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

Caio Vigo Pereira[†]

[†]*Department of Economics - University of Kansas*

This draft: August 24, 2020

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 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.

Corresponding author - 1460 Jayhawk Blvd - Snow Hall, The University of Kansas, USA. email - caiovigo@gmail.com

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 (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.

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.

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:

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

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

$$rx^{(n)}_{t+h/12} \equiv \underbrace{p^{(n-h/12)}_{t+h/12} - p^{(n)}_t}_{\text{holding period return } r^{(n)}_{t+h/12}} - (h/12)y^{(h/12)}_t \tag{4}$$

$$= ny^{(n)}_t - (n-h/12)y^{(n-h/12)}_{t+h/12} - (h/12)y^{(h/12)}_t$$

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

$$\begin{aligned} f^{(n)}_t &\equiv p^{(n-h/12)}_t - \quad p^{(n)}_t \\ &= ny^{(n)}_t - (n-h/12)y^{(n-h/12)}_t \end{aligned} \tag{5}$$

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[1]

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

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 $\frac{1}{n}\mathbb{E}\left(rx^{(n)}_{t+1/12} + rx^{(n-1/12)}_{t+2/12} + rx^{(n-2/12)}_{t+3/12} + \ldots + +rx^{(2/12)}_{t+n-1/12}\right)$, then we can write

$$\begin{aligned} y^{(n)}_t \equiv \; & \frac{1}{n}\mathbb{E}_t\left(y^{(1/12)}_t + y^{(1/12)}_{t+1/12} + \ldots + y^{(1/12)}_{t+n-1/12}\right) + \\ & \frac{1}{n}\mathbb{E}_t\left(rx^{(n)}_{t+1/12} + rx^{(n-1/12)}_{t+2/12} + rx^{(n-2/12)}_{t+3/12} + \ldots + +rx^{(2/12)}_{t+n-1/12}\right) \quad . \end{aligned} \tag{9}$$

[1] 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:

$$y^{(n)}_t \equiv \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1} y^{(h/12)}_{t+j\cdot h/12}\right) \quad + \quad \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1} rx^{(n-j\cdot h/12)}_{t+h/12(j+1)}\right) \tag{6}$$

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

$$y^{(n)}_t \equiv \frac{1}{n}\left(\sum_{j=0}^{n-1} y^{(1)}_{t+j}\right) \quad + \quad \frac{1}{n}\left(\sum_{j=0}^{n-1} rx^{(n-j)}_{t+j+1}\right) \tag{7}$$

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

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

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

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

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 $f\left(\cdot\right)^{-1}$ 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 such a way that each $\boldsymbol{z}_t \in \mathbf{Z}_t$ has its own effect on the yield curve $\boldsymbol{y}_t^{(n)}$. Thus, for a function $g\left(\cdot\right)$ we can write $\mathbb{E}_t\left(y_t^{(n)}\right) = g\left(\boldsymbol{y}_t^{(n)}; N\right)$[2].

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 to expected excess returns from shocks to expected future yields. What boils down to estimating the function $g\left(\cdot\right)$. 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

[2]Which 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.

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

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

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

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

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 ($\hat{g}_{i,t}$)[3]. 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

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

[3] 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 $\widehat{\boldsymbol{G}}_t$ and $\boldsymbol{\lambda}$ are 5×1 vectors given by $\widehat{\boldsymbol{G}}_t \equiv \begin{bmatrix} \hat{g}_{1,t} & \hat{g}^3_{1,t} & \hat{g}_{3,t} & \hat{g}_{5,t} & \hat{g}_{8,t} \end{bmatrix}^\top$, 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 $\widehat{LN}^h_t = \widehat{\boldsymbol{\lambda}}^\top \widehat{\boldsymbol{G}}_t$, the predictive regression in this approach is given by

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

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

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

where $\boldsymbol{Z}_t = \left\{ \boldsymbol{Z}^{\mathrm{y}}_t, \boldsymbol{Z}^{\mathrm{y}^\complement}_t \right\}$ 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 $\boldsymbol{Z}_t = \{\boldsymbol{Z}^{\mathrm{y}}_t\}$ and no macroeconomic variables are used to define the state space. Evidence against the spanning hypothesis is showed when $\boldsymbol{Z}^{\mathrm{y}^\complement}_t \neq \emptyset$.

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)[4] 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 $rx^{(n)}_{t+h/12} = g(\boldsymbol{Z}_t, \boldsymbol{\theta}_t)$ 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 $g(\boldsymbol{Z}_t) = g^{(L)}\left(\ldots\left(g^{(2)}\left(g^{(1)}(\boldsymbol{Z}_t)\right)\right)\right)$. The name *network* comes from this chain and its interconnectedness architecture, and *feedforward* because the information flows in one direction

[4] They are also known as feedforward neural network

from $\boldsymbol{Z}_t$ through these functions, to finally obtain an output $rx^{(n)}_{t+h/12}$. 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 $g^{(L)}(\cdot)$ 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 $rx^{(n)}_{t+h/12} = g\left(\boldsymbol{Z}_t, \boldsymbol{\theta}_t, \boldsymbol{w}\right) = \phi\left(\boldsymbol{Z}_t, \boldsymbol{\theta}_t\right)^{\top} \boldsymbol{w}$.

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 $\boldsymbol{W}$ 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[5] 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.

[5]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

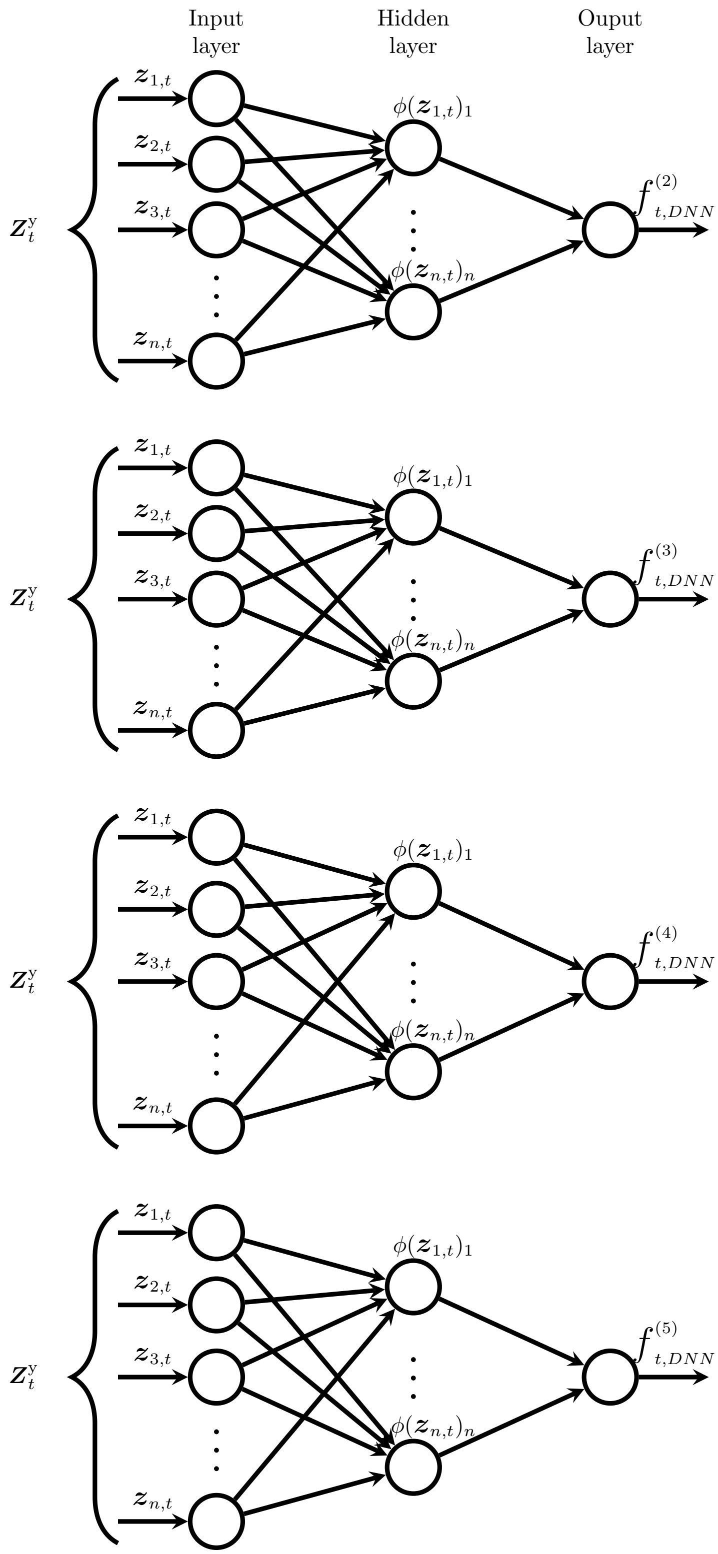

Figure 1 shows the general structure of the deep feed forward designed to obtain the DNN factor $\mathfrak{f}_{DNN}$. There are four groups of networks, each group for $n \in \{1, 2, 3, 4\}$. The inputs layer receives data from $\boldsymbol{Z}_t^{\mathrm{y}} = \{\boldsymbol{z}_{1,t}, \boldsymbol{z}_{2,t},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

