projection of our spanning and unspanning factor(s) in an increasing set of the first five principal components: PC1 (level), PC2 (slope), PC3 (curvature), PC4, and PC5. The principal components are computed from the the monthly set of yields $\boldsymbol{Z}_{t}^{\mathrm{y}} = \left\{ y_{t}^{(1/12)}, y_{t}^{(2/12)}, \ldots, y_{t}^{(60/12)} \right\}$. We also compute the regressions for the Cochrane-Piazzesi and Ludvingson-Ng factors, and for the sake of completeness with the latent DNN factors. In column **(I)**, we regress the factors on PC1, in column **(II)**, we regress the factors on PC1 and PC2, in column **(III)**, we regress the factor on PC1, PC2, and PC3, in column **(IV)**, we regress the factors on PC1, PC2, PC3, and PC4, and finally in column **(V)**, we regress the factors on PC1, PC2, PC3, PC4, and PC5. The period of analysis ranges from 1993:01 to 2017:12.

Figure 11 reports the coefficients estimates of $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_{t}\right)_{t}^{h}$, $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ and $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$. In a recursive approach we seek to show how the estimates of the parameters varies across an expanding sample size and the statistical significance as well. The figure has four panels, each panel representing one maturity. Statistically significant coefficients are presented as green points. Clearly we see that, despite the initial variation in the estimates for the first years, what is expected given the limited sample size, the (i) estimates do not behave erratically with abrupt variations, and (ii) the vast majority of the estimates for each year from 1994 to 2017 is statistically significant.

In figure 12 we plot the estimates obtained in these regressions ranging from 1994 to 2017 across all maturities. The figure shows a clear pattern for the estimates of $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_{t}\right)_{t}^{h}$ is increasing in the maturity (n). Notice that the estimates of this increasing line shifted during both recessions in the period of analysis. Another pattern that can be inferred from this figure is that over time the difference between longer maturities and shorter shrunk over the period 1993 to 2017, what can be seen as the curve of $\left(\boldsymbol{\tau}^{\top}\widehat{\widetilde{\mathfrak{F}}}_{t}\right)_{t}^{h}$ becoming more flat over time.

Similarly, figure 12 plots the regression coefficients of $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-n),h}_{t+h/12}$ as a function of Maturity (n), and figure 15 in Appendix A.1 plots the regression coefficients of $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{h}_{t+h/12}$. The pattern mentioned above maintains for the unspanned factor $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-n),h}_{t+h/12}$. However, for $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{h}_{t+h/12}$ the curve as a function of maturities are much flatter when compared to the other state variables, and analogously to the Cochrane-Piazzesi factor, it has a more clear tent shape. This format becomes more evident during the recessions, when mid levels of maturity have the highest value for this factor, while low and high level of maturities are smaller.

4.7 GMM/GEL Estimation

As an additional robustness check, in a similar fashion to Lee (2018), we can use of GMM to estimate jointly the parameters of our state variables and obtain better standard errors estimates for the inference of our parameters. This is especially important as we have some generated regressors in our analysis, and that the dependent variables $(rx^{(n)}_{t+h})$ have clear cross-sectional correlations among them. The states variables are obtained as:

$$\overline{rx}_t = \kappa_0 + \kappa_1 \xi^{(2)}_t + \kappa_2 \xi^{(3)}_t + \kappa_3 \xi^{(4)}_t + \kappa_4 \xi^{(5)}_t + \overline{u}_t \tag{30}$$

$$(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_t = \delta_0 + \delta_1 \left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t \right)^h_t + e_t \tag{31}$$

where $\overline{rx}_t = \sum_{n=2}^{5} rx^{(n)}_t / 4$ and $(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_t = \widehat{\kappa}_{1,t} \xi^{(2)}_t + \widehat{\kappa}_{2,t} \xi^{(3)}_t + \widehat{\kappa}_{3,t} \xi^{(4)}_t + \widehat{\kappa}_{4,t} \xi^{(5)}_t$, which is obtained in equation (30). Then, in the second stage for the risk premium forecasts we run:

$$rx^{(n)}_{t+h/12} = \beta^{(n)}_0 + \beta^{(n)}_1 \left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t \right)^h_t + \beta^{(n)}_2 \boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_t + \epsilon^{(n)}_{t+h/12} \tag{32}$$

for each of the four maturities $n \in \{2, 3, 4, 5\}$. We can define the vector of moments of our GMM as below:

$$g_T\left(\boldsymbol{\theta}\right) = \begin{bmatrix} \overline{u}_t \otimes \left(1 \quad \xi_t^{(2)} \quad \xi_t^{(3)} \quad \xi_t^{(4)} \quad \xi_t^{(5)}\right) \\ e_t \otimes \left(1 \quad \left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h\right) \\ \epsilon_{t+1}^{(2)} \otimes \left(1 \quad \left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h \quad \boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^h\right) \\ \epsilon_{t+1}^{(3)} \otimes \left(1 \quad \left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h \quad \boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^h\right) \\ \epsilon_{t+1}^{(4)} \otimes \left(1 \quad \left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h \quad \boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^h\right) \\ \epsilon_{t+1}^{(5)} \otimes \left(1 \quad \left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h \quad \boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^h\right) \end{bmatrix} \tag{33}$$

where $\boldsymbol{\theta}$ is the 19×1 vector of all the parameters as in

$$\boldsymbol{\theta} = \begin{bmatrix} \kappa_0 & \kappa_1 & \kappa_2 & \kappa_3 & \kappa_4 & \delta_0 & \delta_1 & \beta_0^{(2)} & \beta_1^{(2)} & \beta_2^{(2)} \dots & \beta_0^{(5)} & \beta_1^{(5)} & \beta_2^{(5)} \end{bmatrix}^\top.$$

Table 8: GMM and GEL estimations of the spanning factor $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ and the unspanned factor $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(h}$

	Variable	Estimate	se(GMM)	se(GEL)	se(OLS)	GMM t-stat	GEL t-stat	OLS t-stat
$rx_{t+h/12}^{(2)}$	$(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h$	0.811	0.130	0.141	0.110	6.259***	5.741***	7.355***
	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$	0.525	0.080	0.087	0.124	6.579***	6.002***	4.228***
$rx_{t+h/12}^{(3)}$	$(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h$	0.943	0.196	0.219	0.173	4.802***	4.307***	5.447***
	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$	0.757	0.075	0.080	0.202	10.089***	9.445***	3.743***
$rx_{t+h/12}^{(4)}$	$(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h$	1.065	0.260	0.294	0.236	4.097***	3.624***	4.519***
	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$	0.945	0.119	0.124	0.282	7.972***	7.62***	3.355***
$rx_{t+h/12}^{(5)}$	$(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h$	1.181	0.320	0.364	0.296	3.695***	3.248**	3.99***
	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_t^h$	1.081	0.194	0.216	0.359	5.558***	5.006***	3.011**

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8 reports the the standard errors and t-statistics from the Generalized Empirical Likelihood (GEL), Generalized Empirical Likelihood (GEL), and Ordinary Least Squares (OLS). The vector of moments for the GMM and GEL is given in equation (33). The standard errors for GMM and OLS use HAC variance-covariance matrix, as in Newey and West (1987).

