Asymmetric Violations of the Spanning Hypothesis

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

[1]Northwestern University
[2]Erasmus University

SoFiE 2024 - PUC-Rio

June 14th, 2024

Intro

- Yield curve dynamics is of major interest:
  - Monetary policy transmission + Fiscal policy assessment;
  - Risk management and long-term investment decisions;
  - Risk premia measurement and portfolio allocation;

- Arbitrage-free Affine Term Structure models: our workhorse, many good properties but generate sharp predictions;

Intro

- Yield curve dynamics is of major interest:
  - Monetary policy transmission + Fiscal policy assessment;
  - Risk management and long-term investment decisions;
  - Risk premia measurement and portfolio allocation;

- Arbitrage-free Affine Term Structure models: our workhorse, many good properties but generate sharp predictions;

- Common *implication* of many term structure models: the "Spanning Hypothesis":
  - The yield curve spans all information necessary to forecast future yields and bond returns;
  - Information about different sources of macroeconomic risks should be embedded in bond prices (and yields);
  - Arises from many full-information models (Wachter (2006), Dewachter and Lyrio (2006), Piazzesi and Schneider (2007), Rudebusch and Wu (2008), Rudebusch and Swanson (2012), Duffee (2013), ...);

This paper

Do macroeconomic variables help forecasting excess bond returns and/or future yields *after* we condition on the current yield curve?

- Literature often offers a binary answer:
  - **Yes**: Cooper and Priestley (2009), Ludvigson and Ng (2009), Joslin et al. (2014), Greenwood and Vayanos (2014), Cieslak and Povala (2015), Fernandes and Vieira (2019);
  - **Probably Not**: Duffee (2013), Bauer and Hamilton (2018);
  - Econometric inference here is challenging: small sample + persistent regressors;

This paper

Do macroeconomic variables help forecasting excess bond returns and/or future yields *after* we condition on the current yield curve?

- Literature often offers a binary answer:
  - **Yes**: Cooper and Priestley (2009), Ludvigson and Ng (2009), Joslin et al. (2014), Greenwood and Vayanos (2014), Cieslak and Povala (2015), Fernandes and Vieira (2019);
  - **Probably Not**: Duffee (2013), Bauer and Hamilton (2018);
  - Econometric inference here is challenging: small sample + persistent regressors;

- We show evidence that the answer is more nuanced: *asymmetric* violations;
- **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;
- Stronger violations when inflation is higher, when the policy maker is more likely to act;

How do we do this in 25 minutes?

1. Design an out-of-sample forecasting exercise for excess bond returns:
   - We use a large panel of macroeconomic variables instead of selecting a few variables
   - Out-of-sample period: 1990-2021
2. Propose a decomposition of excess bond returns based on Nelson-Siegel factors:
   - Reduced-form model for the yield curve with great fit;
   - Predictability of factors gets distributed along the yield curve through a single factor;
   - Study factor predictability using different machine learning methods;
   - All the action comes from the predictability of a single factor;

How do we do this in 25 minutes?

1. Design an out-of-sample forecasting exercise for excess bond returns:
   - We use a large panel of macroeconomic variables instead of selecting a few variables
   - Out-of-sample period: 1990-2021
2. Propose a decomposition of excess bond returns based on Nelson-Siegel factors:
   - Reduced-form model for the yield curve with great fit;
   - Predictability of factors gets distributed along the yield curve through a single factor;
   - Study factor predictability using different machine learning methods;
   - All the action comes from the predictability of a single factor;
3. Why should anyone care? Because it's money on the table ($$$)!
   - Significant Sharpe ratio improvements in a mean-variance allocation strategy ($\approx 0.2 \rightarrow 0.4$);
   - But larger when trading shorter maturity bonds ($\approx$ 2 years);
4. Are the gains of using more complicated models equally present over time? No!
   - Gains are concentrated on periods of higher inflation;

Literature

- Bond returns forecasting and tests of the Spanning Hypothesis
  - 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)
- Nelson-Siegel modeling
  - Nelson and Siegel (1987), Diebold and Li (2006), Diebold et al. (2006), Diebold and Rudebusch (2013), van Dijk et al. (2013), HÃďnnikÃďinen (2017), Fernandes and Vieira (2019)
- Economic value of predictability
  - Thornton and Valente (2012), Sarno et al. (2016), Gargano et al. (2019), Bianchi et al. (2021)
- **Our contribution**: out-of-sample tests of the spanning hypothesis using a novel decomposition of excess bond returns that makes the asymmetry easy to **identify**