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 derived from our deep neural network, as below:

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

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

$\mathfrak{f}^{(n)}_{t,DNN}$ 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 $\boldsymbol{Z}_t = \left\{\boldsymbol{Z}_t^{\mathrm{y}}, \boldsymbol{Z}_t^{\mathrm{y}^{\complement}}\right\}$, we write the dynamics of $\boldsymbol{Z}_t$ that capture all the risks of the economy following a Gaussian VAR process given by:

$$\begin{aligned}\begin{bmatrix}\boldsymbol{Z}_t^{\mathrm{y}}\\ \boldsymbol{Z}_t^{\mathrm{y}^{\complement}}\end{bmatrix} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\begin{bmatrix}\boldsymbol{Z}_{t-1}^{\mathrm{y}}\\ \boldsymbol{Z}_{t-1}^{\mathrm{y}^{\complement}}\end{bmatrix} + \boldsymbol{\Sigma}\epsilon_t\\ \boldsymbol{Z}_t &= \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{Z}_{t-1} + \boldsymbol{\Sigma}\epsilon_t \qquad \epsilon_t \sim N(0,\boldsymbol{I})\end{aligned} \tag{25}$$

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:

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

where $\boldsymbol{M}_{\boldsymbol{Z}_t^{\mathrm{y}}}\boldsymbol{Z}_t^{\mathrm{y}^{\complement}}$ is the annihilator matrix of the space spanned by $\boldsymbol{Z}_t^{\mathrm{y}}$, i.e., $\boldsymbol{M}_{\boldsymbol{Z}_t^{\mathrm{y}}}\boldsymbol{Z}_t^{\mathrm{y}^{\complement}} \equiv \boldsymbol{Z}_t^{\mathrm{y}^{\complement}} - \mathrm{Proj}\left[\boldsymbol{Z}_t^{\mathrm{y}^{\complement}}|\boldsymbol{Z}_t^{\mathrm{y}}\right]$. Previous models have assumed that the $\boldsymbol{Z}_t^{\mathrm{y}^{\complement}}$ was spanned by $\boldsymbol{Z}_t^{\mathrm{y}}$, thus imposing the restriction of $\boldsymbol{Z}_t^{\mathrm{y}^{\mathsf{C}}} = \mathrm{Proj}\left[\boldsymbol{Z}_t^{\mathrm{y}^{\mathsf{C}}} | \boldsymbol{Z}_t^{\mathrm{y}}\right]$ 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 $\boldsymbol{Z}_t^{\mathrm{y}^{\mathsf{C}}}$ 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 $\boldsymbol{Z}_t^{\mathrm{y}^{\mathsf{C}}}$ is correlated with macroeconomic variables and sentiment-based measures.

In our methodology, $\boldsymbol{Z}_t$ is given by the derived factor $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$, and $\boldsymbol{Z}_t^{\mathrm{y}^{\mathsf{C}}}$ by a function of $\boldsymbol{\xi}_{t+h/12}^h$ as $f(\boldsymbol{\xi}_{t+h/12}^h)$. To close our illustrative term-structure model, analogous to Joslin et al. (2014), we argue $f(\boldsymbol{\xi}_{t+h/12}^h)$ 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 full list factors.

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[6]. Thus, we reconstruct the log yield of a zero-coupon with n-period maturity at time t as

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

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

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

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 $n = 2, 3, 4$ and 5 years. In Appendix A, 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

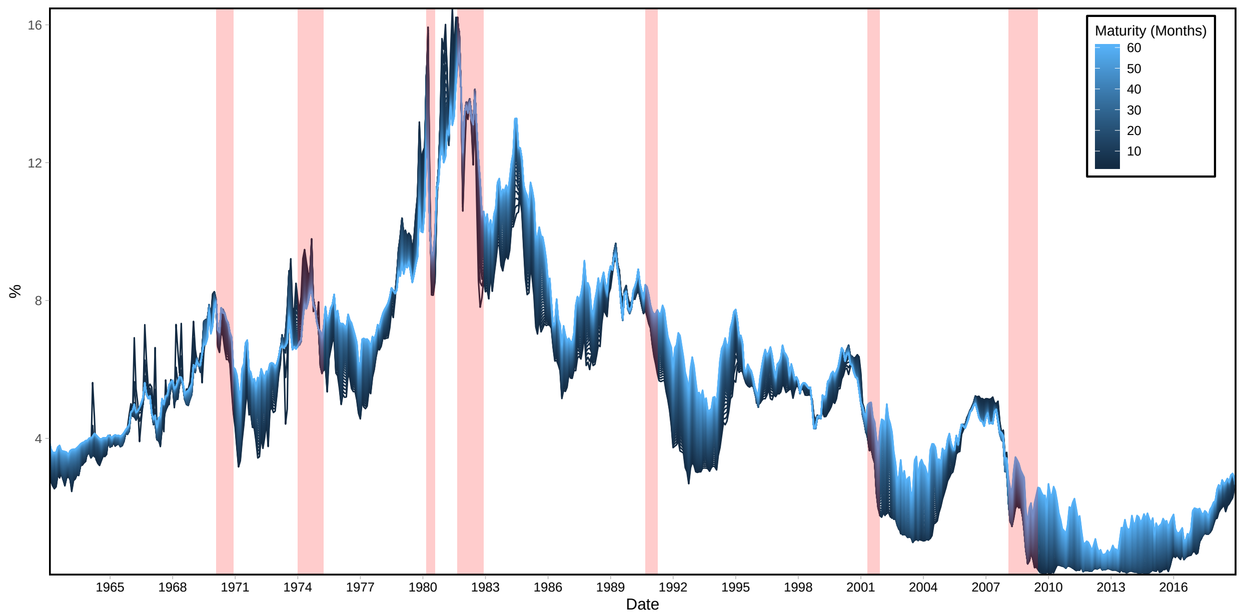

Figure 2 shows the log yields for all maturities we consider: $y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}$. At each month t, there are 60 yields represented by variation of color. The log yields of the zero-coupons bonds are reconstructed with equation 27, using the last day of each month estimated 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[7] 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:

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

Figure 3: 1-Month Bonds Excess Returns

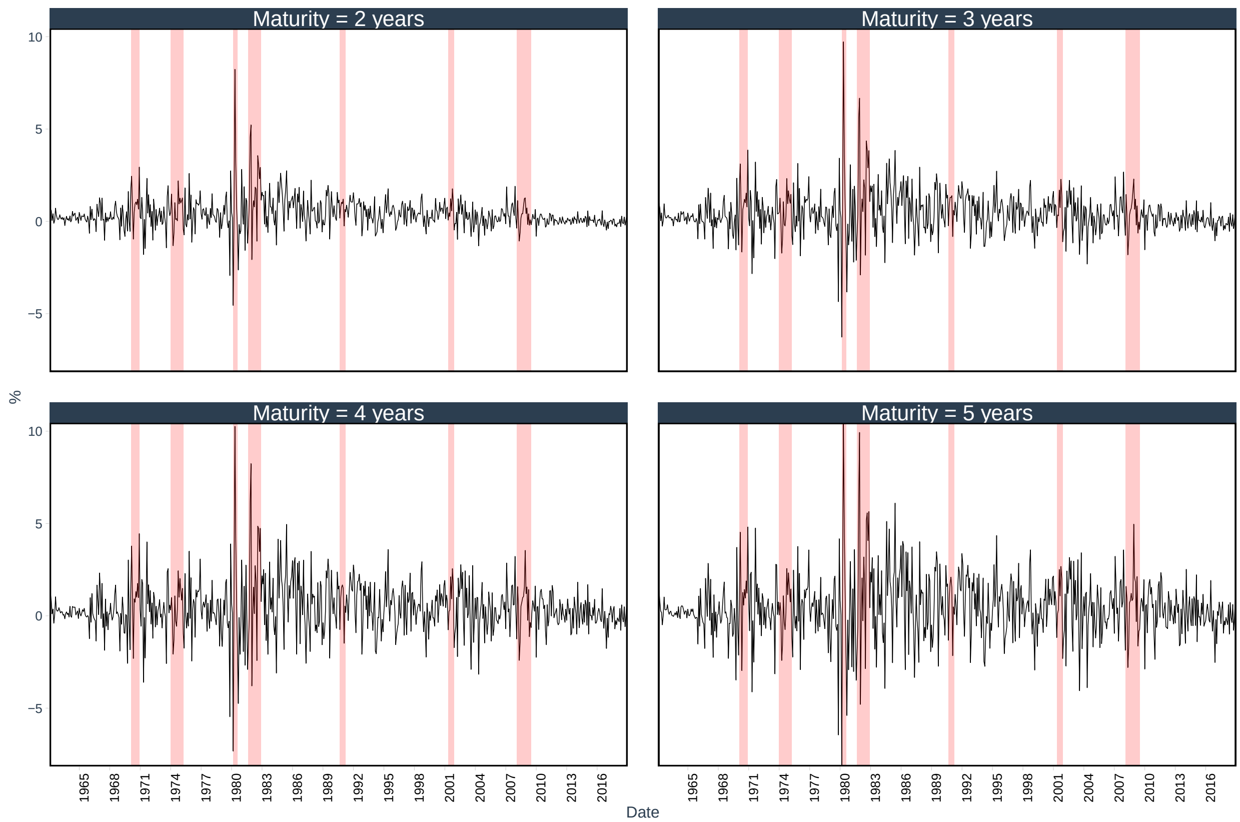

Figure 3 shows the 1-month excess returns for maturities with $n = 2, 3, 4$ and 5 years. The excess returns are calculeted as in equation 4, i.e., $rx^{(n)}_{t+1/12} = ny^{(n)}_t - (n + 1/12)y^{(n-h/12)}_{t+1/12} - y^n_t$. Each panel represents one of the four maturities. The y-axis shows values in percentage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