Table 8 presents the standard error estimates using OLS and GMM for the risk premium forecasts parameters as in equation 32, for the four maturities considered. The takeaway from this table is that even though that the standard errors from the GMM are a little bit larger when compared to the ones generated by OLS, the inference over the interested parameters keeps still being strongly significant. We control the auto-correlation with a HAC

variance-covariance matrix for both methods, as in Newey and West (1987).

To corroborate the robustness of our generated regressors, we also make of use of the Generalized Empirical Likelihood (GEL). As Newey and Smith (2004) and Anatolyev (2005) had shown, GEL has a significant advantage over the GMM estimation, since the bias of the latter does not increase with the number of moment conditions, what does not necessarily holds for GMM. Thus, the efficiency improves when the number of conditions goes up. We can use the same vector of moments as in equation (33) to estimate the risk premium forecasts parameters. Table 8 also reports the standard errors and the t-statistics for the same regressions. We see that the standard errors slightly larger than GMM and OLS ones, but still strongly statistically significant. In short, these results endorse the robustness of our factors in the forecasting the risk premium of the Treasury bonds.

5 Conclusion

In this paper we proposed a novel approach for deriving a single state factor consistent with a dynamic term-structure with unspanned risks. We make use of deep neural networks to uncover relationships in the full set of information from the yield curve. This allows us through an approximation to to derive a single state variable factor that spans the space of all the information from the term-structure. We also introduced a way to obtain unspanned risks from the yield curve that is used to complete our state space.

We show that this parsimonious number of state variables have predictive power for excess returns of bonds over 1-month holding period. Additionally, we provide an intuitive interpretation of derived factors, and show what information from macroeconomic variables and sentiment-based measures they can capture.

References

Anatolyev, S. (2005). GMM, GEL, serial correlation, and asymptotic bias. *Econometrica*, 73(3):983–1002.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*, pages 507–547. University of Chicago Press.

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11:685–725.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636.

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

Bauer, M. D. and Rudebusch, G. D. (2017). Resolving the spanning puzzle in macro-finance term structure models. *Review of Finance*, 21(2):511–553.

Bianchi, D., Büchner, M., and Tamoni, A. (2019). Bond risk premia with machine learning. *USC-INET Research Paper*, (19-11).

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. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Chen, L., Pelger, M., and Zhu, J. (2019). Deep learning in asset pricing. *Available at SSRN 3350138*.

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

Cochrane, J. H. (2015). Comments on "robust bond risk premia" by michael bauer and jim hamilton. *Unpublished working paper. University of Chicago*.

Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1):138–160.

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

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.

Dubiel-Teleszynski, T., Kalogeropoulos, K., and Karouzakis, N. (2019). Predicting bond risk premia via sequential learning.

Duffee, G. (2013). Forecasting interest rates. In *Handbook of economic forecasting*, volume 2, pages 385–426. Elsevier.

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

Farrell, M. H., Liang, T., and Misra, S. (2021). Deep neural networks for estimation and inference. *Econometrica*, 89(1):181–213.

Feng, G., He, J., and Polson, N. G. (2018a). Deep learning for predicting asset returns. *arXiv preprint arXiv:1804.09314*.

Feng, G., Polson, N. G., and Xu, J. (2018b). Deep learning in characteristics-sorted factor models. *arXiv preprint arXiv:1805.01104*.

Freyberger, J., Neuhierl, A., and Weber, M. (2017). Dissecting characteristics nonparametrically. Technical report, National Bureau of Economic Research.

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.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Gu, S., Kelly, B., and Xiu, D. (2018). Empirical asset pricing via machine learning. Technical report, National Bureau of Economic Research.

Gu, S., Kelly, B., and Xiu, D. (2020). Autoencoder asset pricing models. *Journal of Econometrics*.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, 54(8):2291–2304.

Hansen, P. R. and Timmermann, A. (2012). Choice of sample split in out-of-sample forecast evaluation.

Heaton, J., Polson, N. G., and Witte, J. (2016). Deep portfolio theory. *arXiv preprint arXiv:1605.07230*.

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.

Hornik, K., Stinchcombe, M., White, H., et al. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.

Huang, J.-Z. and Shi, Z. (2019). Determinants of bond risk premia: A machine-learning-based resolution of the spanning controversy. Working Paper.

Joslin, S., Priebsch, M., and Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, 69(3):1197–1233.

Kelly, B. and Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, 68(5):1721–1756.

Kozak, S., Nagel, S., and Santosh, S. (2019). Shrinking the cross-section. *Journal of Financial Economics*.

Lee, J. (2018). Risk premium information from treasury-bill yields. *Journal of Financial and Quantitative Analysis*, 53(1):437–454.

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.

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Püttmann, L. (2018). Patterns of panic: Financial crisis language in historical newspapers.

Rossi, B. and Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30(3):432–453.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28.

