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

Caio Vigo Pereira

The University of Kansas
Department of Economics

March 2020

A Machine Learning Factor-Based Interpretation for the Bond Risk Premia in US

Overview (I)

- I study the time variation of the risk premia in U.S. Treasuries bonds.
- I 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, I derive a single state variable factor that provide a better approximation to the spanned space of all the information from the term-structure.

A Machine Learning Factor-Based Interpretation for the Bond Risk Premia in US

Overview (II)

- I also introduce a way to obtain unspanned risks from the yield curve that is used to complete the state space.

- I show that this parsimonious number of state variables have predictive power for excess returns of bonds over 1-month holding period.

- Additionally, I provide an intuitive interpretation of derived factors, and show what information from macroeconomic variables and sentiment-based measures they can capture.

Introduction

- 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?

- 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.

Central Question

- *What is the linear combination of forecasting variables that captures common movement in expected returns across assets?*

Introduction

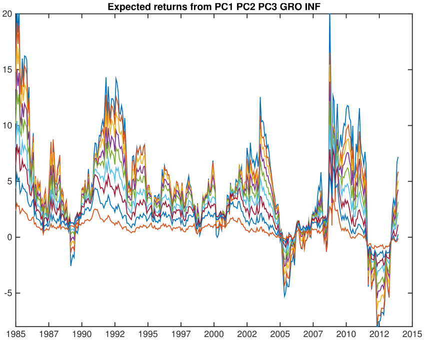


Figure 11: Fitted values of excess return-forecasting regressions for 1-10 year Treasury zeros. The regressions are $rx^{(n)}_{t+1} = a^{(n)} + b^{(n)}[PC1_t\ PC2_t\ PC3_t] + c^{(n)}[GRO_t\ INF_t] + \varepsilon^{(n)}_{t+1}$.

Source: Cochrane (2015)

- There is a clear one-factor structure of expected returns.

- The expected returns on all maturities move in lockstep, with longer maturities moving more than shorter maturities.

Introduction

A Different Route

It is possible that this search for deriving, building and estimating factors that represent state variables in macro-finance models may be limited.

- The process done by financial economists of manually discovering and hand picking this list of factors may be leaving unseen relationships between the state variables out in their derivation.

- To do so, I make use of one of the most powerful approaches in machine learning: **deep neural network** to uncover relationships in the full set of information from the yield curve.

Notation

Log yields

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

where,

$y_t^{(n)}$ denote the log yield of a n-maturity bond at time t

$p_t^{(n)}$ denote the natural logarithm price of this bond

- **holding period returns**

$$\begin{aligned} r_{t+\Delta}^{(n)} &\equiv p_{t+\Delta}^{(n-\Delta)} - p_t^{(n)} \\ r_{t+h/12}^{(n)} &\equiv p_{t+h/12}^{(n-h/12)} - p_t^{(n)} = n y_t^{(n)} - (n - h/12) y_{t+h/12}^{(n-h/12)} \end{aligned} \tag{2}$$

Notation

- **Excess Returns**

$$
\begin{aligned}
rx_{t+h/12}^{(n)} &\equiv \text{holding period return } r_{t+h/12}^{(n)} \quad - \quad \text{1-period yield} \\
&= ny_t^{(n)} - (n - h/12)y_{t+h/12}^{(n-h/12)} - (h/12)y_t^{(h/12)}
\end{aligned} \tag{3}
$$

- **Forward rates** at time t for loans between time $t + n - h/12$ and $t + n$ as

$$
\begin{aligned}
f_t^{(n)} &\equiv p_t^{(n-h/12)} - \quad p_t^{(n)} \\
&= ny_t^{(n)} - (n - h/12)y_t^{(n-h/12)}
\end{aligned} \tag{4}
$$

Expectation Hypothesis

Expectation Hypothesis

- Yields of long maturity bonds should be the average of the future expected yield of short maturity bonds.

$$y_t^{(n)} \equiv \underbrace{\frac{1}{n}\mathbb{E}_t\left(y_t^{(1/12)} + y_{t+1/12}^{(1/12)} + \ldots + y_{t+n-1/12}^{(1/12)}\right)}_{\text{expectations component}} + \quad \text{yield risk premium} \quad . \tag{5}$$

Expectation Hypothesis

Risk Premium: difference between a long rate and the expected average of future short rates.

$$y_t^{(n)} \equiv \underbrace{\frac{1}{n}\mathbb{E}_t\left(y_t^{(1/12)} + y_{t+1/12}^{(1/12)} + \ldots + y_{t+n-1/12}^{(1/12)}\right)}_{\text{expectations component}} + \underbrace{\frac{1}{n}\mathbb{E}_t\left(rx_{t+1/12}^{(n)} + rx_{t+2/12}^{(n-1/12)} + rx_{t+3/12}^{(n-2/12)} + \ldots + + rx_{t+n-1/12}^{(2/12)}\right)}_{\text{yield risk premium}} \tag{6}$$

Expectation Hypothesis

Assuming that the agents' information set at time t can be summarized by a state vector $\mathbf{Z}_t$

$$y_t^{(n)} = \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1} \mathbb{E}\left[y_{t+j\cdot h/12}^{(h/12)} | \mathbf{Z}_t\right]\right) \quad + \quad \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1} \left[rx_{t+h/12(j+1)}^{(n-j\cdot h/12)} | \mathbf{Z}_t\right]\right) \quad . \tag{7}$$

$\mathbf{Z}_t$ should contain all the information used by investors to forecast at time t the excess-returns for all future periods.

