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

Caio Vigo Pereira

April 22, 2021

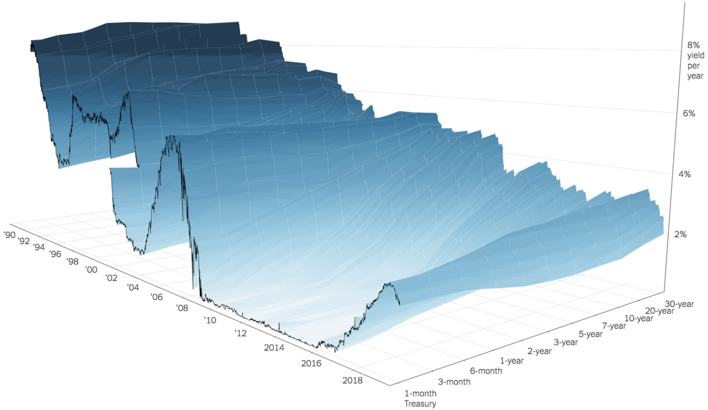
Motivation

- I study the time variation of the **risk premia in U.S. Treasuries bonds**.

- Treasury bonds play an important role in financial markets $\Rightarrow$ its risk and return dynamics is of **central economic importance**.
  - ↳ major importance for monetary policy
  - ↳ strategic in investors' portfolios
  - ↳ understanding of financial events: e.g., zero rates in 2008, 2020

- Understanding pricing of U.S. Treasuries is a central question in the study of bond markets.
  - ↳ The U.S. Treasury market is the largest government debt market in the world with an estimated value of **$14 trillion (2019)**.
  - ↳ $\approx$ **30%** of the entire U.S. bond market (corporate debt + mortgage and municipal bonds + money market instruments + asset-backed securities)

Treasuries Yields



risk premia: *difference between the current long rate and the expected average of future short rates.*

Motivation

- **Bond Premia**
  ↳ Long literature. Back from Fama and Bliss (1987)
  ↳ Nonetheless, never fully answered/understood
  ↳ Many **factors** were proposed in the literature:
    ↳ Fama and Bliss (1987) → *forward spreads*
    ↳ Cochrane and Piazzesi (2005) → *linear combination of forward rates*
    ↳ Ludvigson and Ng (2009) → *linear combination of macro PC loadings*
    ↳ Cieslak and Povala (2015) and Lee (2018) → *trend inflation*

Bauer and Hamilton (2018)

↳ evidence against the use factors not derived from the yield curve (non spanning) → **"spanning puzzle" literature**

↳ raised methodological issues: econometric problems when overlapping returns is used.

Research Question

- 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.
  ↳ *A **parsimonious** number is key here.*

Central Question

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

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.

Contribution

Methodological/Theory

- I propose a novel approach for deriving a **parsimonious number** of state factor consistent with a **dynamic term-structure with unspanned risks** theoretically motivated model.

- I use **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.

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

Empirical Findings

- I show that this parsimonious number of state variables have predictive power for excess returns of bonds over 1-month holding period (in/out-of-sample).

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

Contribution - **Discussion**

- **First,** through DNNs, we can introduce **nonlinearities** when modeling the bond risk premia in our first step of the recursive process.
  ↳ while still making use of a linear combination of the latent factors in the second step,
  ↳ and generating a parsimonious number of factors (state variables).

  ↳ With neural networks we can introduce flexible and complex nonlinear relationships from the inputs while approximating arbitrarily well a rich set of smooth functions.

  ↳ Consistent with recent findings (e.g., Gu et al. (2018); Bianchi et al. (2019)) → importance of allowing for **nonlinearities**.

  ↳ The approach is at the intersection of bond premia and **sequential learning** as in Gargano et al. (2019) and Dubiel-Teleszynski et al. (2019).

- **Second,** the approach avoids hand-picking the variables from the yield curve
  ↳ as through a DNN we are able to recursively learn the best-approximating function that condenses the yield curve into a single latent factor.

Contribution - **Discussion**

- **Third,** I overcome some the issues raised by Bauer and Hamilton (2018)
  ↳ use of non-overlapping returns, as done by the most recent literature (Gargano et al., 2019)
  ↳ I make use of the term structure at the higher frequency of 1-month holding period with maturities ranging up to 60 months ahead.

- **Fourth,** we start our process with only information from the term structure.

- **Fifth,** we take a broader interpretation of the unspanning factor.
  ↳ we can link with other sources of risks (macroeconomics and sentiment-based variables)

Notation

Details

Log yields

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

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**

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

- **Excess Returns**

$$rx_{t+h/12}^{(n)} \equiv \text{holding period return } r_{t+h/12}^{(n)} - \text{1-period yield}$$

Spanning Hypothesis

- **SH is a central issue in macro-finance models** (Gürkaynak et al., 2007; Duffee, 2013; Bauer and Hamilton, 2018) EH

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).
- **Yield curve completely reflects and spans this information.**
- Influential works/factors:
  - ↳ Spanning: **Fama and Bliss (1987)** FB Details and **Cochrane and Piazzesi (2005)** CP Details
  - ↳ Not spanning: **Ludvigson and Ng (2009)** LN Details

A Deep-Learning Structure for Bond Premia

Partition of $\boldsymbol{Z}_t$

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

$\boldsymbol{Z}_t^{y}$ contains only yield curve variables [yields, forward rates]
$\boldsymbol{Z}_t^{y^{\complement}}$ contains any other variable (complement) [e.g., macro and sentiment-based variables]

We can summarize previous approaches with the following predictive regression:

$$rx_{t+h/12}^{(n)} = \beta^{\top} \boldsymbol{Z}_t + \epsilon_{t+h/12} \tag{1}$$

