

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

Caio Vigo Pereira

Federal Reserve Bank of Richmond
April 22, 2021

Motivation

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

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**.

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 police
 - ↳ 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.

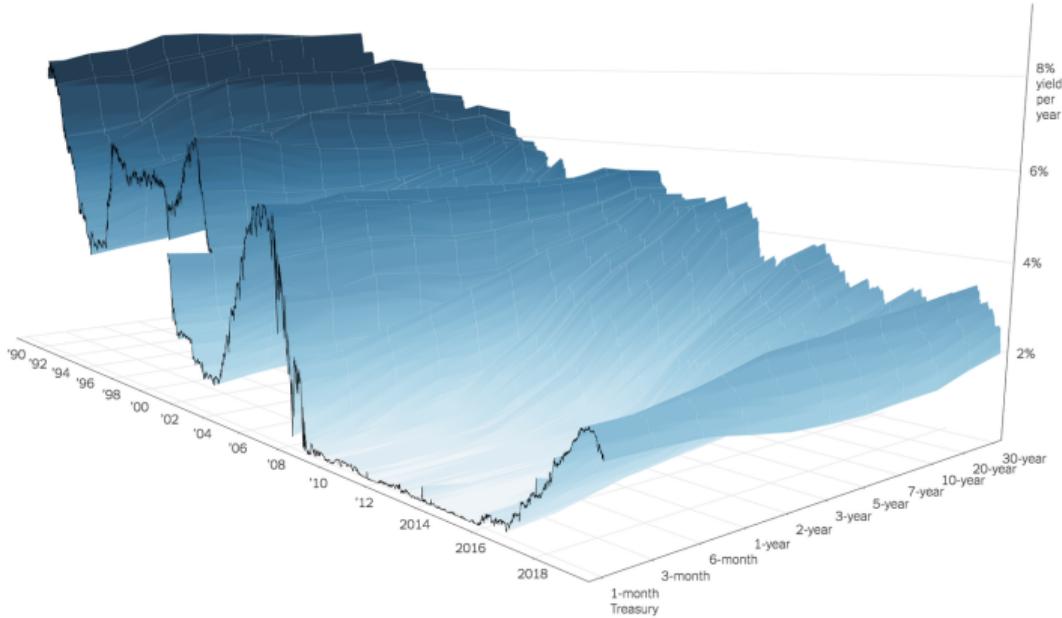
Motivation

- I study the time variation of the **risk premia in U.S. Treasuries bonds**.
- Treasury bonds play an important role in financial markets ⇒ its risk and return dynamics is of **central economic importance**.
 - ↳ major importance for monetary police
 - ↳ 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)**.

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 police
 - ↳ 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)

Motivation

- **Bond Premia**

- ↳ Long literature. Back from Fama and Bliss (1987)
- ↳ Nonetheless, never fully answered/understood
- ↳ Many **factors** were proposed in the literature:

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*

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.

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.

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.

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.

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.

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).

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.

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,

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).

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.

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**.

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).

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

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

- **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)

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)

Contribution - Discussion

- **Third,** I overcome some of 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.

Contribution - Discussion

- **Third,** I overcome some of 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.

Contribution - Discussion

- **Third,** I overcome some of 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.

Contribution - Discussion

- **Third,** I overcome some of 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)

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

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

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

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

- 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 rule out the importance of macro variables (current or future).

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 rule out the importance of macro variables (current or future).
- **Yield curve completely reflects and spans this information.**

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 rule 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

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

Z_t^y contains only yield curve variables [yields, forward rates]

$Z_t^{y^c}$ contains any other variable (complement) [e.g., macro and sentiment-based variables]

A Deep-Learning Structure for Bond Premia

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

Z_t^y contains only yield curve variables [yields, forward rates]

$Z_t^{y^c}$ contains any other variable (complement) [e.g., macro and sentiment-based variables]

We can summarize previous approaches with the following predictive regression:

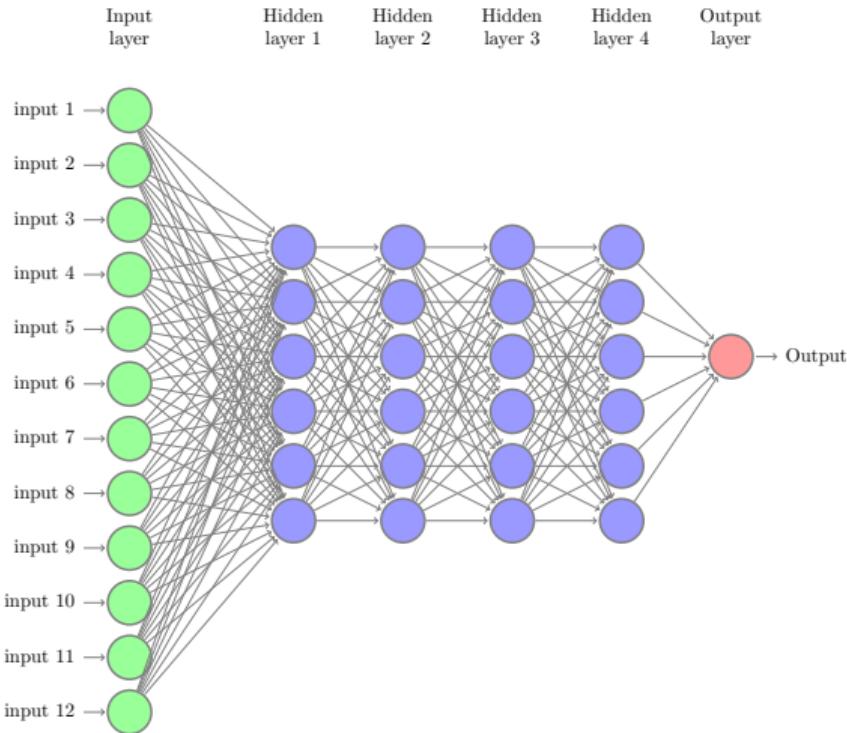
$$rx_{t+h/12}^{(n)} = \beta^\top Z_t + \epsilon_{t+h/12} \quad (1)$$

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

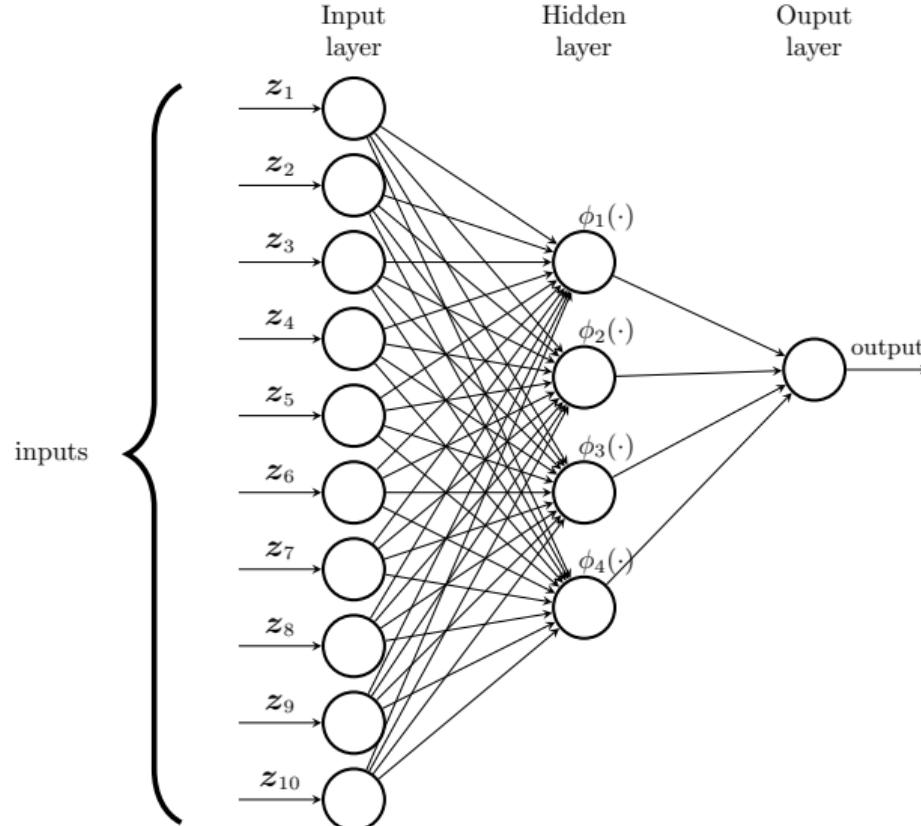
A Deep-Learning Structure for Bond Premia

- 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, $rx_{t+h/12}^{(n)}$) and the inputs (here, Z_t). Supervised Learning
- General structure:
 - (1) Input layer
 - (2) Several hidden layers of **nodes** with predetermined non-linear functions (ϕ)
 - (3) Output layer