A   Appendix

A.1   Data

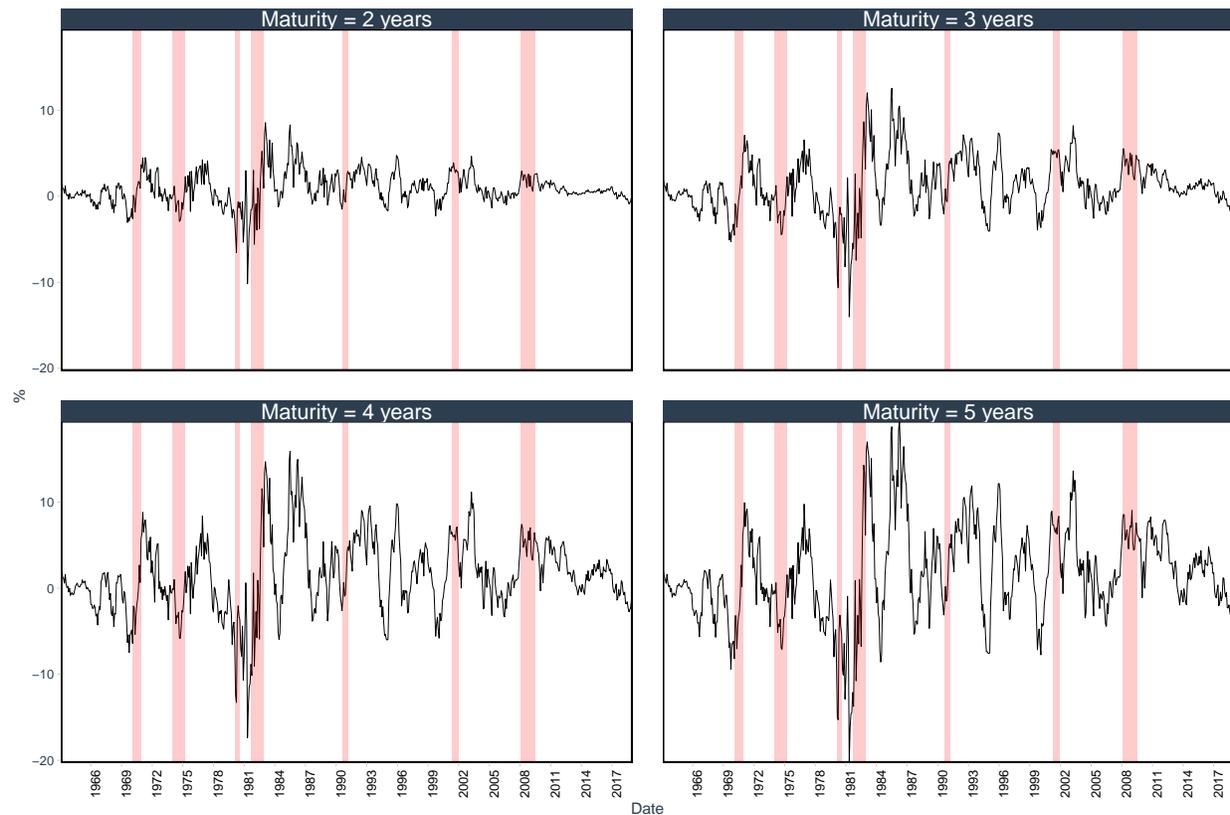Figure 14: 12-Month Bonds Excess Returns

Figure 14 shows the 12-month excess returns for maturities with $n = 2, 3, 4$ and 5 years. The excess returns are calculeted as in equation (4), i.e., $rx_{t+1}^{(n)} = ny_t^{(n)} - (n+1)y_{t+1}^{(n-1)} - y_t^n$. Each panel represents one of the four maturities. The y-axis shows values in percentage (%). NBER-classified recessions are shaded in light red.

Table 9: Descriptive Statistics - DNN Factor $\mathfrak{f}_{t,DNN}^{(n),h}$ by MLP Architecture and Choice of $\boldsymbol{Z}_t^{\mathrm{y}}$

Panel A: $\boldsymbol{Z}_t$: set of forward rates

| | $\mathfrak{f}_{t,DNN}^{(2),h}$ | | | $\mathfrak{f}_{t,DNN}^{(3),h}$ | | | $\mathfrak{f}_{t,DNN}^{(4),h}$ | | | $\mathfrak{f}_{t,DNN}^{(5),h}$ | | |
Model	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$
DNN 1	0.3300	0.0368	0.8099	0.3361	0.0458	0.8372	0.3377	0.0639	0.7711	0.2874	0.1114	0.7993
DNN 2	0.3360	0.0364	0.8964	0.3369	0.0342	0.8780	0.3463	0.0563	0.8708	0.3214	0.0772	0.8462
DNN 3	0.4274	0.0225	0.8092	0.4270	0.0237	0.7397	0.4199	0.0341	0.8338	0.4054	0.0413	0.8563

Panel B: $\boldsymbol{Z}_t$: set of yields

| | $\mathfrak{f}_{t,DNN}^{(2),h}$ | | | $\mathfrak{f}_{t,DNN}^{(3),h}$ | | | $\mathfrak{f}_{t,DNN}^{(4),h}$ | | | $\mathfrak{f}_{t,DNN}^{(5),h}$ | | |
Model	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$
DNN 1	0.3441	0.0373	0.8106	0.3359	0.0393	0.8324	0.3308	0.0534	0.7812	0.2781	0.1122	0.9090
DNN 2	0.3370	0.0360	0.8821	0.3379	0.0327	0.8508	0.3351	0.0432	0.8066	0.3288	0.0628	0.8350
DNN 3	0.4363	0.0204	0.7833	0.4335	0.0226	0.7264	0.4298	0.0305	0.8150	0.4226	0.0403	0.8411

Panel C: $\boldsymbol{Z}_t$: set of yields and forward rates

| | $\mathfrak{f}_{t,DNN}^{(2),h}$ | | | $\mathfrak{f}_{t,DNN}^{(3),h}$ | | | $\mathfrak{f}_{t,DNN}^{(4),h}$ | | | $\mathfrak{f}_{t,DNN}^{(5),h}$ | | |
Model	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$
DNN 1	0.3421	0.0512	0.6709	0.3313	0.0690	0.6546	0.3088	0.0969	0.7587	0.2642	0.1110	0.6448
DNN 2	0.3418	0.0356	0.8732	0.3421	0.0340	0.7917	0.3434	0.0449	0.8122	0.3429	0.0647	0.8116
DNN 3	0.4375	0.0202	0.7627	0.4321	0.0232	0.7543	0.4246	0.0310	0.8092	0.4190	0.0381	0.8319

Table 9 reports the mean, standard deviation and the first autocorrelation ($\hat{\rho}_1$) of the derived $\mathfrak{f}_{t,DNN}^{(n),h}$ for each scenario under consideration. Each panel considers a different set of information from the term-structure to derive the factor $\mathfrak{f}_{t,DNN}^{(n),h}$. Panel A shows the the derived DNN factors for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60)} \right\}$, Panel B for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$ and Panel C for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)}, y_t^{(1)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$. The descriptive statistics is computed for for each group of maturity $n \in \{1, 2, 3, 4\}$. Period of analysis ranges from 1993:01 to 2017:12.