Expectation Hypothesis

Literature: gathered evidence against the expectations hypothesis ($rx_{t+h/12^n}$ is forecastable).

Most influential papers:

- Fama and Bliss (1987) Details
- Cochrane and Piazzesi (2005) Details

Spanning Hypothesis

Spanning Hypothesis

- All relevant information to forecast yields and excess returns can be found on the term-structure.
- The yields curve fully spans all necessary information, and thus, no other variable already present in the term-structure should be necessary.
- It does not rules out the importance of macro variables (current or future).
- $\boxed{\textbf{Yield curve completely reflects and spans this information.}}$
- Most influential work against it: Ludvigson and Ng (2009) Details

A Deep-Learning Structure for Bond Premia

We can summarize previous approaches with the following predictive regression:

$$rx_{t+h/12}^{(n)} = \beta^{\top} \boldsymbol{Z}_t + \epsilon_{t+h/12} \tag{8}$$

where $\boldsymbol{Z}_t = \left\{\boldsymbol{Z}_t^{\mathrm{y}}, \boldsymbol{Z}_t^{\mathrm{y}^{\complement}}\right\}$ is the set of state variables, being:

$\boldsymbol{Z}_t^{\mathrm{y}}$ contains only yield curve variables

$\boldsymbol{Z}_t^{\mathrm{y}^{\complement}}$ contains any other variable (complement), e.g., macro and sentiment based variables

- **Spanning hypothesis** $\Rightarrow \boldsymbol{Z}_t = \{\boldsymbol{Z}_t^{\mathrm{y}}\}$ (only yield curve information).
- Evidence against the **spanning hypothesis** $\Rightarrow \boldsymbol{Z}_t^{\mathrm{y}^{\complement}} \neq \emptyset$.

A Deep-Learning Structure for Bond Premia

- Deep neural networks (DNN) attempt to replicate the brain architecture in a computer: many levels of processing information.

- **Main idea:** to extract complex non-linear combinations of the input data by conditioning on both the target (here, $rx^{(n)}_{t+h/12}$) and the inputs (here, $\boldsymbol{Z}_t$). $\boxed{\text{Supervised Learning}}$

- General structure:
  (1) Input layer
  (2) Several hidden layers of **nodes** with predetermined non-linear functions (ϕ)
  (3) Output layer

A Deep-Learning Structure for Bond Premia

- DNN defines a mapping such as $rx^{(n)}_{t+h/12} = g\left(\boldsymbol{Z}_t, \boldsymbol{\theta}_t\right)$ to *learn* the parameter $\boldsymbol{\theta}_t$ that provides the best function approximation.

- Represented in a direct acyclic graph with a chain of functions $g\left(\boldsymbol{Z}_t\right) = g^{(L)}\left(\ldots\left(g^{(2)}\left(g^{(1)}\left(\boldsymbol{Z}_t\right)\right)\right)\right)$.

- The final model can be represented as $rx^{(n)}_{t+h/12} = g\left(\boldsymbol{Z}_t, \boldsymbol{\theta}_t, \boldsymbol{w}\right) = \phi\left(\boldsymbol{Z}_t, \boldsymbol{\theta}_t\right)^\top \boldsymbol{w}$.

- **Notations:**
    - *network:* comes from this chain and its interconnectedness architecture
    - *depth* (L): number of layers ("deep learning")
    - *first layer:* $g^{(1)}$
    - *output layer:* last layer $g^{(L)}(\cdot)$

A Deep-Learning Structure for Bond Premia

Universal Approximation Theorem (Hornik et al., 1989; Cybenko, 1989)