Data

Yield curve data:

- Taken from Liu and Wu (2021). We focus on the 1973-2021 period.
- Constructed from CRSP data - we have nothing to say about non-US data (yet!)
- Provides longer maturities than Fama and Bliss (1987)
- Lower fitting errors than Gurkaynak et al. (2007)

Macroeconomic data:

- FRED-MD data set, detailed in McCracken and Ng (2016), maintained by St. Louis Fed
- 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 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:

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

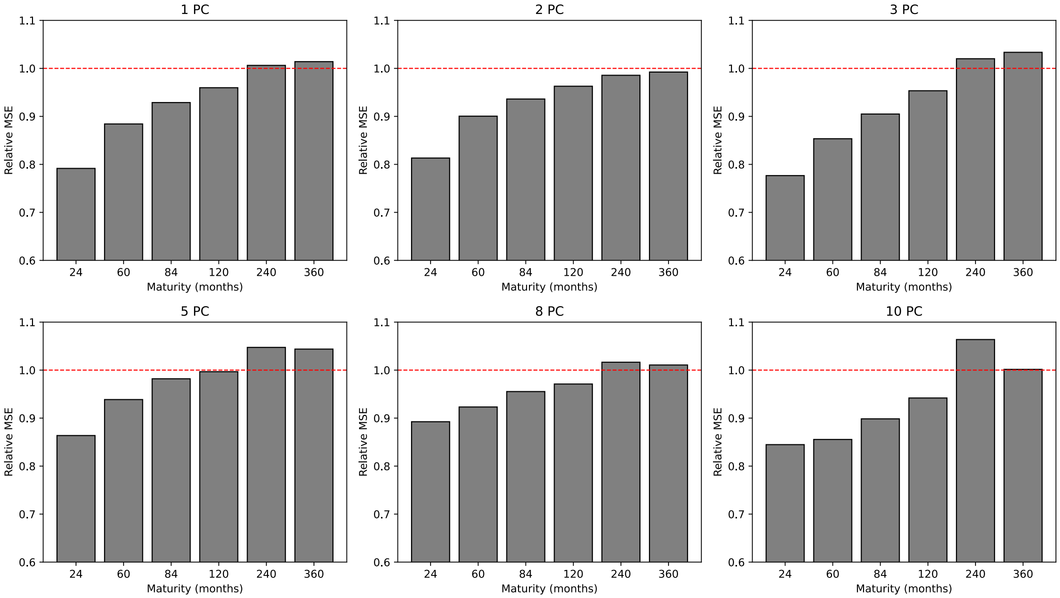
- We estimate a linear model with an expanding sample forecasting design:

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

- 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 Ratios With and Without Macro Data

▸ Controlling by 3 YC PCs ▸ *p*-values ▸ In-sample



Modeling Yields

- Macroeconomic variables improved forecasting for shorter maturities;
- We need a device to more thoroughly assess this asymmetry in the violation of the SH;
- Forecasting returns amounts to forecasting $y_{t+12}^{(n)}$;

Modeling Yields

- Macroeconomic variables improved forecasting for shorter maturities;
- We need a device to more thoroughly assess this asymmetry in the violation of the SH;
- Forecasting returns amounts to forecasting $y^{(n)}_{t+12}$;

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

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

- β_1 is a long-run factor: $\lim_{\tau\to\infty} y_t^{(\tau)} = \beta_{1,t}$;
- β_2 is a short-run factor: its absolute loading decreases with τ (measured in months).
- β_3 is a medium-run factor: its loading is hump-shaped.
- We set $\lambda = 0.0609$ and estimate the model by OLS date by date with $1 \leq \tau \leq 120$.

