

Bond Risk Premiums with Machine Learning

Bianchi, Büchner, and Tamoni (2021)

Key-paper review, corrigendum, and thesis research status

Presented by Soo Jung Kim | 2026.09.11

The Review of Financial Studies · Vol. 34, No. 2 · 2021

Road Map

1. **Background: Bond Prices Basics**
2. **Literature Review: Bianchi et al. (2021)**
3. **Corrigendum: What went wrong?**
4. **My Research Question**
5. **Preliminary Results**
6. **Further Analysis**

Bond Prices, Yields, and Forward Rates

Zero-coupon bond price and continuously compounded yield

$$P_t^{(n)} = e^{-ny_t^{(n)}}, \quad p_t^{(n)} \equiv \log P_t^{(n)} = -ny_t^{(n)}.$$

An increase in yields lowers the price of a fixed future payoff.

One-year excess return

$$xr_{t+12}^{(n)} = p_{t+12}^{(n-12)} - p_t^{(n)} - y_t^{(1)}.$$

Holding-period bond return minus the short-term risk-free return.

Forward rate

$$f_t^{(n)} = p_t^{(n-1)} - p_t^{(n)} = ny_t^{(n)} - (n-1)y_t^{(n-1)}.$$

The rate implied today for borrowing between $n-1$ and n .

The forward-rate curve is summarized by

$$\mathbf{x}_t = (f_t^{(1)}, f_t^{(2)}, \dots, f_t^{(10)})'.$$

Yield Decomposition and Term Structure Model

Yield Decomposition Identity

Under the log bond-pricing accounting identity, the n -period yield can be represented as the sum of two components:

$$y_t^{(n)} = \underbrace{\frac{1}{n} \sum_{j=0}^{n-1} \mathbb{E}_t \left[y_{t+j}^{(1)} \mid \mathbf{x}_t \right]}_{\text{Expected future short rates}} + \underbrace{\frac{1}{n} \sum_{j=0}^{n-1} \mathbb{E}_t \left[x r_{t+j+1}^{(n-j)} \mid \mathbf{x}_t \right]}_{\text{Expected future excess returns}}$$

- ▶ A central question in bond return predictability is whether today's yield curve contains information about future excess bond returns. The goal is to extract predictive signals from the forward-rate curve.
- ▶ Long-term yields reflect both expected future short rates and expected future bond risk premia. This study focuses on the second component (**expected future excess bond returns**).

Literature Review: Research Question

Can machine learning uncover nonlinear and high-dimensional predictability in government bond excess returns?

Traditional approach

$$\mathbb{E}_t[xr_{t+12}] = \alpha + \beta^\top \mathbf{x}_t$$

- ▶ Small predictor set
- ▶ Mostly linear relations

Machine-learning approach

$$\mathbb{E}_t[xr_{t+12}] = f_\theta(\mathbf{x}_t)$$

- ▶ Many predictors
- ▶ Nonlinearities and interactions

Data and Prediction Setting

Monthly U.S. zero-coupon Treasury yields from Liu and Wu (2019) and macroeconomic variables from FRED-MD are used.

Target Variable

The realized one-year holding-period excess return is

$$xr_t^{(n)} = p_{t+12}^{(n-12)} - p_t^{(n)} - y_t^{(1)}.$$

Predictors

$$\mathbf{x}_t = (f_t^{(1)}, \dots, f_t^{(10)})^\top$$

$\mathbf{x}_t$: one-year forward-rate curve

$\mathbf{m}_t$: macroeconomic information

$\mathbf{z}_t = [\mathbf{x}_t, \mathbf{m}_t]$: combined predictor set

Sample and Evaluation

- ▶ Sample: 1971:08–2018:12
- ▶ Maturities: 2, 3, 4, 5, 7, 10 years
- ▶ OOS evaluation begins in January 1989
- ▶ Yield-only vs. Yield + Macro