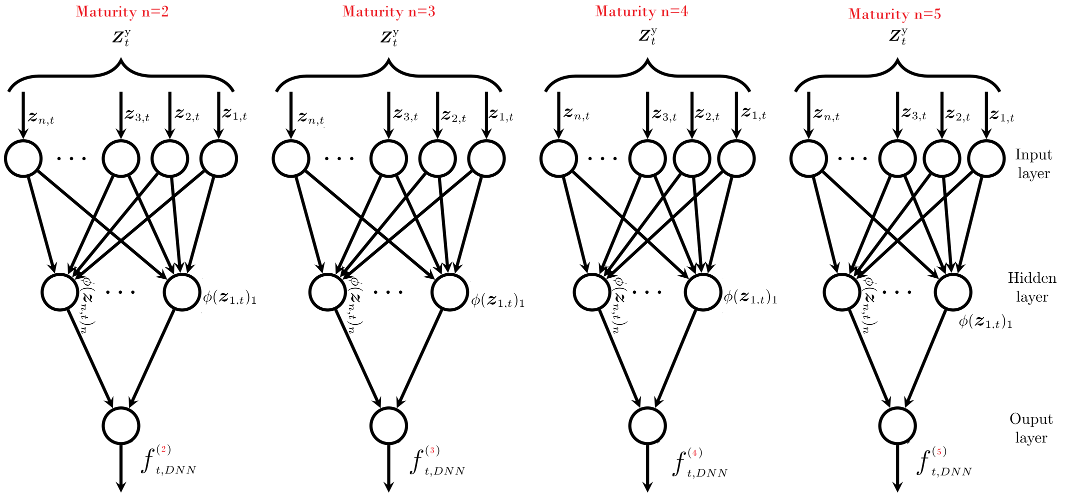
- Feedforward network with a linear output layer and **at least one hidden layer** with any activation function can approximate **any function**[1] from one finite-dimensional space to another with any desired nonzero amount of error.

- In short, 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.

- **Implication:** there exists a network large enough to achieve any degree of accuracy.

A Deep-Learning Structure for Bond Premia



A Deep-Learning Structure for Bond Premia

DNN Factors

$$
\begin{aligned}
\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} &= \tau_0 + \tau_1 \mathfrak{f}_{t,DNN}^{(2),h} + \tau_2 \mathfrak{f}_{t,DNN}^{(3),h} + \tau_3 \mathfrak{f}_{t,DNN}^{(4),h} + \tau_4 \mathfrak{f}_{t,DNN}^{(5),h} + \bar{\epsilon}_{t+h/12} \\
&= \boldsymbol{\tau}^{\top} \widehat{\boldsymbol{\mathfrak{F}}}_t^{h} + \bar{\epsilon}_{t+h/12}
\end{aligned}
\tag{9}
$$

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 $\boldsymbol{\tau} \equiv \begin{bmatrix} \tau_0 & \tau_1 & \tau_2 & \tau_3 & \tau_4 \end{bmatrix}^{\top}$.

A Deep-Learning Structure for Bond Premia

Unspanned Factor

- We orthogonalize the excess returns by the deep neural network factor $\mathfrak{f}_{t,DNN}^{(n)}$, and denote it by $\xi_t^{(n),h}$ as

$$\xi_{t+h/12}^{(n),h} = rx_{t+h/12}^{(n)} - \beta_0 - \beta_1 \mathfrak{f}_{t,DNN}^{(n),h} \quad . \tag{10}$$

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

An Illustrative Term-Structure Model

The no-arbitrage assumption rely on the fundamental asset pricing equation:

$$P_t^{(n)} = \mathbb{E}_t \left(\mathcal{M}_{t+1} P_{t+1}^{(n-1)} \right) \tag{11}$$

where

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

SDF:

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

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 $\epsilon_{t+h/12}$.

An Illustrative Term-Structure Model

- Define $\boldsymbol{Z}_t = \left\{ \boldsymbol{Z}_t^{\mathrm{y}}, \boldsymbol{Z}_t^{\mathrm{y}^{\complement}} \right\}$

- Dynamics of $\boldsymbol{Z}_t$ that capture all the risks of the economy following a Gaussian VAR process given by:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{Z}_t^{\mathrm{y}} \\ \boldsymbol{Z}_t^{\mathrm{y}^{\complement}} \end{bmatrix} &= \boldsymbol{\mu} + \boldsymbol{\Phi} \begin{bmatrix} \boldsymbol{Z}_{t-1}^{\mathrm{y}} \\ \boldsymbol{Z}_{t-1}^{\mathrm{y}^{\complement}} \end{bmatrix} + \boldsymbol{\Sigma}\epsilon_t \\ \boldsymbol{Z}_t &= \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{Z}_{t-1} + \boldsymbol{\Sigma}\epsilon_t \qquad \epsilon_t \sim N(0, \boldsymbol{I}) \end{aligned} \tag{13}$$

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.

An Illustrative Term-Structure Model

- In a similar fashion to Joslin et al. (2014), we can write:

$$\boldsymbol{Z}_t^{\mathrm{y}^{\complement}} = \gamma_0 + \gamma_1 \boldsymbol{Z}_t^{\mathrm{y}} + \boldsymbol{M}_{\boldsymbol{Z}_t^{\mathrm{y}}} \boldsymbol{Z}_t^{\mathrm{y}^{\complement}} \tag{14}$$

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.,

$$\boldsymbol{M}_{\boldsymbol{Z}_t^{\mathrm{y}}} \boldsymbol{Z}_t^{\mathrm{y}^{\complement}} \equiv \boldsymbol{Z}_t^{\mathrm{y}^{\complement}} - \mathrm{Proj}\left[\boldsymbol{Z}_t^{\mathrm{y}^{\complement}} | \boldsymbol{Z}_t^{\mathrm{y}}\right] \tag{15}$$