Table 10: Descriptive Statistics - $\xi_t^{(n),h}$ by MLP Architecture and Choice of $\boldsymbol{Z}_t^{\mathrm{y}}$

Panel A: $\boldsymbol{Z}_t$: set of forward rates

	$\xi_t^{(2),h}$			$\xi_t^{(3),h}$			$\xi_t^{(4),h}$			$\xi_t^{(5),h}$		
Model	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$
DNN 1	-0.0311	0.4799	0.2417	0.0187	0.7713	0.1718	0.0610	1.0581	0.1324	0.1558	1.3409	0.1178
DNN 2	-0.0265	0.4797	0.2427	0.0145	0.7716	0.1645	0.0550	1.0651	0.1323	0.1240	1.3424	0.1099
DNN 3	-0.0394	0.4796	0.2428	-0.0015	0.7729	0.1636	0.0402	1.0630	0.1261	0.0898	1.3419	0.1036

Panel B: $\boldsymbol{Z}_t$: set of yields

	$\xi_t^{(2),h}$			$\xi_t^{(3),h}$			$\xi_t^{(4),h}$			$\xi_t^{(5),h}$		
Model	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$
DNN 1	-0.0344	0.4810	0.2462	0.0144	0.7705	0.1625	0.0701	1.0591	0.1263	0.1804	1.3347	0.1059
DNN 2	-0.0254	0.4793	0.2422	0.0159	0.7721	0.1657	0.0646	1.0600	0.1279	0.1135	1.3410	0.1077
DNN 3	-0.0685	0.4952	0.2903	-0.0032	0.7724	0.1640	0.0366	1.0636	0.1276	0.0874	1.3429	0.1062

Panel C: $\boldsymbol{Z}_t$: set of yields and forward rates

	$\xi_t^{(2),h}$			$\xi_t^{(3),h}$			$\xi_t^{(4),h}$			$\xi_t^{(5),h}$		
Model	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$	Mean	sd	$\hat{\rho}_1$
DNN 1	-0.0240	0.4760	0.2492	0.0336	0.7683	0.1747	0.0999	1.0560	0.1423	0.1932	1.3370	0.1004
DNN 2	-0.0269	0.4795	0.2389	0.0146	0.7711	0.1650	0.0590	1.0612	0.1296	0.1051	1.3428	0.1084
DNN 3	-0.0622	0.4956	0.2867	0.0012	0.7728	0.1603	0.0450	1.0620	0.1221	0.0872	1.3419	0.1030

Table 9 reports the mean, standard deviation and the first autocorrelation ($\hat{\rho}_1$) of the $\xi_t^{(n),h}$ for each scenario under consideration. Each panel considers a different set of information from the term-structure to derive the factor $\mathfrak{f}_{t,DNN}^{(n),h}$. Panel A shows the the derived DNN factors for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60)} \right\}$, Panel B for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$ and Panel C for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)}, y_t^{(1)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$. The descriptive statistics is computed for for each group of maturity $n \in \{1, 2, 3, 4\}$. Period of analysis ranges from 1993:01 to 2017:12.

Figure 15: Regression Coefficients of $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{z}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_{t+h/12}$ Over Time as a Function of Maturity (n)
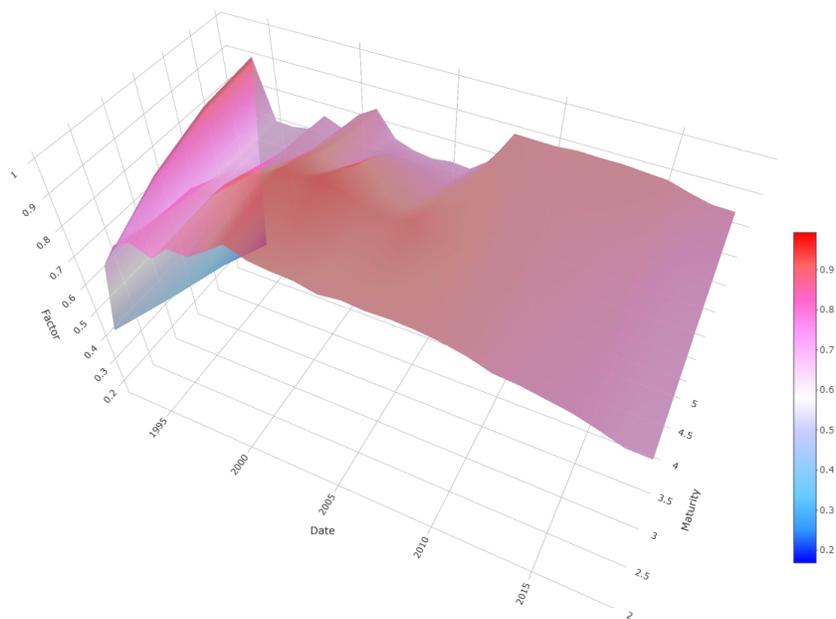
Figure 15 plots the behavior of our unspanned factor $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{z}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_{t+h/12}$ as a function of maturity (n) over the period of analysis (1993-2017).