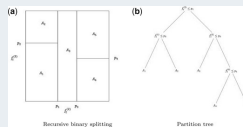
Forecasting Methodologies

Model Families & Intuition

Family	Description
Dimension Reduction (PCA, PLS)	Extracts low-dimensional linear factors; PCA is unsupervised, whereas PLS uses the target to construct predictive components.
Penalized Linear (Ridge, Lasso)	Imposes shrinkage to prevent overfitting in a linear setup.
Regression Trees (RF, Extreme Tree)	Partitions data into state-dependent regimes (nonlinear).
Neural Networks (Feed-forward)	Directly learns complex nonlinear mappings and interactions.

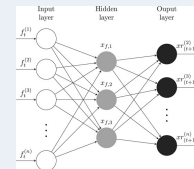
Structural Flow

1. Regression Trees



$$Z_t \xrightarrow[\text{splits}]{\text{partition}} \text{regime}_t \xrightarrow{\text{const}} \hat{x}_{t+1}^{(n)}$$

2. Neural Networks



$$Z_t \xrightarrow{\text{hidden layers}} h_t \xrightarrow{\text{output}} \hat{x}_{t+1}^{(n)}$$

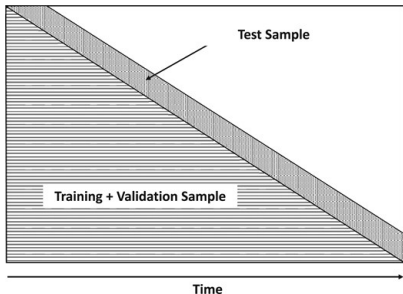
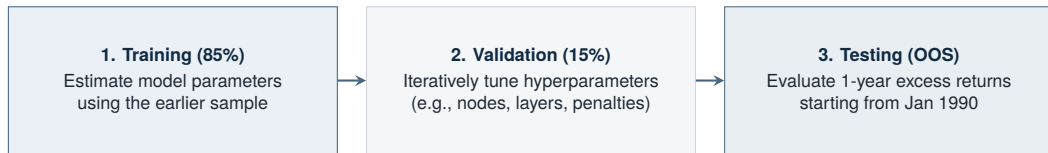
3. Penalized Linear (Ridge / Lasso)

$$\mathcal{L}(\alpha, \beta) = \frac{1}{T} \sum_{\tau \in \mathcal{T}} \left(x_{\tau}^{(n)} - \alpha - \beta^{\top} \mathbf{x}_{\tau} \right)^2 + \text{Penalty}_{\lambda}(\beta)$$

$$\text{Penalty}_{\lambda}(\beta) = \begin{cases} \lambda \|\beta\|_2^2, & \text{Ridge,} \\ \lambda \|\beta\|_1, & \text{Lasso.} \end{cases}$$

Estimation Strategy

To prevent overfitting and preserve time-series dependence, the data is partitioned sequentially using a recursive expanding window.



Expanding Window & Time-Series Dependence

- ▶ **Expanding Window:** Model is re-estimated at each time t by adding one monthly observation (ratio 85%/15% preserved).
- ▶ **Time-Series Dependence:** Chronological order is strictly preserved.

Hypothesis: Testing Out-of-Sample Predictability

Null hypothesis

$$H_0 : R_{\text{OOS}}^2 \leq 0$$

No reduction in forecast error relative to the recursively available historical mean.

Out-of-sample R^2

$$R_{\text{OOS}}^2 = 1 - \frac{\sum_{t \in \mathcal{O}} \left(x r_t^{(n)} - \hat{x} r_{t, \text{model}}^{(n)} \right)^2}{\sum_{t \in \mathcal{O}} \left(x r_t^{(n)} - \hat{x} r_{t, \text{hist}}^{(n)} \right)^2}.$$

$R_{\text{OOS}}^2 > 0$ means that the model produces smaller out-of-sample forecast errors than the historical-mean benchmark.