- Previous models have assumed that the $\boldsymbol{Z}_t^{\mathrm{y}^{\complement}}$ was spanned by $\boldsymbol{Z}_t^{\mathrm{y}}$, thus imposing the restriction:

$$\boldsymbol{Z}_t^{\mathrm{y}^{\complement}} = \mathrm{Proj}\left[\boldsymbol{Z}_t^{\mathrm{y}^{\complement}} | \boldsymbol{Z}_t^{\mathrm{y}}\right] \tag{16}$$

An Illustrative Term-Structure Model

In our methodology,

- $\boldsymbol{Z}_t$ is given by the derived factor $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$
- $\boldsymbol{Z}_t^{\mathrm{y}^{\mathsf{C}}}$ by a function of $\xi^h_{t+h/12}$ as $f(\xi^h_{t+h/12})$
- Analogous to Joslin et al. (2014), we argue $f(\xi^h_{t+h/12})$ complete and fill the unspanned factor in the state space.

Linear Rotation of the State Space

$\left[\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h, f(\xi^h_{t+h/12})\right]$ and $\boldsymbol{Z}_t$ represent linear rotations of the same full list factors.

A Deep-Learning Structure for Bond Premia

- **Intuition:** lies in the fact that at each t, $\widehat{\xi}^{(n)}_{t+h/12}$ is orthogonal to $\mathfrak{f}^{(n),h}_{t,DNN}$.

- 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 $\boldsymbol{Z}^{y}_{t}$, are unspanned and should be in an **orthogonal space**.

- Hence, the unspanned information in $\widehat{\boldsymbol{\xi}}^{h}_{t+h/12}$ could be capturing $\boxed{\text{macroeconomic information}}$ or $\boxed{\text{sentiment measures}}$ not spanned by the term-structure.

A Deep-Learning Structure for Bond Premia

- Consistent with our adapted dynamic term-structure model, the orthogonal vector from Proj $\left[f(\xi_{t+h/12}^{(n)}) | \boldsymbol{Z}_t^{y} \right]$ has predictive power for excess returns.

Alternative (1)

$\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_t^h \equiv$ projection of $\overline{rx}_{t+h/12}$ in $\widehat{\boldsymbol{\xi}}_{t+h/12}$

Yielding: $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$

Alternative (2)

- Similar projection
- For each maturity $n \in \{2, 3, 4, 5\}$ we regress $rx_{t+h/12}^{(n)}$ on $\widehat{\boldsymbol{\xi}}_{t+h/12}^{(-n),h} \equiv \widehat{\boldsymbol{\xi}}_{t+h/12}^{h} \setminus \widehat{\xi}_{t+h/12}^{(n),h}$

Yielding: $\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\boldsymbol{\mathfrak{F}}}}\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)_{t+h/12}^{(-n),h}$

Data & Strategy

- 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).

$$
\begin{aligned}
y_t^{(n)} = \beta_{0,t} \quad & +\beta_{1,t}\left(\frac{1-\exp\left(-n/\tau_1\right)}{n/\tau_1}\right) \\
& +\beta_{2,t}\left(\frac{1-\exp\left(-n/\tau_1\right)}{n/\tau_1} - \exp\left(-n/\tau_1\right)\right) \\
& +\beta_{3,t}\left(\frac{1-\exp\left(-n/\tau_2\right)}{n/\tau_2} - \exp\left(-n/\tau_2\right)\right)
\end{aligned}
\tag{17}
$$

- Full period of data ranges from 1962:01 to 2017:12.

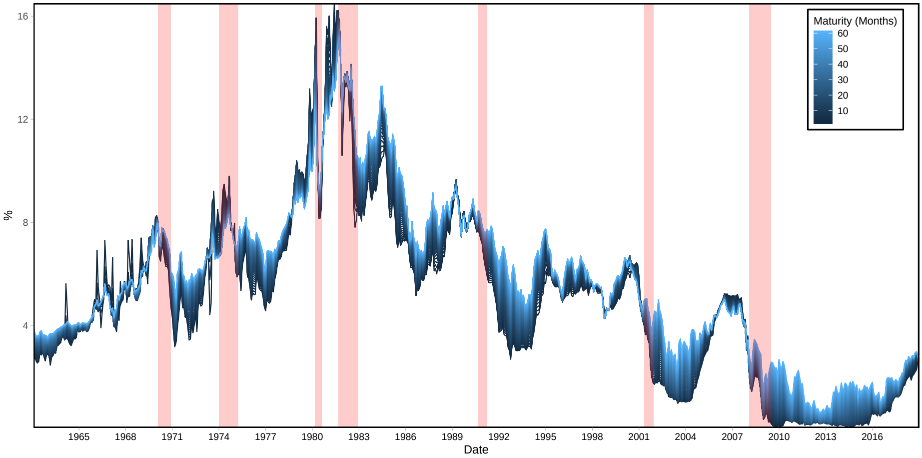
- We use these estimated parameters from the last day of each month to construct a monthly derived zero-coupon bonds log yields with maturities up to 60 months from each t.