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

Artificial Neural Network: A Primer

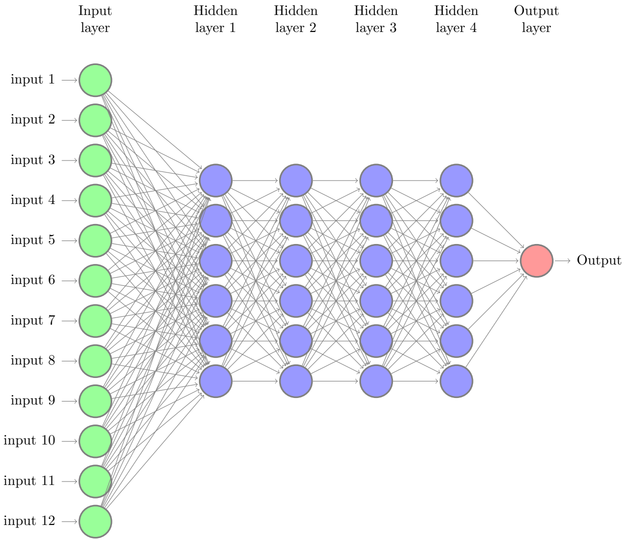
Idea: Attempt to replicate the brain architecture

↳ Many levels of processing information

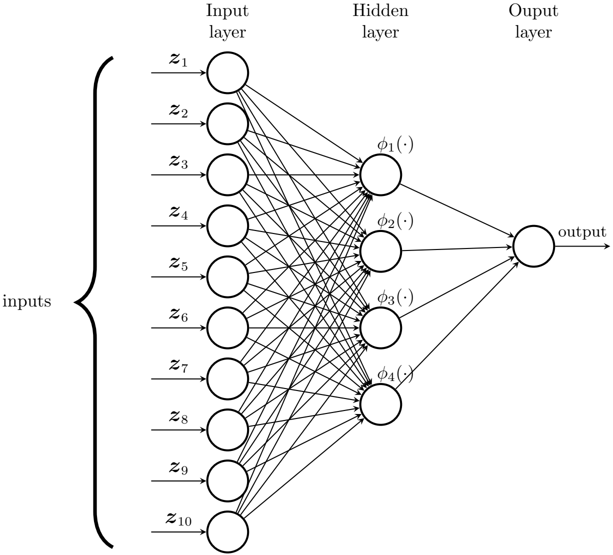
Goal: Extract complex non-linear combinations of the input

↳ Supervised Learning

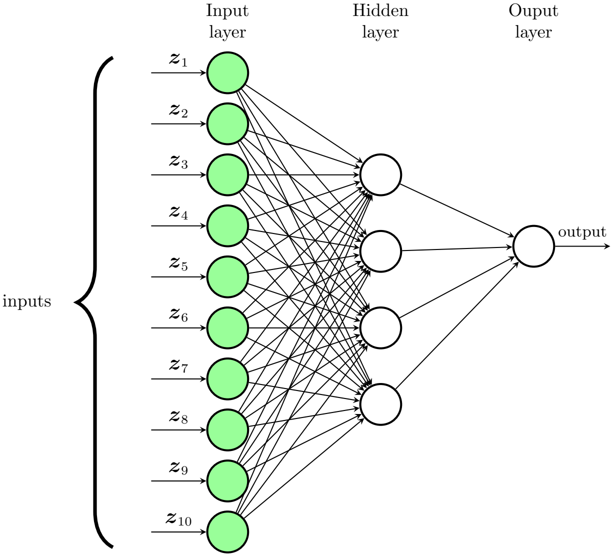
↳ Conditioning on target (here, $rx^{(n)}_{t+h/12}$) and the inputs (here, $\boldsymbol{Z}_t$)