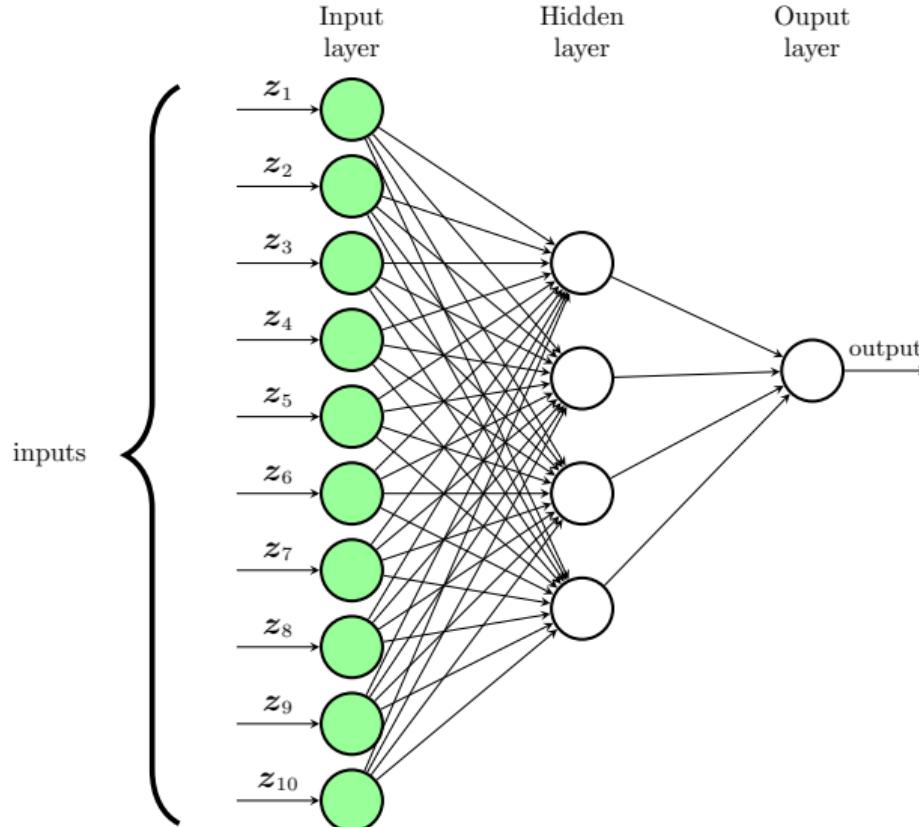
Artificial Neural Network: A Primer



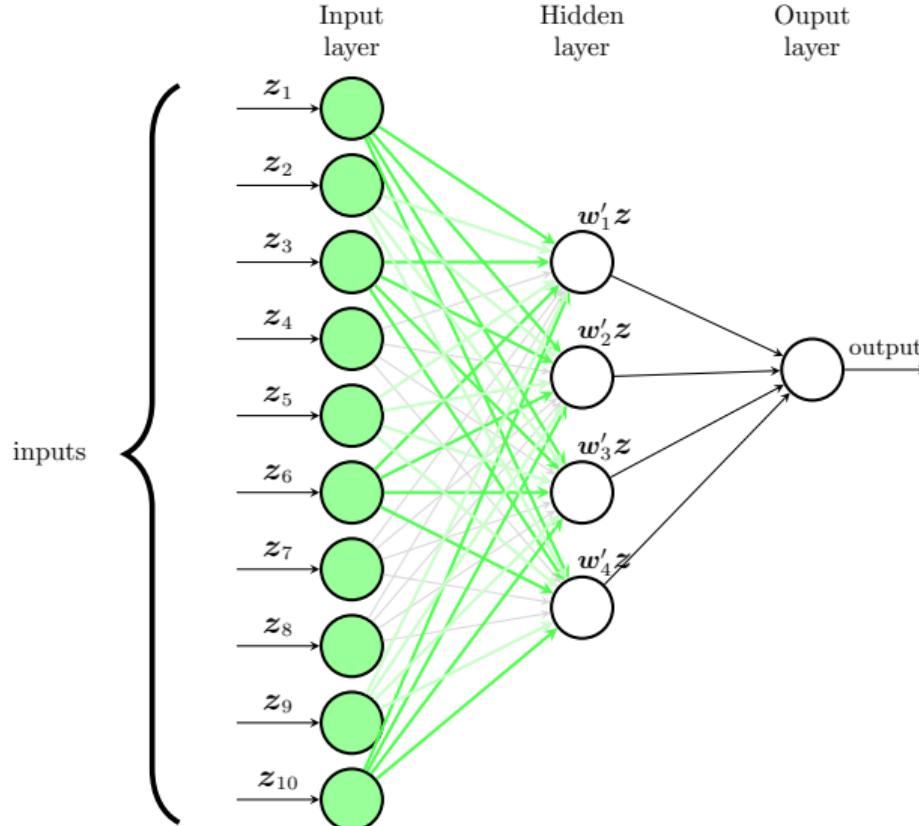
Artificial Neural Network: A Primer



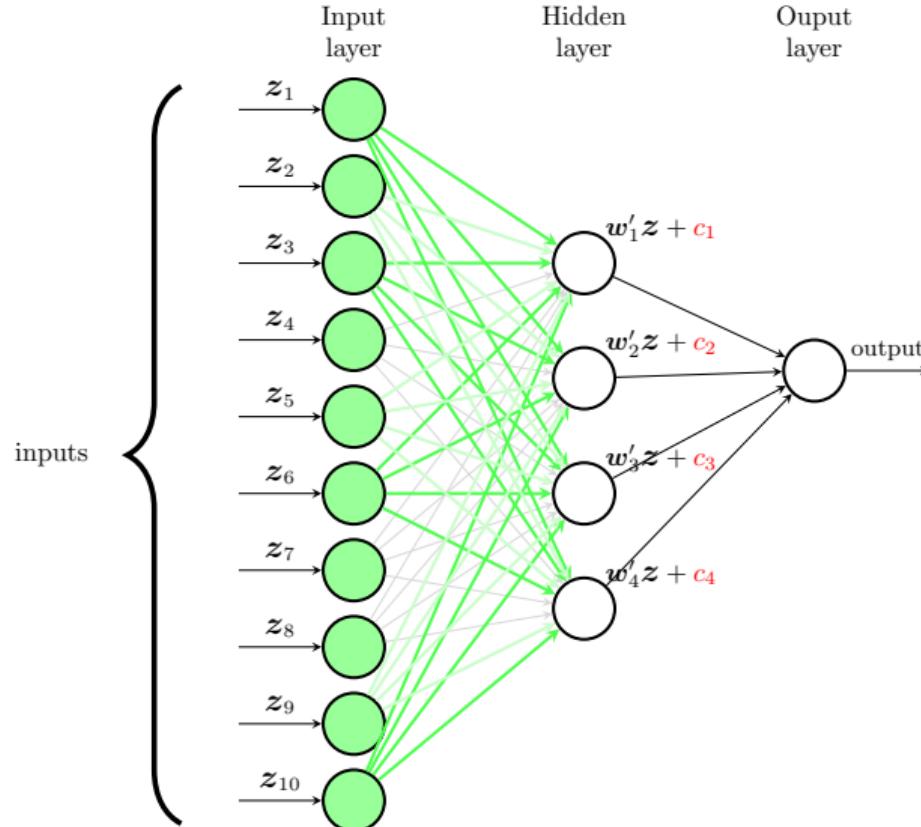
Artificial Neural Network: A Primer



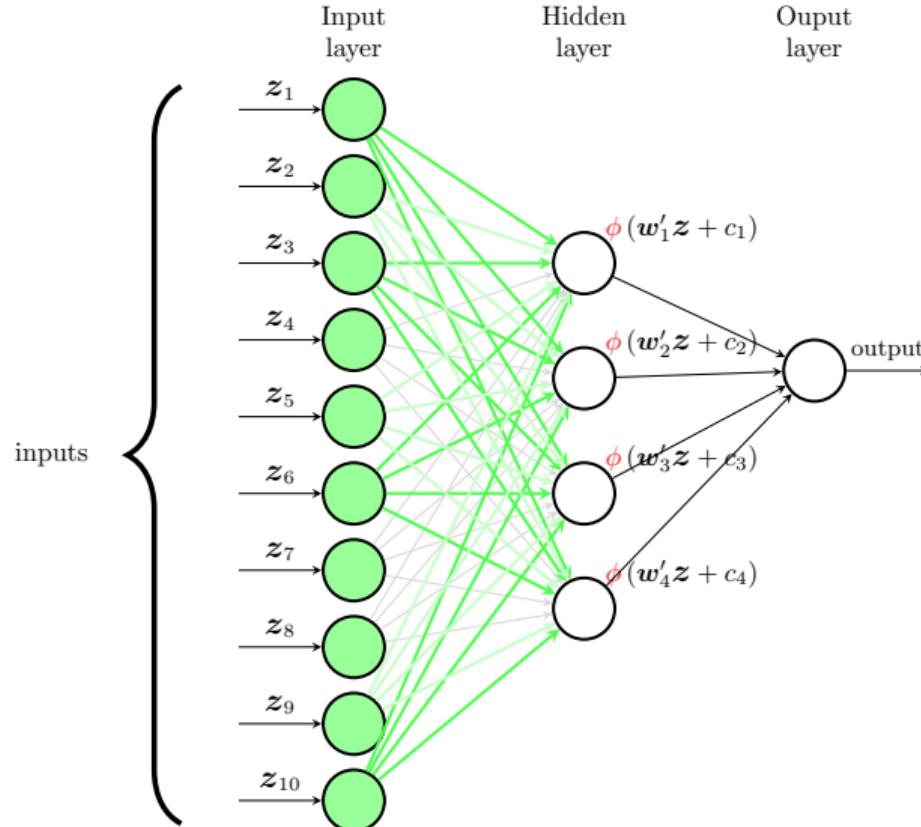
Artificial Neural Network: A Primer



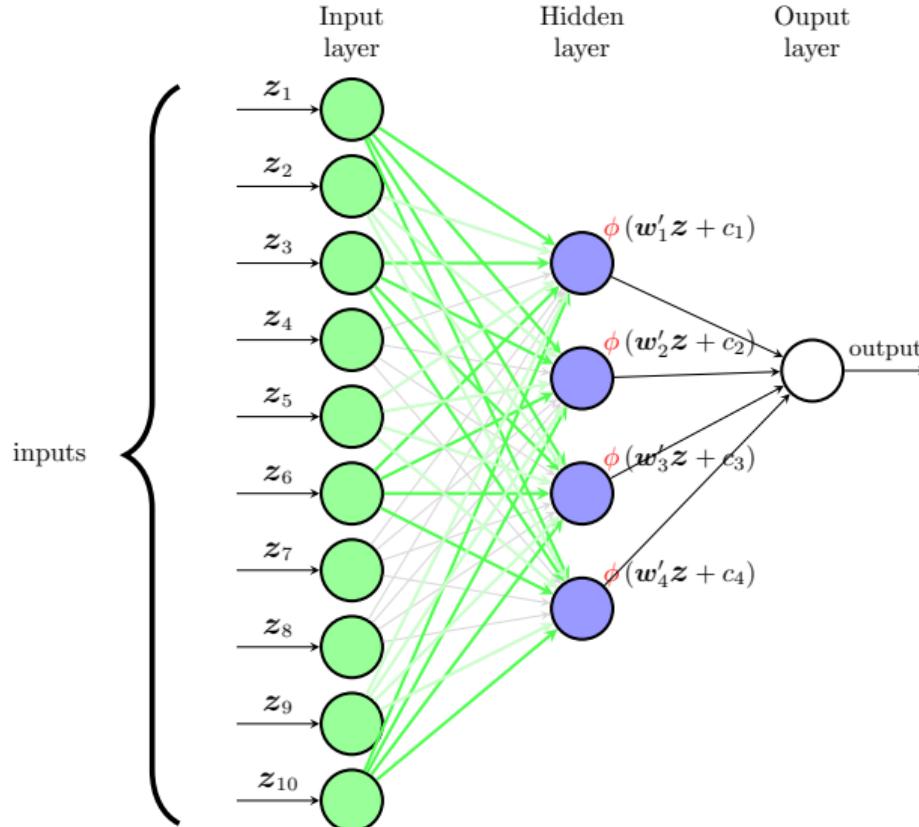
Artificial Neural Network: A Primer



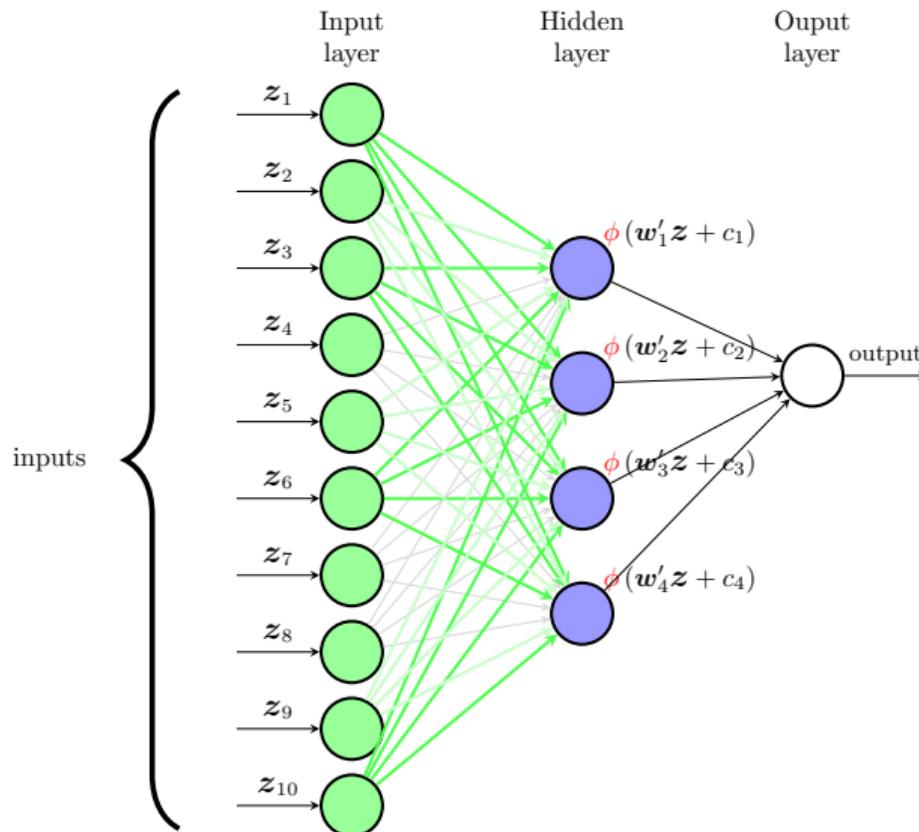
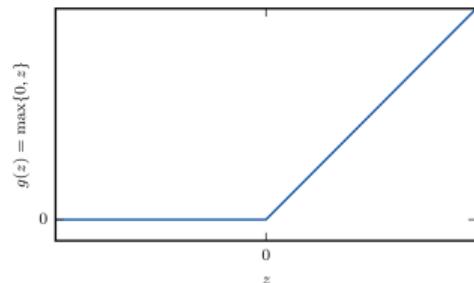
Artificial Neural Network: A Primer



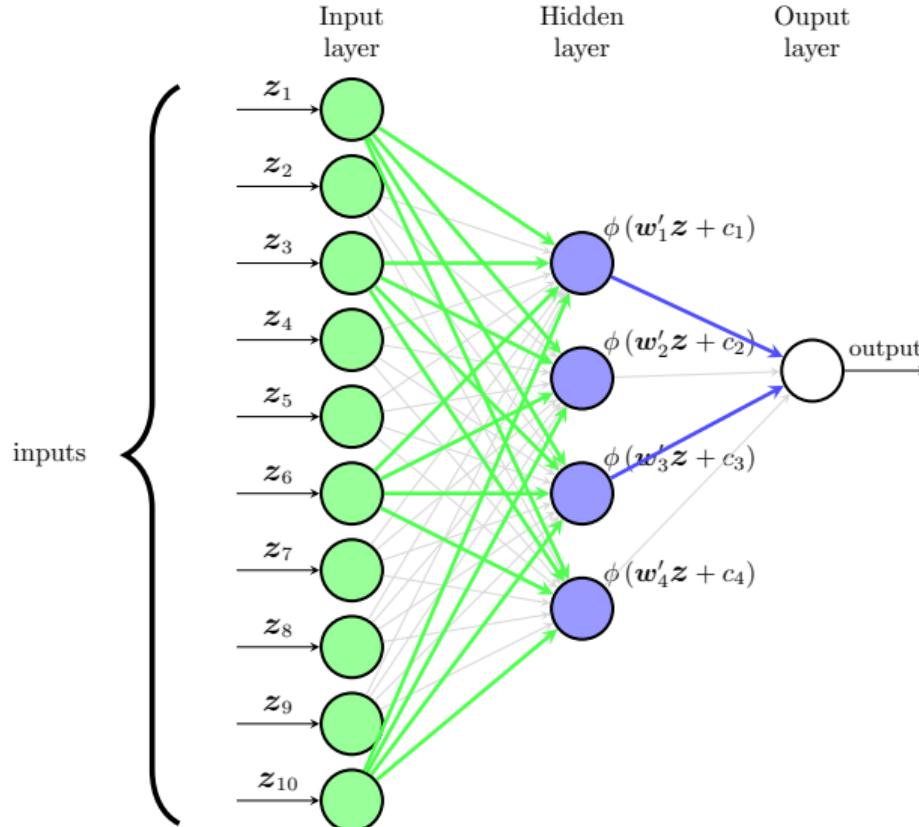