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

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 1**. 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 $\widehat{\boldsymbol{\mathfrak{F}}}_t^h$ from equation 19 at each t can be summarized in following way. First, for each set of $\boldsymbol{Z}_t^{\mathrm{y}}$ in consideration, we feed each one of the three DNNs architectures with the entire past information of each variable in $\boldsymbol{Z}_t^{\mathrm{y}}$. We use the 10% most recent data in each $\boldsymbol{z}_t^{\mathrm{y}} \in \boldsymbol{Z}_t^{\mathrm{y}}$ for validation. After the set of weights are chosen, with the final set of weights and the final approximated function, we use $\boldsymbol{Z}_t^{\mathrm{y}}$ to predict $rx_{t+h/12}^{(n)}$ in each group of maturity $n \in \{1, 2, 3, 4\}$. Thus, we form the 4×1 vector of $\widehat{\boldsymbol{\mathfrak{F}}}_t^h$ as the factor at t generated by each DNN. As a final step we run univariate regressions to obtain $\boldsymbol{\xi}_{t+h/12}^h$, as shown in equation 21. Similarly, we build the 4×1 vector using the observation t residuals as $\widehat{\boldsymbol{\xi}}^{h}_{t+h/12} = \begin{bmatrix} \widehat{\xi}^{(2),h}_{t+h/12} & \widehat{\xi}^{(3),h}_{t+h/12} & \widehat{\xi}^{(4),h}_{t+h/12} & \widehat{\xi}^{(5),h}_{t+h/12} \end{bmatrix}$.

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

Consistent with our adapted dynamic term-structure model, the orthogonal vector from $\text{Proj}\left[f(\boldsymbol{\xi}^{(n)}_{t+h/12})|\boldsymbol{Z}^{\mathrm{y}}_{t}\right]$ has predictive power for excess returns. Thus, we use the projection error $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})^{h}_{t+h/12}$ for alternative 1 and $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)^{(-n),h}_{t+h/12}$ for alternative 2 in our predictive analysis in the following section.

The intuition that motivates our construction of $\boldsymbol{Z}^{\mathrm{y}^{\complement}}_{t}$ lies in the fact that at each t, $\widehat{\xi}^{(n)}_{t+h/12}$ is orthogonal to $\mathfrak{f}^{(n),h}_{t,DNN}$, 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 $\boldsymbol{Z}^{\mathrm{y}}_{t}$, are unspanned and should be in an orthogonal space. Hence, the unspanned information in $\widehat{\boldsymbol{\xi}}^{h}_{t+h/12}$ 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 $\boldsymbol{Z}^{\mathrm{y}^{\complement}}_{t}$ 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 $\mathfrak{f}^{(n),h}_{t,DNN}$ 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 $\mathfrak{f}^{(n),h}_{t,DNN}$. Column (1) shows the the derived DNN factors when we feed the MLP with all the forward rates, i.e., $\boldsymbol{Z}^{\mathrm{y}}_{t} = \left\{f^{(2/12)}_{t}, f^{(3/12)}_{t}, \ldots, f^{(60/12)}_{t}\right\}$, column (2) when the set of yields is used, i.e.,

$\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$ and column (3) when both previous sets are combined, i.e., $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60)}, y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$. 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 $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h$ as a linear combination of the derived factors $\mathfrak{f}_{t,DNN}^{(n),h}$. Figure 6 plots $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h$ for each DNN architecture and the three different sets of $\boldsymbol{Z}_t^{\mathrm{y}}$, 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 $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h$ 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 $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h$, 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: $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t \right)_t^h$ 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

Figure 4: DNN Factor $\mathfrak{f}_{t,DNN}^{(n),h}$ by MLP Architecture and Choice of $\boldsymbol{Z}_t^{\mathrm{y}}$

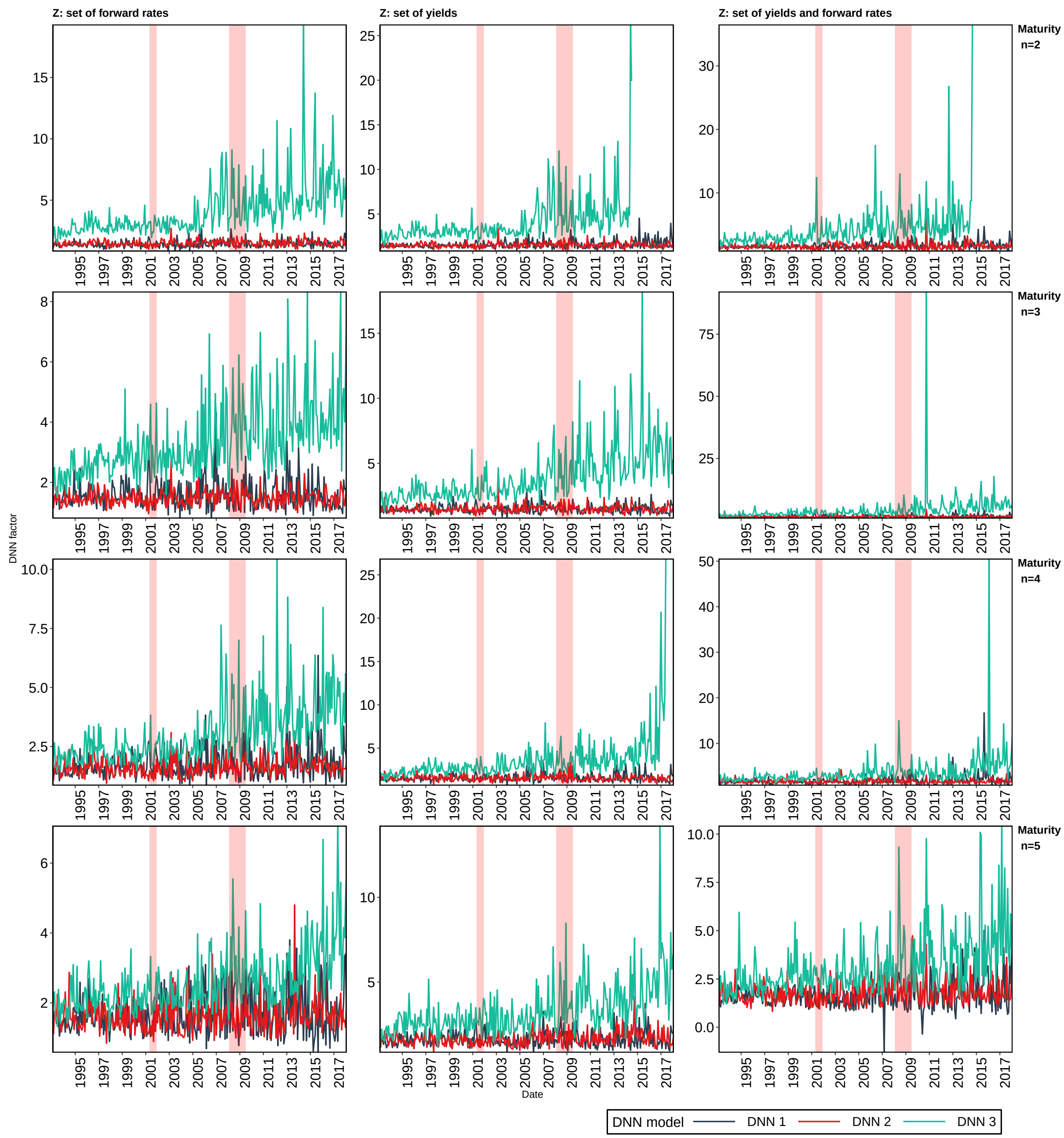

Figure 4 shows the derived $\mathfrak{f}_{t,DNN}^{(n),h}$ for each scenario under consideration. Each column uses a different set of information from the term-structure to derive the factor $\mathfrak{f}_{t,DNN}^{(n),h}$. Column (1) shows the the derived DNN factors for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60)}\right\}$, column (2) for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}\right\}$ and column (3) for $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)}, y_t^{(1)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}\right\}$. 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