Decomposing Returns

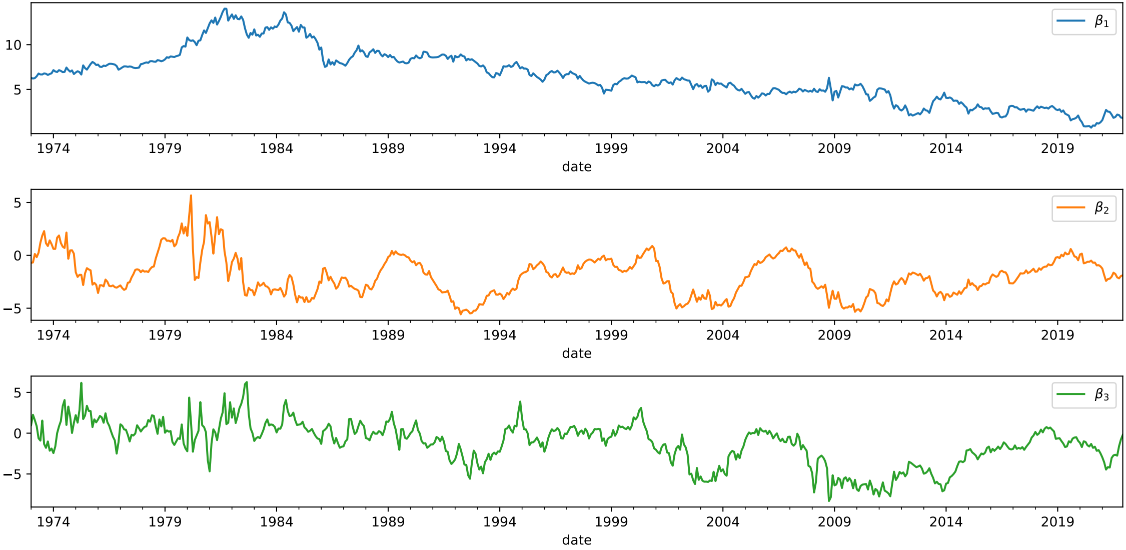
Proposition 1

Suppose the yield curve follows the Nelson-Siegel representation and assume that the decay parameter is a positive constant $\lambda_t = \lambda > 0$. Define $\theta \equiv 12\lambda$. Then, the one-year excess bond return for a maturity of n years is given by

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

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

Factor Realizations (1973-2021)



Estimation details | Alternative Estimation Procedures | Polynomial Model | Quality of Fit

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 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
- Result holds if we allow for even more PCs, but we lose statistical power

Regularization Methods - Notation

- PCA is not "supervised": dimensionality reduction decoupled from prediction
- Regularization works by penalizing a model for using too many variables
- Statistical trade-off: model "size" vs model flexibility

Regularization Methods - Notation

- PCA is not "supervised": dimensionality reduction decoupled from prediction
- Regularization works by penalizing a model for using too many variables
- Statistical trade-off: model "size" vs model flexibility

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

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

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

Regularization Methods - Notation

- PCA is not "supervised": dimensionality reduction decoupled from prediction
- Regularization works by penalizing a model for using too many variables
- Statistical trade-off: model "size" vs model flexibility

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

- We estimate 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

- We target both factors and their innovations due to time-series persistence

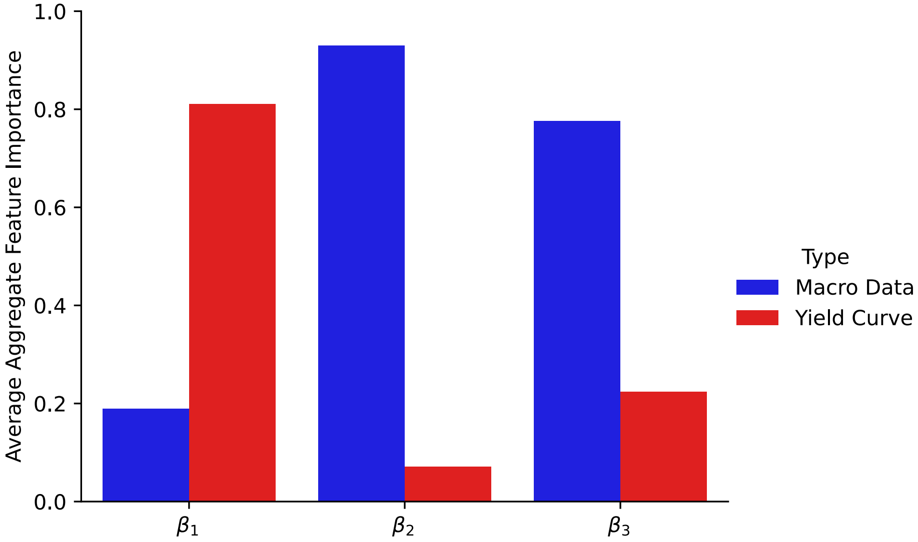
Other Controls Stationarity Matters Chosen variables

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)



Individual Feature Importance

Does it matter that much?