Artificial Neural Network: A Primer



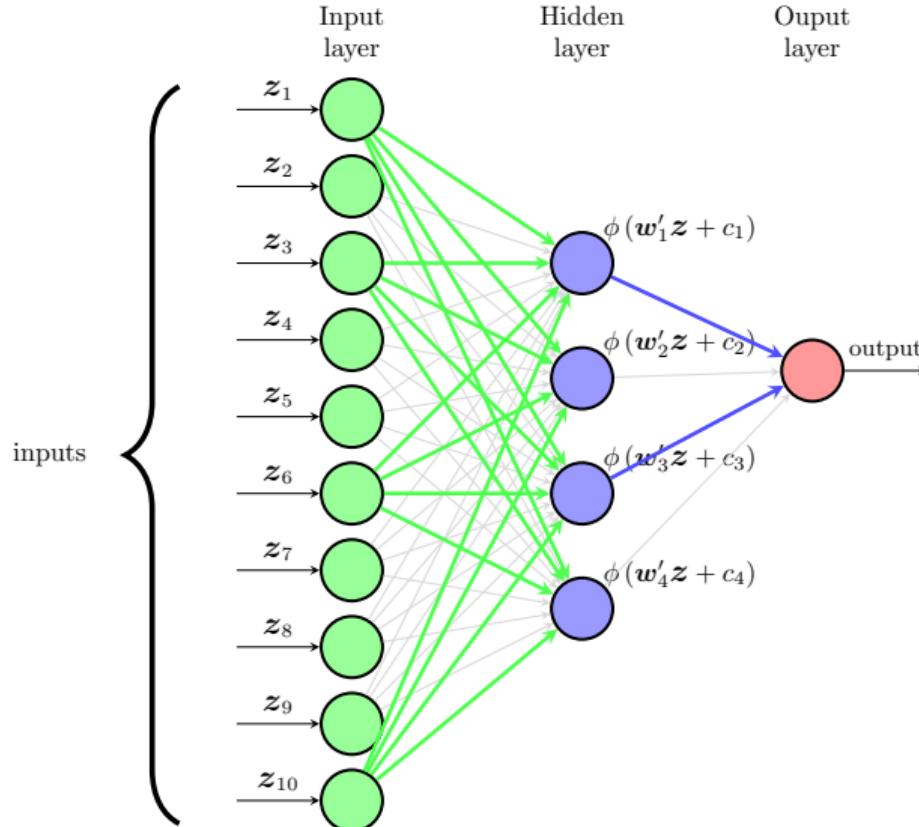
Artificial Neural Network: A Primer



Artificial Neural Network: A Primer



Artificial Neural Network: A Primer



A Deep-Learning Structure for Bond Premia

- DNN defines a mapping such as $rx_{t+h/12}^{(n)} = g(\mathbf{Z}_t, \theta_t)$ to *learn* the parameter θ_t that provides the best function approximation.
- Represented in a direct acyclic graph with a chain of functions
$$g(\mathbf{Z}_t) = g^{(L)} \left(\dots \left(g^{(2)} \left(g^{(1)}(\mathbf{Z}_t) \right) \right) \right).$$