Results (1): Yield-Only Models

Model	2Y	3Y	4Y	5Y	7Y	10Y	EW
<i>Dimension reduction</i>							
PCA: 10 components	-53.0	-36.2	-27.5	-20.3	-12.4	-0.7	-17.7
PCA: 5 components	-54.9	-38.8	-30.4	-20.3	-15.3	-3.1	-20.0
PCA: 3 components	-17.6	-10.2	-5.3	1.0	2.9	10.2	1.0
PCA-Squared: 5 components	-55.0	-42.1	-32.5	-22.6	-16.4	-5.5	-22.4
PCA-Squared: 3 components	-46.8	-38.8	-30.6	-21.0	-17.0	-8.9	-22.5
PLS: 5 components	-56.3	-40.5	-33.4	-25.7	-18.6	-9.5	-24.8
PLS: 3 components	-57.9	-40.7	-31.9	-22.9	-14.6	-1.8	-20.2
<i>Penalized linear models</i>							
Ridge	-40.1	-25.8	-18.9	-11.4	-6.0	4.6	-9.9
Lasso	-11.5	-8.1	-2.5	0.2	2.3	9.3	2.3
Elastic Net	-10.7	-8.4	-5.1	0.4	0.3	7.0	0.7
<i>Regression trees</i>							
Gradient Boosted Tree	4.9	4.9	7.0	10.8	7.1	11.2	8.9
Random Forest	12.2	13.1	14.3	16.9	19.7	24.6	16.9
Extreme Tree	13.9	10.1	14.3	16.9	19.7	24.7	19.7
<i>Neural networks</i>							
NN: 1 layer, 3 nodes	12.7	16.4	19.5	21.6	23.1	26.4	23.0
NN: 1 layer, 5 nodes	7.0	10.7	14.6	17.2	18.9	22.9	18.2
NN: 1 layer, 7 nodes	-3.2	3.0	7.1	12.1	13.6	17.1	11.5
NN: 2 layers, 3 nodes	7.4	10.8	12.9	14.7	16.3	19.3	16.4
NN: 2 layers, 5 nodes	8.9	13.0	16.0	18.5	20.4	23.6	20.2
NN: 2 layers, 7 nodes	9.0	15.2	18.1	21.2	23.7	27.4	23.1
NN: 3 layers, 3 nodes	3.4	8.9	10.9	11.2	12.7	15.1	12.9
NN: 3 layers, 5 nodes	3.6	7.4	8.9	10.6	11.6	13.7	11.7
NN: 3 layers, 7 nodes	3.9	6.7	8.3	8.9	11.1	13.2	10.8
NN: 3 layers, (5, 4, 3) nodes	-0.9	2.8	5.6	6.5	8.3	10.6	8.0
NN: 1 layer, 7 nodes + lagged rates	6.4	14.4	17.0	19.5	20.3	24.4	20.6

Notes: Values are R^2_{OOS} in percent. EW is the equal-weighted average across maturities.