Data & Strategy

Derived zero-coupon bonds log yields for maturities (n) up to 60 months

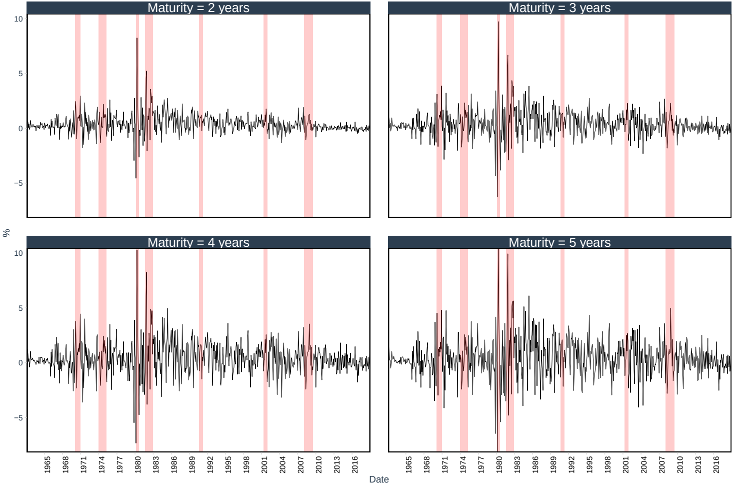
Data & Strategy

Derived zero-coupon bonds log yields for maturities (n) up to 60 months



Data & Strategy

1-Month Bonds Excess Returns



Empirical Strategy

- Period of evaluation from 1993:01 to 2017:12.
- Feed our deep neural network with 3 different sets of information from the **term-structure**:

$$Z_t^y = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)} \right\}$$
$$Z_t^y = \left\{ y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}$$
$$Z_t^y = \left\{ f_t^{(2/12)}, f_t^{(3/12)}, \ldots, f_t^{(60/12)}, y_t^{(1/12)}, y_t^{(2/12)}, \ldots, y_t^{(60/12)} \right\}.$$

Empirical Strategy

Activation Function: rectified linear activation function (ReLU)

$$ReLU(x) = \begin{cases} 0 & \text{, if } x \leq 0 \\ x & \text{, otherwise} \end{cases} . \tag{18}$$

- Default and recommended for use with the majority of feedforward neural networks (Goodfellow et al., 2016).

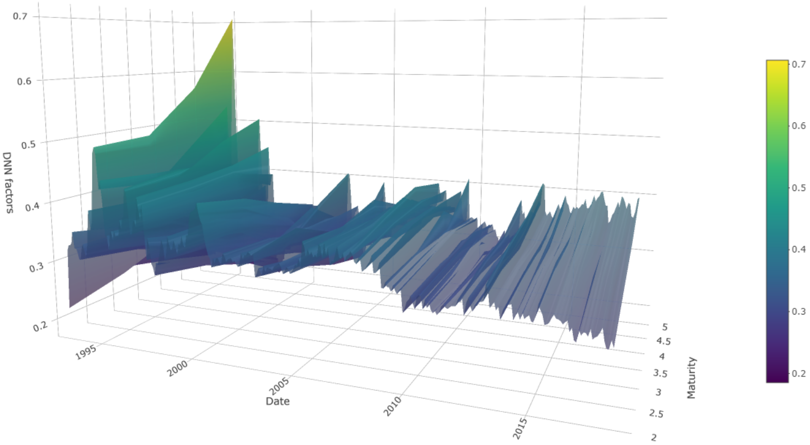
- Yields a nonlinear transformation.

- ReLU units are nearly linear.

- **Additional advantages:** Retains 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.

Empirical Results

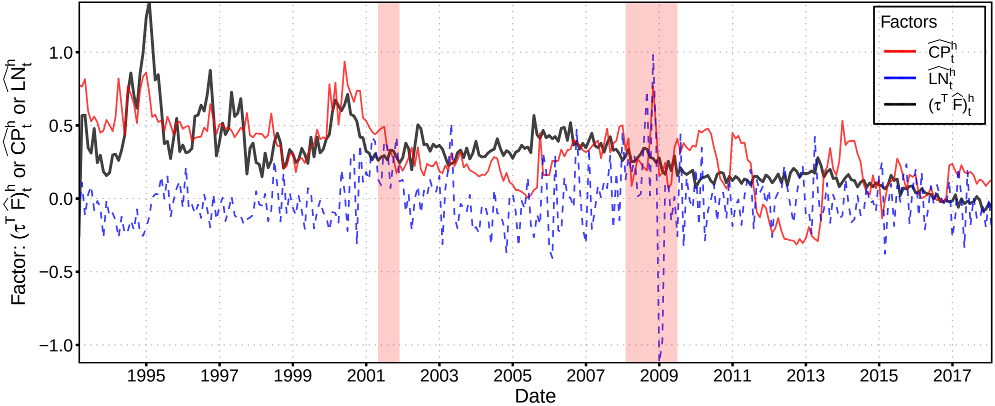
Derived Factors $\mathfrak{f}_{t,DNN}^{(n),h}$ for **DNN 2** Generated Using the Set of Yields



