Asymmetric Violations of the Spanning Hypothesis

Raul Riva[1]
(joint with Gustavo Freire[2])

[1]Northwestern University
[2]Erasmus University

Northwestern Job Market Preparation

September 6th, 2024

Intro

- Yield curve dynamics is of major interest both for policy makers and market participants:
  - Monetary policy transmission and fiscal policy assessment;
  - Risk management and long-term investment decisions;
  - Risk premia measurement and portfolio allocation;

- Workhorse: arbitrage-free Dynamic Term Structure Models;

Intro

- Yield curve dynamics is of major interest both for policy makers and market participants:
  - Monetary policy transmission and fiscal policy assessment;
  - Risk management and long-term investment decisions;
  - Risk premia measurement and portfolio allocation;

- Workhorse: arbitrage-free Dynamic Term Structure Models;

- Pervasive feature: the Spanning Hypothesis;
  - Evolution of yields reveals the evolution of underlying risk factors;
  - Today's yield curve is **all you need** to forecast future yields (and bond returns!);

Intro

- Yield curve dynamics is of major interest both for policy makers and market participants:
  - Monetary policy transmission and fiscal policy assessment;
  - Risk management and long-term investment decisions;
  - Risk premia measurement and portfolio allocation;
- Workhorse: arbitrage-free Dynamic Term Structure Models;
- Pervasive feature: the Spanning Hypothesis;
  - Evolution of yields reveals the evolution of underlying risk factors;
  - Today's yield curve is **all you need** to forecast future yields (and bond returns!);
- Leads to important repercussions:
  - Yield forecasting and portfolio allocation through simple models should be preferred;
  - Hedging interest-rate derivatives with few assets;
  - Model-implied prices of risk are empirically very different;

The Current Debate

Can other state variables help forecast bond returns and yields *given* current yields?

The Current Debate

Can other state variables help forecast bond returns and yields *given* current yields?

- Literature so far often offers a binary answer:
  - **Yes**: Cooper and Priestley (2009), Ludvigson and Ng (2009), Joslin et al. (2014), Cieslak and Povala (2015), Greenwood and Vayanos (2014), Fernandes and Vieira (2019), Huang and Shi (2023)
  - **Probably Not**: Duffee (2013), Bauer and Hamilton (2018) $\implies$ inference is usually hard!

The Current Debate

Can other state variables help forecast bond returns and yields *given* current yields?

- Literature so far often offers a binary answer:
  - **Yes**: Cooper and Priestley (2009), Ludvigson and Ng (2009), Joslin et al. (2014), Cieslak and Povala (2015), Greenwood and Vayanos (2014), Fernandes and Vieira (2019), Huang and Shi (2023)
  - **Probably Not**: Duffee (2013), Bauer and Hamilton (2018) $\Longrightarrow$ inference is usually hard!

This paper: more nuanced answer - *asymmetric* violations;

- Out-of-sample forecasts + data-rich environment;
- **Stronger** violations at the **shorter end** of the yield curve;
- No evidence of violations at the longer end of the yield curve;
- Violations are economically meaningful for a mean-variance investor;
- Evidence on one possible mechanism: monetary policy;

Flight Plan

1. Literature Review and Data
2. Forecast bond excess returns
   - Can we improve risk premium measurement by incorporating macro data?
3. Decompose bond risk premia and forecast factors
   - Novel decomposition of bond returns using Nelson-Siegel factors;
   - Different machine-learning methods showcase where predictability comes from;
4. How much are we leaving on the table?
   - Macro signals provide asymmetric Sharpe Ratio improvements when trading bonds;
5. Time-varying violations
   - Violations are predicted by deviation from Taylor Rule;

Literature Review and Data

Literature

- Bond return forecasting and testing of different theories about the yield curve:
  - Cochrane and Piazzesi (2005), Cooper and Priestley (2009), Ludvigson and Ng (2009), Joslin et al. (2014), Greenwood and Vayanos (2014), Cieslak and Povala (2015), Bauer and Hamilton (2018), Bianchi et al. (2021), Hoogteijling et al. (2021), van der Wel and Zhang (2021), Borup et al. (2023), Huang and Shi (2023)
- Nelson-Siegel modeling
  - Nelson and Siegel (1987), Diebold and Li (2006), Diebold et al. (2006), Moench (2008), Diebold and Rudebusch (2013), van Dijk et al. (2013), Hännikäinen (2017), Fernandes and Vieira (2019)
- Machine Learning methods for forecasting in Finance and Economics
  - Gu et al. (2020), Medeiros et al. (2021), Bianchi et al. (2021), Giannone et al. (2021), Goulet-Coulombe (2023), Goulet-Coulombe et al. (2023), Filippou et al. (2023), Shen and Xiu (2024)
- **Contribution**: asymmetry in out-of-sample forecasting in data-rich environments

Data

Yield curve data:

- Taken from Liu and Wu (2021) - Full sample: 1973-2021; Out-of-sample: 1990-2021;
- Constructed from CRSP data;
- Provides longer maturities than Fama and Bliss (1987)
- Lower fitting errors than Gurkaynak et al. (2007)

Macroeconomic data:

- FRED-MD, by St. Louis Fed [McCracken and Ng (2016)]
- Monthly frequency, a total of 126 variables covering different groups of variables
- Price indexes, output and unemployment measures, real estate market indicators, exchange rates, monetary aggregates, inventories and investment measures, credit spreads...

Forecasting Bond Returns

Forecasting Excess Bond Returns

- Let $y_t^{(n)}$ be the n-year zero-coupon rate at month t;
- The 1-year excess bond returns for a maturity of n years are given by:

$$xr_{t+12}(n) \equiv n \cdot y_t^{(n)} - (n-1) \cdot y_{t+12}^{(n-1)} - y_t^{(1)} \tag{1}$$

Forecasting Excess Bond Returns

- Let $y_t^{(n)}$ be the n-year zero-coupon rate at month t;
- The 1-year excess bond returns for a maturity of n years are given by:

$$xr_{t+12}(n) \equiv n \cdot y_t^{(n)} - (n-1) \cdot y_{t+12}^{(n-1)} - y_t^{(1)} \tag{1}$$

- Estimate a linear model with an expanding sample forecasting design:

$$xr_{t+12}(n) = \alpha_n + \theta_n' C_t + \gamma_n' PC_t + \epsilon_{t+12,n} \tag{2}$$

- C_t controls for the yield curve using forward rates $f_t(n) = n \cdot y_t^{(n)} - (n-1) \cdot y_t^{(n-1)}$;
- PC_t are principal components extracted from the FRED-MD data set;

Forecasting Excess Bond Returns

- Let $y_t^{(n)}$ be the n-year zero-coupon rate at month t;
- The 1-year excess bond returns for a maturity of n years are given by:

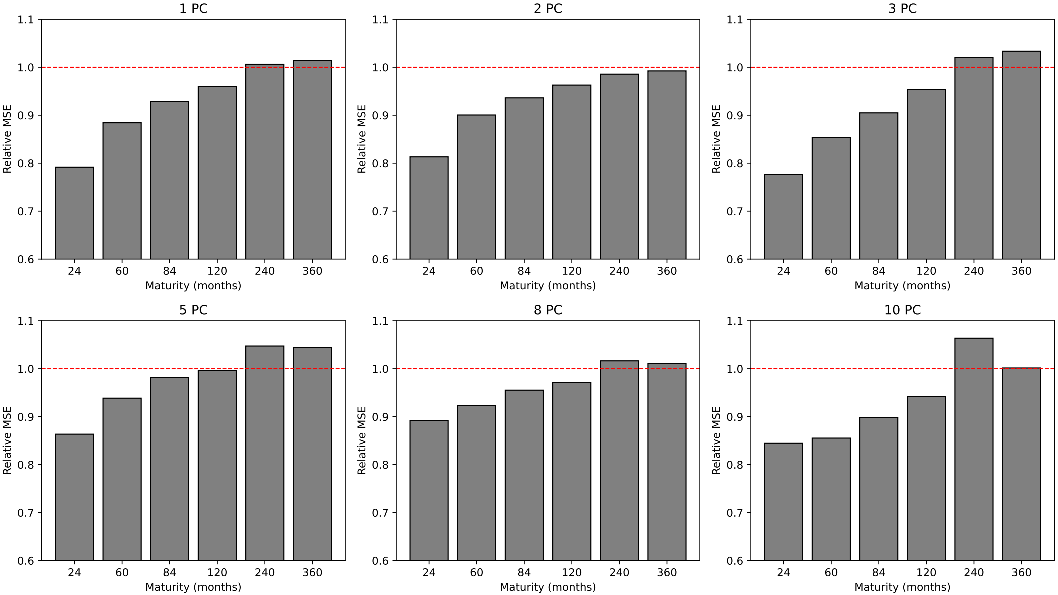
$$xr_{t+12}(n) \equiv n \cdot y_t^{(n)} - (n-1) \cdot y_{t+12}^{(n-1)} - y_t^{(1)} \tag{1}$$

- Estimate a linear model with an expanding sample forecasting design:

$$xr_{t+12}(n) = \alpha_n + \theta_n' C_t + \gamma_n' PC_t + \epsilon_{t+12,n} \tag{2}$$

- C_t controls for the yield curve using forward rates $f_t(n) = n \cdot y_t^{(n)} - (n-1) \cdot y_t^{(n-1)}$;
- PC_t are principal components extracted from the FRED-MD data set;
- Spanning hypothesis: allowing for $\gamma_n \neq 0$ should not improve the forecast of $xr_{t+12}(n)$;
- Previous literature focuses on testing $\gamma_n = 0$. We focus directly on $\widehat{xr}_{t+12}(n)$;

MSE With/Without Macro Data



Other Controls | p-values | In-sample | Ragged Edge | GSW Yields

Decomposing Bond Risk Premia

Modeling Yields

- Forecasting returns amounts to forecasting $y_{t+12}^{(n)}$;
- Yields are well described by a low-rank factor structure (Litterman and Scheinkman (1991))