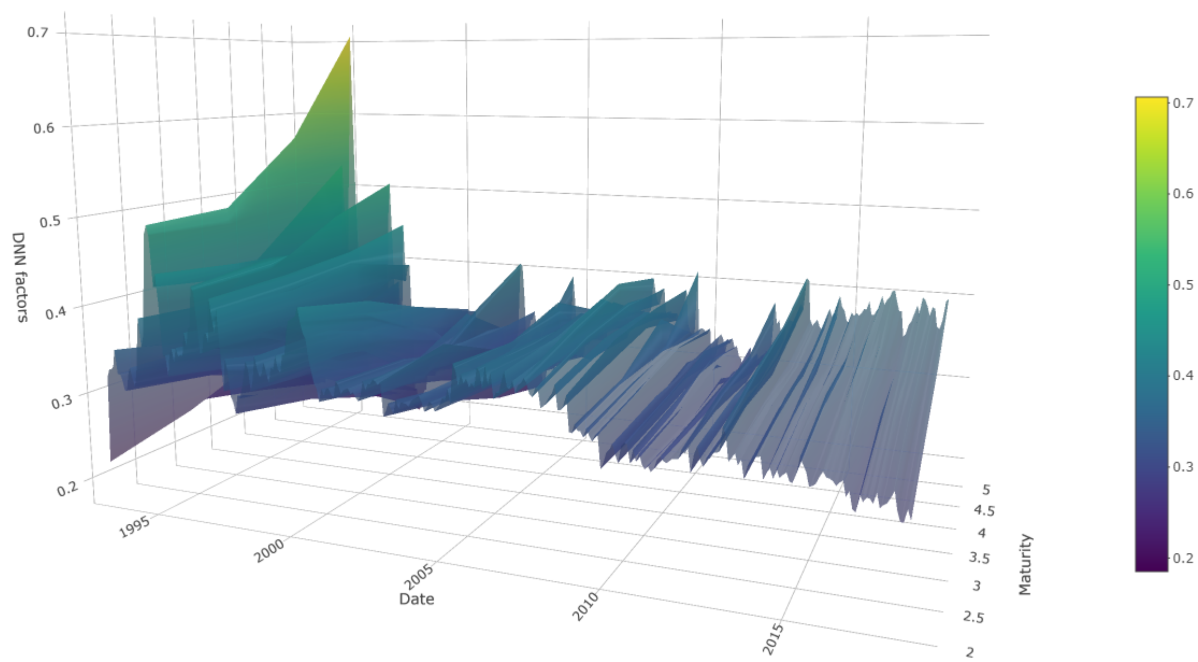

Figure 5 shows a 3D representation of $\mathfrak{f}_{t,DNN}^{(n),h}$ generated by the MLP architecture **DNN 2** using the set of yields $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$ in terms of maturity for the period of analysis. **DNN 2** is a 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 $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ Series by DNN Architecture and Choice of $\boldsymbol{Z}_t^{\mathrm{y}}$

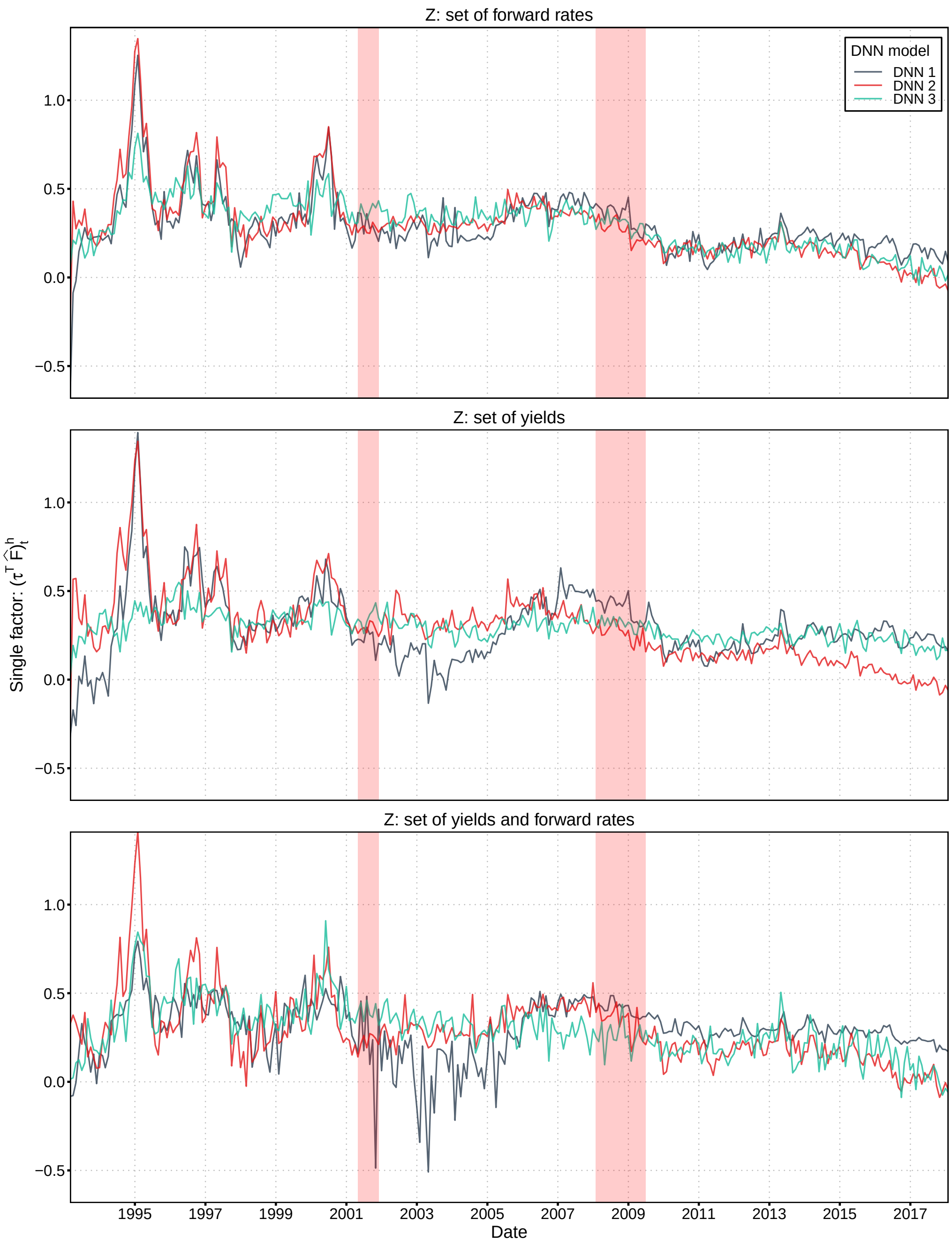

Figure 6 shows $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ for each DNN architecture and the three different sets of $\boldsymbol{Z}_t^{\mathrm{y}}$. The first panel plots $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ when $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60)}\right\}$ is used to obtain $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ 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 $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ 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.

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

	DNN 1			DNN 2			DNN 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: $rx_{t+h/12}^{(2)}$									
$(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-2),h}$		0.760***			0.779***			0.875***	
		(0.204)			(0.180)			(0.211)	
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$			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: $rx_{t+h/12}^{(3)}$									
$(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-3),h}$		0.799***			0.789***			0.984***	
		(0.234)			(0.219)			(0.236)	
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$			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
Panel C: $rx_{t+h/12}^{(4)}$									
$(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-4),h}$		0.795***			0.807***			1.038***	
		(0.291)			(0.288)			(0.289)	
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$			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: $rx_{t+h/12}^{(5)}$									
$(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-5),h}$		0.854**			0.848***			1.069***	
		(0.336)			(0.318)			(0.339)	
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$			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 $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$, $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_t^h$ and $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$ as 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 $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ 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.

only use as the $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$ 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 $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$. The two alternatives for the single factor that captures the unspanned information from the yield curve are the 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.