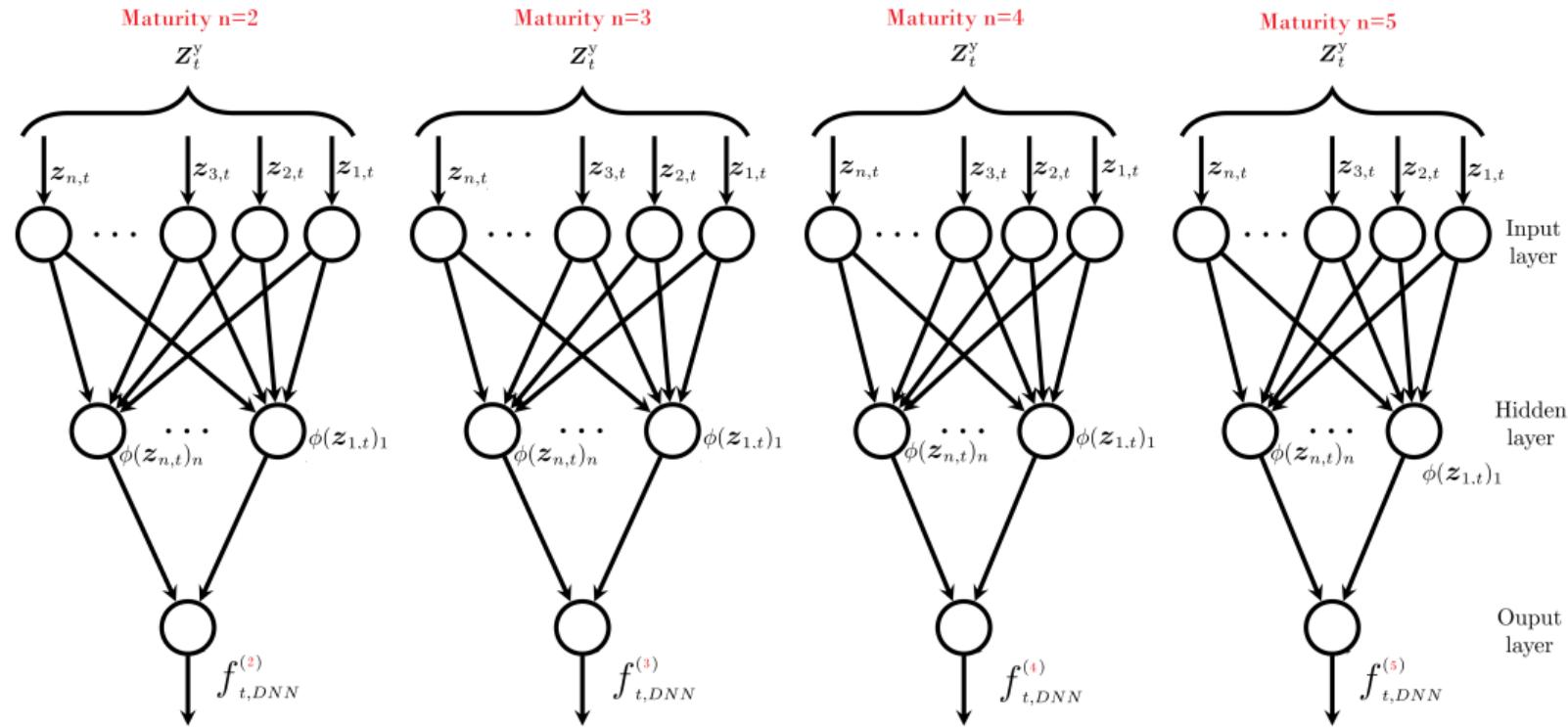
A Deep-Learning Structure for Bond Premia

- DNN defines a mapping such as $rx_{t+h/12}^{(n)} = g(\mathbf{Z}_t, \theta_t)$ to learn the parameter θ_t that provides the best function approximation.
- Represented in a direct acyclic graph with a chain of functions
$$g(\mathbf{Z}_t) = g^{(L)} \left(\dots \left(g^{(2)} \left(g^{(1)}(\mathbf{Z}_t) \right) \right) \right).$$

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.
 - ↳ In short, a simple neural network can represent a wide variety of functions.

A Deep-Learning Structure for Bond Premia



DNN Factors

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

DNN Factors

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

where $\widehat{\mathfrak{F}}_t$ and $\boldsymbol{\tau}$ are 5×1 vectors given by $\widehat{\mathfrak{F}}_t \equiv [1 \quad f_{t,DNN}^{(2),h} \quad f_{t,DNN}^{(3),h} \quad f_{t,DNN}^{(4),h} \quad f_{t,DNN}^{(5),h}]^\top$, and $\boldsymbol{\tau} \equiv [\tau_0 \quad \tau_1 \quad \tau_2 \quad \tau_3 \quad \tau_4]^\top$.

- We **recursively** orthogonalize the excess returns generated by the deep neural network factor $f_{t,DNN}^{(n)}$, and denote it by $\xi_t^{(n),h}$.
- The factor $\xi_{t+h/12}^{(n),h}$ that lies in an orthogonal vector to the space spanned by $f_{t,DNN}^{(n)}$, can be seen as all the information not spanned by the term-structure captured by $f_{t,DNN}^{(n)}$.

A Deep-Learning Structure for Bond Premia

Linear Rotation of the State Space

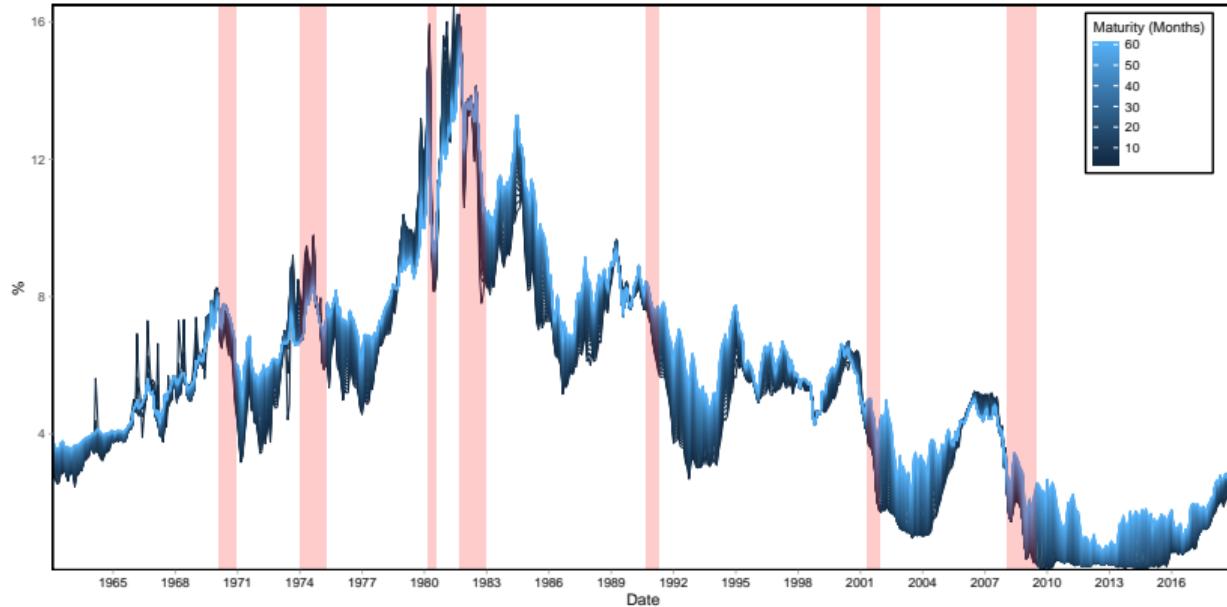
Proposition 2. As in the dynamic term structure model of Joslin et al. (2014), $f(\xi_{t+h/12}^h)$ complete and fill the unspanned factor in the state space, in a such a way that $\left[\left(\tau^\top \widehat{\mathfrak{F}}_t \right)_t^h, f(\xi_{t+h/12}^h) \right]$ and Z_t represent linear rotations 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 $\widehat{\xi}_{t+h/12}^h$ could be capturing macroeconomic information or sentiment measures not spanned by the term-structure.