Modeling Yields

- Forecasting returns amounts to forecasting $y_{t+12}^{(n)}$;
- Yields are well described by a low-rank factor structure (Litterman and Scheinkman (1991))

Assume a Nelson-Siegel model for yields as in Diebold and Li (2006):

$$y_t^{(\tau)} = \beta_{1,t} + \beta_{2,t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_{3,t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{3}$$

- $\lambda > 0$ is a decay parameter, assumed constant;
- β_1 is a long-run factor: $\lim_{\tau\to\infty} y_t^{(\tau)} = \beta_{1,t}$;
- β_2 is a short-run factor: loading decreases with τ;
- β_3 is a medium-run factor: hump-shaped loadings;

Modeling Yields

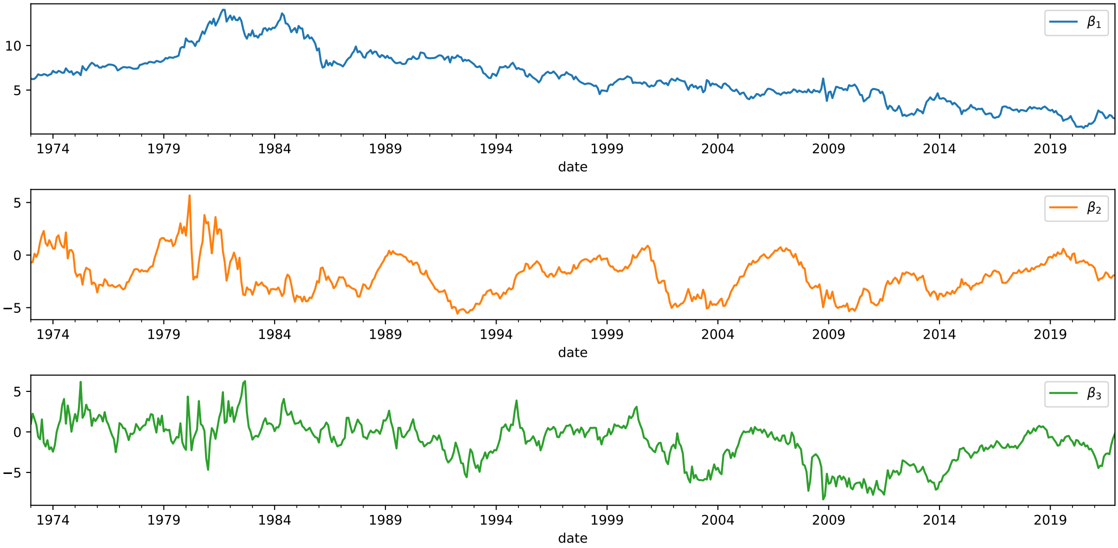
- Forecasting returns amounts to forecasting $y_{t+12}^{(n)}$;
- Yields are well described by a low-rank factor structure (Litterman and Scheinkman (1991))

Assume a Nelson-Siegel model for yields as in Diebold and Li (2006):

$$y_t^{(\tau)} = \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \tag{3}$$

- $\lambda > 0$ is a decay parameter, assumed constant;
- β_1 is a long-run factor: $\lim_{\tau\to\infty} y_t^{(\tau)} = \beta_{1,t}$;
- β_2 is a short-run factor: loading decreases with τ;
- β_3 is a medium-run factor: hump-shaped loadings;
- Set $\lambda = 0.0609$; Estimate using OLS date by date, $1 \leq \tau \leq 120$;
- OLS estimates $\implies$ β's are linear combinations of yields;

Factor Realizations (1973-2021)



▸ Estimation details ▸ Alternative Estimation Procedures ▸ Polynomial Model

Decomposing Returns

Proposition 1

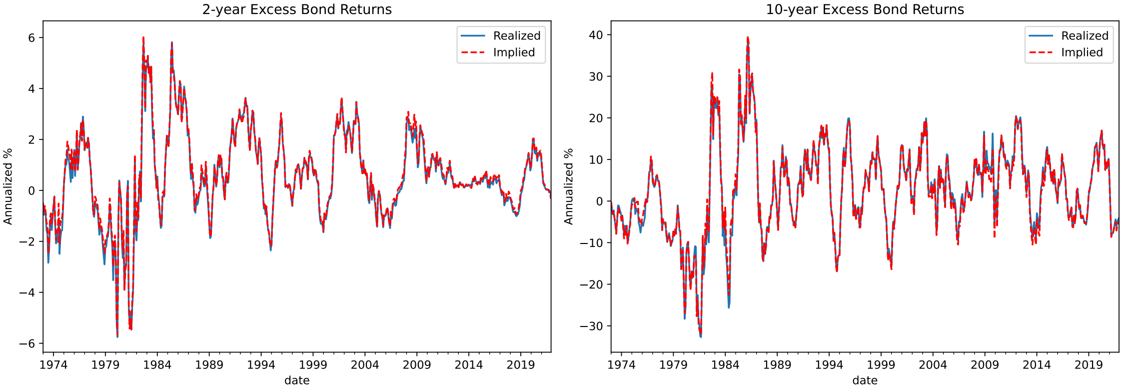
Assume the Nelson-Siegel representation with $\lambda > 0$. Then, $xr_{t+12}(n)$ can be written as:

$$
\begin{aligned}
xr_{t+12}(n) = {} & (n-1)\Big[\beta_{1,t} - \beta_{1,t+12}\Big] \\
& + \left(\frac{1-e^{-\theta(n-1)}}{\theta}\right)\Big[e^{-\theta}\beta_{2,t} - \beta_{2,t+12}\Big] \\
& + \left(\frac{1-e^{-\theta(n-1)}}{\theta} - ne^{-\theta(n-1)} + 1\right)\Big[e^{-\theta}\beta_{3,t} - \beta_{3,t+12}\Big] + \left(1-e^{-\theta(n-1)}\right)\beta_{3,t+12}
\end{aligned}
\tag{4}
$$

where $\theta \equiv 12 \cdot \lambda$.

- Terms in parentheses are **not** time-varying and brackets **do not** depend on the maturity

Is this approximation any good?



- Blue: $xr_t(n)$ observed from data for $n = 2$ and $n = 10$;
- Red: $xr_t(n)$ implied by our estimates of the factors + Proposition 1;
- The Fed actually uses a variant of the NS model to report their yield curve

Forecasting Nelson-Siegel Factors

- OLS factor estimation implies that β's are **linear combinations of yields**;
- Under the spanning hypothesis: macro data should not be helpful to forecast factors:

$$\beta_{i,t+12} = \alpha_i + \theta_i' C_t + \gamma_i' PC_t + \epsilon_{i,t+12}, \quad i \in \{1,2,3\} \tag{5}$$

Forecasting Nelson-Siegel Factors

- OLS factor estimation implies that β's are **linear combinations of yields**;
- Under the spanning hypothesis: macro data should not be helpful to forecast factors:

$$\beta_{i,t+12} = \alpha_i + \theta_i' C_t + \gamma_i' PC_t + \epsilon_{i,t+12}, \quad i \in \{1, 2, 3\} \tag{5}$$

- We use the out-of-sample R^2 to measure the forecasting ability:

$$R^2_{oos} = 1 - \frac{\sum_{t=t_0}^{T} \left(\beta_{i,t} - \widehat{\beta}_{i,t}\right)^2}{\sum_{t=t_0}^{T} \left(\beta_{i,t} - \overline{\beta}_{i,t}\right)^2} \tag{6}$$

- $\overline{\beta}_{i,t}$ is a benchmark model: for example a random walk;
- OOS period: 1990-2021, with a recursive forecasting approach (384 total forecasts);
- We use a Diebold-Mariano test to make inference about any forecasting improvement;

Out-of-Sample PCA-based Forecasts

Innovations + other controls | In-sample analysis

Table: R^2 out-of-sample against a random walk and Diebold-Mariano p-values

		Number of Macro PCs					p-values				
Target	No Macro	1	2	3	4	5	1	2	3	4	5
β_1	-0.21	-0.17	-0.19	-0.15	-0.11	-0.09	0.18	0.33	0.13	0.11	0.10
β_2	-0.08	-0.08	**0.17**	**0.22**	**0.21**	**0.23**	0.49	0.01	0.02	0.02	0.02
β_3	-0.12	-0.15	-0.06	-0.07	-0.07	-0.07	0.92	0.07	0.19	0.20	0.21

- Improving over a random walk is hard, but possible for (and *only* for) β_2
- Asymmetry in bond return predictability happens because a single factor can be predicted;
- Result holds if we allow for even more PCs, but we lose statistical power

Leveraging Regularization Methods

- PCA is not "supervised": dimensionality reduction decoupled from prediction;
- Regularization: penalize complex models $\implies$ bias-variance trade-off;
- Model selection tells exactly what variables were chosen to compose forecasts;

Leveraging Regularization Methods

- PCA is not "supervised": dimensionality reduction decoupled from prediction;
- Regularization: penalize complex models $\Longrightarrow$ bias-variance trade-off;
- Model selection tells exactly what variables were chosen to compose forecasts;

Let $\psi_1, \psi_2 \geq 0$ be scalars and let $||.||_p$ be the L^p norm. Consider the minimization:

$$\min_{\alpha_i, \gamma_i} \left\{ \frac{1}{T - 12 - t_0} \sum_{t=t_0}^{T-12} \left(\beta_{i,t+12} - \alpha_i - \gamma_i' X_t \right)^2 + \underbrace{\psi_1 \cdot ||\gamma_i||_1 + \psi_2 \cdot ||\gamma_i||_2}_{\text{model complexity penalty}} \right\} \tag{7}$$

$$\widehat{\beta}_{i,t+12} = \widehat{\alpha}_i + \widehat{\gamma}_i' X_t \tag{8}$$

Leveraging Regularization Methods

- PCA is not “supervised”: dimensionality reduction decoupled from prediction;
- Regularization: penalize complex models $\implies$ bias-variance trade-off;
- Model selection tells exactly what variables were chosen to compose forecasts;

Let $\psi_1, \psi_2 \geq 0$ be scalars and let $||.||_p$ be the L^p norm. Consider the minimization:

$$\min_{\alpha_i, \gamma_i} \left\{ \frac{1}{T - 12 - t_0} \sum_{t=t_0}^{T-12} \left(\beta_{i,t+12} - \alpha_i - \gamma_i' X_t\right)^2 + \underbrace{\psi_1 \cdot ||\gamma_i||_1 + \psi_2 \cdot ||\gamma_i||_2}_{\text{model complexity penalty}} \right\} \tag{7}$$

$$\widehat{\beta}_{i,t+12} = \widehat{\alpha}_i + \widehat{\gamma}_i' X_t \tag{8}$$

1. $\psi_1 = 0, \psi_2 > 0 \implies$ Ridge
2. $\psi_1 > 0, \psi_2 = 0 \implies$ Lasso
3. $\psi_1, \psi_2 > 0 \implies$ Elastic Net

- Calibrate ψ_1, ψ_2 using a 80-20 split validation set for each date t using grid search.