Table 11: FRED-MD

Group	FRED Code	Description	Group	FRED Code	Description
Output and income	RPI	Real Personal Income	Money and credit	M1SL	M1 Money Stock
	W875RX1	Real personal income ex transfer receipts		M2SL	M2 Money Stock
	INDPRO	IP Index		M2REAL	Real M2 Money Stock
	IPFPNSS	IP: Final Products and Nonindustrial Supplies		AMBSL	St. Louis Adjusted Monetary Base
	IPFINAL	IP: Final Products (Market Group)		TOTRESNS	Total Reserves of Depository Institutions
	IPCONGD	IP: Consumer Goods		NONBORRES	Reserves Of Depository Institutions
	IPDCONGD	IP: Durable Consumer Goods		BUSLOANS	Commercial and Industrial Loans
	IPNCONGD	IP: Nondurable Consumer Goods		REALLN	Real Estate Loans at All Commercial Banks
	IPBUSEQ	IP: Business Equipment		NONREVSL	Total Nonrevolving Credit
	IPMAT	IP: Materials		CONSPI	Nonrevolving Consumer Credit to Personal Income
	IPDMAT	IP: Durable Materials		MZMSL	MZM Money Stock
	IPNMAT	IP: Nondurable Materials		DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding
	IPMANSICS	IP: Manufacturing (SIC)		DTCTHFNM	Total Consumer Loans and Leases Outstanding
	IPB51222s	IP: Residential Utilities		INVEST	Securities in Bank Credit at All Commercial Banks
	IPFUELS	IP: Fuels	Interest and exchange rates	FEDFUNDS	Effective Federal Funds Rate
	CUMFNS	Capacity Utilization: Manufacturing		CP3Mx	3-Month AA Financial Commercial Paper Rate
Labor market	HWI	Help-Wanted Index for United States		TB3MS	3-Month Treasury Bill
	HWIURATIO	Ratio of Help Wanted/No. Unemployed		TB6MS	6-Month Treasury Bill
	CLF16OV	Civilian Labor Force		GS1	1-Year Treasury Rate
	CE16OV	Civilian Employment		GS5	5-Year Treasury Rate
	UNRATE	Civilian Unemployment Rate		GS10	10-Year Treasury Rate
	UEMPMEAN	Average Duration of Unemployment (Weeks)		AAA	Moody's Seasoned Aaa Corporate Bond
	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks		BAA	Moody's Seasoned Baa Corporate Bond
	UEMP5TO14	Civilians Unemployed for 43599 Weeks		COMPAPFFx	3-Month Commercial Paper Minus
	UEMP15OV	Civilians Unemployed - 15 Weeks & Over		TB3SMFFM	3-Month Treasury C Minus FEDFUNDS
	UEMP15T26	Civilians Unemployed for 15-26 Weeks		TB6SMFFM	6-Month Treasury C Minus FEDFUNDS
	UEMP27OV	Civilians Unemployed for 27 Weeks and Over		T1YFFM	1-Year Treasury C Minus FEDFUNDS
	CLAIMSx	Initial Claims		T5YFFM	5-Year Treasury C Minus FEDFUNDS
	PAYEMS	All Employees: Total Nonfarm		T10YFFM	10-Year Treasury C Minus FEDFUNDS
	USGOOD	All Employees: Goods-Producing Industries		AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS
	CES1021000001	All Employees: Mining and Logging: Mining		BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS
	USCONS	All Employees: Construction		TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies
	MANEMP	All Employees: Manufacturing		EXSZUSx	Switzerland / U.S. Foreign Exchange Rate
	DMANEMP	All Employees: Durable goods		EXJPUSx	Japan / U.S. Foreign Exchange Rate
	NDMANEMP	All Employees: Nondurable goods		EXUSUKx	U.S. / U.K. Foreign Exchange Rate
	SRVPRD	All Employees: Service-Providing Industries		EXCAUSx	Canada / U.S. Foreign Exchange Rate
	USTPU	All Employees: Trade, Transportation & Utilities	Prices	WPSFD49207	PPI: Finished Goods
	USWTRADE	All Employees: Wholesale Trade		WPSFD49502	PPI: Finished Consumer Goods
	USTRADE	All Employees: Retail Trade		WPSID61	PPI: Intermediate Materials
	USFIRE	All Employees: Financial Activities		WPSID62	PPI: Crude Materials
	USGOVT	All Employees: Government		OILPRICEx	Crude Oil, Spliced WTI and Cushing
	CES0600000007	Avg Weekly Hours : Goods-Producing		PPICMM	PPI: Metals and Metal Products:
	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing		CPIAUCSL	CPI : All Items
	AWHMAN	Avg Weekly Hours : Manufacturing		CPIAPPSL	CPI : Apparel
	CES0600000008	Avg Hourly Earnings : Goods-Producing		CPITRNSL	CPI : Transportation
	CES2000000008	Avg Hourly Earnings : Construction		CPIMEDSL	CPI : Medical Care
	CES3000000008	Avg Hourly Earnings : Manufacturing		CUSR0000SAC	CPI : Commodities
Housing	HOUST	Housing Starts: Total New Privately Owned		CUSR0000SAD	CPI : Durables
	HOUSTNE	Housing Starts, Northeast		CUSR0000SAS	CPI : Services
	HOUSTMW	Housing Starts, Midwest		CPIULFSL	CPI : All Items Less Food
	HOUSTS	Housing Starts, South		CUSR0000SA0L2	CPI : All Items Less Shelter
	HOUSTW	Housing Starts, West		CUSR0000SA0L5	CPI : All Items Less Medical Care
	PERMIT	New Private Housing Permits (SAAR)		PCEPI	Personal Cons. Expend.: Chain Index
	PERMITNE	New Private Housing Permits, Northeast (SAAR)		DDURRG3M086SBEA	Personal Cons. Exp: Durable goods
	PERMITMW	New Private Housing Permits, Midwest (SAAR)		DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods
	PERMITS	New Private Housing Permits, South (SAAR)		DSERRG3M086SBEA	Personal Cons. Exp: Services
	PERMITW	New Private Housing Permits, West (SAAR)	Stock market	S&P 500	S&P's Common Stock Price Index: Composite
Consumption, orders, and inventories	DPCERA3M086SBEA	Real Personal Consumption Expenditures		S&P div yield	S&P's Composite Common Stock: Dividend Yield
	CMRMTSPLx	Real Manu. and Trade Industries Sales		S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio
	RETAILx	Retail and Food Services Sales		VXOCLSx	VXO
	ACOGNO	New Orders for Consumer Goods			
	AMDMNOx	New Orders for Durable Goods			
	ANDENOx	New Orders for Nondefense Capital Goods			
	AMDMUOx	Unfilled Orders for Durable Goods			
	BUSINVx	Total Business Inventories			
	ISRATIOx	Total Business: Inventories to Sales Ratio			
	UMCSENTx	Consumer Sentiment Index			

Table 11 lists the 128 macroeconomic and financial variables from the FRED-MD dataset. The table reports the group, FRED code and a description of each variable. The variables are split in one of the 8 groups: (1) output and income, (2) labor market, (3) housing, (4) consumption, orders, and inventories, (5) money and credit, (6) interest and exchange rates, (7) prices, and (8) stock market.