- Do these asymmetric violations matter in practice?
- If there is additional predictability in bond returns, traders should take advantage of that!
- We study the problem of a investor similar to Thornton and Valente (2012);
  - One-year fixed investment horizon;
  - Monthly trading decisions;
  - Mean-variance utility function;
  - At time t, she can either invest in the risk-free 1-year bond rate or in a risky n-year bond;

Does it matter that much?

- Do these asymmetric violations matter in practice?
- If there is additional predictability in bond returns, traders should take advantage of that!
- We study the problem of a investor similar to Thornton and Valente (2012);
  - One-year fixed investment horizon;
  - Monthly trading decisions;
  - Mean-variance utility function;
  - At time t, she can either invest in the risk-free 1-year bond rate or in a risky n-year bond;
- $R_{p,t+12}$ is the gross return of her portfolio: $R_{p,t+12} = 1 + y_t^{(1)} + w_t' \mathbf{xr}_{t+12}$;

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

- $\boldsymbol{\mu}_{t+12|t} \equiv \mathbb{E}_t \left[\boldsymbol{xr}_{t+12} \right]$ and $\Sigma_{t+12|t} \equiv \mathbb{E}_t \left[\left(\boldsymbol{xr}_{t+12} - \boldsymbol{\mu}_{t+12|t} \right) \left(\boldsymbol{xr}_{t+12} - \boldsymbol{\mu}_{t+12|t} \right)' \right]$;

How to form expectations?

- Optimal solution: $\boldsymbol{w}_t^* = \frac{1}{\gamma} \cdot \Sigma_{t+12|t}^{-1} \boldsymbol{\mu}_{t+12|t}$, and we let $\gamma = 3$;
- Our methodology delivers estimates of $\boldsymbol{\mu}_{t+12|t}$ with and without macro data;

How to form expectations?

- Optimal solution: $\boldsymbol{w}_t^* = \frac{1}{\gamma} \cdot \Sigma_{t+12|t}^{-1} \boldsymbol{\mu}_{t+12|t}$, and we let $\gamma = 3$;
- Our methodology delivers estimates of $\boldsymbol{\mu}_{t+12|t}$ with and without macro data;
- We follow Thornton and Valente (2012) to allow for time-varying volatility:

$$\widehat{\Sigma}_{t+12|t} \equiv \sum_{i=0}^{\infty} \boldsymbol{\epsilon}_{t-i} \boldsymbol{\epsilon}_{t-i}' \odot \Omega_{t-i}, \quad \Omega_{t-i} \equiv \alpha \cdot e^{-\alpha \cdot i} \mathbf{1}\mathbf{1}'$$

  where $\boldsymbol{\epsilon}_t$ is the 12-month ahead forecasting error;
- As time goes by, the past is exponentially less important. We set $\alpha = 0.05$;

How to form expectations?

- Optimal solution: $\boldsymbol{w}_t^* = \frac{1}{\gamma} \cdot \Sigma_{t+12|t}^{-1} \boldsymbol{\mu}_{t+12|t}$, and we let $\gamma = 3$;
- Our methodology delivers estimates of $\boldsymbol{\mu}_{t+12|t}$ with and without macro data;
- We follow Thornton and Valente (2012) to allow for time-varying volatility:

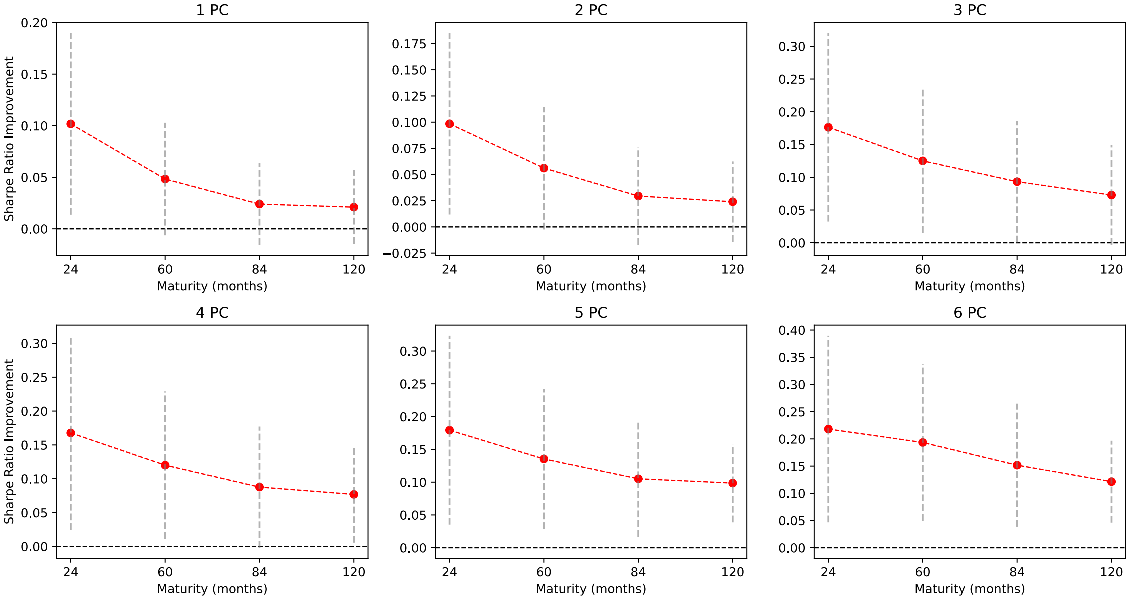
$$\widehat{\Sigma}_{t+12|t} \equiv \sum_{i=0}^{\infty} \boldsymbol{\epsilon}_{t-i} \boldsymbol{\epsilon}_{t-i}' \odot \Omega_{t-i}, \quad \Omega_{t-i} \equiv \alpha \cdot e^{-\alpha \cdot i} \mathbf{1}\mathbf{1}'$$

  where $\boldsymbol{\epsilon}_t$ is the 12-month ahead forecasting error;

- As time goes by, the past is exponentially less important. We set $\alpha = 0.05$;
- Leverage? Two flavors: $-1 \leq w_t^{(n)} \leq 2$ (unconstrained) or $0 \leq w_t^{(n)} \leq 1$ (constrained);
- Our metric: Sharpe ratio = average risk premium over its volatility (1990-2021);
- Focus on the Sharpe ratio *improvement* from using macro data across maturities;

Baseline Sharpe Ratio ≈ 0.2 (Constrained Case)

Unconstrained Case



Are Violations Happening All the Time?

- When is it more valuable to use a more complicated model?
- Conditional predictive ability test from Giacomini and White (2006);

Are Violations Happening All the Time?

- When is it more valuable to use a more complicated model?
- Conditional predictive ability test from Giacomini and White (2006);
- Define the loss function $L_{i,t} \equiv \left(\beta_{i,t} - \widehat{\beta}_{i,t}\right)^2$ and the difference $D_{i,t} \equiv L_{i,t}^{(\text{SH})} - L_{i,t}^{(\text{Macro})}$;
- If $D_{i,t} > 0$, baseline loss was higher $\implies$ macro data was useful;
- Focus on forecasts using Random Forrests (best overall model) + rolling windows;

Are Violations Happening All the Time?

- When is it more valuable to use a more complicated model?
- Conditional predictive ability test from Giacomini and White (2006);
- Define the loss function $L_{i,t} \equiv \left(\beta_{i,t} - \widehat{\beta}_{i,t}\right)^2$ and the difference $D_{i,t} \equiv L_{i,t}^{(\mathsf{SH})} - L_{i,t}^{(\mathsf{Macro})}$;
- If $D_{i,t} > 0$, baseline loss was higher $\implies$ macro data was useful;
- Focus on forecasts using Random Forrests (best overall model) + rolling windows;
- We are interested in testing $H_0 : \mathbb{E}\left[D_{i,t+12} | \mathcal{G}_t\right] = 0$; $\mathcal{G}_t$ is chosen by the econometrician;
- We study different state variables $\mathbf{x}_t$ and take $\mathcal{G}_t$ as the natural filtration of $\mathbf{x}_t$;

Are Violations Happening All the Time?