Regularization Methods - Performance

Table: R^2 out-of-sample of regularized linear models

Target	No Macro Data			All Macro Data			p-value		
	Ridge	Lasso	Elastic Net	Ridge	Lasso	Elastic Net	Ridge	Lasso	Elastic Net
β_1	-4.84	-4.82	-4.69	-4.06	-4.30	-4.18	0.00	0.00	0.00
β_2	-0.08	-0.13	-0.19	**0.07**	**0.07**	**0.06**	0.05	0.00	0.01
β_3	-0.41	-0.59	-0.59	-0.47	-0.46	-0.45	0.78	0.04	0.03
$\Delta\beta_1$	0.12	0.12	0.09	0.01	0.12	0.12	0.96	0.50	0.27
$\Delta\beta_2$	0.01	-0.02	-0.01	**0.15**	**0.22**	**0.19**	0.02	0.00	0.00
$\Delta\beta_3$	0.04	-0.02	-0.03	-0.13	-0.09	-0.08	1.00	0.95	0.95

- No matter the target and the method, predictability through β_2;

▸ Other Controls ▸ Stationarity Matters ▸ Chosen variables

What about non-linearities? Maybe important for the long end?

- So far: linear methods... but theory implies a richer restriction;

What about non-linearities? Maybe important for the long end?

- So far: linear methods... but theory implies a richer restriction;
- Flexibility vs computational cost + low signal-to-noise ratio environment + small sample;
- Our choice: a **Random Forest** with the CART algorithm from Breiman (1995)
  - Sequentially pick the variable and splitting point that minimizes the in-sample MSE
  - Procede until each observation belongs to a single final node
  - Each tree is a forecaster with low bias and high variance
- We fit 500 trees for each date and predict using the average - an *ensemble* estimator

What about non-linearities? Maybe important for the long end?

- So far: linear methods... but theory implies a richer restriction;
- Flexibility vs computational cost + low signal-to-noise ratio environment + small sample;
- Our choice: a **Random Forest** with the CART algorithm from Breiman (1995)
  - Sequentially pick the variable and splitting point that minimizes the in-sample MSE
  - Procede until each observation belongs to a single final node
  - Each tree is a forecaster with low bias and high variance
- We fit 500 trees for each date and predict using the average - an *ensemble* estimator

Convenient way of tracking what variables matter: "feature importance":

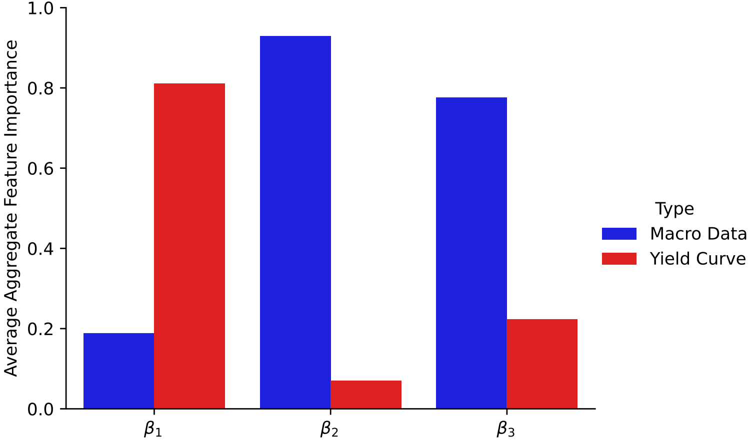
- For each split, track the reduction in MSE from a naive benchmark (sample mean);
- Compute the fraction of total MSE reduction due to each variable, for each tree and date;
- Average over trees and dates: how important is each variable for total MSE reduction?

What about non-linearities? Random Forests to the rescue!

	Lagged Factors			Forward Rates		
Target	No Macro	All Macro	p-value	No Macro	All Macro	p-value
β_1	-1.48	-1.93	0.87	-0.76	-0.72	0.39
β_2	-0.08	**0.27**	0.01	-0.34	**0.23**	0.00
β_3	-0.41	-0.16	0.02	-0.58	-0.22	0.01
$\Delta\beta_1$	-0.17	0.00	0.05	-0.53	-0.04	0.00
$\Delta\beta_2$	-0.08	**0.32**	0.00	-0.42	**0.32**	0.00
$\Delta\beta_3$	-0.37	-0.01	0.02	-0.33	-0.25	0.25

- This is the best method so far with $R^2 > 30\%$ for the first time
- Main result is **not** due to linear forecasting methods
- Forecasting innovations is usually better than forecasting factors directly

Average Feature Importance (Macro Variables vs Yield Curve)



- Heavy lifting is done mainly by the yield curve for β_1;
- Macro signals command almost all of the improvement for β_2;

▸ Individual Feature Importance

Economic Significance

Does it matter that much?

- Do these asymmetric violations matter in practice? Is this a statistical glitch?

- If there is additional predictability in bond returns, agents should take advantage of that!

Does it matter that much?

- Do these asymmetric violations matter in practice? Is this a statistical glitch?
- If there is additional predictability in bond returns, agents should take advantage of that!
- We model portfolio choice for a mean-variance investor:
  - One-year fixed investment horizon;
  - Monthly trading decisions: how much to invest in risky bonds vs earn the risk-free rate;
  - Mean-variance utility function: likes expected returns, dislikes return volatility;

Does it matter that much?

- Do these asymmetric violations matter in practice? Is this a statistical glitch?
- If there is additional predictability in bond returns, agents should take advantage of that!
- We model portfolio choice for a mean-variance investor:
  - One-year fixed investment horizon;
  - Monthly trading decisions: how much to invest in risky bonds vs earn the risk-free rate;
  - Mean-variance utility function: likes expected returns, dislikes return volatility;
- $R_{p,t+12}$ is the gross return of her portfolio: $R_{p,t+12} = 1 + y_t^{(1)} + \boldsymbol{w}_t'[\boldsymbol{xr_{t+12}}]$;

$$\max_{\boldsymbol{w}_t} \left\{ \mathbb{E}_t \left[R_{p,t+12}(\boldsymbol{w}_t) \right] - \frac{\gamma}{2} \cdot \mathrm{Var}_t \left[R_{p,t+12}(\boldsymbol{w}_t) \right] \right\}$$

Does it matter that much?

- Do these asymmetric violations matter in practice? Is this a statistical glitch?
- If there is additional predictability in bond returns, agents should take advantage of that!
- We model portfolio choice for a mean-variance investor:
  - One-year fixed investment horizon;
  - Monthly trading decisions: how much to invest in risky bonds vs earn the risk-free rate;
  - Mean-variance utility function: likes expected returns, dislikes return volatility;
- $R_{p,t+12}$ is the gross return of her portfolio: $R_{p,t+12} = 1 + y_t^{(1)} + \boldsymbol{w}_t'[\boldsymbol{xr}_{\boldsymbol{t}+12}]$;

$$\max_{\boldsymbol{w}_t} \left\{ \mathbb{E}_t \left[R_{p,t+12}(\boldsymbol{w}_t) \right] - \frac{\gamma}{2} \cdot \mathrm{Var}_t \left[R_{p,t+12}(\boldsymbol{w}_t) \right] \right\}$$

- $\boldsymbol{\mu}_{t+12|t} \equiv \mathbb{E}_t \left[\boldsymbol{xr}_{t+12} \right]$ and $\Sigma_{t+12|t} \equiv \mathbb{E}_t \left[\left(\boldsymbol{xr}_{t+12} - \boldsymbol{\mu}_{t+12|t} \right) \left(\boldsymbol{xr}_{t+12} - \boldsymbol{\mu}_{t+12|t} \right)' \right]$;
- Optimal solution: $\boldsymbol{w}_t^* = \frac{1}{\gamma} \cdot \Sigma_{t+12|t}^{-1} \boldsymbol{\mu}_{t+12|t}$, and we let $\gamma = 3$;

Conditional Risk Premia and Volatility

- Our methodology delivers estimates of $\boldsymbol{\mu}_{t+12|t}$ with and without macro signals;
- We follow Thornton and Valente (2012) to allow for time-varying volatility:

$$\widehat{\Sigma}_{t+12|t} \equiv \sum_{i=0}^{\infty} \boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}'_{t-i} \odot \Omega_{t-i}, \quad \Omega_{t-i} \equiv \alpha \cdot e^{-\alpha \cdot i} 11'$$

  where $\boldsymbol{\epsilon}_t$ is the 12-month ahead forecasting error; set $\alpha = 0.05$.
- Riva (2024): Spanning Hypothesis also constrains $\Sigma_{t+12|t}$, out of scope here;