Results (2): Yield and Macro Models

Model	2Y	3Y	4Y	5Y	7Y	10Y	EW
<i>Dimension reduction</i>							
PCA: 8 macro PCs	-9.8	-2.9	0.3	3.0	3.3	4.5	1.8
PCA: Ludvigson–Ng	-3.4	0.2	1.6	1.6	-1.4	-4.7	-1.3
PLS: 8 components	-40.7	-19.7	-12.0	-8.2	-2.7	3.4	-6.4
<i>Penalized linear models</i>							
Ridge: CP factor	-45.3	-23.6	-16.7	-13.2	-3.1	5.3	-5.6
Lasso: CP factor	6.4	11.2	12.9	14.4	19.6	23.7	21.0
Elastic Net: CP factor	6.4	11.0	14.3	15.7	21.7	29.1	22.0
Ridge: forward rates	-52.2	-28.7	-22.7	-18.3	-13.1	-3.5	-15.4
Lasso: forward rates	11.0	12.0	12.3	16.4	19.9	23.6	20.7
Elastic Net: forward rates	10.2	14.2	16.0	13.2	19.9	23.6	21.0
<i>Regression trees</i>							
Gradient Boosted Tree	13.1	15.9	18.1	23.8	22.5	25.5	26.2
Random Forest	26.7	21.5	22.0	24.4	20.0	25.0	26.6
Extreme Tree	23.0	23.4	22.3	23.7	29.9	29.6	29.2
<i>Neural networks</i>							
NN: 1 layer, forward rates direct	6.0	13.6	17.8	22.0	22.5	26.5	22.4
NN: 2 layers, forward rates direct	16.6	21.5	24.9	28.0	29.4	32.3	29.3
NN: 3 layers, forward rates direct	24.8	26.3	29.7	32.0	31.7	33.7	32.5
NN: 1 layer, forward-rate net	8.4	19.0	23.8	25.6	27.2	29.5	27.2
NN: 2 layers, forward-rate net	12.1	15.7	20.0	23.5	25.4	28.1	25.1
NN: 3 layers, forward-rate net	7.6	16.3	20.2	23.7	25.0	28.1	25.0
NN: 1 layer, group ensemble direct	12.6	17.3	21.6	24.2	25.9	29.6	25.9
NN: 1 layer, group ensemble net	20.0	25.6	29.5	31.2	33.6	36.3	34.0
NN: 2 layers, group ensemble net	17.3	23.6	27.8	29.8	31.0	33.0	31.6
NN: 3 layers, group ensemble net	13.6	20.0	23.7	26.1	27.5	30.7	27.9

Main result: Nonlinear neural networks perform best, especially when yield and macroeconomic information are modeled through group ensembling.

Corrigendum: Yield-Only Models

What Was Corrected?

- ▶ **The Flaw:** The original analysis prematurely fed future-overlapping 1-year excess returns ($h = 12$ months) into the training set during the expanding window.
- ▶ **Look-Ahead Leakage:** Due to high interest rate autocorrelation, models implicitly peek at upcoming yield paths, leading to artificially inflated out-of-sample R^2 . The corrigendum enforces a 12-month target-availability cutoff.

Models / Maturities	2Y	3Y	4Y	5Y	7Y	10Y	EW
<i>Selected corrected results: R^2_{OOS} (percent)</i>							
PCA: 10 components	-92.8	-71.0	-59.5	-49.9	-37.4	-20.8	-45.2
Ridge	-67.0	-49.3	-40.8	-31.4	-24.5	-10.8	-29.5
Gradient boosted tree	-64.9	-73.8	-75.1	-64.4	-62.4	-53.2	-64.2
NN: 1 layer, 3 nodes	2.6	4.3	4.5	4.4	4.4	3.8	4.4
NN: 2 layers, 7 nodes	6.8	6.8	6.4	5.5	5.6	4.6	5.9
Inverse validation-loss weighted model	4.8	5.5	5.5	4.9	4.8	4.0	5.0

Main result: Despite correction, nonlinear neural networks and ensemble models still retain positive and economically meaningful out-of-sample performance (e.g., up to 5.0% EW R^2_{OOS}).

Limitation: Once future information leakage was blocked, static neural network gains dropped materially, proving that rigorous time-series alignment is critical.