Empirical Results

Comparison with Other Factors from the Literature

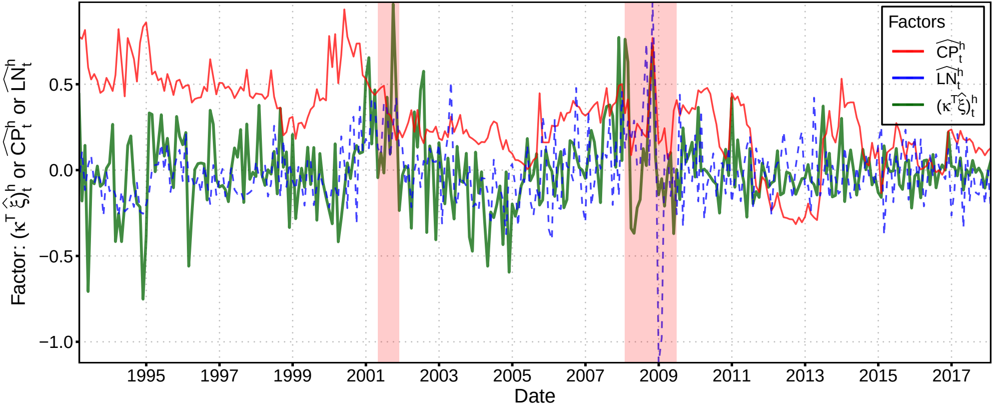
Figure 1: Time Series of our Derived Factor $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$, along with $\widehat{CP}_t^h$ and $\widehat{LN}_t^h$



Empirical Results

Comparison with Other Factors from the Literature

Figure 2: Time Series of our Derived Factor $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^h$, along with $\widehat{CP}_t^h$ and $\widehat{LN}_t^h$



Empirical Results - 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

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

	DNN 1			DNN 2			DNN 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}})_t^h$	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

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Empirical Results - 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

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}})_t^h$	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

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

Empirical Results - 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

Panel C:

	$rx^{(4)}_{t+h/12}$								
$(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}})_t^h$	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

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

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

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

$(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-5),h}$		0.854**			0.848***			1.069***	
		(0.336)			(0.318)			(0.339)	
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{h}$			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: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Empirical Results

Correlation Matrix

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

Empirical Results - Predictive Regressions with $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$ 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: $rx_{t+h/12}^{(2)}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_{t+h/12}^{(-2),h}$		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

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Empirical Results - Predictive Regressions with $\left(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}_t\right)_t^h$ 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 B: $rx_{t+h/12}^{(3)}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^\top\widehat{\mathfrak{F}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-3),h}$		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: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Empirical Results - Predictive Regressions with $\left(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}_t\right)_t^h$ 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 C: $rx_{t+h/12}^{(4)}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_{t+h/12}^{(-4),h}$		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

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

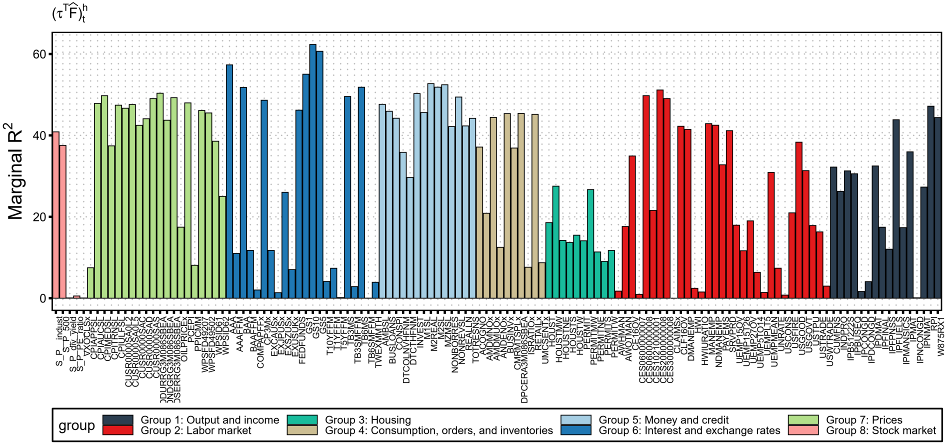
Empirical Results - Predictive Regressions with $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$ and $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$, along with the Cochrane-Piazzesi and Ludvingson-Ng factors, and Fama-Bliss Regressions with Forward Spreads

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}})_t^h$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-5),h}_{t+h/12}$		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