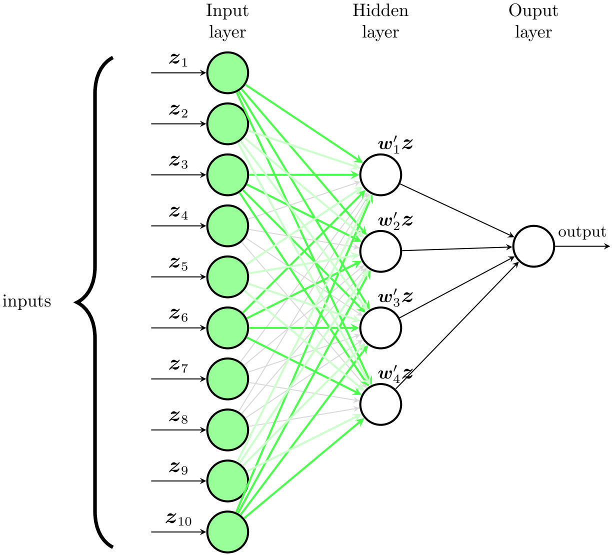
Artificial Neural Network: A Primer



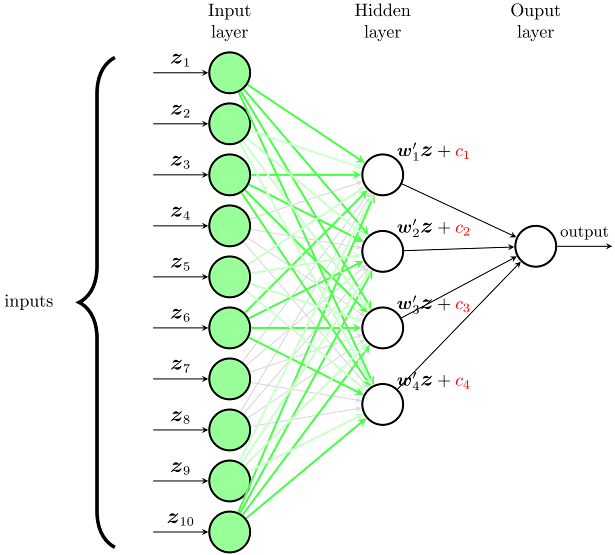
Artificial Neural Network: A Primer



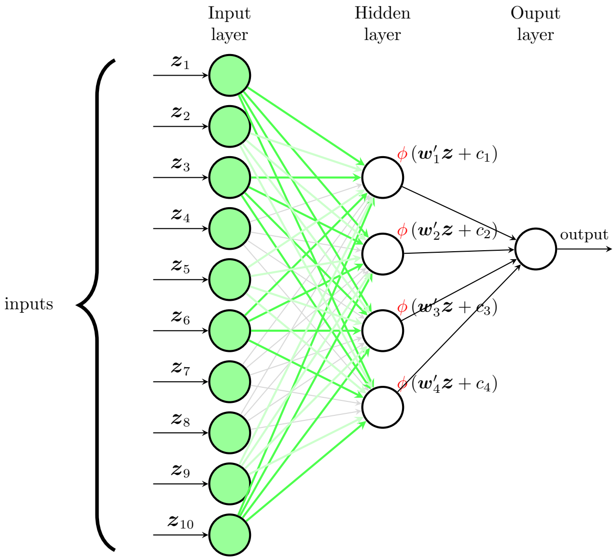
Artificial Neural Network: A Primer



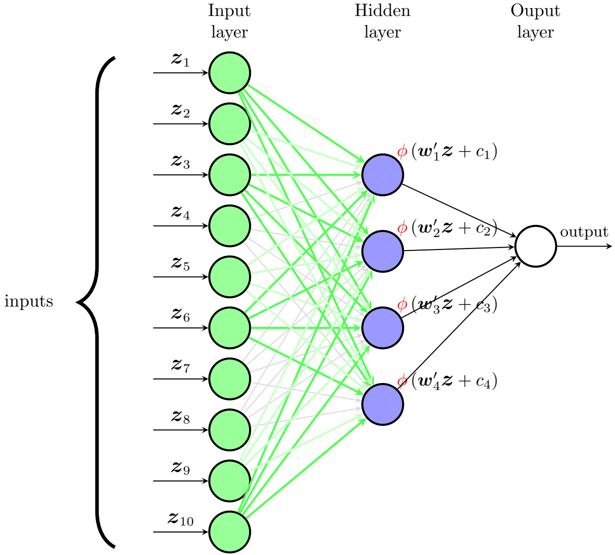
Artificial Neural Network: A Primer



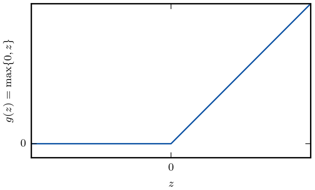
Artificial Neural Network: A Primer

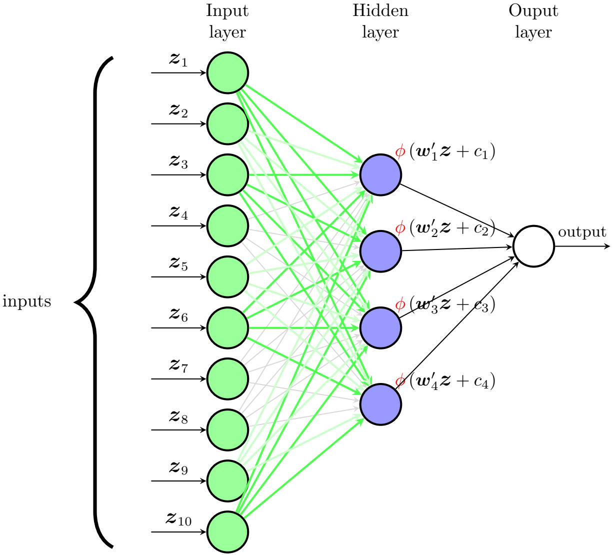


Artificial Neural Network: A Primer

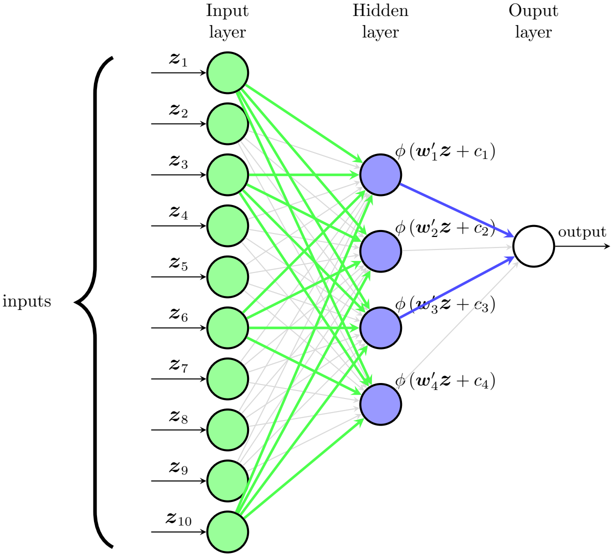


Artificial Neural Network: A Primer

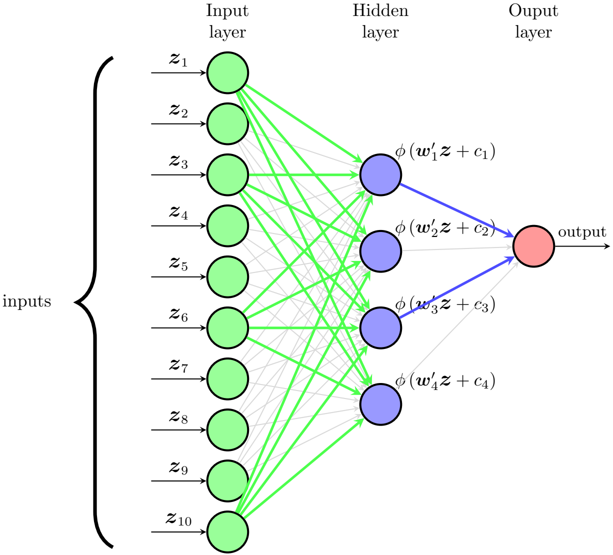




Artificial Neural Network: A Primer



Artificial Neural Network: A Primer



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.