Details

Data & Strategy

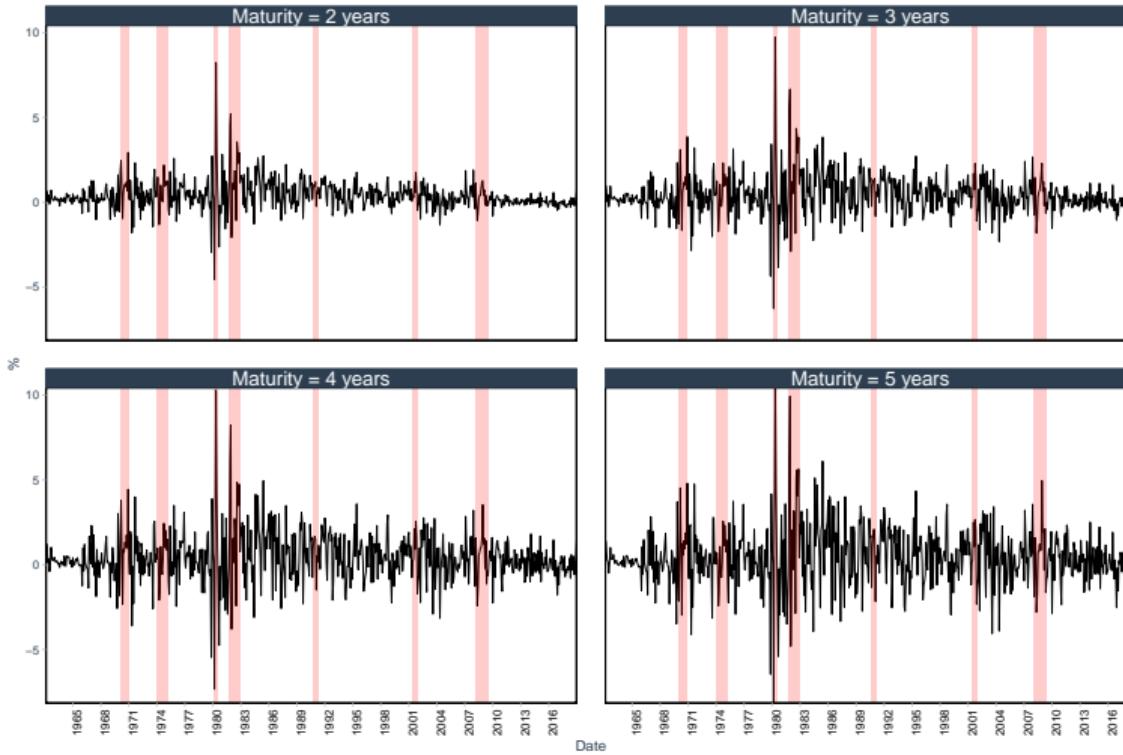
Derived zero-coupon bonds log yields for maturities (n) up to 60 months



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

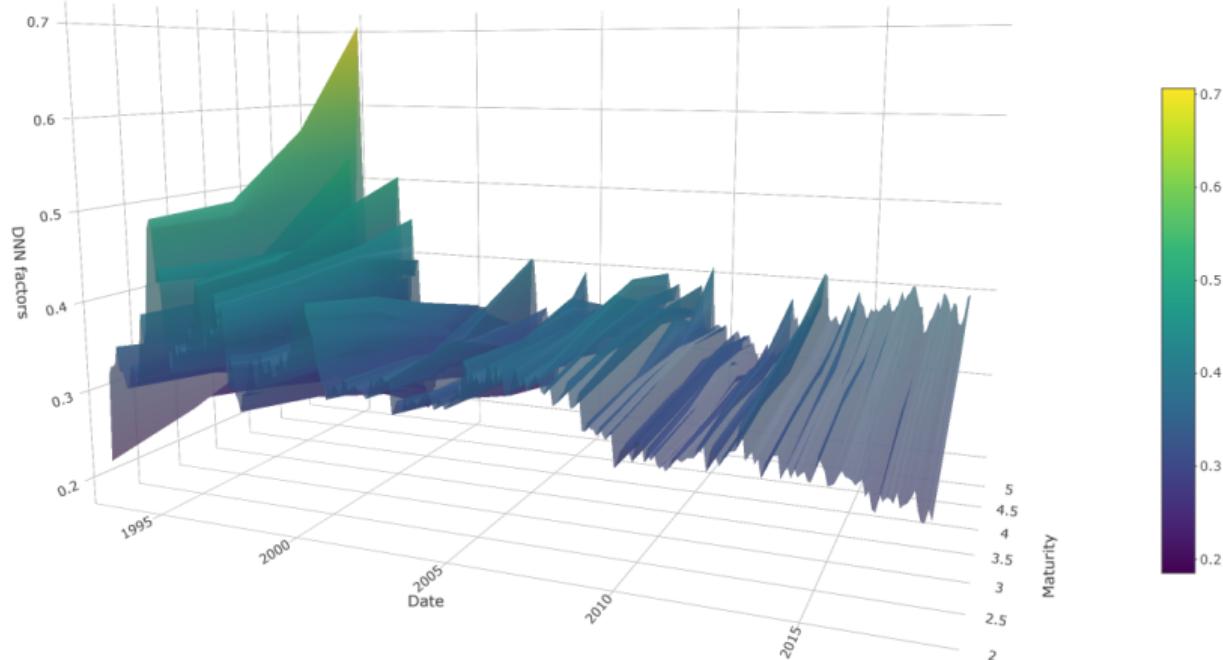
Treasuries Excess Returns

1-Month Bonds Excess Returns (1962-2017)



Empirical Results

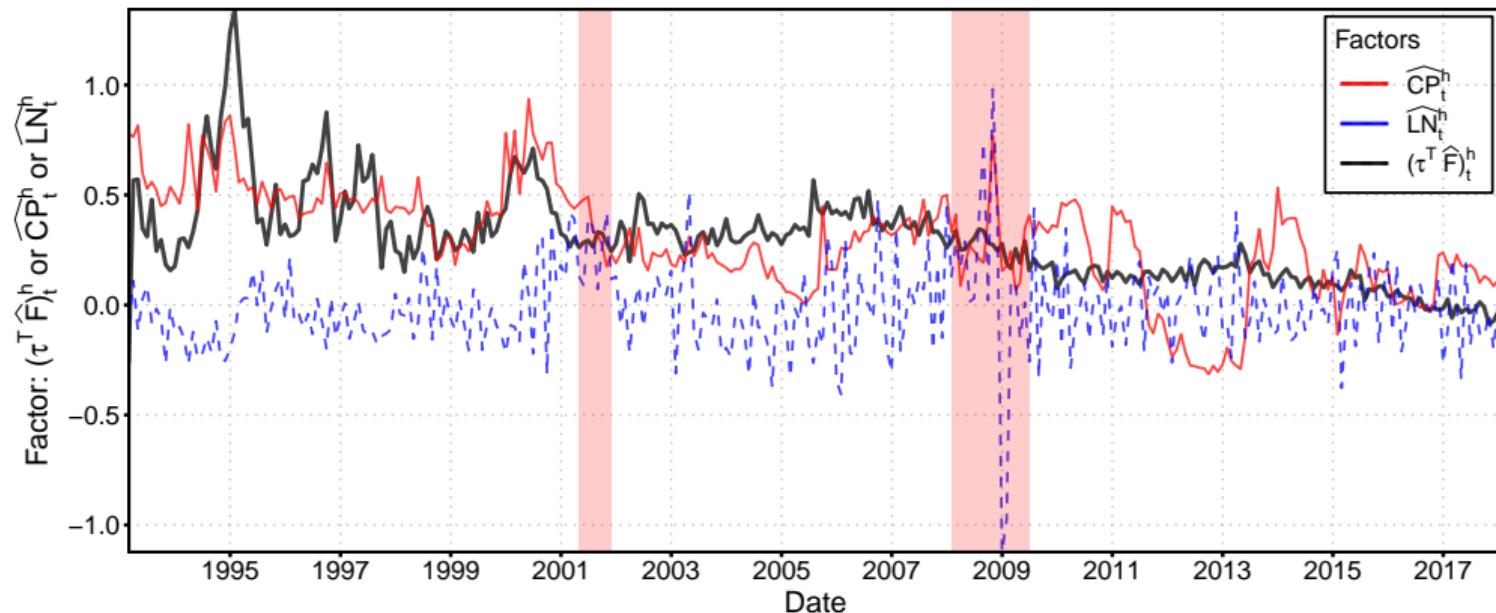
Derived Factors $f_{t,DNN}^{(n),h}$ for **DNN 2** Generated Using the Set of Yields



Empirical Results

Comparison with Other Factors from the Literature

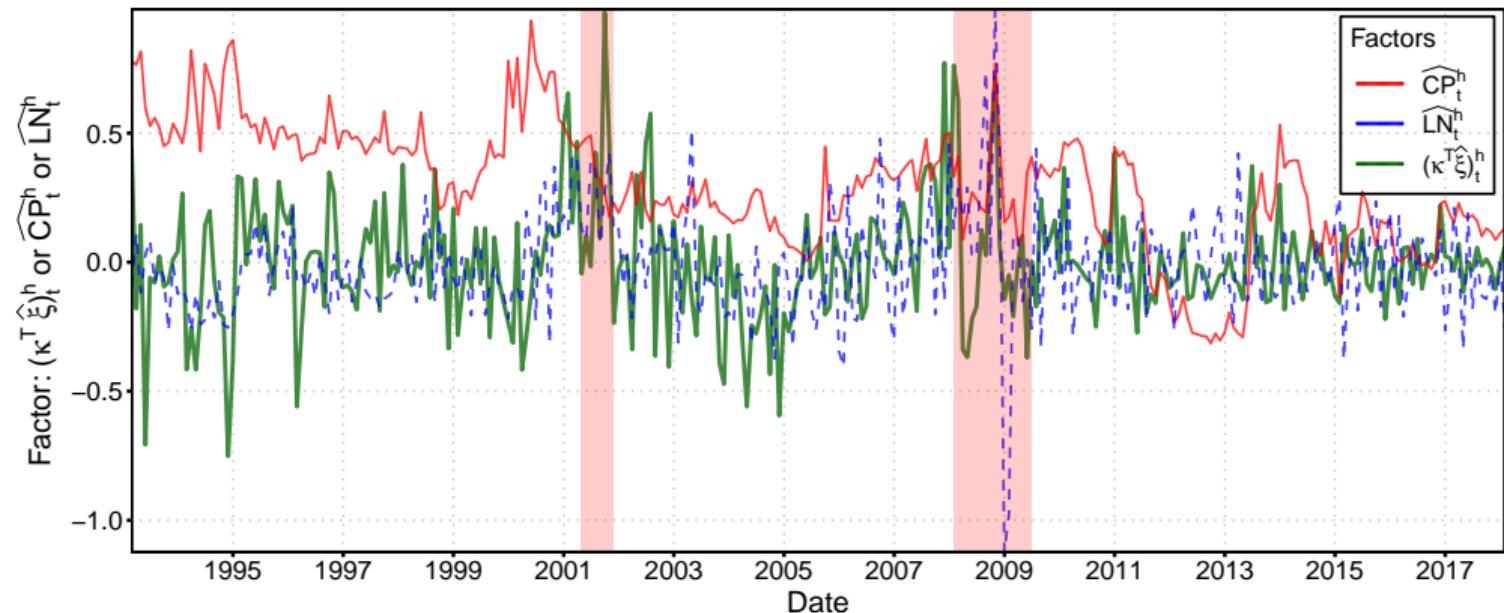
Figure 1: Time Series of our Derived Factor $(\tau^\top \widehat{\mathbf{F}}_t)^h$, along with \widehat{CP}_t^h and \widehat{LN}_t^h