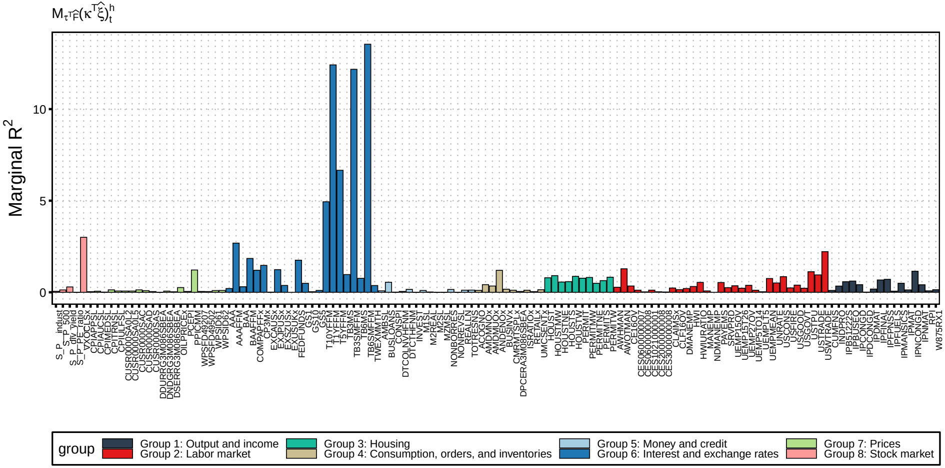
Marginal R^2 of the factor $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$



Empirical Results - Economic Interpretation

Marginal R^2 of the factor $\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_{t+h/12}$

Additional Results



Out-of-Sample Forecasting Performance

- 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.

- At each $\tau \in \tau_{OoS}$, we use all the previous information up to $\tau - 1$ to obtain the point forecast of $rx^{(n)}$ for the month τ.

Out-of-Sample R^2 (Campbell and Thompson, 2007; Gargano et al., 2019)

The out-of-sample R^2 is computed as

$$R^{2(n)}_{OoS,i} = 1 - \frac{\sum_{\tau \in \tau_{OoS}} \left(rx^{(n)}_{t+h/12|t} - \widehat{rx}^{(n)}_{t+h/12|t} \right)^2}{\sum_{\tau \in \tau_{OoS}} \left(rx^{(n)}_{t+h/12|t} - \overline{rx}^{(n)}_{t+h/12|t} \right)^2} \tag{19}$$

Out-of-Sample Forecasting Performance (R^2)

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

Conclusion

- I proposed a novel approach for deriving a **single state factor** consistent with a dynamic term-structure with unspanned risks.

- Making use of **deep neural networks** to uncover relationships in the term-structure, I build 2 state factors that provides a good approximation to the space that spans all the information from the term-structure.

- I also introduced a way to obtain unspanned risks from the yield curve that is used to complete the state space.

- I show that this parsimonious number of state variables have predictive power for excess returns of bonds over 1-month holding period.

- Additionally, I provide an intuitive interpretation of derived factors, and show what information from macroeconomic variables and sentiment-based measures they can capture.

Next Steps

- Use convolutional Neural Networks
- Retrain DNN $\rightarrow$ 1992:01 - 1972:01 and on
- Test of Exogeneity with GMM.

References I

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.

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.

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.

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.

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.

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

Fama and Bliss (1987)

Return

- Fama and Bliss (1987) builds forward rates spreads and use these variables as covariates.
- Forward rate spread between of a n-year maturity bond: $fs_t^{(n,h)} \equiv f_t^{(n)} - \quad y_t^{(h/12)}(h/12)$.

Predictive Regression

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \epsilon_{t+h/12} \quad . \tag{20}$$

Cochrane and Piazzesi (2005)

Return

- Cochrane and Piazzesi (2005) derive a single factor to use as predictor (CP_t^h).

- First, they estimate (CP_t^h) as

$$\begin{aligned} \frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} &= \gamma_0 + \gamma_1 f_t^{(1)} + \gamma_2 f_t^{(2)} + \gamma_3 f_t^{(3)} + \gamma_4 f_t^{(4)} + \gamma_5 f_t^{(5)} + \bar{\epsilon}_{t+h/12} \\ \overline{rx}_{t+h/12} &= \underbrace{\boldsymbol{\gamma}^\top \mathbf{f}_t}_{CP_t^h} + \bar{\epsilon}_{t+h/12} \end{aligned} \tag{21}$$

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 $\boldsymbol{\gamma} \equiv [\gamma_0 \quad \gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4 \quad \gamma_5]^\top$.

Predictive Regression

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{CP}_t^h + \epsilon_{t+h/12} \quad . \tag{22}$$

Ludvigson and Ng (2009)

Return

- Ludvigson and Ng (2009) use a large panel of macro variables, and build a single linear combination (LN_t^h) out of the first i estimated principal components ($\hat{g}_{i,t}$).