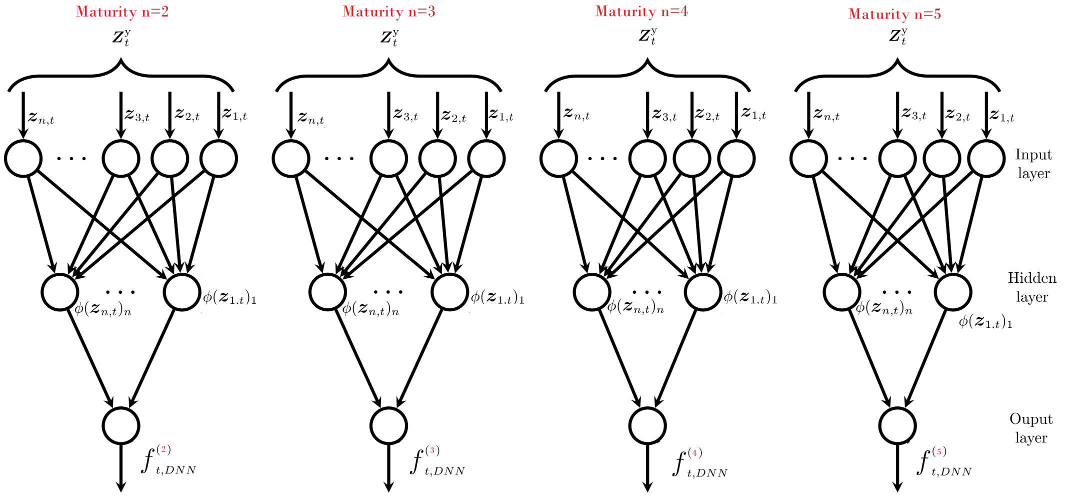
- Represented in a direct acyclic graph with a chain of functions $g\left(\boldsymbol{Z}_t\right) = g^{(L)}\left(\ldots\left(g^{(2)}\left(g^{(1)}\left(\boldsymbol{Z}_t\right)\right)\right)\right)$.

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.

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

A Deep-Learning Structure for Bond Premia



DNN Factors

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

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

- We **recursively** orthogonalize the excess returns generated by the deep neural network factor $\mathfrak{f}_{t,DNN}^{(n)}$, and denote it by $\boldsymbol{\xi}_t^{(n),h}$.

- 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 $\mathfrak{f}_{t,DNN}^{(n)}$.

A Deep-Learning Structure for Bond Premia

Linear Rotation of the State Space

Proposition 2. *As in the dynamic term structure model of Joslin et al. (2014), $f(\xi^h_{t+h/12})$ complete and fill the unspanned factor in the state space, in a such a way that* $\left[\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)^h_t, f(\xi^h_{t+h/12})\right]$ *and* $\boldsymbol{Z}_t$ *represent linear rotations of the same economy-wide risks underlying all tradable assets available to agents in the economy.*

A Deep-Learning Structure for Bond Premia

Linear Rotation of the State Space

Proposition 2. *As in the dynamic term structure model of Joslin et al. (2014), $f(\xi^h_{t+h/12})$ complete and fill the unspanned factor in the state space, in a such a way that [**Spanning Factor**, **Unspanning Factor**] and $\boldsymbol{Z}_t$ represent linear rotations of the same economy-wide risks underlying all tradable assets available to agents in the economy.*

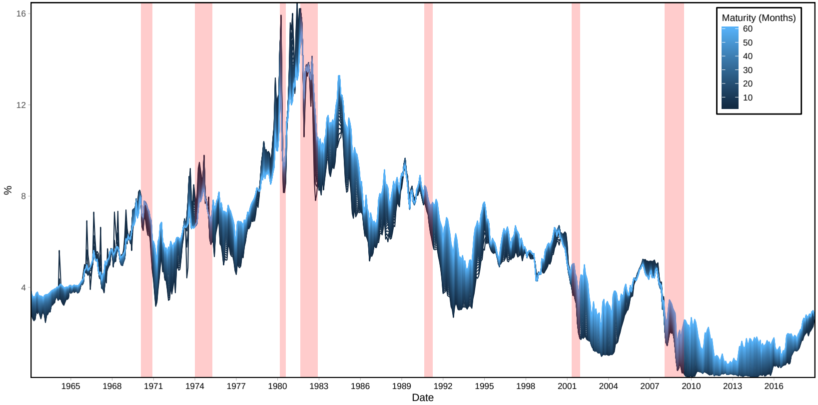
- Analogous to Joslin et al. (2014), we argue

  ↳ that the unspanned information in $\widehat{\xi}^h_{t+h/12}$ could be capturing macroeconomic information or sentiment measures not spanned by the term-structure.