Corrigendum: Yield + Macro Models

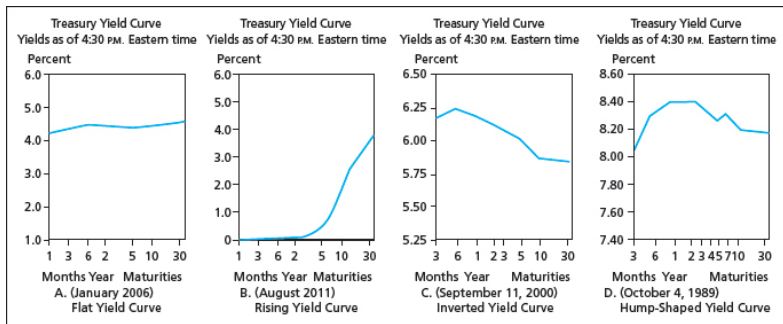
Model	2Y	3Y	4Y	5Y	7Y	10Y	EW
<i>Selected corrected results: R^2_{OOS} (percent)</i>							
PCA: first 8 PCs	-24.3	-12.8	-7.8	-3.8	-2.7	-1.8	-4.7
Ridge: CP factor	-122.6	-82.8	-64.4	-51.7	-35.2	-18.7	-41.3
Ridge: forward rates	-146.9	-111.6	-99.8	-87.7	-73.5	-52.9	-79.5
Gradient boosted tree	-17.4	-14.0	-9.6	-6.3	-7.7	-11.4	-7.9
NN 1 layer, fwd rates direct	3.4	3.2	8.6	16.7	21.1	28.2	19.6
NN 1 layer, fwd rates net	0.1	15.0	15.5	16.7	16.0	22.1	18.2
NN 1 layer group ensemble, net	3.6	8.9	11.3	10.9	12.9	17.0	13.3
Equally weighted model	7.1	15.1	16.2	18.4	20.1	24.0	20.3
Inverse validation-loss weighted model	6.0	14.0	15.2	17.4	19.3	23.3	19.4

Main result: Macroeconomic variables add information beyond the yield curve, with macro-aware neural-network combinations outperforming yield-only models.

Limitation: The apparent macro value is difficult to interpret structurally because it is intertwined with flexible architecture choices and forecast-weighting rules.

But is today's yield curve the whole story?

[Figure 1] Term Structure



Level Parallel shifts: common movement in discount rates across maturities.

Slope Long-short spread: reflecting expectations about future short rates.

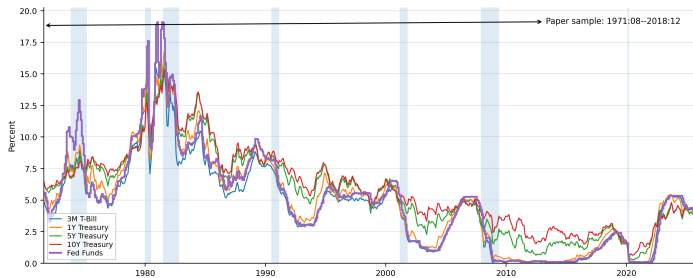
Curvature Middle maturities: compact signal of nonlinear term-structure shifts.

Why history may matter: The term structure changes every time; the same current level, slope, and curvature can arise after very different tightening, easing, or recession paths.



Eldrich, E. (<https://ealdrich.github.io/Teaching/Econ133/LectureNotes/termStructure.html>).

[Figure 2] Economic Intuition: Interest Rate Paths



- ▶ U.S. interest rates and macroeconomic conditions followed very different paths across tightening, easing, and recession episodes.
- ▶ This motivates using an RNN to summarize historical yield-curve and macroeconomic states, rather than relying only on the current state.

Source: Federal Reserve Bank of St. Louis, FRED-MD monthly database, 2026-04 vintage. Recession shading: NBER recession dates.

New Research Question

Bianchi et al. (2021) use a feed-forward NN as a flexible nonlinear mapping from the current forward-rate curve to expected excess returns. But is the current curve a sufficient state?

My Research Question Once the nonlinear current-state mapping is controlled for, does the **path by which the current yield curve was reached** contain incremental predictive information?

$$\mathbb{E}[xr_{t \rightarrow t+12}^{(n)} \mid \mathbf{x}_t, \mathcal{H}_t^x] \stackrel{?}{=} \mathbb{E}[xr_{t \rightarrow t+12}^{(n)} \mid \mathbf{x}_t], \quad \mathcal{H}_t^x = \{\mathbf{x}_{t-L}, \dots, \mathbf{x}_{t-1}\}.$$

If the current curve is a sufficient state, history should add nothing.