Conditional Risk Premia and Volatility

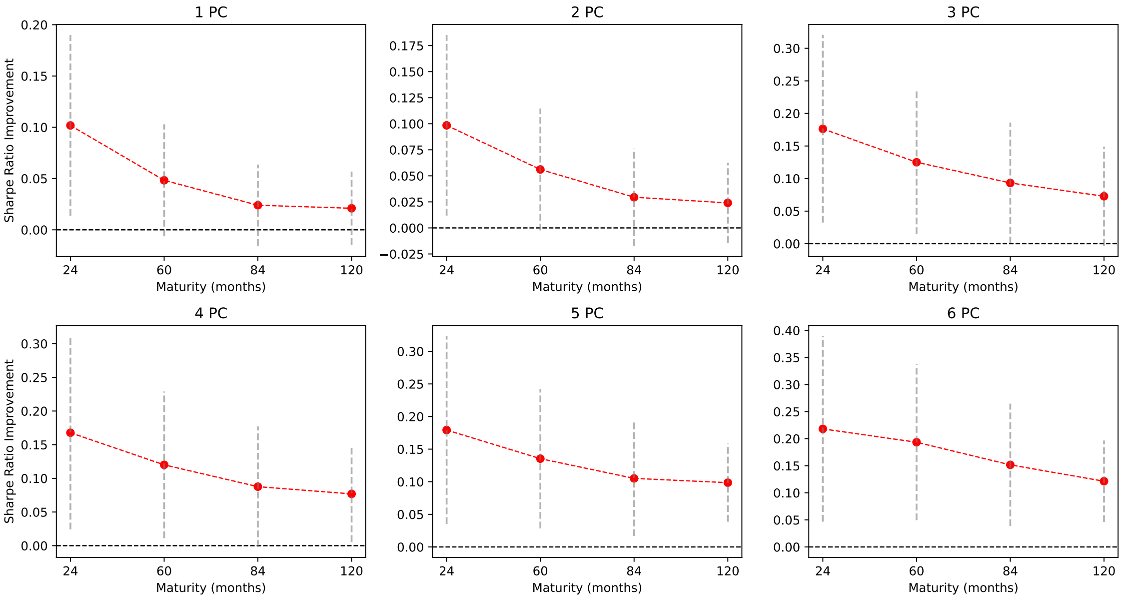
- Our methodology delivers estimates of $\boldsymbol{\mu}_{t+12|t}$ with and without macro signals;
- We follow Thornton and Valente (2012) to allow for time-varying volatility:

$$\widehat{\Sigma}_{t+12|t} \equiv \sum_{i=0}^{\infty} \boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}'_{t-i} \odot \Omega_{t-i}, \quad \Omega_{t-i} \equiv \alpha \cdot e^{-\alpha \cdot i} 11'$$

  where $\boldsymbol{\epsilon}_t$ is the 12-month ahead forecasting error; set $\alpha = 0.05$.
- Riva (2024): Spanning Hypothesis also constrains $\Sigma_{t+12|t}$, out of scope here;
- Leverage? Trading costs? Let $0 \leq w_t^{(n)} \leq 1$ (but also $-1 \leq w_t \leq 2$ in the paper);
- Our metric: Sharpe ratio = average risk premium over its volatility;
- Focus on the Sharpe ratio *improvement*;

Baseline Sharpe Ratio ≈ 0.2 (Constrained Case)

Unconstrained Case



Time-varying violations

Where do violations come from?

- Model selection techniques: unemployment and inflation are frequent choices;
- Market participants: macro data is helpful for monetary policy;
  - Larger influence on shorter maturities;

Where do violations come from?

- Model selection techniques: unemployment and inflation are frequent choices;
- Market participants: macro data is helpful for monetary policy;
  - Larger influence on shorter maturities;
- How does monetary policy appear in many DTSMs?
  - i_t = short rate; X_t = underlying risk factors;
  - Known reaction function: $i_t = \delta_0 + \delta_1' X_t$
  - No scope for deviations;

Where do violations come from?

- Model selection techniques: unemployment and inflation are frequent choices;
- Market participants: macro data is helpful for monetary policy;
  - Larger influence on shorter maturities;
- How does monetary policy appear in many DTSMs?
  - i_t = short rate; X_t = underlying risk factors;
  - Known reaction function: $i_t = \delta_0 + \delta_1' X_t$
  - No scope for deviations;
- In practice:
  - δ_0, δ_1 are unknown to everyone;
  - Many reasons to deviate: time-varying loadings, large shocksx, other objectives, etc;

Where do violations come from?

- Model selection techniques: unemployment and inflation are frequent choices;
- Market participants: macro data is helpful for monetary policy;
  - Larger influence on shorter maturities;
- How does monetary policy appear in many DTSMs?
  - i_t = short rate; X_t = underlying risk factors;
  - Known reaction function: $i_t = \delta_0 + \delta_1' X_t$
  - No scope for deviations;
- In practice:
  - δ_0, δ_1 are unknown to everyone;
  - Many reasons to deviate: time-varying loadings, large shocksx, other objectives, etc;
- **Conjecture**: deviations predict future precision for β_2 when we allow for macro signals;

How to test this conjecture?

- Fit a Taylor Rule with inflation and unemployment:

$$i_t = \delta_0 + \delta_\pi \cdot \text{Inflation}_t + \delta_u \cdot \text{Unemployment}_t$$

- Rolling-window estimation: 60 months;
- Deviation: $\phi_t \equiv i_t - \widehat{i_t}$;

How to test this conjecture?

- Fit a Taylor Rule with inflation and unemployment:

$$i_t = \delta_0 + \delta_\pi \cdot \text{Inflation}_t + \delta_u \cdot \text{Unemployment}_t$$

- Rolling-window estimation: 60 months;
- Deviation: $\phi_t \equiv i_t - \widehat{i_t}$;
- Loss function: $L_{i,t} \equiv \left(\widehat{\beta}_{i,t} - \beta_{i,t}\right)^2, \quad i = 1, 2, 3;$
- Lower loss $\iff$ more precision;

How to test this conjecture?

- Fit a Taylor Rule with inflation and unemployment:

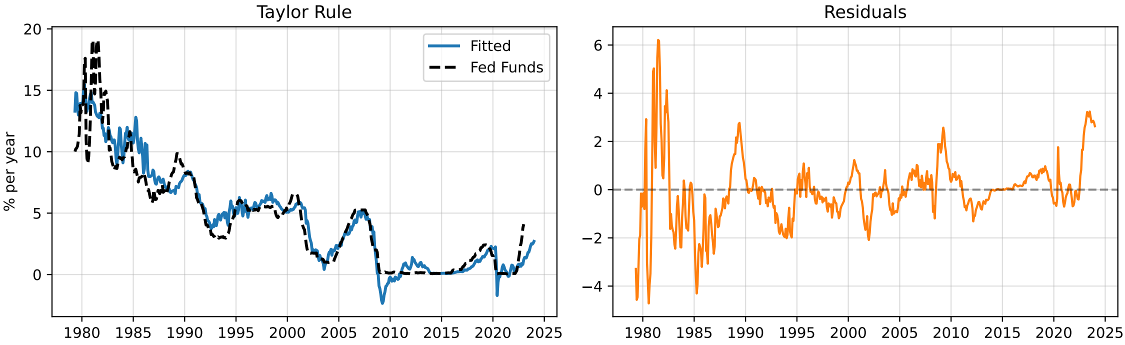
$$i_t = \delta_0 + \delta_\pi \cdot \text{Inflation}_t + \delta_u \cdot \text{Unemployment}_t$$

- Rolling-window estimation: 60 months;
- Deviation: $\phi_t \equiv i_t - \widehat{i_t}$;
- Loss function: $L_{i,t} \equiv \left(\widehat{\beta}_{i,t} - \beta_{i,t}\right)^2, \quad i = 1, 2, 3;$
- Lower loss $\iff$ more precision;
- Can deviations predict precision?

$$L_{i,t+12} = a + b \cdot |\phi_t| + \eta_{i,t+12}$$

- Any evidence that $b < 0$ for β_2? What about β_1?

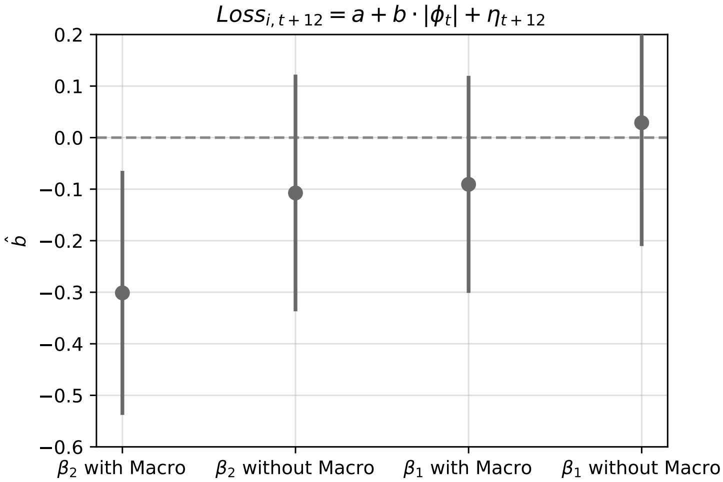
Quality of fit



- Rolling-window Taylor Rule fits well the Fed Funds rate;
- Robustness: headline inflation, monthly inflation;

▸ Other Rules

Improvements in Forecasting



- Deviations $\implies$ better predictions **with** macro signals;
- Only for short-run factor β_2;
- True across Taylor Rules and controls; ▸ Robustness
- In the paper:
  $\uparrow$ inflation $\implies$ $\uparrow$ violations;
  ▸ Extra Results

Wrap-Up

Main takeaways:

- The shorter end of the American nominal yield curve violates the Spanning Hypothesis;
- Extra predictability $\implies$ Sharpe ratio $\uparrow \approx 0.1 - 0.2$;
- Violations connected to monetary policy;

Wrap-Up

Main takeaways:

- The shorter end of the American nominal yield curve violates the Spanning Hypothesis;
- Extra predictability $\implies$ Sharpe ratio $\uparrow \approx 0.1 - 0.2$;
- Violations connected to monetary policy;

And now so what?