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 sepa-

Figure 7: Time Series of our Derived Factors $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ and $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_t^h$, along with $\widehat{CP}_t^h$ and $\widehat{LN}_t^h$

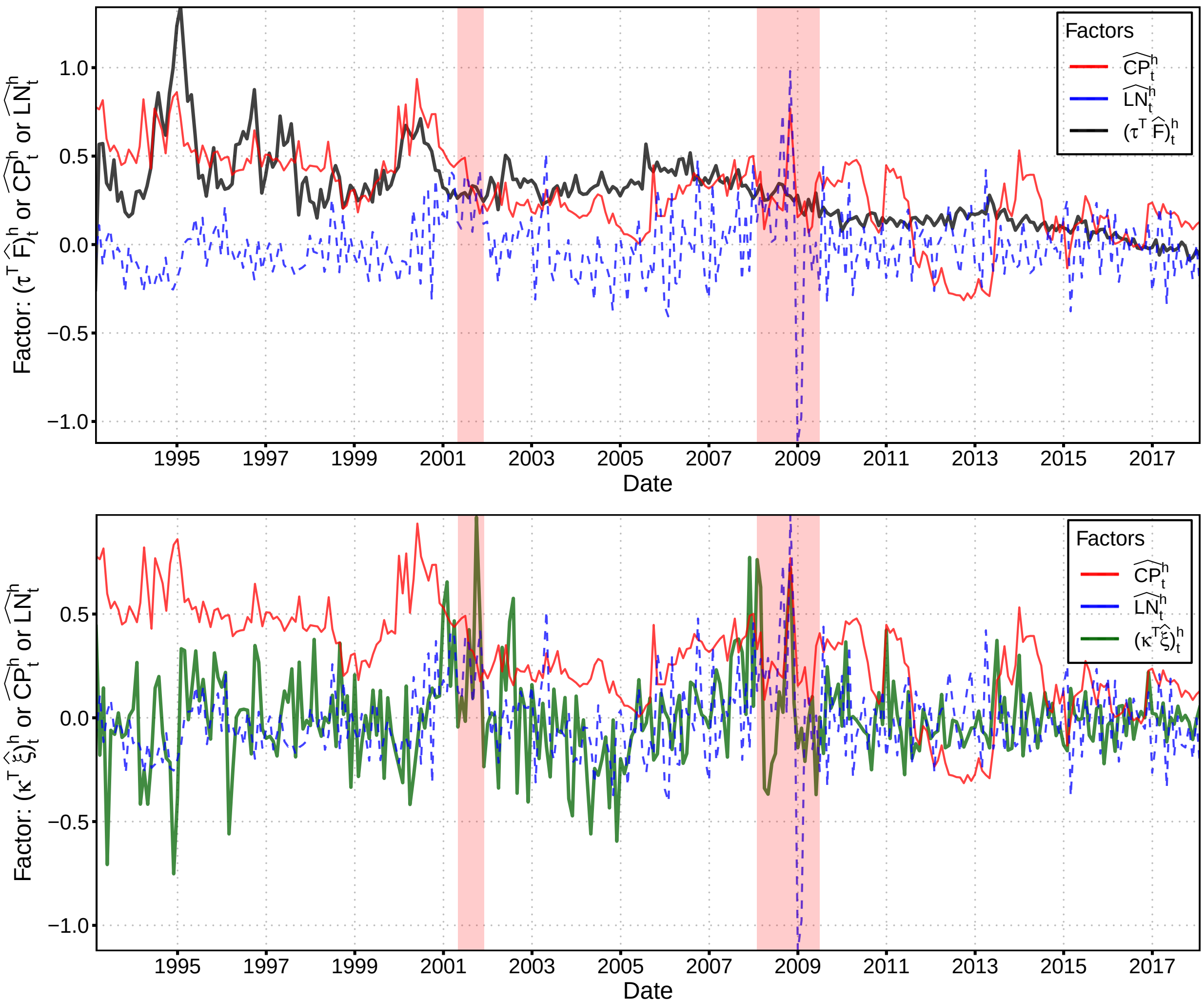

Figure 7 plots in two separated panels our single factor that spans the information from the term-structure, as well as the factor with the spanned risks (alternative 1). The graph in the top plots $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ along with the Cochrane-Piazzesi and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

Table 2: Correlation Matrix

	$(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}})^h_t$	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^h_{t+h/12}$	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^{(-2),h}_{t+h/12}$	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^{(-3),h}_{t+h/12}$	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^{(-4),h}_{t+h/12}$	$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^{(-5),h}_{t+h/12}$	$\widehat{CP}^h_t$	$\widehat{LN}^h_t$
$(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}})^h_t$	1	0	0	0	0	0	0.556	-0.059
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^h_{t+h/12}$	0	1	0.995	0.912	0.904	0.919	0.129	0.171
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^{(-2),h}_{t+h/12}$	0	0.995	1	0.938	0.900	0.888	0.135	0.174
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^{(-3),h}_{t+h/12}$	0	0.912	0.938	1	0.947	0.849	0.170	0.203
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^{(-4),h}_{t+h/12}$	0	0.904	0.900	0.947	1	0.959	0.173	0.204
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})^{(-5),h}_{t+h/12}$	0	0.919	0.888	0.849	0.959	1	0.146	0.178
$\widehat{CP}^h_t$	0.556	0.129	0.135	0.170	0.173	0.146	1	-0.007
$\widehat{LN}^h_t$	-0.059	0.171	0.174	0.203	0.204	0.178	-0.007	1

Table 2 reports the correlation between our single factor $\left(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}_t\right)^h_t$, with the factors that complete the state space in our dynamic term-structure model. The first alternative is $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)^h_t$, which is the unique factor obtained as the projection of $\overline{rx}_{t+h/12}$ in $\widehat{\boldsymbol{\xi}}_{t+h/12}$. The second alternative is a similar projection, however for each maturity $n \in \{2,3,4,5\}$ we regress $rx^{(n)}_{t+h/12}$ on $\widehat{\boldsymbol{\xi}}^{(-n),h}_{t+h/12} \equiv \widehat{\boldsymbol{\xi}}^h_{t+h/12} \setminus \widehat{\xi}^{(n),h}_{t+h/12}$. We use orthogonal vector from the projection of each one of them on $\left(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}_t\right)^h_t$ 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.

rated panels our single factor that spans the information from the term-structure, $\left(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}_t\right)^h_t$, as well as the factor with the spanned risks from alternative 1, $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)^h_t$, 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, $\left(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}_t\right)^h_t$ 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 $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)^h_t$, or $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)^{(-n),h}_t$. Given that Ludvingson-Ng factor is solely based in macroeconomic variables information, we see that the correlation of $\widehat{LN}^h_t$ 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 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 $\left(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}_t\right)^h_t$, $\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)^h_t$ 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 $(h = 1)$. For each maturity (in each one of the four panels), there

Table 3: 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 Regressions with Forward Spreads

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-2),h}_{t+h/12}$		0.658***		0.745***		0.704***		0.558***
		(0.172)		(0.182)		(0.182)		(0.185)
$\widehat{LN}_t^h$	0.617***	0.529***					0.559***	0.518***
	(0.127)	(0.120)					(0.110)	(0.110)
$fs_t^{(n,h)}$			−0.746	−0.225			−0.570	−0.172
			(0.476)	(0.438)			(0.437)	(0.429)
$\widehat{CP}_t^h$					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)
$(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-3),h}_{t+h/12}$		0.620***		0.852***		0.692***		0.585**
		(0.209)		(0.228)		(0.226)		(0.237)
$\widehat{LN}_t^h$	0.921***	0.800***					0.900***	0.823***
	(0.209)	(0.201)					(0.194)	(0.191)
$fs_t^{(n,h)}$			−0.215	0.410			−0.053	0.394
			(0.554)	(0.532)			(0.525)	(0.542)
$\widehat{CP}_t^h$					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 $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ 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 A reports the predictive regressions for maturity $n = 2$ years. Panel B reports the predictive regressions for maturity $n = 3$ years. The state factor $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ 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.

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)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-4),h}$		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 R^2	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-5),h}$		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 R^2	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 $\boldsymbol{Z}_t^{\mathrm{y}} = \left\{y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)}\right\}$ to feed the MLP. We use Newey-West robust standard errors. Sample ranges from 1993 : 01 to 2017 : 12.

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[8] 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, table 6 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 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.

[8] https://research.stlouisfed.org/econ/mccracken/fred-databases/

Figure 8: Marginal R^2 of the factors $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ and $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^h$