Empirical Results

Comparison with Other Factors from the Literature

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



Empirical Results - Predictive Regressions with $(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t)_t^h$ and $(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-n),h}$, 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)
$(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}})_t^h$	0.847*** (0.124)	0.842*** (0.115)	0.853*** (0.128)	0.824*** (0.117)	0.525*** (0.154)	0.582*** (0.140)	0.582*** (0.145)	0.614*** (0.135)
$M_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}} (\boldsymbol{\kappa}^\top \boldsymbol{\xi})_t^{(-2),h}$		0.658*** (0.172)		0.745*** (0.182)		0.704*** (0.182)		0.558*** (0.185)
\bar{LN}_t^h	0.617*** (0.127)	0.529*** (0.120)					0.559*** (0.110)	0.518*** (0.110)
$f_{\mathfrak{s}t}^{(n,h)}$			-0.746 (0.476)	-0.225 (0.438)			-0.570 (0.437)	-0.172 (0.429)
\bar{CP}_t^h					0.454*** (0.126)	0.364*** (0.112)	0.465*** (0.112)	0.375*** (0.109)
Constant	-0.013 (0.037)	-0.012 (0.034)	0.031 (0.051)	0.002 (0.047)	-0.060 (0.039)	-0.050 (0.036)	-0.031 (0.045)	-0.044 (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

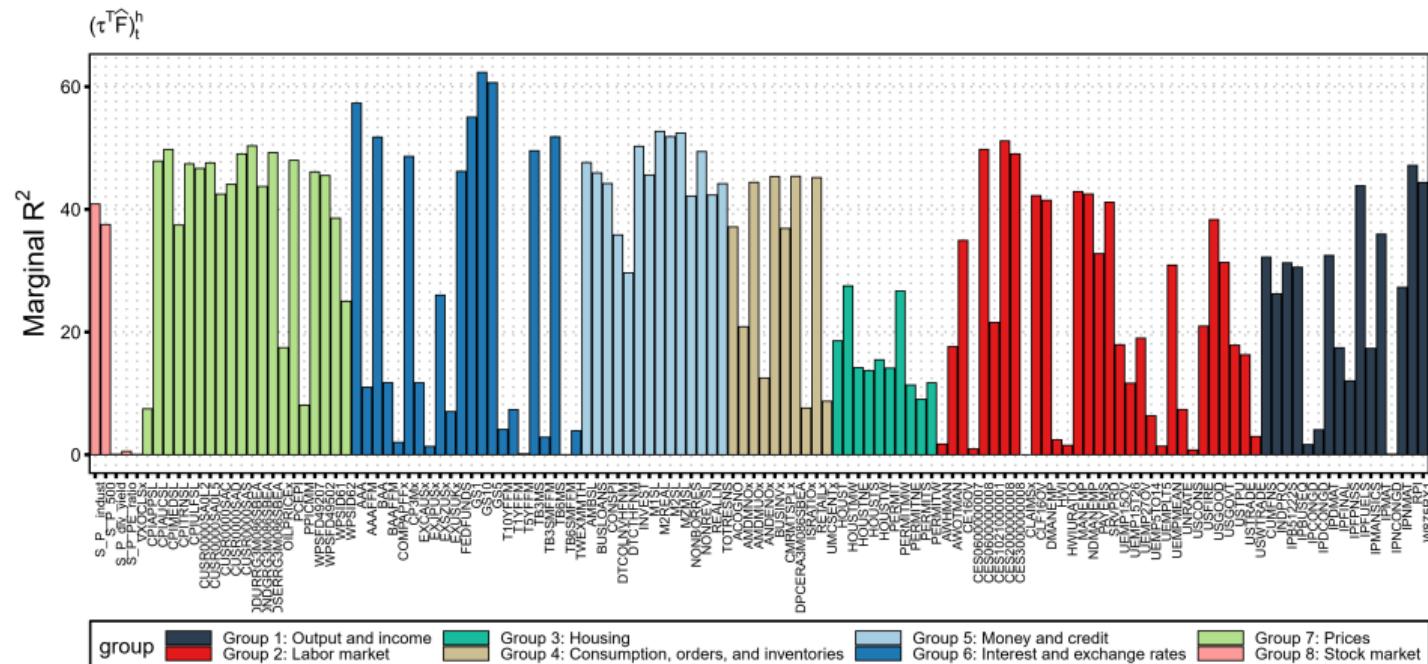
Correlation Matrix

	$(\tau^\top \widehat{\mathfrak{F}})_t^h$	$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^h$	$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^{(-2),h}$	$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^{(-3),h}$	$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^{(-4),h}$	$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^{(-5),h}$	\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 - Economic Interpretation

Empirical Results - Economic Interpretation

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

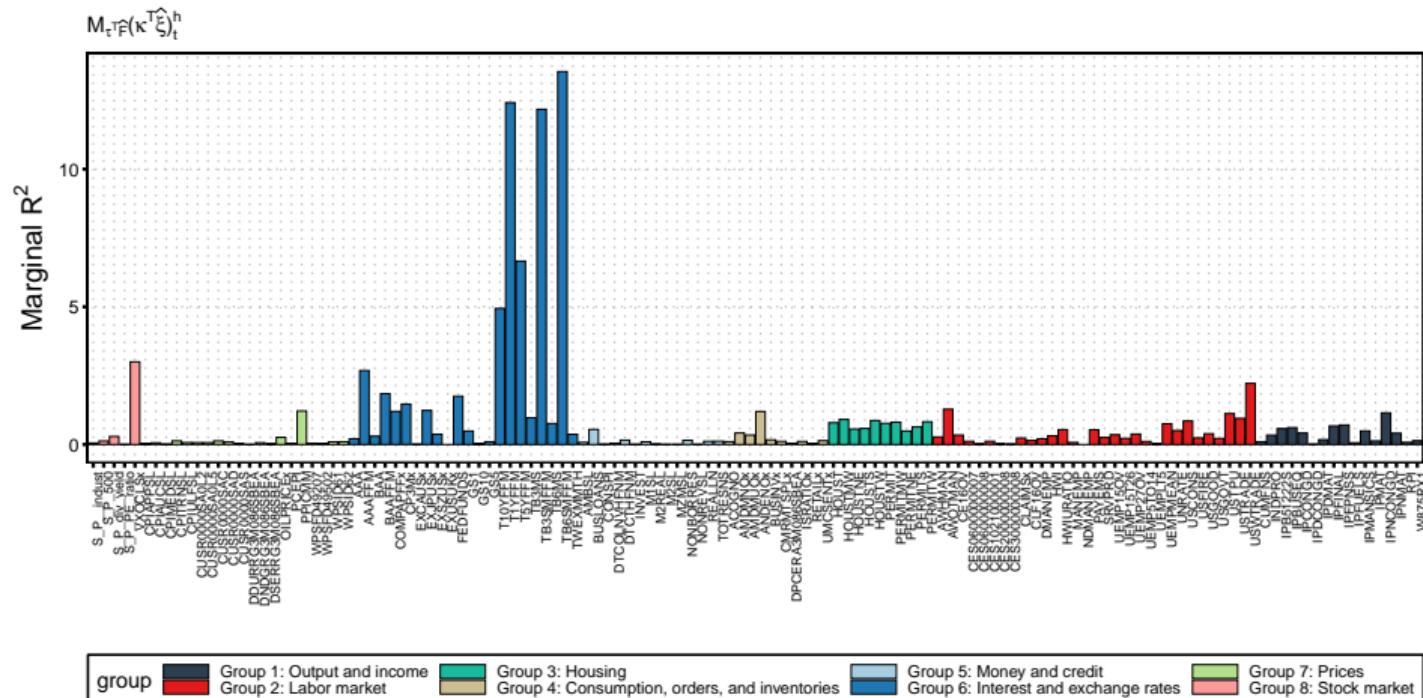


Empirical Results - Economic Interpretation

Sentiment-based Results

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

Additional Results



Overview

Introduction

Framework

Data & Empirical Strategy

Empirical Results

References

Appendix

Empirical Results

Out-of-Sample Forecasting Performance

Empirical 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 .