- First, they estimate (LN_t^h) as

$$\begin{aligned} \frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} &= \lambda_0 + \lambda_1\hat{g}_{1,t} + \lambda_2\hat{g}_{1,t}^3 + \lambda_3\hat{g}_{3,t} + \lambda_4\hat{g}_{4,t} + \lambda_5\hat{g}_{8,t} + \bar{\epsilon}_{t+h/12} \\ \overline{rx}_{t+h/12} &= \underbrace{\boldsymbol{\lambda}^\top \widehat{\boldsymbol{G}}_t}_{LN_t^h} + \bar{\epsilon}_{t+h/12} \end{aligned} \tag{23}$$

where $\widehat{\boldsymbol{G}}_t$ and $\boldsymbol{\lambda}$ are 5×1 vectors given by $\widehat{\boldsymbol{G}}_t \equiv \left[\hat{g}_{1,t} \quad \hat{g}_{1,t}^3 \quad \hat{g}_{3,t} \quad \hat{g}_{5,t} \quad \hat{g}_{8,t}\right]^\top$, and $\boldsymbol{\lambda} \equiv [\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5]^\top$.

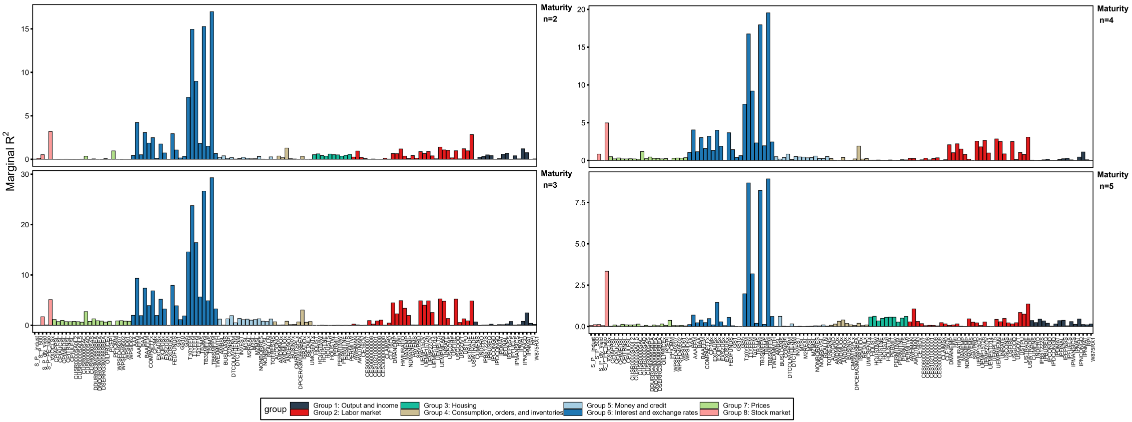
Predictive Regression

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{LN}_t^h + \epsilon_{t+h/12} \quad . \tag{24}$$

Empirical Results - Economic Interpretation

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

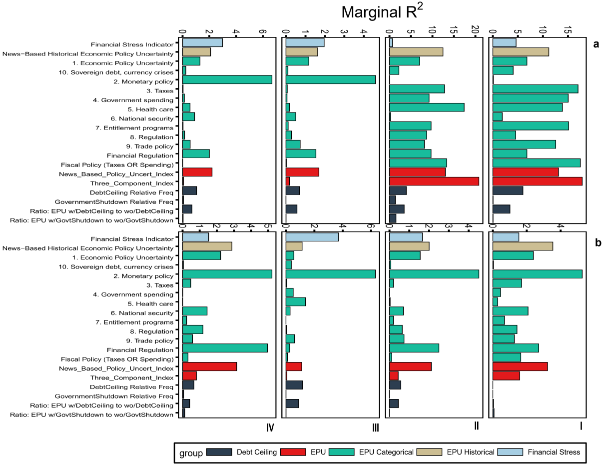
Return



Empirical Results - Economic Interpretation

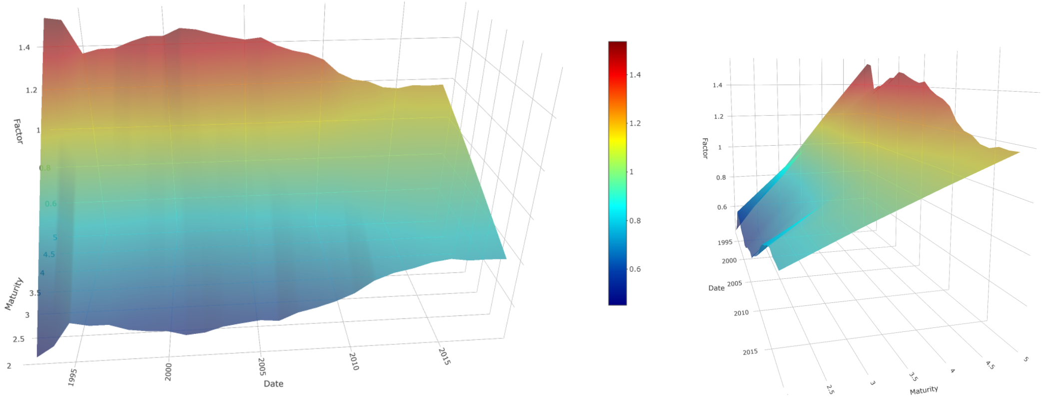
Marginal R^2 Using Sentiment-Based Measures

Return



Empirical Results

Regression Coefficients of $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$ Over Time as a Function of Maturity (n)



Empirical Results

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