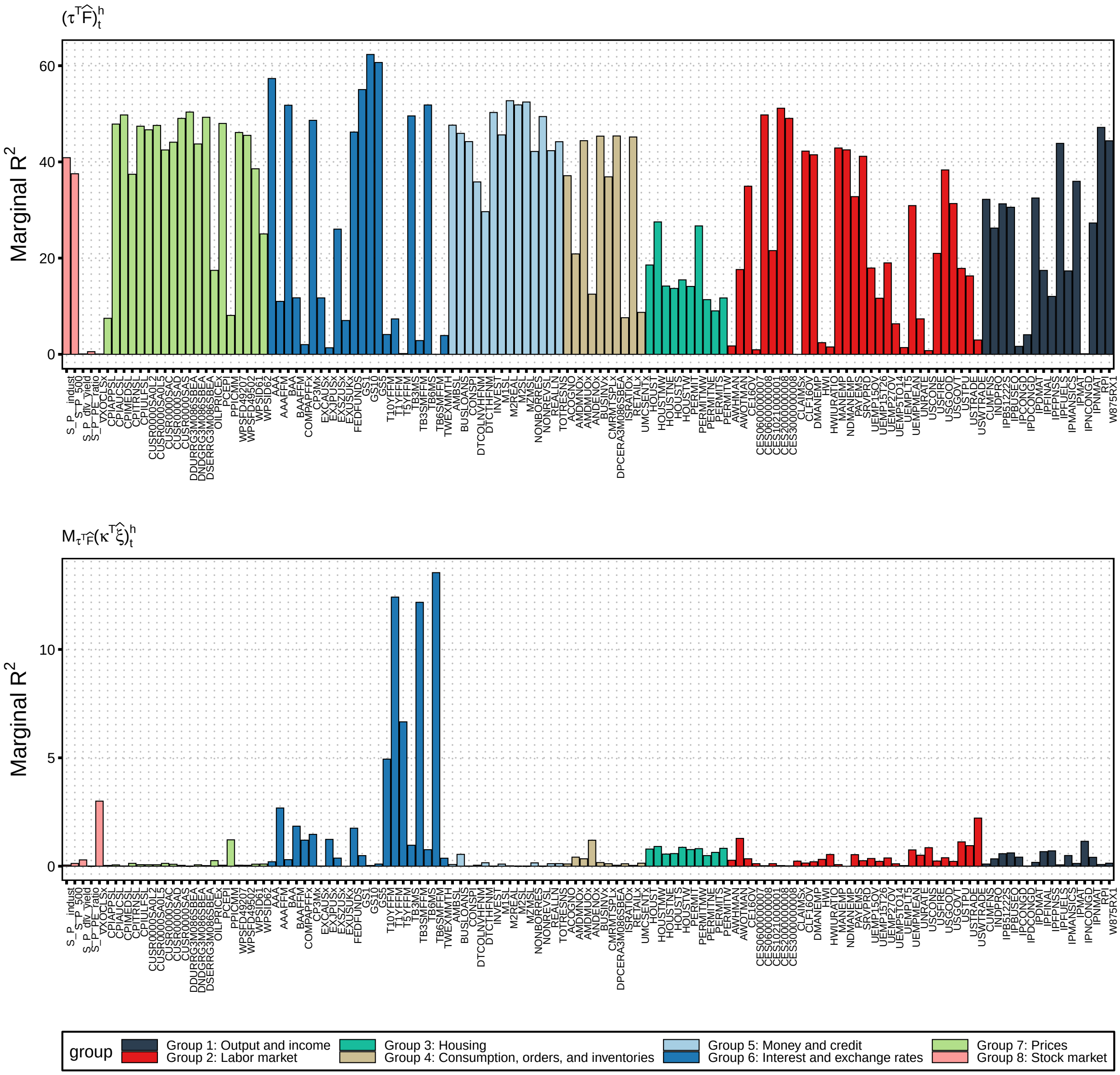

Figure 8 reports the marginal R^2 of the factor $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ in the top panel, and $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^h$ in the bottom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

Figure 9: Marginal R^2 of the factors $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$

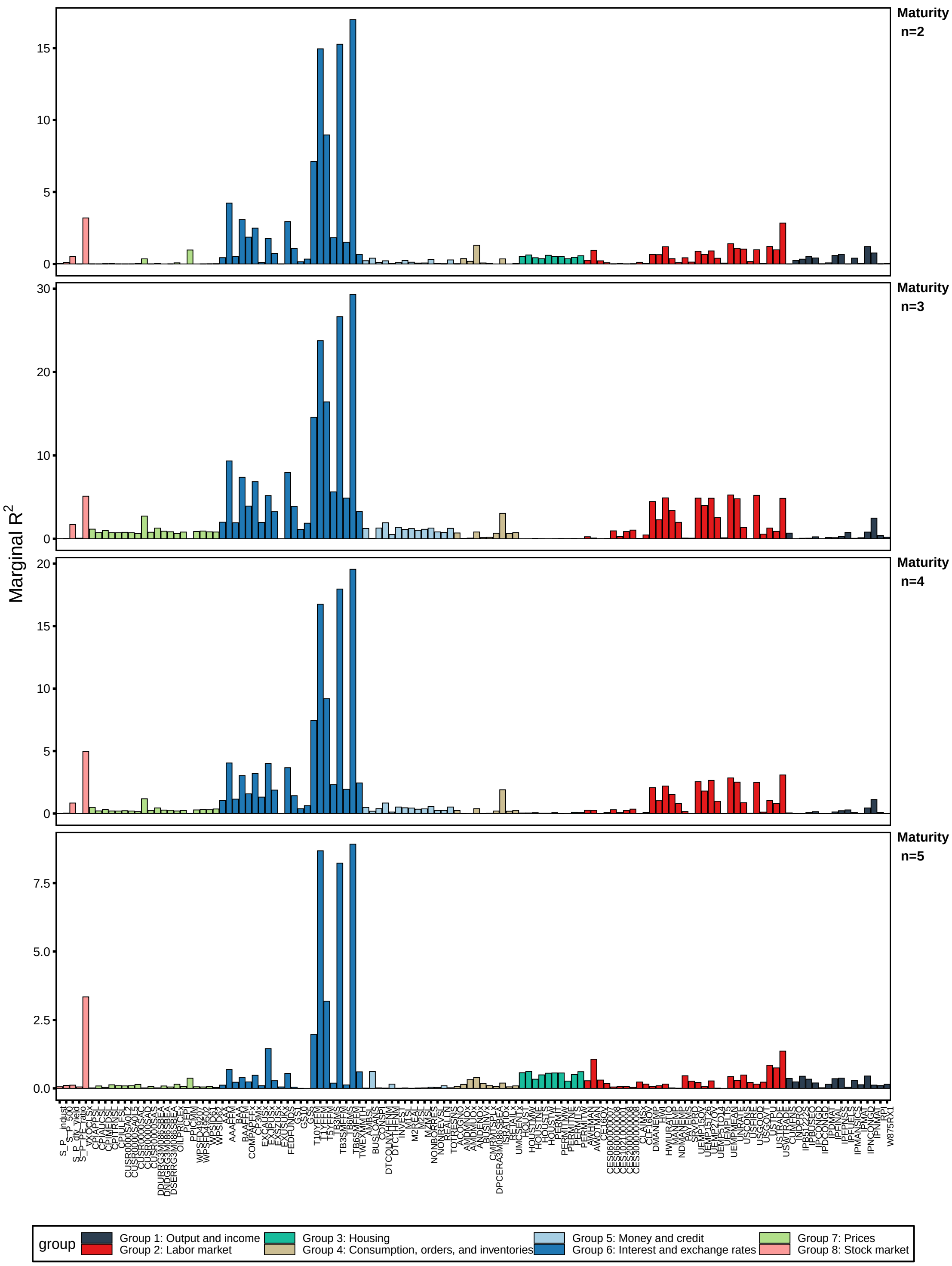

Figure 9 reports the marginal R^2 of the factors $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ by maturity. The marginal R^2 is obtained with the regression of each one the 128 variables from the FRED-MD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