Details

Data & Strategy

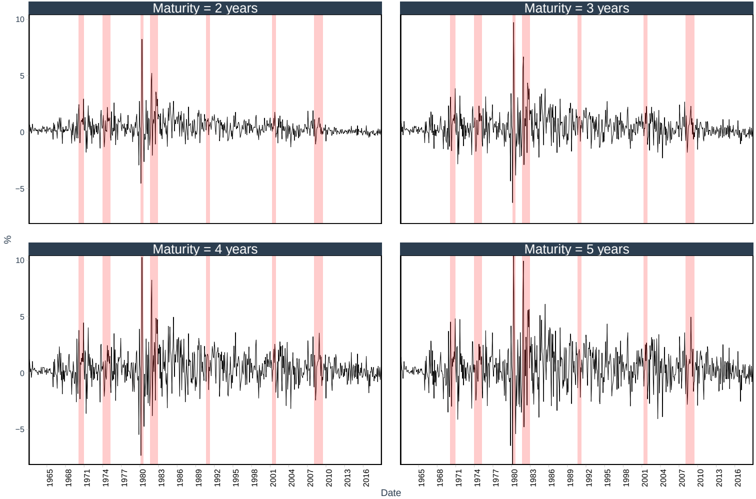
Derived zero-coupon bonds log yields for maturities (n) up to 60 months



- Full period: 1962:01 to 2017:12
- Period of evaluation: 1993:01 to 2017:12

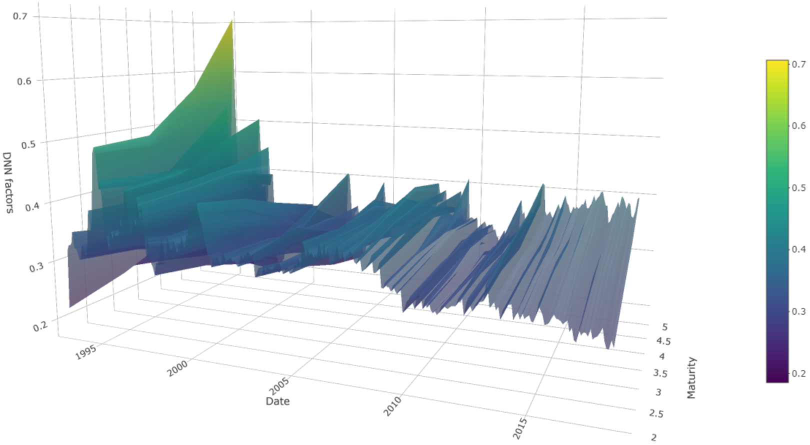
Treasuries Excess Returns

1-Month Bonds Excess Returns (1962-2017)



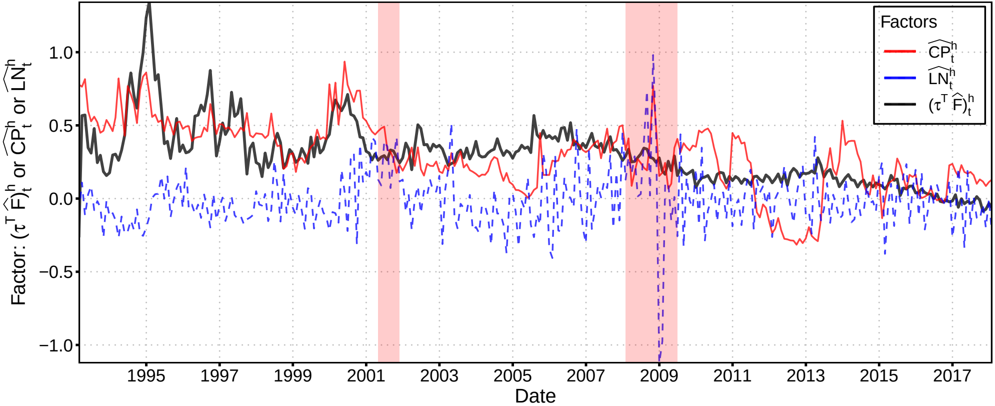
Empirical Results

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



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 $\widehat{LN}_t^h$

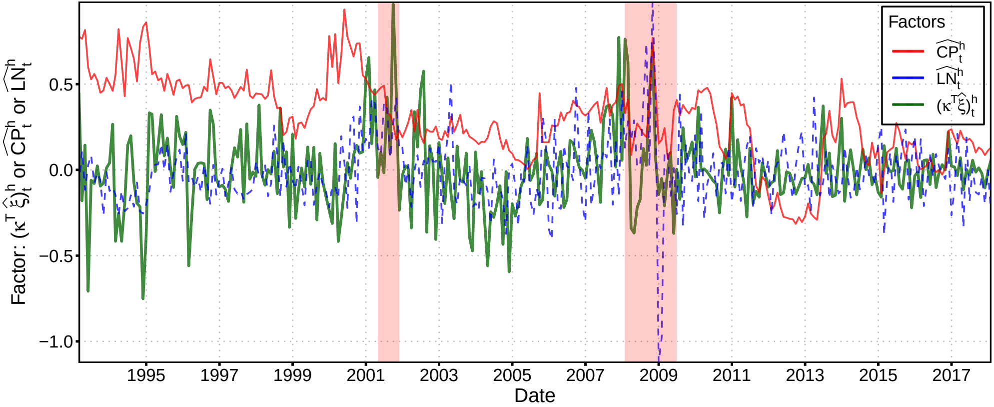


Empirical Results

Comparison with Other Factors from the Literature

Correlation

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



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

Details

	$rx^{(2)}_{t+h/12}$		$rx^{(3)}_{t+h/12}$		$rx^{(4)}_{t+h/12}$		$rx^{(5)}_{t+h/12}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}})_t^h$	0.811^{***}	0.811^{***}	0.943^{***}	0.943^{***}	1.065^{***}	1.065^{***}	1.181^{***}	1.181^{***}
	(0.131)	(0.119)	(0.199)	(0.188)	(0.264)	(0.253)	(0.325)	(0.312)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-n),h}$		0.779^{***}		0.789^{***}		0.807^{***}		0.848^{***}
		(0.180)		(0.219)		(0.288)		(0.318)
Constant	−0.010	−0.010	−0.003	−0.003	0.004	0.004	0.010	0.010
	(0.039)	(0.035)	(0.063)	(0.060)	(0.088)	(0.086)	(0.114)	(0.111)
Observations	300	300	300	300	300	300	300	300
Adjusted R^2	0.119	0.178	0.063	0.100	0.042	0.069	0.032	0.060