- Different frameworks for shorter and longer maturities;
- DTSMs + unknown reaction functions = uncharted territory;
  - Future work!

Wrap-Up

Main takeaways:

- The shorter end of the American nominal yield curve violates the Spanning Hypothesis;
- Extra predictability $\implies$ Sharpe ratio $\uparrow \approx 0.1 - 0.2$;
- Violations connected to monetary policy;

And now so what?

- Different frameworks for shorter and longer maturities;
- DTSMs + unknown reaction functions = uncharted territory;
  - Future work!

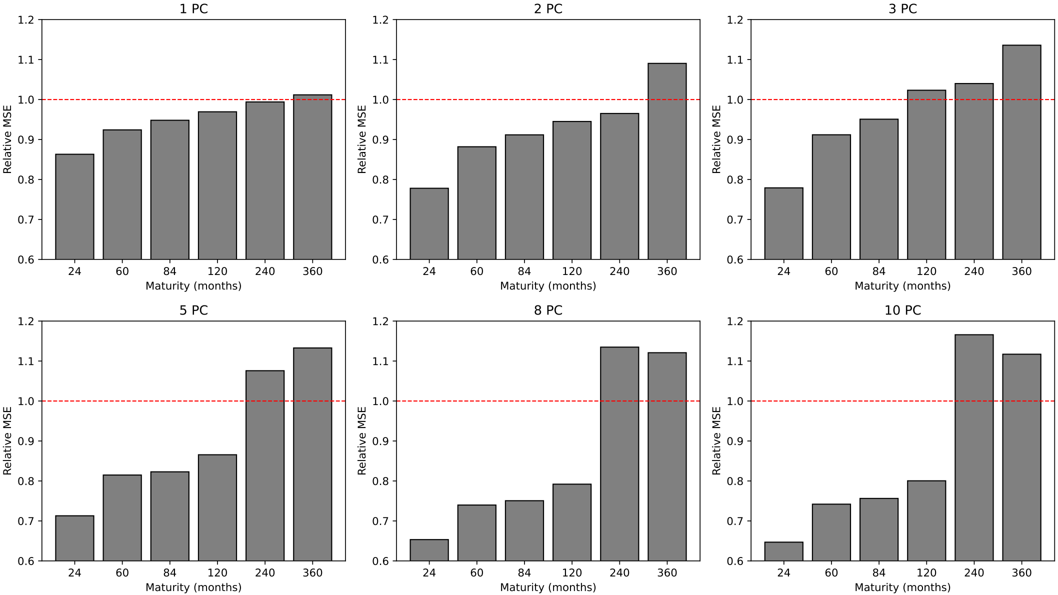
Thank you!

Appendix

(Thank you!)

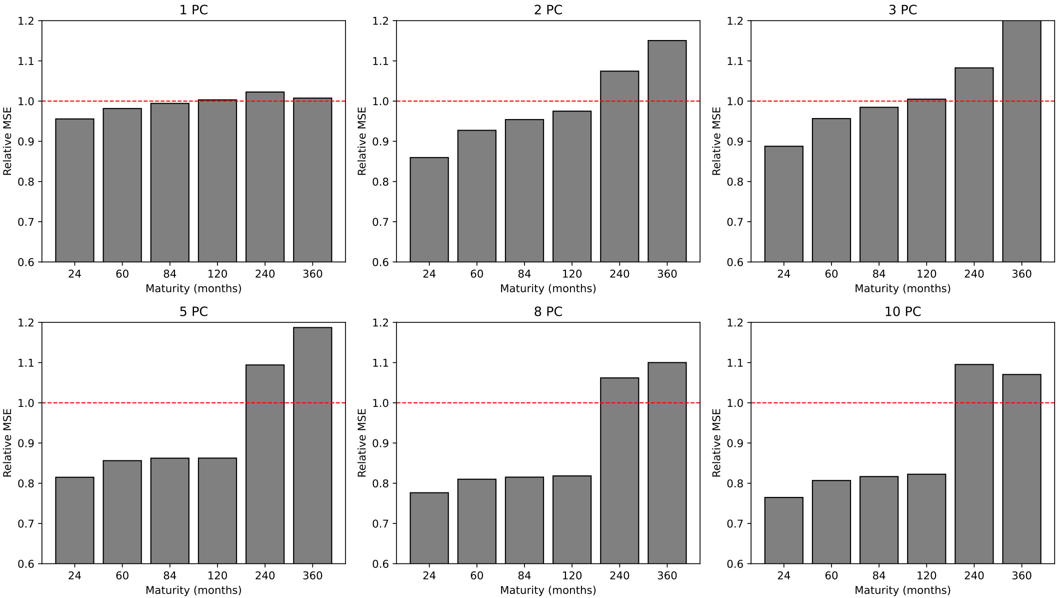
Excess Bond Returns Relative MSE Ratios

Back



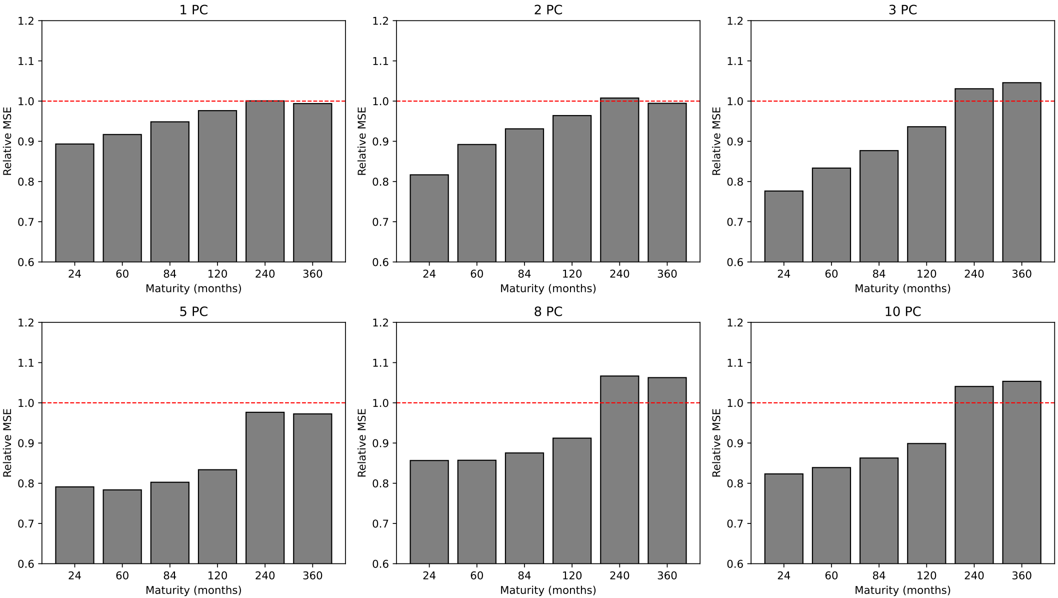
MSE Ratios: GSW Yields (NY Fed)

▸ Back



MSE Ratios: Ragged Edge Case

Back



p-Values for MSE Ratios of Excess Bond Returns

Back

	Maturity in months					
	24	60	84	120	240	360
1 PC	0.00	0.01	0.02	0.05	0.74	0.92
2 PC	0.00	0.01	0.01	0.04	0.16	0.32
3 PC	0.02	0.01	0.04	0.13	0.81	0.96
4 PC	0.04	0.06	0.13	0.24	0.55	0.65
5 PC	0.18	0.28	0.42	0.48	0.80	0.84
6 PC	0.21	0.25	0.35	0.38	0.69	0.66
7 PC	0.16	0.09	0.13	0.16	0.34	0.28
8 PC	0.24	0.23	0.32	0.37	0.59	0.57
9 PC	0.12	0.11	0.19	0.33	0.75	0.80
10 PC	0.15	0.12	0.19	0.28	0.79	0.51

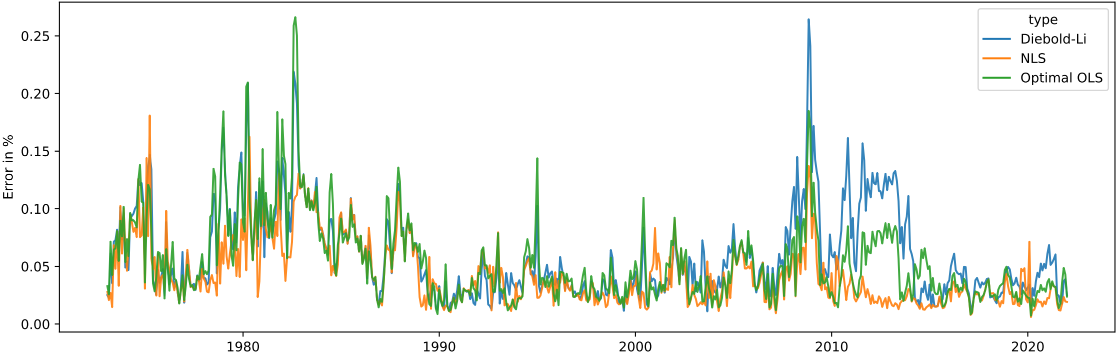
In-Sample Evidence Forecasting Returns

▸ Back

	2-year			10-year			20-year			30-year		
PC 1	0.09***	0.12***	0.13***	0.04**	0.07***	0.07***	-0.01	-0.00	0.00	-0.03	-0.02	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)
PC 2		-0.07**	-0.07**		-0.07***	-0.06**		-0.01	0.00		0.00	0.02
		(0.03)	(0.03)		(0.02)	(0.02)		(0.04)	(0.05)		(0.05)	(0.06)
PC 3		0.11***	0.11***		0.08***	0.08***		0.05**	0.05*		0.04	0.03
		(0.03)	(0.02)		(0.03)	(0.02)		(0.03)	(0.03)		(0.03)	(0.03)
PC 4		-0.02	-0.02		-0.05***	-0.06***		-0.06***	-0.06***		-0.09***	-0.08***
		(0.02)	(0.03)		(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)
PC 5		-0.04	-0.04		-0.09***	-0.08***		-0.08**	-0.08*		-0.09**	-0.09*
		(0.03)	(0.03)		(0.03)	(0.03)		(0.04)	(0.05)		(0.05)	(0.05)
PC 6			0.03			0.07***			0.04			0.06
			(0.03)			(0.03)			(0.04)			(0.05)
PC 7			0.06*			0.04			0.01			0.01
			(0.03)			(0.03)			(0.03)			(0.03)
PC 8			-0.08***			-0.08***			-0.04			-0.04
			(0.03)			(0.03)			(0.04)			(0.05)
N	588	588	588	588	588	588	422	422	422	422	422	422
R2 Adj.	0.28	0.36	0.40	0.28	0.36	0.40	0.16	0.23	0.24	0.15	0.22	0.23
R2 Adj. (No Macro Data)	0.15	0.15	0.15	0.25	0.25	0.25	0.16	0.16	0.16	0.14	0.14	0.14