Table 12: Descriptive Statistics - FRED-MD

Fred Code	tcode	Group	Description	Full Sample Mean	sd	1962-1992 Mean	sd	1993-2017 Mean	sd
RPI	5	Output and income	Real Personal Income	0.0030	0.0060	0.0030	0.0040	0.0020	0.0070
W875RX1	5	Output and income	Real personal income ex transfer receipts	0.0020	0.0060	0.0030	0.0040	0.0020	0.0070
INDPRO	5	Output and income	IP Index	0.0020	0.0070	0.0030	0.0080	0.0020	0.0060
IPFPNSS	5	Output and income	IP: Final Products and Nonindustrial Supplies	0.0020	0.0070	0.0030	0.0080	0.0010	0.0060
IPFINAL	5	Output and income	IP: Final Products (Market Group)	0.0020	0.0080	0.0030	0.0080	0.0010	0.0070
IPCONGD	5	Output and income	IP: Consumer Goods	0.0020	0.0090	0.0020	0.0090	0.0010	0.0070
IPDCONGD	5	Output and income	IP: Durable Consumer Goods	0.0020	0.0210	0.0030	0.0220	0.0020	0.0190
IPNCONGD	5	Output and income	IP: Nondurable Consumer Goods	0.0010	0.0070	0.0020	0.0080	0.0000	0.0070
IPBUSEQ	5	Output and income	IP: Business Equipment	0.0040	0.0120	0.0040	0.0120	0.0030	0.0120
IPMAT	5	Output and income	IP: Materials	0.0020	0.0090	0.0020	0.0100	0.0020	0.0080
IPDMAT	5	Output and income	IP: Durable Materials	0.0030	0.0130	0.0030	0.0150	0.0030	0.0110
IPNMAT	5	Output and income	IP: Nondurable Materials	0.0020	0.0110	0.0030	0.0100	0.0000	0.0110
IPMANSICS	5	Output and income	IP: Manufacturing (SIC)	0.0020	0.0080	0.0030	0.0090	0.0020	0.0070
IPB51222s	5	Output and income	IP: Residential Utilities	0.0020	0.0350	0.0030	0.0280	0.0010	0.0420
IPFUELS	5	Output and income	IP: Fuels	0.0010	0.0190	0.0010	0.0200	0.0010	0.0180
CUMFNS	2	Output and income	Capacity Utilization: Manufacturing	-0.0060	0.6230	-0.0050	0.7110	-0.0080	0.5000
HWI	2	Labor market	Help-Wanted Index for United States	8.2700	169.6090	3.1960	106.1450	14.3210	222.8650
HWIURATIO	2	Labor market	Ratio of Help Wanted/No. Unemployed	0.0010	0.0330	0.0000	0.0320	0.0030	0.0340
CLF16OV	5	Labor market	Civilian Labor Force	0.0010	0.0030	0.0020	0.0030	0.0010	0.0020
CE16OV	5	Labor market	Civilian Employment	0.0010	0.0030	0.0020	0.0030	0.0010	0.0020
UNRATE	2	Labor market	Civilian Unemployment Rate	-0.0030	0.1730	0.0040	0.1880	-0.0110	0.1550
UEMPMEAN	2	Labor market	Average Duration of Unemployment (Weeks)	0.0090	0.6000	0.0090	0.4740	0.0090	0.7230
UEMPLT5	5	Labor market	Civilians Unemployed - Less Than 5 Weeks	0.0000	0.0550	0.0020	0.0480	-0.0010	0.0610
UEMP5TO14	5	Labor market	Civilians Unemployed for 43599 Weeks	0.0010	0.0540	0.0020	0.0560	-0.0010	0.0520
UEMP15OV	5	Labor market	Civilians Unemployed - 15 Weeks & Over	0.0010	0.0510	0.0030	0.0550	-0.0020	0.0440
UEMP15T26	5	Labor market	Civilians Unemployed for 15-26 Weeks	0.0010	0.0750	0.0030	0.0780	-0.0020	0.0710
UEMP27OV	5	Labor market	Civilians Unemployed for 27 Weeks and Over	0.0010	0.0700	0.0030	0.0780	-0.0010	0.0580
CLAIMSx	5	Labor market	Initial Claims	0.0000	0.0480	0.0000	0.0530	-0.0010	0.0400
PAYEMS	5	Labor market	All Employees: Total nonfarm	0.0010	0.0020	0.0020	0.0020	0.0010	0.0020
USGOOD	5	Labor market	All Employees: Goods-Producing Industries	0.0000	0.0040	0.0000	0.0050	0.0000	0.0040
CES1021000001	5	Labor market	All Employees: Mining and Logging: Mining	0.0000	0.0180	0.0000	0.0230	0.0010	0.0090
USCONS	5	Labor market	All Employees: Construction	0.0010	0.0090	0.0010	0.0100	0.0020	0.0060
MANEMP	5	Labor market	All Employees: Manufacturing	0.0000	0.0040	0.0000	0.0050	-0.0010	0.0030
DMANEMP	5	Labor market	All Employees: Durable goods	0.0000	0.0060	0.0000	0.0070	-0.0010	0.0040
NDMANEMP	5	Labor market	All Employees: Nondurable goods	0.0000	0.0030	0.0000	0.0030	-0.0010	0.0020
SRVPRD	5	Labor market	All Employees: Service-Providing Industries	0.0020	0.0020	0.0020	0.0020	0.0010	0.0010
USTPU	5	Labor market	All Employees: Trade, Transportation & Utilities	0.0010	0.0020	0.0020	0.0020	0.0010	0.0020
USWTRADE	5	Labor market	All Employees: Wholesale Trade	0.0010	0.0020	0.0020	0.0020	0.0010	0.0020
USTRADE	5	Labor market	All Employees: Retail Trade	0.0020	0.0030	0.0020	0.0030	0.0010	0.0020
USFIRE	5	Labor market	All Employees: Financial Activities	0.0020	0.0020	0.0020	0.0020	0.0010	0.0020
USGOVT	5	Labor market	All Employees: Government	0.0010	0.0020	0.0020	0.0020	0.0010	0.0020
CES0600000007	1	Labor market	Avg Weekly Hours : Goods-Producing	40.2940	0.6340	39.9960	0.4900	40.6490	0.6020
AWOTMAN	2	Labor market	Avg Weekly Overtime Hours : Manufacturing	0.0020	0.1350	0.0030	0.1510	0.0010	0.1130
AWHMAN	1	Labor market	Avg Weekly Hours : Manufacturing	40.7920	0.7240	40.4110	0.5750	41.2460	0.6130
CES0600000008	6	Labor market	Avg Hourly Earnings : Goods-Producing	0.0000	0.0040	0.0000	0.0050	0.0000	0.0030
CES2000000008	6	Labor market	Avg Hourly Earnings : Construction	0.0000	0.0080	0.0000	0.0100	0.0000	0.0050
CES3000000008	6	Labor market	Avg Hourly Earnings : Manufacturing	0.0000	0.0050	0.0000	0.0050	0.0000	0.0030
HOUST	4	Housing	Housing Starts: Total New Privately Owned	7.2230	0.3190	7.3070	0.2360	7.1240	0.3730
HOUSTNE	4	Housing	Housing Starts, Northeast	5.0590	0.4130	5.2750	0.3400	4.8020	0.3390
HOUSTMW	4	Housing	Housing Starts, Midwest	5.5580	0.4240	5.6950	0.3130	5.3960	0.4800
HOUSTS	4	Housing	Housing Starts, South	6.4150	0.3030	6.4390	0.2600	6.3860	0.3450
HOUSTW	4	Housing	Housing Starts, West	5.7800	0.3870	5.8540	0.3210	5.6910	0.4370
PERMIT	4	Housing	New Private Housing Permits (SAAR)	7.1750	0.3120	7.2020	0.2570	7.1440	0.3650
PERMITNE	4	Housing	New Private Housing Permits, Northeast (SAAR)	5.0780	0.3900	5.2540	0.3460	4.8690	0.3340
PERMITMW	4	Housing	New Private Housing Permits, Midwest (SAAR)	5.5070	0.3880	5.5810	0.3200	5.4190	0.4410
PERMITS	4	Housing	New Private Housing Permits, South (SAAR)	6.3070	0.3370	6.2350	0.3210	6.3930	0.3360
PERMITW	4	Housing	New Private Housing Permits, West (SAAR)	5.7960	0.3860	5.8590	0.3280	5.7200	0.4340
DPCERA3M086SBEA	5	Consumption, orders, and inventories	Real personal consumption expenditures	0.0030	0.0050	0.0030	0.0060	0.0020	0.0030
CMRMTSPLx	5	Consumption, orders, and inventories	Real Manu. and Trade Industries Sales	0.0020	0.0100	0.0030	0.0120	0.0020	0.0080
RETAILx	5	Consumption, orders, and inventories	Retail and Food Services Sales	0.0050	0.0120	0.0060	0.0130	0.0030	0.0100
ACOGNO	5	Consumption, orders, and inventories	New Orders for Consumer Goods	0.0010	0.0130	0.0000	0.0040	0.0030	0.0180
AMDMNOx	5	Consumption, orders, and inventories	New Orders for Durable Goods	0.0040	0.0380	0.0050	0.0340	0.0020	0.0410
ANDENOx	5	Consumption, orders, and inventories	New Orders for Nondefense Capital Goods	0.0030	0.0780	0.0050	0.0750	0.0020	0.0810
AMDMUOx	5	Consumption, orders, and inventories	Unfilled Orders for Durable Goods	0.0050	0.0100	0.0060	0.0110	0.0030	0.0100
BUSINVx	5	Consumption, orders, and inventories	Total Business Inventories	0.0040	0.0060	0.0060	0.0060	0.0030	0.0050
ISRATIOx	2	Consumption, orders, and inventories	Total Business: Inventories to Sales Ratio	0.0000	0.0170	0.0000	0.0200	0.0000	0.0140
UMCSENTx	2	Consumption, orders, and inventories	Consumer Sentiment Index	0.0210	3.2860	0.0200	2.7990	0.0230	3.7910