1. Static NN

$$\mathbf{x}_t \rightarrow xr$$

Tests whether a flexible nonlinear mapping of the **current curve** predicts returns.

2. Lagged FFNN

$$(\mathbf{x}_{t-L}, \dots, \mathbf{x}_t) \rightarrow xr$$

Tests whether adding more historical observations is enough, **without recurrence**.

3. NN + RNN

$$\mathbf{x}_t \rightarrow \text{NN}, \quad \Delta \mathbf{x}_{t-L+1:t} \rightarrow \text{RNN}$$

Tests whether **ordered historical dynamics** add information beyond the current-state NN.

Model Extension: NN + RNN

Static Current-State NN

Current-state predictors:

$$\hat{x}r_{t,\text{NN}}^{(n)} = g_n(\mathbf{x}_t)$$

Macro extension:

$$\hat{x}r_{t,\text{NN}}^{(n)} = g_n(\mathbf{x}_t, \mathbf{m}_t)$$

Historical-Path RNN

Yield-path specification:

$$\mathbf{h}_t^x = \Phi_\theta(\Delta \mathbf{x}_{t-L+1:t})$$

Macro-path extension:

$$\mathbf{h}_t^{xm} = \Phi_\theta(\Delta \mathbf{x}_{t-L+1:t}, \mathbf{m}_{t-L+1:t})$$

Path contribution:

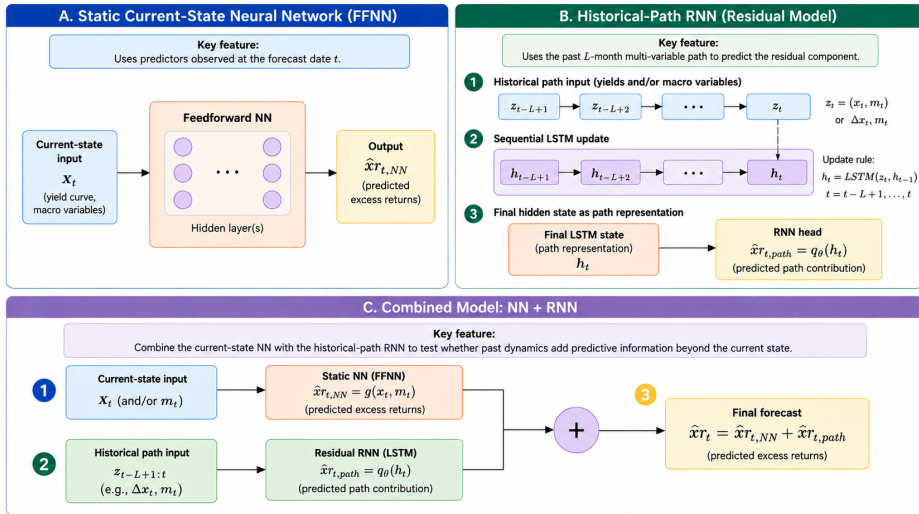
$$\hat{x}r_{t,\text{path}}^{(n)} = q_{n,\theta}(\mathbf{h}_t)$$

Combined NN + RNN Forecast

$$\hat{x}r_t^{(n)} = \underbrace{\hat{x}r_{t,\text{NN}}^{(n)}}_{\text{current state}} + \underbrace{\hat{x}r_{t,\text{path}}^{(n)}}_{\text{incremental path information}}$$

The static NN captures nonlinear information in the current state, while the RNN tests whether historical paths contain additional predictive information beyond the current state.