- When is it more valuable to use a more complicated model?
- Conditional predictive ability test from Giacomini and White (2006);
- Define the loss function $L_{i,t} \equiv \left(\beta_{i,t} - \widehat{\beta}_{i,t}\right)^2$ and the difference $D_{i,t} \equiv L_{i,t}^{(\text{SH})} - L_{i,t}^{(\text{Macro})}$;
- If $D_{i,t} > 0$, baseline loss was higher $\implies$ macro data was useful;
- Focus on forecasts using Random Forrests (best overall model) + rolling windows;
- We are interested in testing $H_0 : \mathbb{E}\left[D_{i,t+12} | \mathcal{G}_t\right] = 0$; $\mathcal{G}_t$ is chosen by the econometrician;
- We study different state variables $\mathbf{x}_t$ and take $\mathcal{G}_t$ as the natural filtration of $\mathbf{x}_t$;
- We also study the associated regression:

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

▸ Math Details ▸ Time-Series for $D_{i,t}$ ▸ Trivial $\mathcal{G}_t$

Conditional Predictive Ability

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

	(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
R2	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

Wrap-Up

Main takeaways:

- The shorter end of the American nominal yield curve violates the Spanning Hypothesis;
- The longer end behaves very much as many affine DTSMs predict!
- This extra predictability can create a Sharpe ratio improvement of $\approx 0.1 - 0.2$;
- Using a more complicated model pays off when one faces higher inflation rates;

Wrap-Up

Main takeaways:

- The shorter end of the American nominal yield curve violates the Spanning Hypothesis;
- The longer end behaves very much as many affine DTSMs predict!
- This extra predictability can create a Sharpe ratio improvement of $\approx 0.1 - 0.2$;
- Using a more complicated model pays off when one faces higher inflation rates;

And now so what?

- Shorter and longer rates should probably be modeled within different frameworks;
- Why do we think this asymmetry is happening? Our conjecture:
  - Shorter end is more heavily influenced by monetary policy... and fund managers know that!
  - Macro data may help market participants to anticipate monetary policy decisions;
- Models with spanning assume that the central bank's reaction function is **known**!
  - How would a DTSM with an unknown reaction function look like? Future work!

Wrap-Up

Main takeaways:

- The shorter end of the American nominal yield curve violates the Spanning Hypothesis;
- The longer end behaves very much as many affine DTSMs predict!
- This extra predictability can create a Sharpe ratio improvement of $\approx 0.1 - 0.2$;
- Using a more complicated model pays off when one faces higher inflation rates;

And now so what?

- Shorter and longer rates should probably be modeled within different frameworks;
- Why do we think this asymmetry is happening? Our conjecture:
  - Shorter end is more heavily influenced by monetary policy... and fund managers know that!
  - Macro data may help market participants to anticipate monetary policy decisions;
- Models with spanning assume that the central bank's reaction function is **known**!
  - How would a DTSM with an unknown reaction function look like? Future work!

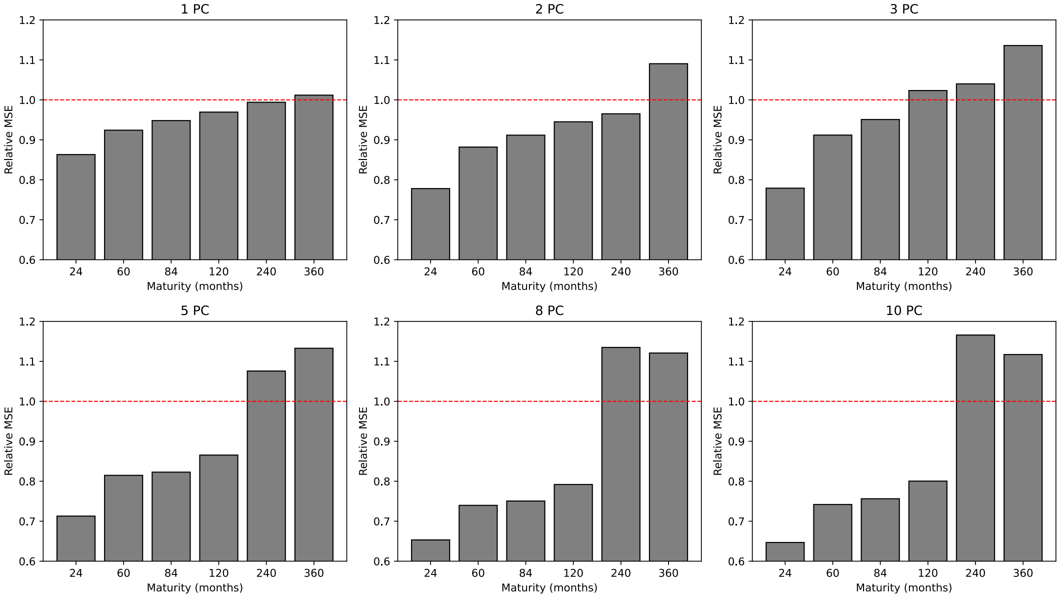
Thank you! (By the way, I'll be on the job market this year!)