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

Empirical Results - 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

Details

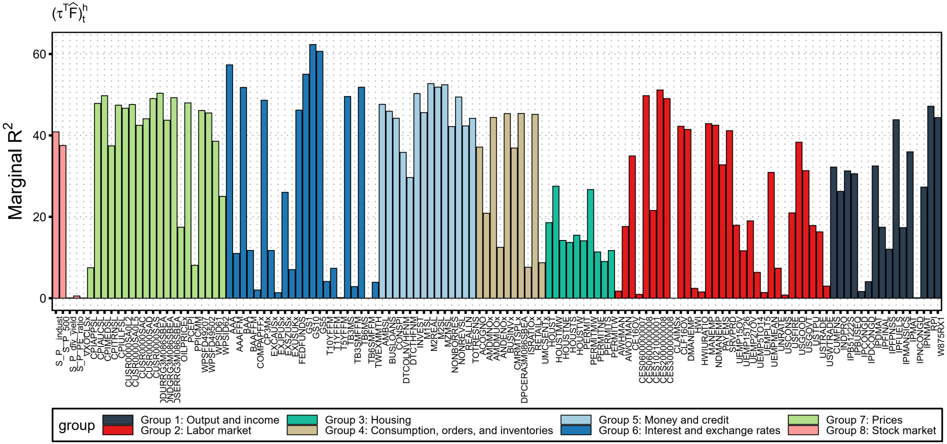
Panel A: $rx_{t+h/12}^{(2)}$

	(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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-2),h}$		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: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

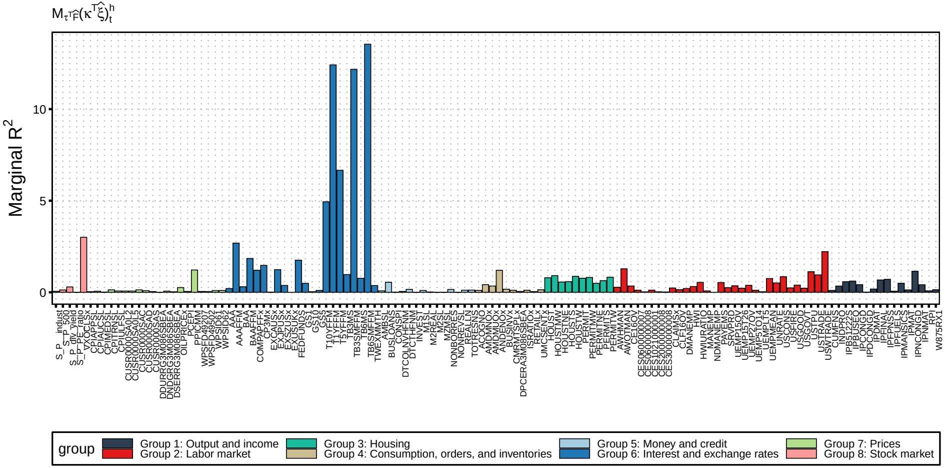
Empirical Results - Economic Interpretation

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



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

$$R^{2(n)}_{OoS,i} = 1 - \frac{\sum_{\tau \in \tau_{OoS}} \left(rx^{(n)}_{t+h/12|t} - \widehat{rx}^{(n)}_{t+h/12|t} \right)^2}{\sum_{\tau \in \tau_{OoS}} \left(rx^{(n)}_{t+h/12|t} - \overline{rx}^{(n)}_{t+h/12|t} \right)^2} \tag{3}$$

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

Additional Results

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
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}} (\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^h + \epsilon_{t+h/12}$	0.22	0.05	-0.01	-0.03
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{LN}_t^h + \epsilon_{t+h/12}$	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

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

Additional Results

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

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 **a single state factor** 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.

References I

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 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.

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

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

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

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

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

References II

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.

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.

Notation

Return

- **holding period returns**

$$\begin{aligned} r_{t+\Delta}^{(n)} &\equiv p_{t+\Delta}^{(n-\Delta)} - p_t^{(n)} \\ r_{t+h/12}^{(n)} &\equiv p_{t+h/12}^{(n-h/12)} - p_t^{(n)} = ny_t^{(n)} - (n-h/12)y_{t+h/12}^{(n-h/12)} \end{aligned} \quad (4)$$

- **Excess Returns**

$$\begin{aligned} rx_{t+h/12}^{(n)} &\equiv \text{holding period return } r_{t+h/12}^{(n)} \quad - \quad \text{1-period yield} \\ &= ny_t^{(n)} - (n-h/12)y_{t+h/12}^{(n-h/12)} - (h/12)y_t^{(h/12)} \end{aligned} \quad (5)$$

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

$$\begin{aligned} f_t^{(n)} &\equiv p_t^{(n-h/12)} - \quad p_t^{(n)} \\ &= ny_t^{(n)} - (n-h/12)y_t^{(n-h/12)} \end{aligned} \quad (6)$$