[Figure 3] Empirical Design: NN vs. RNN



New Hypothesis and Model Selection

Null hypothesis

$$H_0 : R_{\text{OOS}, NN+RNN|NN}^{2,(n)} \leq 0 \quad \text{vs.} \quad H_1 : R_{\text{OOS}, NN+RNN|NN}^{2,(n)} > 0.$$

Out-of-sample R^2

$$R_{\text{OOS}, NN+RNN|NN}^{2,(n)} = 1 - \frac{\sum_{t \in \mathcal{O}} (xr_t^{(n)} - \widehat{xr}_{t, NN+RNN}^{(n)})^2}{\sum_{t \in \mathcal{O}} (xr_t^{(n)} - \widehat{xr}_{t, NN}^{(n)})^2}.$$

Model selection

- ▶ 100 random initializations
- ▶ Rank models using validation loss only
- ▶ Average the best 10 forecasts
- ▶ Hyperparameters fixed before viewing test performance

[Table 1] Preliminary Results

These results replicate the authors' original results (before corrigendum).

Model	R^2_{OOS}						p -value					
	$xr^{(2)}$	$xr^{(3)}$	$xr^{(4)}$	$xr^{(5)}$	$xr^{(7)}$	$xr^{(10)}$	$xr^{(2)}$	$xr^{(3)}$	$xr^{(4)}$	$xr^{(5)}$	$xr^{(7)}$	$xr^{(10)}$
<i>Panel A: Yield-Only Models</i>												
Static NN: $\mathbf{x}_t \rightarrow xr_t$	2.7%	2.0%	5.7%	5.6%	8.8%	3.9%	0.09	0.12	0.04	0.02	0.02	0.05
Lagged NN: $\mathbf{x}_{t-12:t-1} \rightarrow xr_t$	1.9%	-2.6%	-2.2%	-1.7%	-2.9%	3.0%	0.07	—	—	—	—	0.05
State-transition RNN: $\mathbf{x}_{t-12:t-1} \rightarrow \mathbf{h}_t \rightarrow xr_t$	1.8%	-8.1%	3.3%	6.3%	9.3%	6.9%	0.01	—	0.03	0.02	0.01	0.01
<i>Panel B: Yield + Macro Models</i>												
Static NN all-together: $(\mathbf{x}_t, \mathbf{m}_t) \rightarrow xr_t$	15.4%	19.4%	20.7%	19.8%	20.8%	20.8%	0.00	0.00	0.00	0.00	0.00	0.00
Static NN f-net: $\mathbf{x}_t \rightarrow \mathbf{z}_t^x, \mathbf{m}_t \rightarrow \mathbf{z}_t^m$	16.7%	17.0%	19.9%	20.8%	20.0%	21.8%	0.01	0.01	0.01	0.01	0.01	0.01
RNN all-together: $(\mathbf{h}_t, \mathbf{m}_t) \rightarrow xr_t$	-10.3%	-10.5%	-8.5%	-4.6%	-1.4%	0.0%	—	—	—	—	—	—
RNN current macro f-net: $\mathbf{h}_t \rightarrow \mathbf{z}_t^h, \mathbf{m}_t \rightarrow \mathbf{z}_t^m$	-4.1%	-6.9%	-7.3%	-9.1%	-5.7%	2.6%	—	—	—	—	—	0.01
RNN macro path f-net: $(\mathbf{h}_t^x, \mathbf{h}_t^m) \rightarrow xr_t$	24.5%	22.9%	22.1%	22.4%	17.9%	21.3%	0.00	0.00	0.00	0.00	0.00	0.00
RNN joint path f-net: $\mathbf{h}_t^{xm} \rightarrow xr_t$	35.0%	32.6%	30.2%	27.8%	26.6%	22.6%	0.00	0.00	0.00	0.01	0.01	0.01

Main result: Recurrent Neural Networks demonstrate superior out-of-sample predictability relative to standard NN models.

Limitation: Because this replicates the authors' exact baseline setup, this preliminary finding remains subject to look-ahead bias.