Appendix

(Thank you!)

Excess Bond Returns Relative MSE Ratios

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

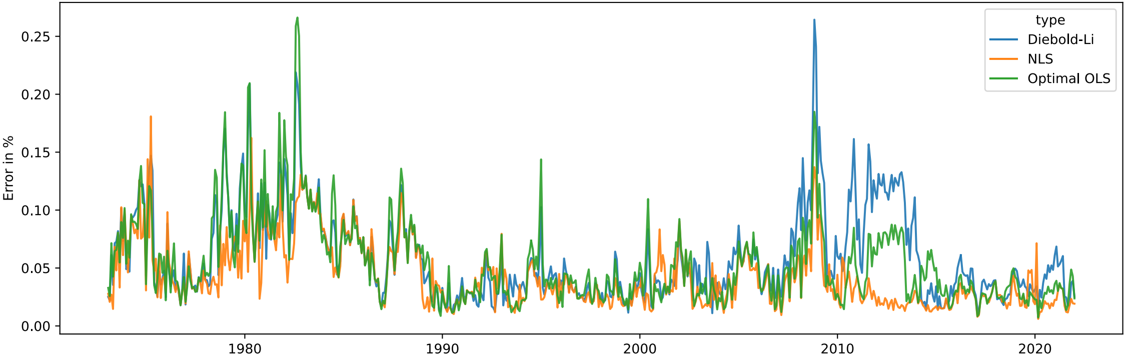
In-Sample Evidence Forecasting Returns

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



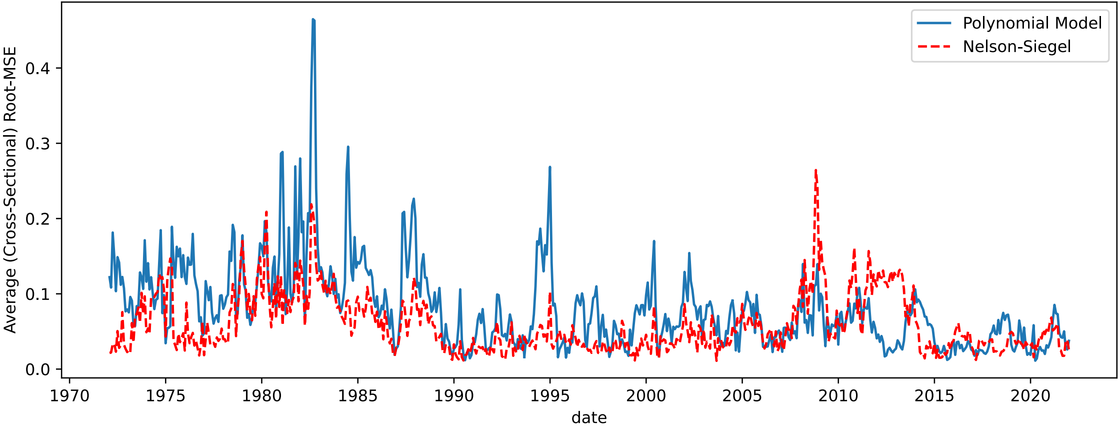
- NLS stands for Non-Linear Least Squares Date by Date
- Optimal OLS is the in-sample best OLS-implied decay fit