Alternative Estimation Procedures

Back



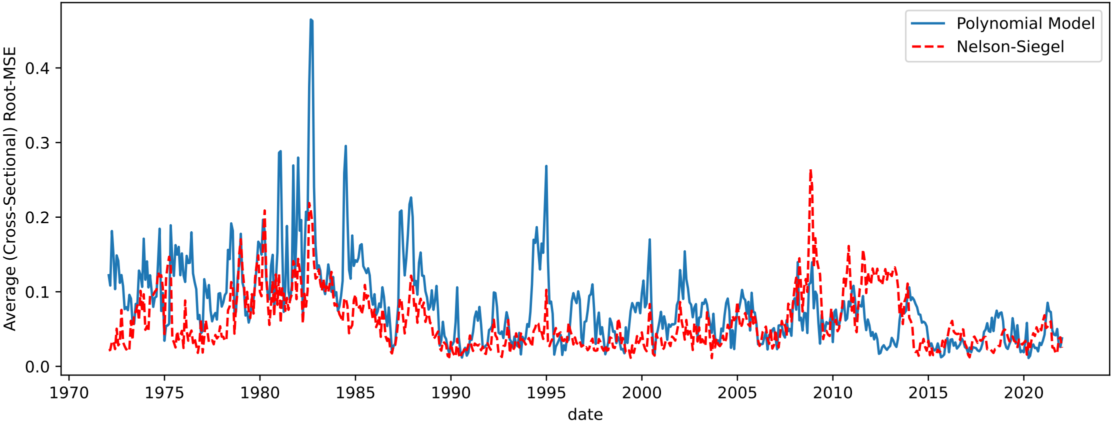
- NLS stands for Non-Linear Least Squares Date by Date
- Optimal OLS is the in-sample best OLS-implied decay fit

Alternative Estimation Procedures

▸ Back

A quadratic polynomial model:

$$y_t^{(\tau)} = c_{1,t} + c_{2,t} \cdot \tau + c_{3,t} \cdot \tau^2 \tag{9}$$



Estimation Details

Back

Define the following matrices for each time t:

$$X \equiv \begin{bmatrix} 1 & \left(\frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1}\right) & \left(\frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1}\right) \\ \vdots & \vdots & \vdots \\ 1 & \left(\frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N}\right) & \left(\frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N}\right) \end{bmatrix}, \quad Y_t = \begin{bmatrix} y_t^{(\tau_1)} \\ \vdots \\ y_t^{(\tau_N)} \end{bmatrix} \tag{10}$$

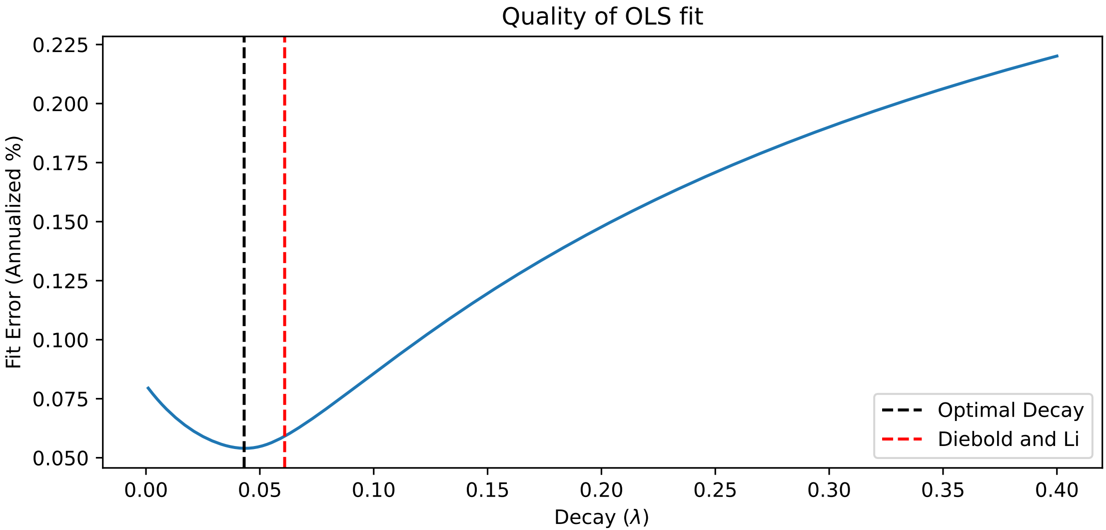
Now estimate betas using OLS:

$$\begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{bmatrix} = \left(X'X\right)^{-1} X' Y_t \tag{11}$$

Notice that X does not depend on t.

Fitting the Decay

Back



- For each λ, fit the model by OLS over the entire sample and compute the average squared fitting error

Out-of-sample PCA-based Forecast of Innovations

▸ Back

Target	No Macro	Number of Macro PCs						p-values					
		1	2	3	4	5	8	1	2	3	4	5	8
Predicting Innovations - Controlling for Forward Rates													
$\Delta\beta_1$	-0.19	-0.15	-0.17	-0.14	-0.10	-0.08	0.05	0.19	0.32	0.17	0.12	0.10	0.01
$\Delta\beta_2$	-0.11	-0.12	0.14	0.18	0.17	0.19	0.18	0.52	0.00	0.02	0.02	0.02	0.05
$\Delta\beta_3$	-0.10	-0.12	-0.06	-0.05	-0.05	-0.06	-0.08	0.93	0.17	0.25	0.26	0.31	0.41
Predicting Factor Levels - Controlling for Lagged Betas													
β_1	-0.10	-0.10	-0.11	-0.14	-0.11	-0.07	0.06	0.51	0.67	0.83	0.56	0.36	0.04
β_2	0.06	0.07	0.21	0.20	0.20	0.20	0.17	0.31	0.01	0.15	0.16	0.18	0.28
β_3	-0.11	-0.14	-0.06	-0.05	-0.05	-0.06	-0.08	0.89	0.16	0.19	0.20	0.23	0.39

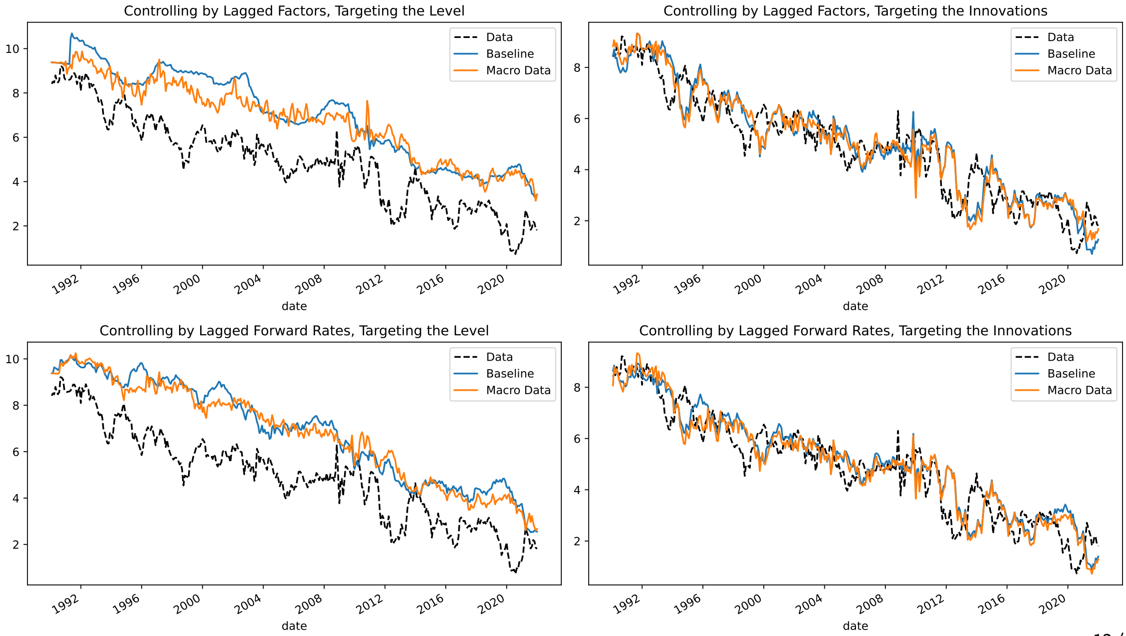
Regularization Methods - Controlling by Lagged Betas