(Continued)

55

Fred Code	tcode	Group	Description	Full Sample Mean	sd	1962-1992 Mean	sd	1993-2017 Mean	sd
M1SL	6	Money and credit	M1 Money Stock	0.000	0.009	0.000	0.005	0.000	0.012
M2SL	6	Money and credit	M2 Money Stock	0.000	0.003	0.000	0.002	0.000	0.004
M2REAL	5	Money and credit	Real M2 Money Stock	0.002	0.005	0.002	0.005	0.003	0.005
AMBSL	6	Money and credit	St. Louis Adjusted Monetary Base	0.000	0.018	0.000	0.005	0.000	0.025
TOTRESNS	6	Money and credit	Total Reserves of Depository Institutions	0.000	0.064	0.000	0.032	0.000	0.089
NONBORRES	7	Money and credit	Reserves Of Depository Institutions	0.000	1.089	0.000	0.035	0.000	1.613
BUSLOANS	6	Money and credit	Commercial and Industrial Loans	0.000	0.006	0.000	0.006	0.000	0.007
REALLN	6	Money and credit	Real Estate Loans at All Commercial Banks	0.000	0.005	0.000	0.003	0.000	0.007
NONREVSL	6	Money and credit	Total Nonrevolving Credit	0.000	0.008	0.000	0.008	0.000	0.008
CONSPI	2	Money and credit	Nonrevolving consumer credit to Personal Income	0.000	0.001	0.000	0.001	0.000	0.001
MZMSL	6	Money and credit	MZM Money Stock	0.000	0.006	0.000	0.006	0.000	0.005
DTCOLNVHFNM	6	Money and credit	Consumer Motor Vehicle Loans Outstanding	0.000	0.025	0.000	0.021	0.000	0.029
DTCTHFNM	6	Money and credit	Total Consumer Loans and Leases Outstanding	0.000	0.021	0.000	0.013	0.000	0.027
INVEST	6	Money and credit	Securities in Bank Credit at All Commercial Banks	0.000	0.011	0.000	0.010	0.000	0.012
FEDFUNDS	2	Interest and exchange rates	Effective Federal Funds Rate	0.000	0.521	0.002	0.691	-0.002	0.162
CP3Mx	2	Interest and exchange rates	3-Month AA Financial Commercial Paper Rate	-0.001	0.508	0.001	0.667	-0.003	0.192
TB3MS	2	Interest and exchange rates	3-Month Treasury Bill	0.000	0.431	0.002	0.562	-0.003	0.173
TB6MS	2	Interest and exchange rates	6-Month Treasury Bill	-0.001	0.405	0.001	0.527	-0.003	0.170
GS1	2	Interest and exchange rates	1-Year Treasury Rate	-0.001	0.423	0.001	0.548	-0.003	0.188
GS5	2	Interest and exchange rates	5-Year Treasury Rate	-0.002	0.326	0.006	0.389	-0.011	0.230
GS10	2	Interest and exchange rates	10-Year Treasury Rate	-0.002	0.282	0.007	0.326	-0.013	0.218
AAA	2	Interest and exchange rates	Moody's Seasoned Aaa Corporate Bond	-0.001	0.224	0.010	0.257	-0.013	0.177
BAA	2	Interest and exchange rates	Moody's Seasoned Baa Corporate Bond	0.000	0.216	0.010	0.228	-0.012	0.201
COMPAPFFx	1	Interest and exchange rates	3-Month Commercial Paper Minus	0.084	0.419	0.031	0.518	0.148	0.241
TB3SMFFM	1	Interest and exchange rates	3-Month Treasury C Minus	-0.482	0.711	-0.729	0.860	-0.189	0.263
TB6SMFFM	1	Interest and exchange rates	6-Month Treasury C Minus	-0.349	0.765	-0.573	0.941	-0.081	0.311
T1YFFM	1	Interest and exchange rates	1-Year Treasury C Minus	0.026	0.770	-0.074	0.964	0.145	0.409
T5YFFM	1	Interest and exchange rates	5-Year Treasury C Minus	0.723	1.371	0.447	1.611	1.051	0.915
T10YFFM	1	Interest and exchange rates	10-Year Treasury C Minus	1.061	1.658	0.595	1.834	1.617	1.205
AAAFFM	1	Interest and exchange rates	Moody's Aaa Corporate Bond Minus	2.080	1.982	1.245	1.993	3.077	1.433
BAAFFM	1	Interest and exchange rates	Moody's Baa Corporate Bond Minus	3.100	2.084	2.323	2.096	4.026	1.646
TWEXMMTH	5	Interest and exchange rates	Trade Weighted U.S. Dollar Index: Major Currencies	0.000	0.015	0.000	0.014	0.000	0.016
EXSZUSx	5	Interest and exchange rates	Switzerland / U.S. Foreign Exchange Rate	-0.002	0.026	-0.003	0.026	-0.001	0.025
EXJPUSx	5	Interest and exchange rates	Japan / U.S. Foreign Exchange Rate	-0.002	0.024	-0.003	0.023	0.000	0.026
EXUSUKx	5	Interest and exchange rates	U.S. / U.K. Foreign Exchange Rate	-0.001	0.022	-0.002	0.023	-0.001	0.021
EXCAUSx	5	Interest and exchange rates	Canada / U.S. Foreign Exchange Rate	0.000	0.014	0.001	0.009	0.000	0.018
WPSFD49207	6	Prices	PPI: Finished Goods	0.000	0.007	0.000	0.006	0.000	0.008
WPSFD49502	6	Prices	PPI: Finished Consumer Goods	0.000	0.009	0.000	0.007	0.000	0.010
WPSID61	6	Prices	PPI: Intermediate Materials	0.000	0.007	0.000	0.006	0.000	0.009
WPSID62	6	Prices	PPI: Crude Materials	0.000	0.041	0.000	0.030	0.000	0.052
OILPRICEx	6	Prices	Crude Oil, spliced WTI and Cushing	0.000	0.095	0.000	0.090	0.000	0.100
PPICMM	6	Prices	PPI: Metals and metal products:	0.000	0.033	0.000	0.025	0.000	0.040
CPIAUCSL	6	Prices	CPI : All Items	0.000	0.003	0.000	0.003	0.000	0.003
CPIAPPSL	6	Prices	CPI : Apparel	0.000	0.005	0.000	0.005	0.000	0.006
CPITRNSL	6	Prices	CPI : Transportation	0.000	0.011	0.000	0.006	0.000	0.015
CPIMEDSL	6	Prices	CPI : Medical Care	0.000	0.003	0.000	0.003	0.000	0.002
CUSR0000SAC	6	Prices	CPI : Commodities	0.000	0.005	0.000	0.004	0.000	0.007
CUSR0000SAD	6	Prices	CPI : Durables	0.000	0.003	0.000	0.003	0.000	0.002
CUSR0000SAS	6	Prices	CPI : Services	0.000	0.003	0.000	0.003	0.000	0.001
CPIULFSL	6	Prices	CPI : All Items Less Food	0.000	0.003	0.000	0.003	0.000	0.003
CUSR0000SA0L2	6	Prices	CPI : All items less shelter	0.000	0.004	0.000	0.003	0.000	0.004
CUSR0000SA0L5	6	Prices	CPI : All items less medical care	0.000	0.003	0.000	0.003	0.000	0.003
PCEPI	6	Prices	Personal Cons. Expend.: Chain Index	0.000	0.002	0.000	0.002	0.000	0.002
DDURRG3M086SBEA	6	Prices	Personal Cons. Exp: Durable goods	0.000	0.003	0.000	0.003	0.000	0.003
DNDGRG3M086SBEA	6	Prices	Personal Cons. Exp: Nondurable goods	0.000	0.006	0.000	0.004	0.000	0.008
DSERRG3M086SBEA	6	Prices	Personal Cons. Exp: Services	0.000	0.002	0.000	0.002	0.000	0.002
S&P 500	5	Stock market	S&P's Common Stock Price Index: Composite	0.005	0.036	0.005	0.036	0.006	0.036
S&P: indust	5	Stock market	S&P's Common Stock Price Index: Industrials	0.006	0.036	0.005	0.037	0.006	0.035
S&P div yield	2	Stock market	S&P's Composite Common Stock: Dividend Yield	-0.001	0.121	0.000	0.148	-0.002	0.077
S&P PE ratio	5	Stock market	S&P's Composite Common Stock: Price-Earnings Ratio	0.000	0.046	0.000	0.039	-0.001	0.053
VXOCLSx	1	Stock market	VXO	18.929	7.390	18.299	5.981	19.679	8.730