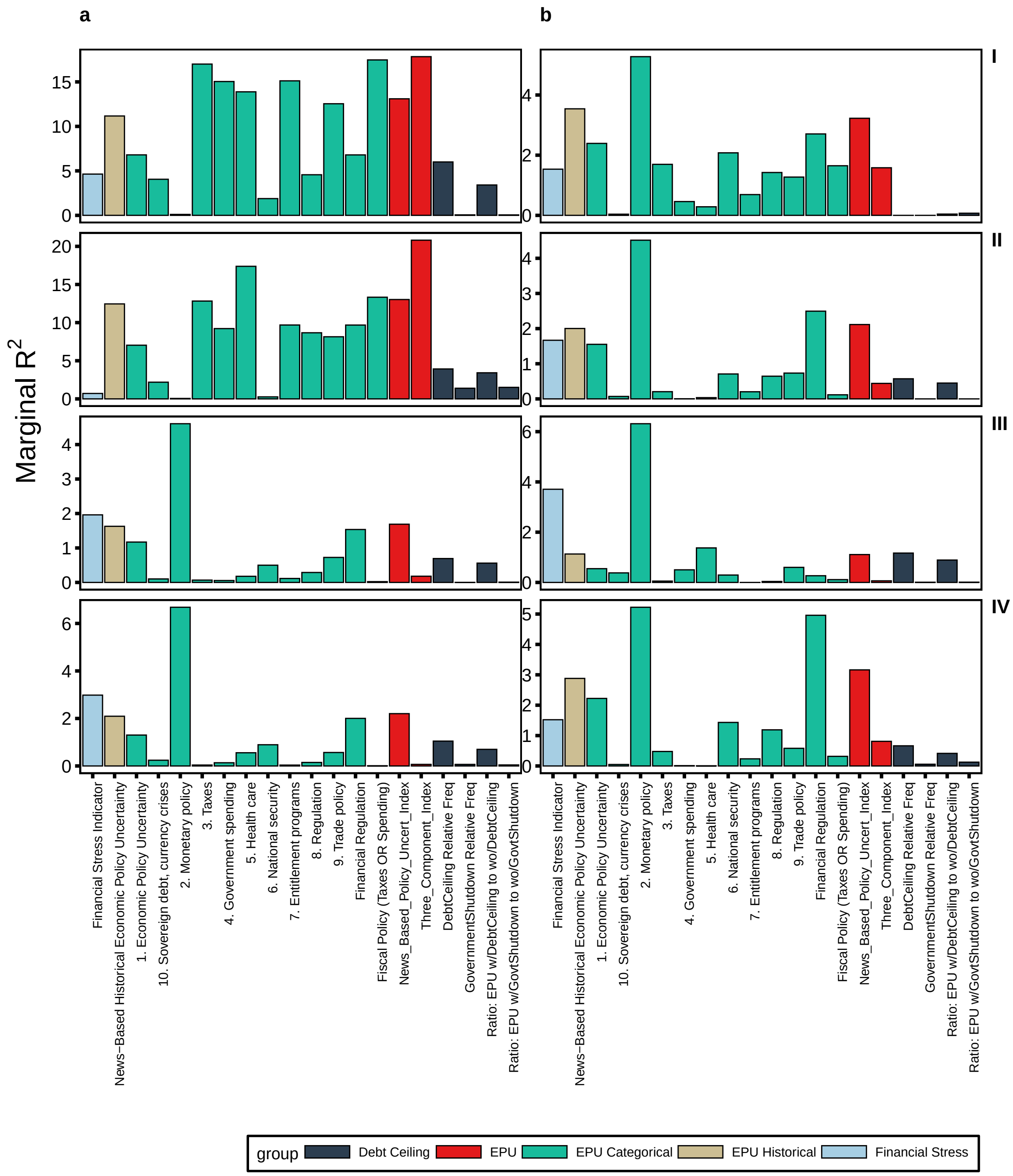

Figure 10 reports the marginal R^2 obtained from sentiment-based measures. It is obtained with the regression of each one these indexes on our state variables. For comparison, we also report for the Cochrane-Piazzesi and Ludvingson-Ng factors. Row (I), panel (a) shows the marginal R^2 for $\widehat{CP}_t^h$,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

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 *3-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[9].

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[10]. 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 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

[9]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

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

$\left(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}}\right)_t^{(-4),h}$, 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

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
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h} + \epsilon_{t+h/12}$	0.21	0.05	-0.01	-0.02
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{h} + \epsilon_{t+h/12}$	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
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h} + \beta_3\widehat{LN}_t^h + \epsilon_{t+h/12}$	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
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h} + \beta_3 fs_t^{(n,h)} + \epsilon_{t+h/12}$	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
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h} + \beta_3\widehat{CP}_t^h + \epsilon_{t+h/12}$	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
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}})_t^h + \beta_2\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h} + \beta_3\widehat{LN}_t^h + \beta_4 fs_t^{(n,h)} + \beta_5\widehat{CP}_t^h + \epsilon_{t+h/12}$	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.

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

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

where $\widehat{rx}^{(n)}_{t+h/12|t}$ is the estimate of the conditional mean of the excess returns for the bond with maturity (n), and $\overline{rx}^{(n)}_{t+h/12|t}$ 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 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 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 reports the coefficients estimates of $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t \right)^h_t$, $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}} (\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-n),h}_{t+h/12}$ and $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}} (\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{h}_{t+h/12}$. In a recursive approach we seek to show how the estimates of the parameters varies across

Figure 11

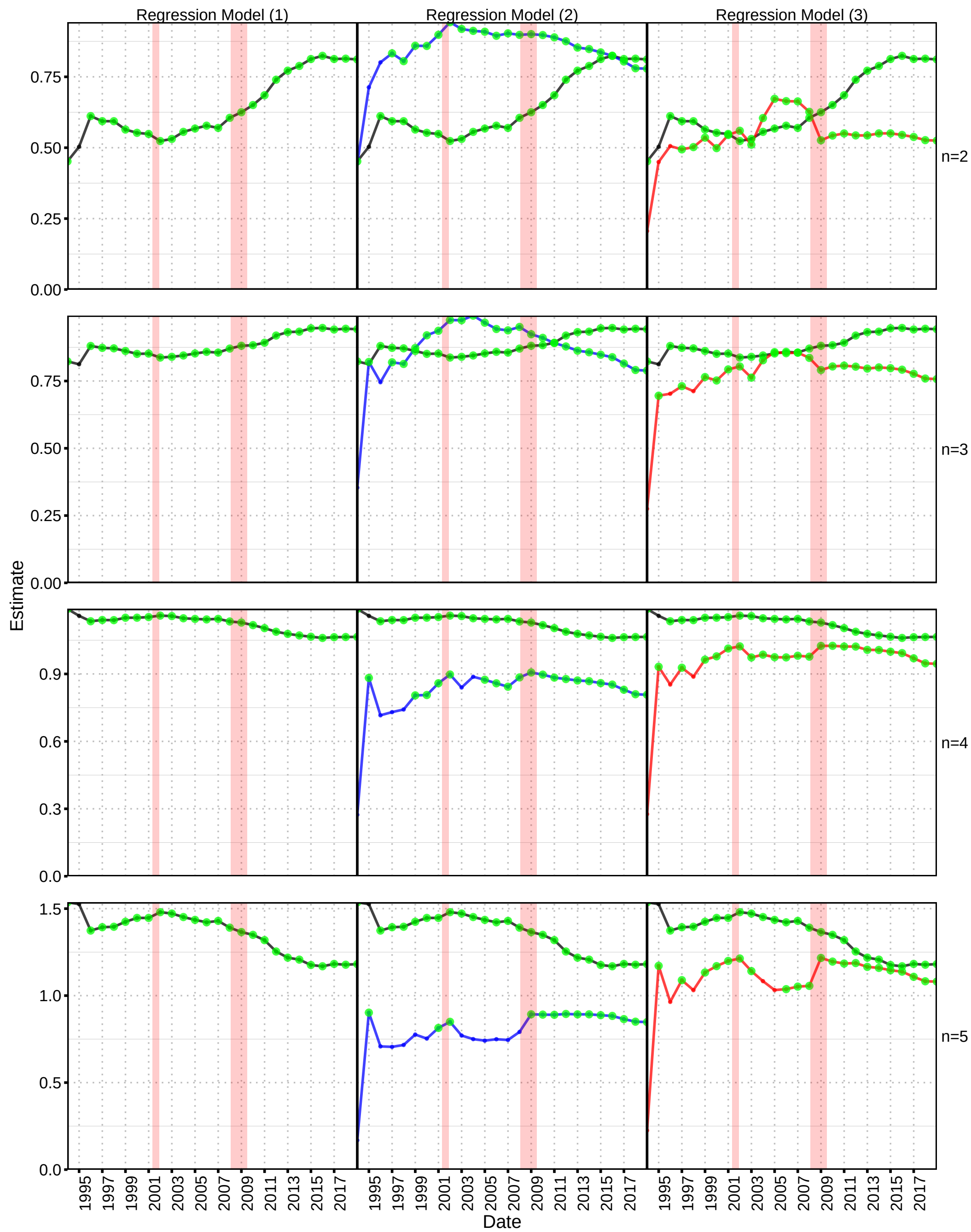

Figure 11 reports the estimates of each one of the three regressions models reported similiar to those summarized in table 1, where the x-axis defines the end of the sample. All samples start in 1993 : 01. Regression Model (1) is given by $rx^{(n)}_{t+h/12} = \beta_0 + \beta_1(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t)^h_t + \epsilon_{t+h/12}$. Regression Model (2) is given by $rx^{(n)}_{t+h/12} = \beta_0 + \beta_1 \boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-n),h}_{t+h/12} + \epsilon_{t+h/12}$. Regression Model (3) is given by $rx^{(n)}_{t+h/12} = \beta_0 + \beta_1 \boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_{t+h/12} + \epsilon_{t+h/12}$. The black line represents the estimates of $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)^h_t$. The blue line represents the estimates of $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-n),h}_{t+h/12}$. Finally, the red represents the estimates of $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_{t+h/12}$. 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 $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$ Over Time as a Function of Maturity (n)

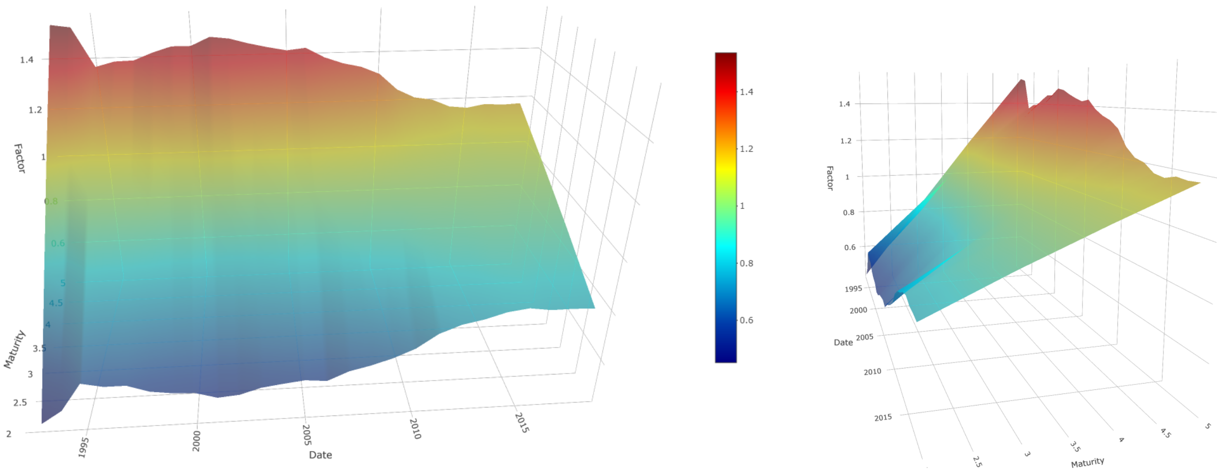

Figure 12 plots the behavior of our spanning factor $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$ as a function of maturity (n) over the period of analysis (1993-2017).