Back

Target	No Macro Data			All Macro Data			p-value		
	Ridge	Lasso	Elastic Net	Ridge	Lasso	Elastic Net	Ridge	Lasso	Elastic Net
Beta 1	-4.91	-4.73	-4.81	-3.76	-5.08	-4.53	0.00	0.97	0.10
Beta 2	0.00	-0.12	-0.12	0.08	0.07	0.02	0.16	0.00	0.08
Beta 3	-0.41	-0.47	-0.49	-0.45	-0.35	-0.39	0.71	0.04	0.09
Innovation 1	0.12	-0.00	0.11	-0.29	0.04	0.08	1.00	0.30	0.84
Innovation 2	0.10	0.08	0.12	0.18	0.25	0.24	0.11	0.00	0.01
Innovation 3	0.08	0.04	0.02	0.00	0.03	0.07	0.95	0.70	0.02

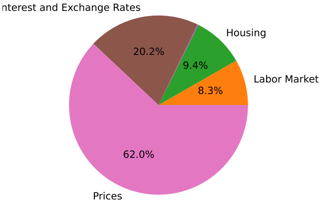
Regularization Failure for β_1

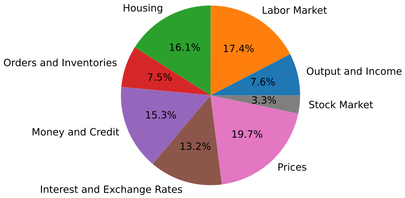
Back

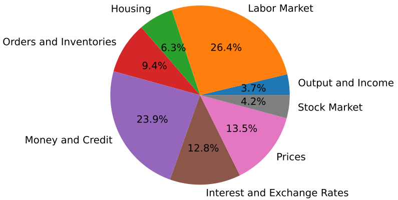


Model Selection - Lasso

How frequently are variables from each group chosen?





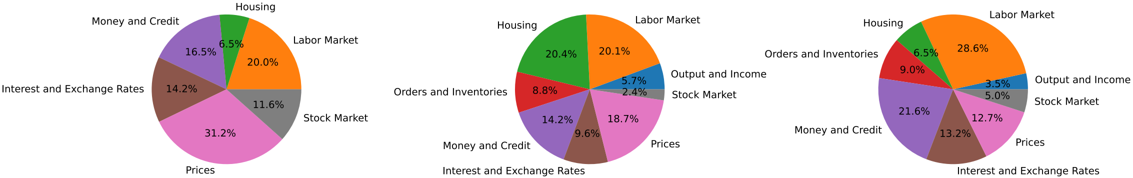


- Typical number of chosen variables is around 10-15
- Price measures are the leading predictor for β_1 - echoes Joslin et al (2014)
- Short and medium run: the "illusion of sparsity" - Giannone et al (2021)

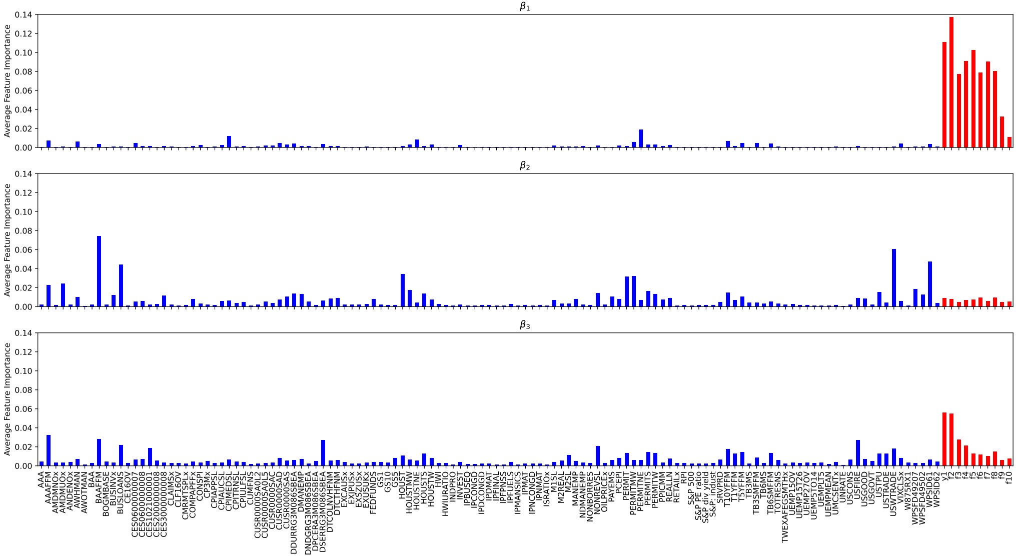
Elastic Net Back

Model Selection - Elastic Net

Back

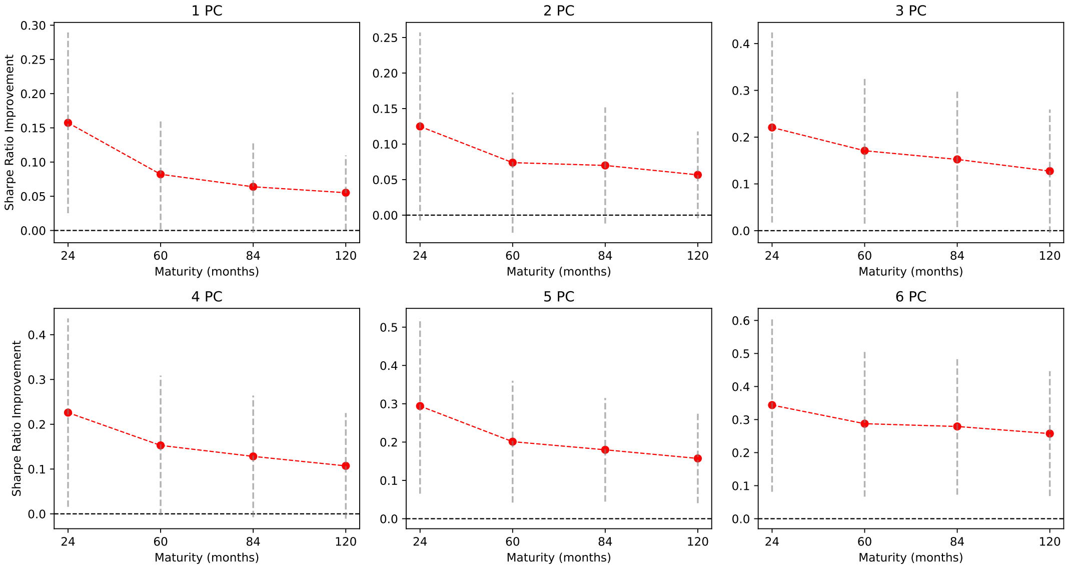


Feature Importance

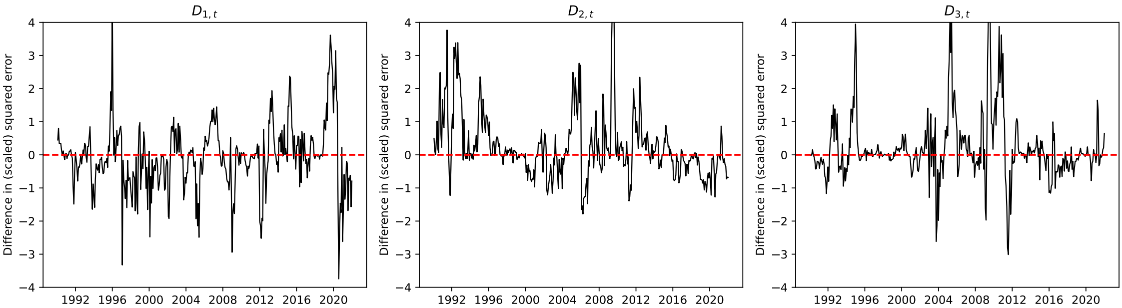


Unconstrained Sharpe Ratio Improvement

Back



Time series of scaled $D_{i,t}$



▸ Back

Random Forrest with Rolling Window (180 months)

	Lagged Factors			Forward Rates		
Target	No Macro	All Macro	p-value	No Macro	All Macro	p-value
β_1	-1.20	-1.32	0.72	-0.63	-0.98	0.98
β_2	-0.07	0.20	0.02	-0.30	0.19	0.00
β_3	-0.47	-0.24	0.04	-0.67	-0.23	0.00

Back

Math Details - Giacomini and White (2006)

Back

- Let x_t be a $q \times 1$ random vector with variables chosen by the econometrician
- Let $\mathrm{z}_{t+h} \equiv \mathrm{x}_t \left(L^{m'}_{t+h} - L^{m}_{t+h} \right)$ for a given forecasting horizon h
- Define

$$\bar{\mathrm{z}}_T \equiv \frac{1}{T-h-t_0} \sum_{t=t_0}^{T-h} \mathrm{z}_{t+h}$$

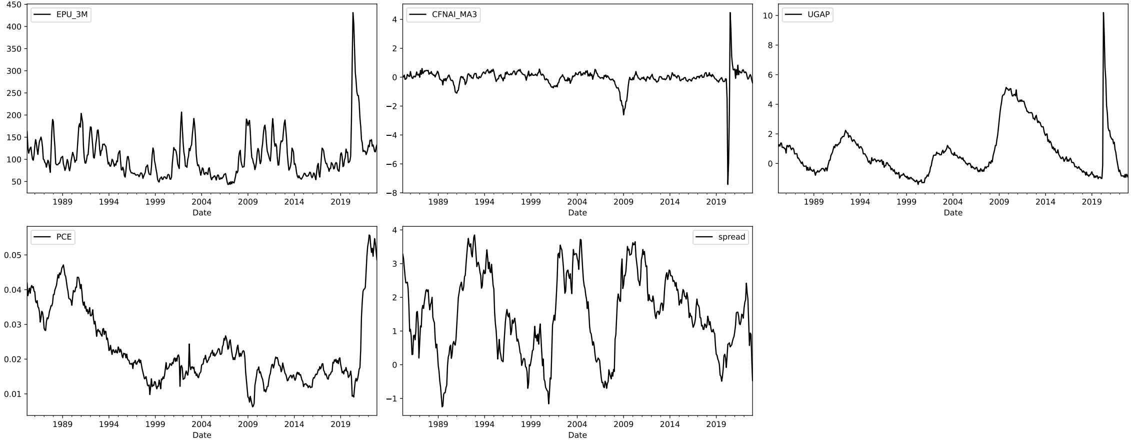
$$\widehat{\Omega}_T \equiv \frac{1}{T-h-t_0} \sum_{t=t_0}^{T-h} \mathrm{z}_{t+h}\mathrm{z}'_{t+h} + \frac{1}{T-h-t_0} \sum_{j=1}^{h-1} w_{j,T} \sum_{t=t_0+j}^{T-h} \left(\mathrm{z}_{t+h-j}\mathrm{z}'_{t+h} + \mathrm{z}_{t+h}\mathrm{z}'_{t+h-j} \right)$$

$$w_{j,T} \to 1, \qquad \text{as } T \to \infty \text{ for each } j \in \{1, ..., h-1\}$$