Expectation Hypothesis

Return

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

$$y_t^{(n)} \equiv \underbrace{\frac{1}{n}\mathbb{E}_t\left(y_t^{(1/12)} + y_{t+1/12}^{(1/12)} + \ldots + y_{t+n-1/12}^{(1/12)}\right)}_{\text{expectations component}} + \underbrace{\frac{1}{n}\mathbb{E}_t\left(rx_{t+1/12}^{(n)} + rx_{t+2/12}^{(n-1/12)} + rx_{t+3/12}^{(n-2/12)} + \ldots + rx_{t+n-1/12}^{(2/12)}\right)}_{\text{yield risk premium}} \tag{7}$$

Expectation Hypothesis

Return

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

$$y_t^{(n)} = \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1} \mathbb{E}\left[y_{t+j\cdot h/12}^{(h/12)} | \mathbf{Z}_t\right]\right) + \frac{1}{n}\left(\sum_{j=0}^{12\cdot n/h-1}\left[rx_{t+h/12(j+1)}^{(n-j\cdot h/12)} | \mathbf{Z}_t\right]\right) . \quad (8)$$

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

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

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \epsilon_{t+h/12} \quad . \tag{9}$$

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

$$
\begin{aligned}
\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} &= \gamma_0 + \gamma_1 f_t^{(1)} + \gamma_2 f_t^{(2)} + \gamma_3 f_t^{(3)} + \gamma_4 f_t^{(4)} + \gamma_5 f_t^{(5)} + \bar{\epsilon}_{t+h/12} \\
\overline{rx}_{t+h/12} &= \underbrace{\gamma^\top \mathbf{f}_t}_{CP_t^h} + \bar{\epsilon}_{t+h/12}
\end{aligned} \tag{10}
$$

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

$$
rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{CP}_t^h + \epsilon_{t+h/12} \quad . \tag{11}
$$

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

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

where $\widehat{\boldsymbol{G}}_t$ and $\boldsymbol{\lambda}$ are 5×1 vectors given by $\widehat{\boldsymbol{G}}_t \equiv \begin{bmatrix}\hat{g}_{1,t} & \hat{g}_{1,t}^3 & \hat{g}_{3,t} & \hat{g}_{5,t} & \hat{g}_{8,t}\end{bmatrix}^\top$, and $\boldsymbol{\lambda} \equiv \begin{bmatrix}\lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5\end{bmatrix}^\top$.

Predictive Regression

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \widehat{LN}_t^h + \epsilon_{t+h/12} \quad . \tag{13}$$

Algorithm 1: Recursively generated factors with updated parameters

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

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

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

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

    Obtain the learned parameters;

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

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

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

  end

end

Return

Algorithm 2: Recursively generated factors with updated parameters

Result:

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

And,

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

An Illustrative Term-Structure Model

Return

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

$$P_t^{(n)} = \mathbb{E}_t \left(\mathcal{M}_{t+1} P_{t+1}^{(n-1)} \right) \tag{14}$$

where

- $P_t^{(n)}$ is the price of a bond,

- $\mathcal{M}_{t+h/12}$ is the stochastic discount factor (SDF).

SDF:

$$\mathcal{M}_{t+h/12} = \exp^{-r_t \frac{1}{2} \Lambda_t^\top \Lambda_t - \Lambda_t^\top \epsilon_{t+h/12}} \tag{15}$$

where Λ_t is the market prices of the risks, i.e., the amount of compensation required by investors to face the unit normal shock $\epsilon_{t+h/12}$.

$$r_t = \rho_0 + \rho_1 \boldsymbol{Z_t} \quad . \tag{16}$$

An Illustrative Term-Structure Model

Return

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

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

$$\begin{aligned} \begin{bmatrix} \boldsymbol{Z}_t^{\mathsf{y}} \\ \boldsymbol{Z}_t^{\mathsf{y}^{\complement}} \end{bmatrix} &= \boldsymbol{\mu} + \boldsymbol{\Phi} \begin{bmatrix} \boldsymbol{Z}_{t-1}^{\mathsf{y}} \\ \boldsymbol{Z}_{t-1}^{\mathsf{y}^{\complement}} \end{bmatrix} + \boldsymbol{\Sigma}\epsilon_t \\ \boldsymbol{Z}_t &= \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{Z}_{t-1} + \boldsymbol{\Sigma}\epsilon_t \qquad \epsilon_t \sim N(0, \boldsymbol{I}) \end{aligned} \tag{17}$$

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

Details

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

$$\boldsymbol{Z}_t^{y^{\complement}} = \gamma_0 + \gamma_1 \boldsymbol{Z}_t^{y} + \boldsymbol{M}_{\boldsymbol{Z}_t^{y}} \boldsymbol{Z}_t^{y^{\complement}} \tag{18}$$

where $\boldsymbol{M}_{\boldsymbol{Z}_t^{y}} \boldsymbol{Z}_t^{y^{\complement}}$ is the annihilator matrix of the space spanned by $\boldsymbol{Z}_t^{y}$, i.e.,

$$\boldsymbol{M}_{\boldsymbol{Z}_t^{y}} \boldsymbol{Z}_t^{y^{\complement}} \equiv \boldsymbol{Z}_t^{y^{\complement}} - \text{Proj}\left[\boldsymbol{Z}_t^{y^{\complement}} | \boldsymbol{Z}_t^{y}\right] \tag{19}$$

In our methodology,

- $\boldsymbol{Z}_t^{y}$ is given by the derived factor $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$
- $\boldsymbol{Z}_t^{y^{\complement}}$ by a function of $\xi_{t+h/12}^h$ as $f(\xi_{t+h/12}^h)$