Empirical 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

$$R_{OoS,i}^{2(n)} = 1 - \frac{\sum_{\tau \in \tau_{OoS}} \left(rx_{t+h/12|t}^{(n)} - \hat{rx}_{t+h/12|t}^{(n)} \right)^2}{\sum_{\tau \in \tau_{OoS}} \left(rx_{t+h/12|t}^{(n)} - \bar{rx}_{t+h/12|t}^{(n)} \right)^2} \quad (3)$$

Empirical Results

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

Additional Results

Regression	Maturity $n = 2$	Maturity $n = 3$	Maturity $n = 4$	Maturity $n = 5$
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1(\tau^\top \widehat{\mathbf{F}}_t)^h + \epsilon_{t+h/12}$	0.17	0.03	-0.02	-0.04
$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 \mathbf{M}_{\tau^\top \widehat{\mathbf{F}}} (\kappa^\top \widehat{\boldsymbol{\xi}})_t^h + \epsilon_{t+h/12}$	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

Empirical Results

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

[Additional Results](#)

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

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.

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.

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.

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

- **holding period returns**

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

- **Excess Returns**

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

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

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

Expectation Hypothesis

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

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

Expectation Hypothesis

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

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

\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)

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

Predictive Regression

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

Cochrane and Piazzesi (2005)

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

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

where \mathbf{f} and γ are 6×1 vectors given by $\mathbf{f} \equiv [1 \quad f_t^{(1)} \quad f_t^{(2)} \quad f_t^{(3)} \quad f_t^{(4)} \quad f_t^{(5)}]^\top$, and $\gamma \equiv [\gamma_0 \quad \gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4 \quad \gamma_5]^\top$.

Predictive Regression

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

Ludvigson and Ng (2009)

- 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

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

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

Predictive Regression

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

Algorithm 1: Recursively generated factors with updated parameters

[Return](#)

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

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

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

 Feed DNN_i with lagged data $Z_{t-1}^y = \{z_{t_0}^y, z_{t_0+1}^y, \dots, z_{t-1}^y\}$ to learn/approximate with output $rx_t^{(n)}$,
 and use the last 10% of the data for validation;

 Obtain the learned parameters;

$$\hat{f}_{t,DNN}^{(n),h} \leftarrow g(Z_{t-1}^y, \theta_{t-1})$$

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

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

end

end

Algorithm 2: Recursively generated factors with updated parameters

Result:

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

And,

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

An Illustrative Term-Structure Model

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

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

where

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

SDF:

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

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

$$r_t = \rho_0 + \rho_1 Z_t \quad . \quad (16)$$

An Illustrative Term-Structure Model

- Define $\mathbf{Z}_t = \{\mathbf{Z}_t^y, \mathbf{Z}_t^{y^c}\}$
- Dynamics of \mathbf{Z}_t that capture all the risks of the economy following a Gaussian VAR process given by:

$$\begin{bmatrix} \mathbf{Z}_t^y \\ \mathbf{Z}_t^{y^c} \end{bmatrix} = \boldsymbol{\mu} + \boldsymbol{\Phi} \begin{bmatrix} \mathbf{Z}_{t-1}^y \\ \mathbf{Z}_{t-1}^{y^c} \end{bmatrix} + \boldsymbol{\Sigma} \boldsymbol{\epsilon}_t \quad (17)$$

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

where $\boldsymbol{\mu}$ is 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:

$$\mathbf{Z}_t^{y^C} = \gamma_0 + \gamma_1 \mathbf{Z}_t^y + \mathbf{M}_{\mathbf{Z}_t^y} \mathbf{Z}_t^{y^C} \quad (18)$$

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

$$\mathbf{M}_{\mathbf{Z}_t^y} \mathbf{Z}_t^{y^C} \equiv \mathbf{Z}_t^{y^C} - \text{Proj} \left[\mathbf{Z}_t^{y^C} | \mathbf{Z}_t^y \right] \quad (19)$$

In our methodology,

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

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

[Return](#)

Panel A:

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

	DNN 1		DNN 2		DNN 3				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$(\boldsymbol{\tau}^\top \widehat{\boldsymbol{\delta}})^h_t$	0.810*** (0.160)	0.810*** (0.149)	0.810*** (0.147)	0.811*** (0.131)	0.811*** (0.119)	0.811*** (0.119)	1.419*** (0.414)	1.419*** (0.377)	1.419*** (0.356)
$M_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\delta}}} (\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^{(-2),h}_t$		0.760*** (0.204)			0.779*** (0.180)			0.875*** (0.211)	
$M_{\boldsymbol{\tau}^\top \widehat{\boldsymbol{\delta}}} (\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})^h_t$			0.591*** (0.139)			0.525*** (0.126)			0.679*** (0.138)
Constant	-0.010 (0.054)	-0.010 (0.050)	-0.010 (0.049)	-0.010 (0.039)	-0.010 (0.035)	-0.010 (0.035)	-0.189* (0.110)	-0.189* (0.101)	-0.189** (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 $(\tau^\top \widehat{\mathfrak{F}}_t)^h$, $(\kappa^\top \widehat{\xi})^h_t$ and $(\kappa^\top \widehat{\xi})^{(-n),h}_t$ as State Variables

[Return](#)

Panel B:

	$rX_{t+h/12}^{(3)}$									
$(\tau^\top \widehat{\mathfrak{F}}_t)^h$	0.959*** (0.248)	0.959*** (0.234)	0.959*** (0.233)	0.943*** (0.199)	0.943*** (0.188)	0.943*** (0.184)	1.175* (0.630)	1.175** (0.566)	1.175** (0.559)	
$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_{t+h/12}^{(-3),h}$		0.799*** (0.234)			0.789*** (0.219)			0.984*** (0.236)		
$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_{t+h/12}^h$			0.765*** (0.225)			0.757*** (0.205)			0.929*** (0.224)	
Constant	-0.008 (0.087)	-0.008 (0.082)	-0.008 (0.082)	-0.003 (0.063)	-0.003 (0.060)	-0.003 (0.059)	-0.072 (0.169)	-0.072 (0.153)	-0.072 (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 $(\tau^\top \widehat{\mathfrak{F}}_t)^h$, $(\kappa^\top \widehat{\xi})^h_t$ and $(\kappa^\top \widehat{\xi})^{(-n),h}_t$ as State Variables

[Return](#)

Panel C:

	$rX_{t+h/12}^{(4)}$								
$(\tau^\top \widehat{\mathfrak{F}})_t^h$	1.073*** (0.334)	1.073*** (0.320)	1.073*** (0.317)	1.065*** (0.264)	1.065*** (0.253)	1.065*** (0.248)	0.864 (0.835)	0.864 (0.759)	0.864 (0.755)
$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^{(-4),h}$		0.795*** (0.291)			0.807*** (0.288)			1.038*** (0.289)	
$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^h$			0.902*** (0.312)			0.945*** (0.284)			1.144*** (0.313)
Constant	0.002 (0.120)	0.002 (0.116)	0.002 (0.115)	0.004 (0.088)	0.004 (0.086)	0.004 (0.085)	0.063 (0.228)	0.063 (0.209)	0.063 (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 $(\tau^\top \widehat{\mathfrak{F}}_t)^h$, $(\kappa^\top \widehat{\xi})^h_t$ and $(\kappa^\top \widehat{\xi})^{(-n),h}_t$ as State Variables

[Return](#)

Panel D:

	$rx_{t+h/12}^{(5)}$								
$(\tau^\top \widehat{\mathfrak{F}})_t^h$	1.158*** (0.415)	1.158*** (0.395)	1.158*** (0.398)	1.181*** (0.325)	1.181*** (0.312)	1.181*** (0.309)	0.542 (1.025)	0.542 (0.949)	0.542 (0.939)
$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^{(-5),h}$		0.854** (0.336)			0.848*** (0.318)			1.069*** (0.339)	
$M_{\tau^\top \widehat{\mathfrak{F}}}(\kappa^\top \widehat{\xi})_t^h$			1.000** (0.398)			1.081*** (0.363)			1.322*** (0.404)
Constant	0.017 (0.152)	0.017 (0.146)	0.017 (0.147)	0.010 (0.114)	0.010 (0.111)	0.010 (0.111)	0.198 (0.284)	0.198 (0.267)	0.198 (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 $(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t)_t^h$ and $(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-n),h}$, along with the Cochrane-Piazzesi and Ludvingson-Ng factors, and Fama-Bliss Regressions with Forward Spreads

[Return](#)

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*** (0.190)	0.989*** (0.184)	0.940*** (0.199)	0.947*** (0.188)	0.559** (0.245)	0.648*** (0.234)	0.626*** (0.238)	0.719*** (0.237)
$M_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}} (\boldsymbol{\kappa}^\top \boldsymbol{\xi})_t^{(-3),h}$		0.620*** (0.209)		0.852*** (0.228)		0.692*** (0.226)		0.585** (0.237)
\bar{LN}_t^h	0.921*** (0.209)	0.800*** (0.201)					0.900*** (0.194)	0.823*** (0.191)
$f_{st}^{(n,h)}$			-0.215 (0.554)	0.410 (0.532)			-0.053 (0.525)	0.394 (0.542)
\bar{CP}_t^h					0.608*** (0.205)	0.467** (0.195)	0.583*** (0.188)	0.437** (0.198)
Constant	-0.007 (0.060)	-0.006 (0.059)	0.021 (0.091)	-0.049 (0.087)	-0.070 (0.063)	-0.054 (0.061)	-0.064 (0.082)	-0.098 (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 $(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t)_t^h$ and $(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-n),h}$, along with the Cochrane-Piazzesi and Ludvingson-Ng factors, and Fama-Bliss Regressions with Forward Spreads