- Under some regularity conditions, they show that as T diverges to ∞:

$$W \equiv T \cdot \mathrm{z}'_{t+h} \widehat{\Omega}_T^{-1} \mathrm{z}_{t+h} \xrightarrow{d} \chi^2_q \tag{12}$$

Conditioning Variables - Time Series



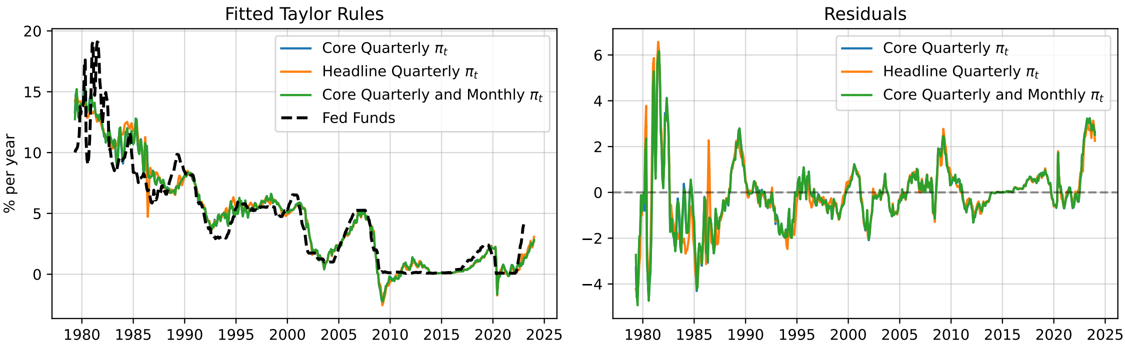
Back

Non-Parametric Evidence on Conditional Predictive Ability

Inflation Tercile	PCE	D_1	D_2	D_3	Control
Low	0.013	-0.152	0.496	2.386	Forward Rates
Medium	0.018	-0.754	0.788	1.923	Forward Rates
High	0.028	0.039	2.430	1.526	Forward Rates
Low	0.013	-0.204	-0.023	0.803	Lagged Factors
Medium	0.018	-0.114	0.120	0.850	Lagged Factors
High	0.028	0.048	1.963	1.492	Lagged Factors

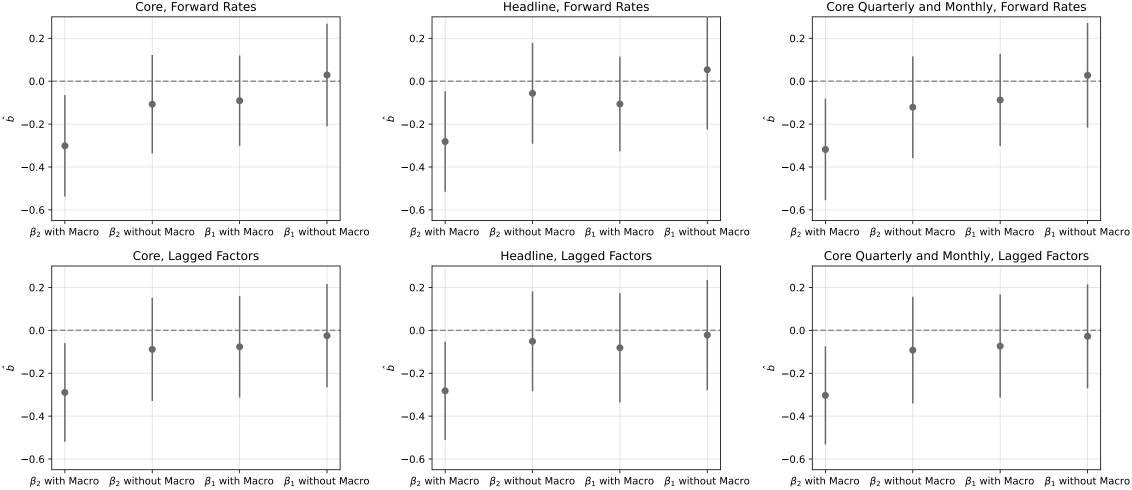
Back

Different Taylor Rules



Robustness on Loss Regression

$$\text{Loss}_{i,t+12} = a + b \cdot |\phi_t| + \eta_{t+12}$$



Conditional Predictive Ability

$$D_{2,t+12} = a + \mathrm{b}'\mathrm{x}_t + u_{t+12}$$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
EPU	-0.08					-0.11				-0.19*
	(0.10)					(0.10)				(0.10)
CFNAI		-0.06				-0.10		-0.09		-0.14
		(0.08)				(0.07)		(0.09)		(0.08)
UGAP			-0.02				0.04	0.01	-0.03	0.05
			(0.10)				(0.09)	(0.09)	(0.08)	(0.13)
PCE				0.30**			0.31**	0.31**	0.30**	0.33***
				(0.12)			(0.12)	(0.12)	(0.12)	(0.11)
Slope					0.09				0.12	0.10
					(0.12)				(0.11)	(0.12)
N	384	384	384	384	384	384	384	384	384	384
R^2	0.01	0.00	0.00	0.09	0.01	0.02	0.09	0.10	0.10	0.13
GW p-values	0.51	0.38	0.84	**0.00**	0.45	0.50	**0.00**	**0.00**	**0.01**	**0.01**

▸ Conditioning Variables ▸ Non-Parametric Evidence

References I

Bauer, M. D. and Hamilton, J. D. (2018). Robust bond risk premia. *Review of Financial Studies*, 31(2):399–448.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *Review of Financial Studies*, 34(2):1046–1089.

Borup, D., Eriksen, J. N., Kjær, M. M., and Thyrsgaard, M. (2023). Predicting bond return predictability. *Management Science*.

Cieslak, A. and Povala, P. (2015). Expected returns in treasury bonds. *Review of Financial Studies*, 28(10):2859–2901.

Cochrane, J. H. and Piazzesi, M. (2005). Bond Risk Premia. *American Economic Review*, 95(1):138–160.

Cooper, I. and Priestley, R. (2009). Time-varying risk premiums and the output gap. *Review of Financial Studies*, 22(7):2601–2633.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F. X. and Rudebusch, G. D. (2013). *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*. The Econometric and Tinbergen Institutes lectures. Princeton University Press, Princeton.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1-2):309–338.

Duffee, G. (2013). Forecasting interest rates. In *Handbook of Economic Forecasting*, pages 385–426. Elsevier.

Fama, E. and Bliss, R. R. (1987). The information in long-maturity forward rates. *American Economic Review*, 77(4):680–92.

Fernandes, M. and Vieira, F. (2019). A dynamic nelson–siegel model with forward-looking macroeconomic factors for the yield curve in the US. *Journal of Economic Dynamics and Control*, 106:103720.

Filippou, I., Rapach, D., Taylor, M. P., and Zhou, G. (2023). The rise and fall of the carry trade: Links to exchange rate predictability. *SSRN Electronic Journal*.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437.

Goulet-Coulombe, P. (2023). The macroeconomy as a random forest. *Journal of Applied Econometrics*, forthcoming.

References II

Goulet-Coulombe, P., Rapach, D., Schütte, E. C. M., and Schwenk-Nebbe, S. (2023). The anatomy of machine learning-based portfolio performance. *SSRN Electronic Journal*.

Greenwood, R. and Vayanos, D. (2014). Bond supply and excess bond returns. *Review of Financial Studies*, 27(3):663–713.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273.

Gurkaynak, R. S., Sack, B., and Wright, J. H. (2007). The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304.

Hoogteijling, T., Martens, M. P., and van der Wel, M. (2021). Forecasting bond risk premia using stationary yield factors. *SSRN Electronic Journal*.

Huang, J.-Z. and Shi, Z. (2023). Machine-learning-based return predictors and the spanning controversy in macro-finance. *Management Science*, 69(3):1780–1804.

Hännikäinen, J. (2017). When does the yield curve contain predictive power? evidence from a data-rich environment. *International Journal of Forecasting*, 33(4):1044–1064.

Joslin, S., Priebsch, M., and Singleton, K. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance*, 69(3):1197–1233.

Litterman, R. B. and Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1(1):54–61.

Liu, Y. and Wu, C. (2021). Reconstructing the yield curve. *Journal of Financial Economics*, 142(3):1395–1425.

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

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

References III

Medeiros, M. C., Vasconcelos, G. F. R., Álvaro Veiga, and Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119.

Moench, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, 146(1):26–43.

Nelson, C. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60(4):473–89.

Riva, R. (2024). How much unspanned volatility can different shocks explain?

Shen, Z. and Xiu, D. (2024). Can machines learn weak signals? *SSRN Electronic Journal*.

Thornton, D. L. and Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *Review of Financial Studies*, 25(10):3141–3168.

van der Wel, M. and Zhang, Y. (2021). Global evidence on unspanned macro risks in dynamic term structure models. *SSRN Electronic Journal*.

van Dijk, D., Koopman, S. J., van der Wel, M., and Wright, J. H. (2013). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics*, 29(5):693–712.