Figure 13: Regression Coefficients of $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ Over Time as a Function of Maturity (n)

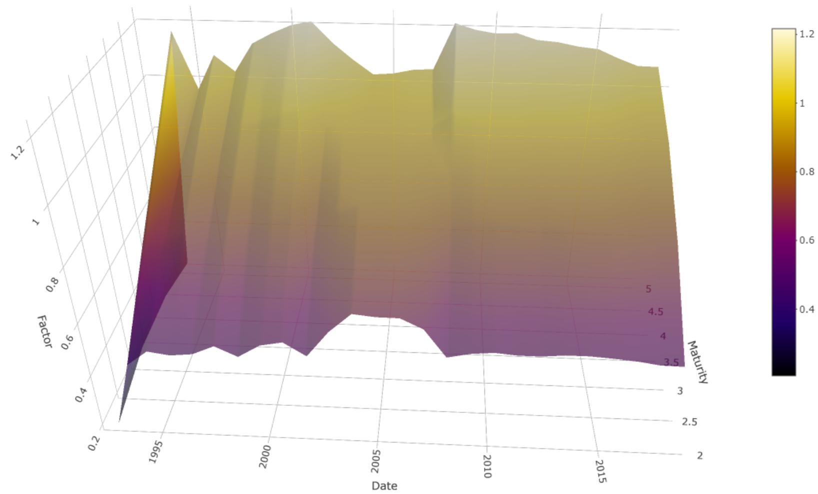

Figure 13 plots the behavior of our unspanned factor $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ as a function of maturity (n) over the period of analysis (1993-2017).

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 $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ 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 $\left(\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ becoming more flat over time.

Similarly, figure 12 plots the regression coefficients of $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$ as a function of Maturity (n), and figure 15 in Appendix A plots the regression coefficients of $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$. The pattern mentioned above maintains for the unspanned factor $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-n),h}$. However, for $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$ 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.

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

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.

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.

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.

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

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.

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2019). Bond return predictability: Economic value and links to the macroeconomy. *Management Science*, 65(2):508–540.

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.

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.

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

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.

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.

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

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.

A Appendix

Figure 14: 12-Month Bonds Excess Returns

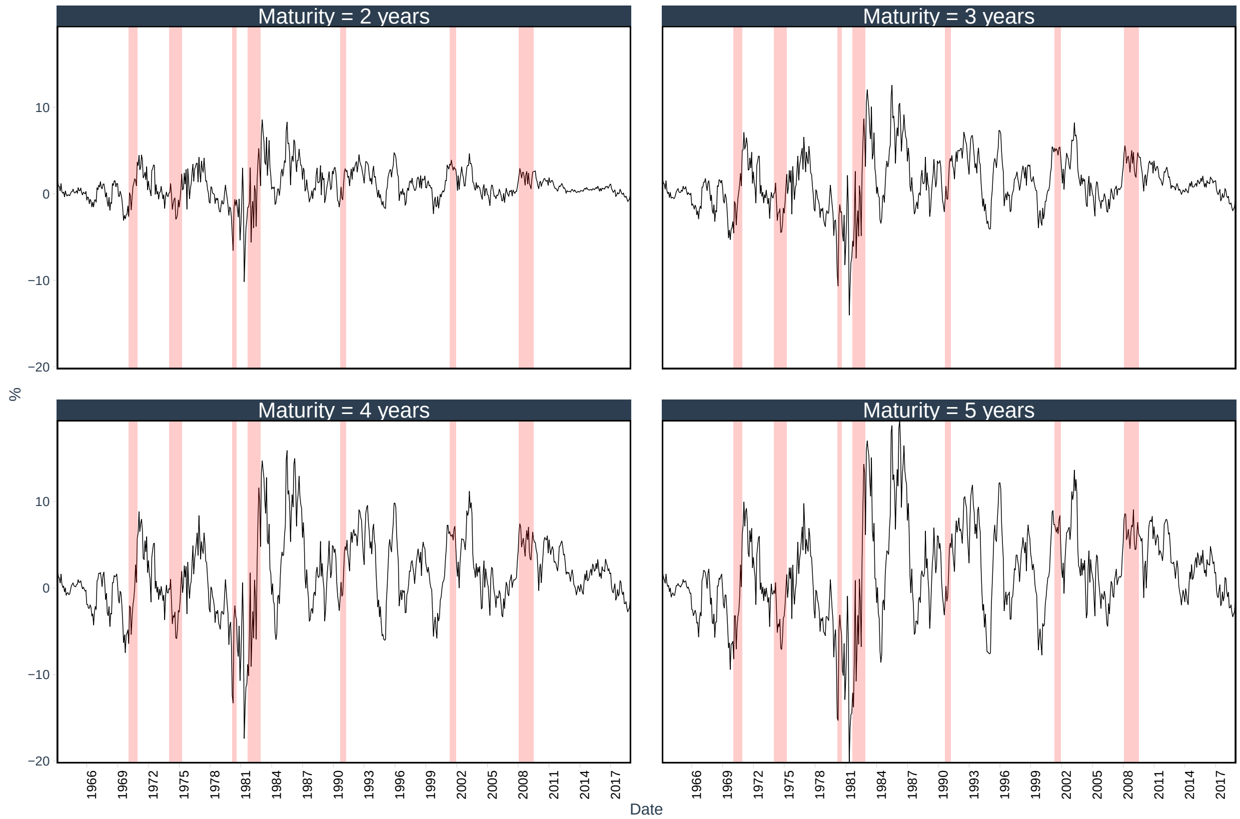

Figure 14 shows the 12-month excess returns for maturities with $n = 2, 3, 4$ and 5 years. The excess returns are calculeted as in equation 4, i.e., $rx^{(n)}_{t+1} = ny^{(n)}_t - (n+1)y^{(n-1)}_{t+1} - y^n_t$. Each panel represents one of the four maturities. The y-axis shows values in percentage (%).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

Figure 15: Regression Coefficients of $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})^{h}_{t+h/12}$ Over Time as a Function of Maturity (n)

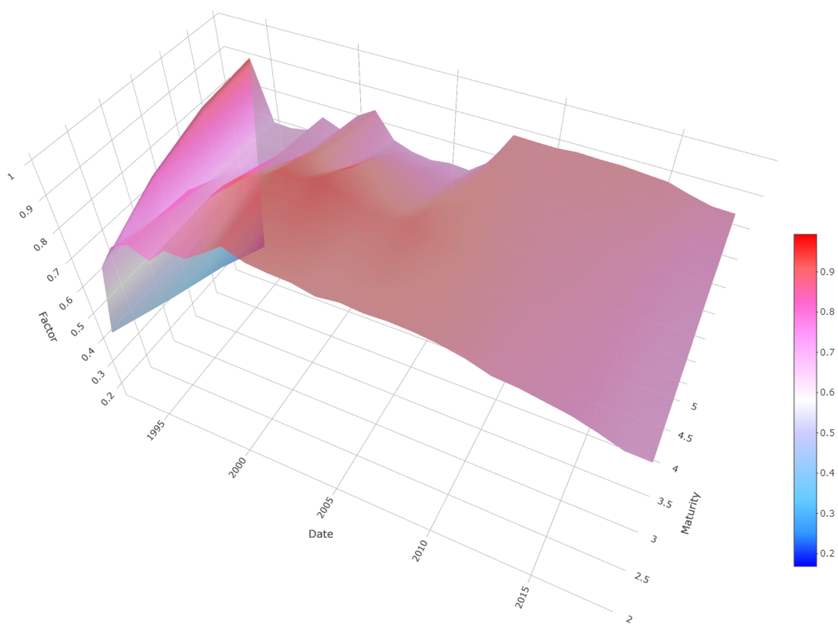

Figure 15 plots the behavior of our unspanned factor $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})^{h}_{t+h/12}$ as a function of maturity (n) over the period of analysis (1993-2017).

Table 6: FRED-MD

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

Table 6 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.