[Return](#)

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*** (0.254)	1.127*** (0.247)	1.082*** (0.270)	1.108*** (0.257)	0.547 (0.335)	0.651** (0.323)	0.685** (0.329)	0.790** (0.329)
$M_{\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}} (\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-4),h}$		0.609** (0.262)		0.872*** (0.289)		0.688** (0.291)		0.555** (0.274)
\bar{LN}_t^h	1.218*** (0.307)	1.079*** (0.287)					1.222*** (0.285)	1.118*** (0.273)
$f_{\mathfrak{s}_t}^{(n,h)}$			0.260 (0.622)	0.665 (0.595)			0.386 (0.593)	0.655 (0.587)
\bar{CP}_t^h					0.822*** (0.290)	0.657** (0.276)	0.755*** (0.265)	0.606** (0.272)
Constant	-0.0003 (0.085)	0.0002 (0.084)	-0.038 (0.130)	-0.103 (0.124)	-0.085 (0.089)	-0.068 (0.087)	-0.144 (0.121)	-0.171 (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 $(\boldsymbol{\tau}^\top \widehat{\mathfrak{F}}_t)_t^h$ and $(\boldsymbol{\kappa}^\top \widehat{\boldsymbol{\xi}})_t^{(-n),h}$, along with the Cochrane-Piazzesi and Ludvingson-Ng factors, and Fama-Bliss Regressions with Forward Spreads

[Return](#)

Panel D:

$rX_{t+h/12}^{(5)}$

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

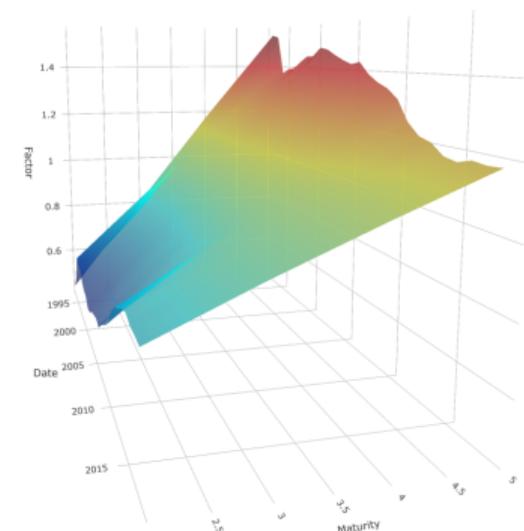
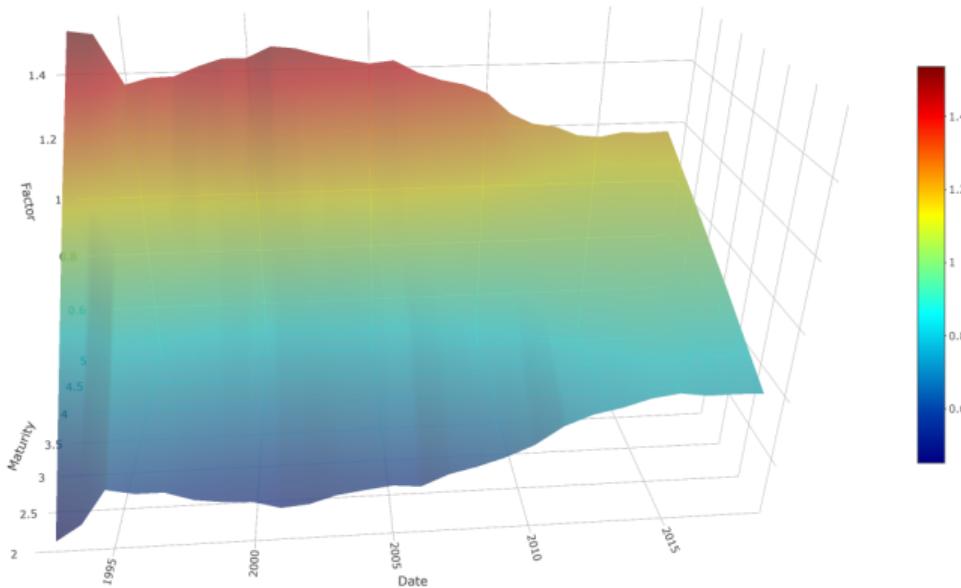
Note:

* p<0.1; ** p<0.05; *** p<0.01

Empirical Results

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

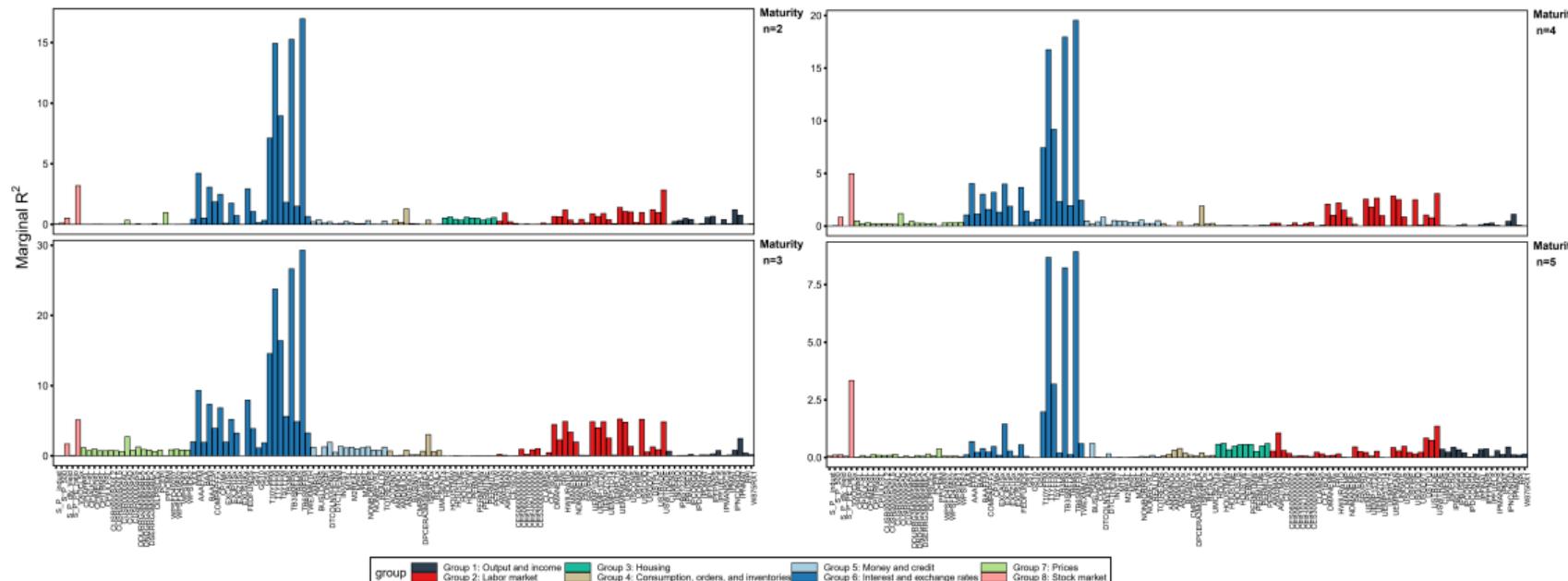
[Return](#)



Empirical Results - Economic Interpretation

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

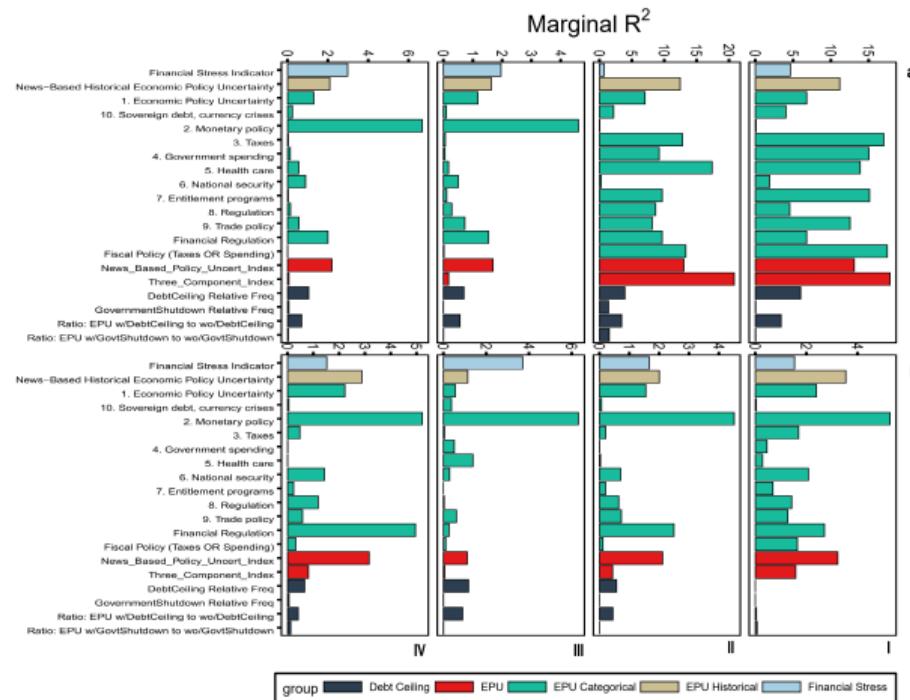
Return



Empirical Results - Economic Interpretation

Marginal R^2 Using Sentiment-Based Measures

[Return](#)



Empirical Results

Regression Coefficients of $M_{\tau^+} \hat{\delta} (\kappa^\top \hat{\xi})_{t+h/12}^{(-n), h}$ Over Time as a Function of Maturity (n)

[Return](#)