As in McCracken and Ng (2016) we transform the variables following the code presented in column 'tcode'. The transformation for a series x are: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $log(x_t)$, (5) $\Delta log(x_t)$, (6) $\Delta^2 log(x_t)$, and (7) $\Delta(x_t/x_{t-1} - 1)$. The column 'gsi' and 'gsi:description' present the comparable series in Global Insight.

Table 13: Descriptive Statistics - Sentiment-Based Measures

Description	Full Sample Mean	Full Sample sd	1985-1992 Mean	1985-1992 sd	1993-2018 Mean	1993-2018 sd
Three Component Index	107.76	32.68	114.78	20.65	105.19	35.78
News-Based Policy Uncert Index	108.06	39.55	106.42	28.04	108.66	43.03
News-Based Historical Economic Policy Uncertainty	146.45	41.47	145.15	30.45	146.93	44.88
1. Economic Policy Uncertainty	101.56	41.96	120.50	39.45	94.63	40.77
2. Monetary policy	97.11	59.30	119.85	64.74	88.78	55.00
Fiscal Policy (Taxes OR Spending)	106.62	65.85	122.12	59.68	100.94	67.19
3. Taxes	106.08	65.32	113.14	55.13	103.49	68.58
4. Government spending	110.91	101.51	152.86	101.14	95.54	97.40
5. Health care	115.76	91.37	68.91	57.97	132.93	95.36
6. National security	96.20	81.23	127.60	85.12	84.69	76.77
7. Entitlement programs	112.41	85.69	79.55	67.82	124.45	88.47
8. Regulation	105.40	55.31	106.27	47.06	105.08	58.13
Financial Regulation	105.80	119.82	123.88	109.99	99.17	122.76
9. Trade policy	93.43	107.95	107.88	70.54	88.14	118.42
10. Sovereign debt, currency crises	116.59	202.91	101.42	164.86	122.14	215.20
DebtCeiling Relative Frequency	0.00	0.00	0.00	0.00	0.00	0.00
GovernmentShutdown Relative Frequency	0.00	0.00	0.00	0.00	0.00	0.00
Ratio: EPU w/DebtCeiling to wo/DebtCeiling	1.00	0.01	1.00	0.00	1.00	0.01
Ratio: EPU w/GovtShutdown to wo/GovtShutdown	1.00	0.01	1.00	0.00	1.00	0.02
Financial Stress Indicator	101.23	0.73	101.05	0.37	101.29	0.81

Table 9 reports the mean and standard deviation of the sentiment-based variables used in section 4.3 for the full sample (1895:01-2017:12) and the period of analysis (1993:01-2017:12). The first one is the economic policy uncertainty measure (EPU) from Baker et al. (2016). The sentiment-based variables are: *Three Component Index, News-Based Policy Uncertainty Index, News-Based Historical Economic Policy Uncertainty.* The categorical EPU considers a range of sub-indexes based solely in news data from *Access World News* of over 2,000 US newspapers. The categories are: *Economic Policy Uncertainty, Monetary policy, Fiscal Policy (Taxes or Spending), Taxes, Government Spending, Health Care, National Security, Entitlement Programs, Regulation, Financial Regulation, Trade Policy, and Sovereign Debt, Currency Crises.* Finally, is the Financial Stress Indicator (FSI) for the U.S from Püttmann (2018).