Correlation Matrix

Return

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

Empirical Results - 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 $\left(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}}\right)^{(-n),h}_t$ as State Variables

Return

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

	DNN 1			DNN 2			DNN 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}})^h_t$	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)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})^{(-2),h}_t$		0.760***			0.779***			0.875***	
		(0.204)			(0.180)			(0.211)	
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})^{h}_t$			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

Return

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

Return

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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-4),h}$		0.795***			0.807***			1.038***	
		(0.291)			(0.288)			(0.289)	
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^h$			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 $\left(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t\right)_t^h$, $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^h$ and $\left(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}}\right)_t^{(-n),h}$ as State Variables

Return

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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-5),h}$		0.854**			0.848***			1.069***	
		(0.336)			(0.318)			(0.339)	
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^h$			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 - 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

Return

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

	(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)
$\boldsymbol{M}_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-3),h}$		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 $\left(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\mathfrak{F}}}_t\right)_t^h$ 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

Return

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

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

Return

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

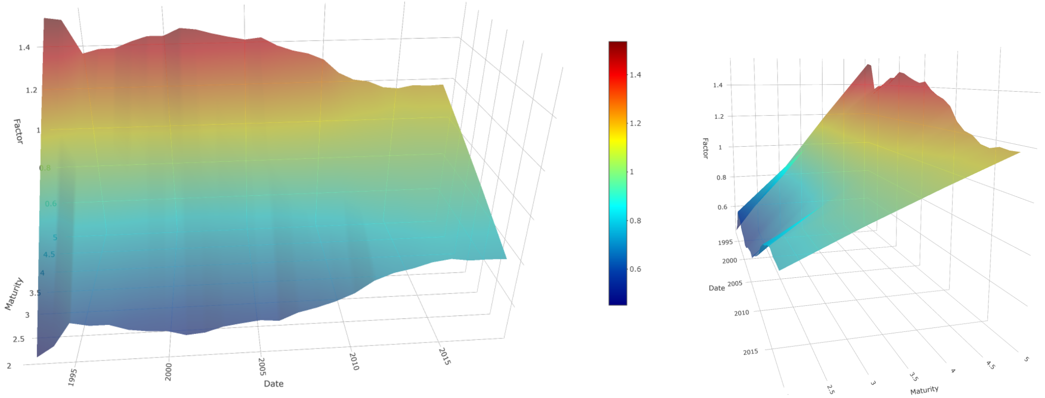
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}})_t^h$	1.268***	1.258***	1.247***	1.263***	0.511	0.626	0.736*	0.834**
	(0.315)	(0.305)	(0.334)	(0.318)	(0.422)	(0.401)	(0.409)	(0.400)
$\boldsymbol{M}_{\boldsymbol{\tau}^{\top}\widehat{\mathfrak{F}}}(\boldsymbol{\kappa}^{\top}\widehat{\boldsymbol{\xi}})_t^{(-5),h}$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

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

Empirical Results

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

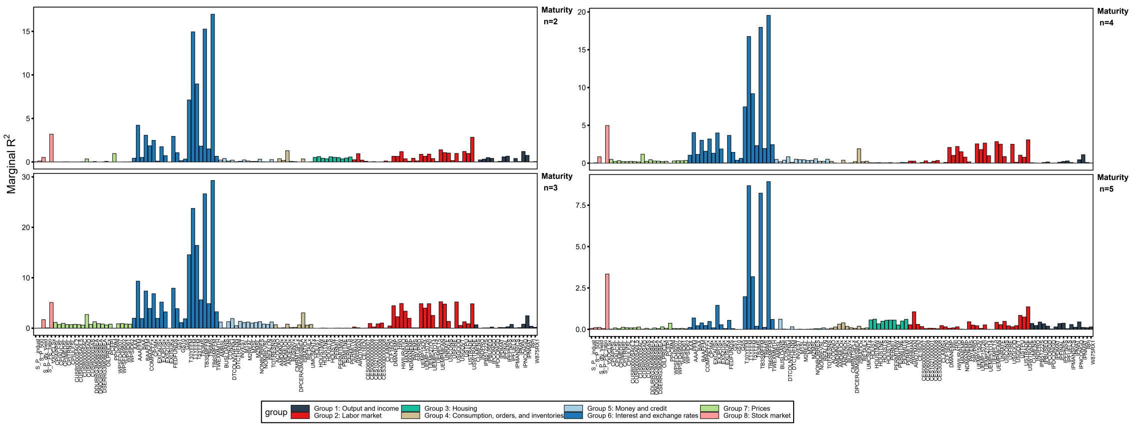
Return



Empirical Results - Economic Interpretation

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

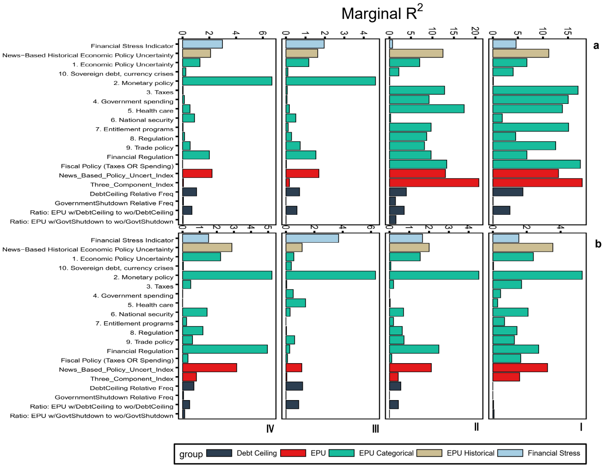
Return



Empirical Results - Economic Interpretation

Marginal R^2 Using Sentiment-Based Measures

Return



Empirical Results

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

Return