[Table 2] Preliminary Results (Corrigendum)

Model	R^2_{OOS}							p -value						
	$xr^{(2)}$	$xr^{(3)}$	$xr^{(4)}$	$xr^{(5)}$	$xr^{(7)}$	$xr^{(10)}$	EW	$xr^{(2)}$	$xr^{(3)}$	$xr^{(4)}$	$xr^{(5)}$	$xr^{(7)}$	$xr^{(10)}$	EW
Bianchi NN (1L–3)	2.6%	4.3%	4.5%	4.4%	4.4%	3.8%	4.4%	0.168	0.101	0.095	0.088	0.106	0.126	0.104
Our Static NN (1L–3)	1.1%	3.0%	4.0%	4.1%	4.2%	3.9%	4.1%	–	–	–	–	–	–	–
Pure sequence RNN (10-seed diagnostic)	-15.0%	-2.8%	-24.7%	-20.7%	-6.1%	-5.9%	-11.7%	–	–	–	–	–	–	–
Static NN (pilot-matched baseline)	6.0%	6.6%	6.7%	6.3%	6.2%	5.3%	6.4%	–	–	–	–	–	–	–
Static NN + Historical-path RNN	-9.0%	-5.7%	-2.8%	3.5%	9.8%	17.0%	7.7%	–	–	–	–	–	–	–
Incremental $R^2_{\text{OOS}, \text{NN}+\text{RNN} \text{NN}}$	-15.9%	-13.2%	-10.2%	-3.0%	3.9%	12.4%	1.4%	–	–	–	–	–	–	–

Preliminary pattern: A pure RNN model performs poorly, while retaining the static NN and adding a historical-path RNN appears more promising, particularly at longer maturities, where the pilot forecast becomes positive and the EW portfolio reaches 7.7% R^2_{OOS} .

Important: This is a **pilot architecture**, not the final hypothesis test. It was produced before the final 100-seed / best-10 protocol and before all NN and RNN training conventions were fully harmonized. I therefore do not interpret the short-maturity failures or the long-maturity gains as final evidence.

Economic Interpretation Challenge

Hypothesis test

$$H_0 : R_{\text{OOS,RNN|NN}}^{2,(n)} \leq 0 \quad \text{vs.} \quad H_1 : R_{\text{OOS,RNN|NN}}^{2,(n)} > 0.$$

The pilot NN+RNN results are **partially consistent with the alternative hypothesis**.

Main empirical findings

- ▶ **Current-state benchmark:** The static NN retains positive out-of-sample predictability after correcting the information set.
- ▶ **NN + RNN:** Pilot results suggest that historical yield-curve dynamics may contain **incremental predictive information**, particularly at longer maturities.

Challenge: Promising predictions, but what economic state does the RNN extract from the historical path?

Possible mechanisms

- ▶ Monetary-policy cycle
- ▶ Time-varying risk premia

Alternative explanations

- ▶ Proxy for omitted variables
- ▶ Flexibility of architecture

Further Analysis

Planned Extensions and Robustness Checks

1. **Model Comparison:** Test robustness across alternative model specifications, architectures, and training procedures.
 - **Model 1:** NN + RNN
 - **Model 2:** State-transition RNN (with an auxiliary task that predicts the current yield curve from historical paths)
 - **Model 3:** Pure sequence RNN (varying the number of layers and hidden units, with 100-seed model averaging)
2. **Macroeconomic Extension:** Add macroeconomic information to Models 1–3 and compare current macro variables with historical macro paths.
3. **Extended Sample:** Extend the sample beyond 2018 to include recent monetary tightening and easing cycles.

Conclusion

1. Bianchi et al. show that a **nonlinear mapping of the current term structure** can predict bond excess returns, although corrected OOS gains are much smaller.
2. The corrected pure-RNN diagnostic shows that **history should not replace the current curve** ; the current curve remains a strong forecasting anchor.
3. My main test therefore asks whether an RNN can extract **incremental historical-path information** after conditioning on the static NN forecast.
4. The central empirical question becomes whether the current yield curve is a **sufficient state** for bond return predictability.

NN: What does today's yield curve imply?

RNN: After conditioning on today, does how we got here still matter?

Thank you

Questions and discussion