Alternative Estimation Procedures

Back

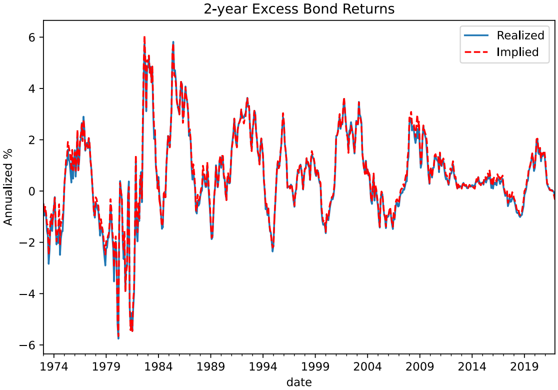
A quadratic polynomial model:

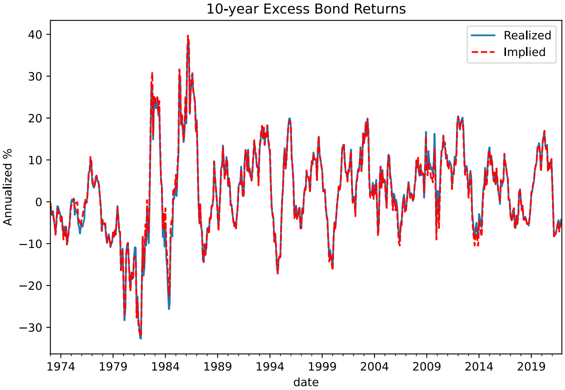
$$y_t^{(\tau)} = c_{1,t} + c_{2,t} \cdot \tau + c_{3,t} \cdot \tau^2 \tag{9}$$



Is this a reasonable model for the US Nominal Yield Curve?

Back





- Blue: $rx_t(n)$ observed from data for $n = 2$ and $n = 10$
- Red: $rx_t(n)$ that would have been implied by our estimates of the factors
- A Nelson-Siegel model fits well the American nominal yield curve
- The Fed actually uses a variant of the NS model to report their yield curve

Estimation Details

Back

Define the following matrices for each time t:

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

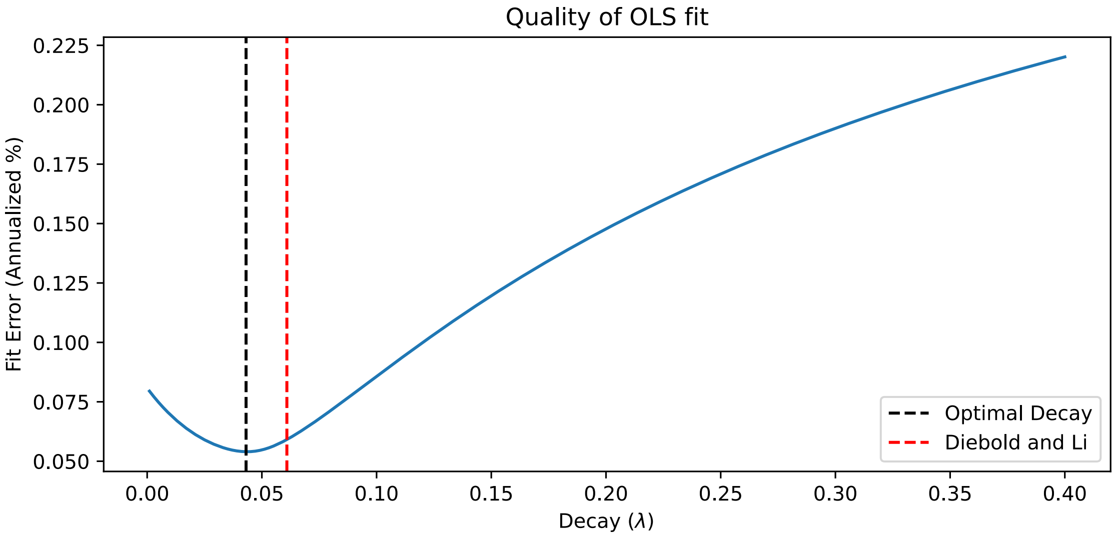
Now estimate betas using OLS:

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

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

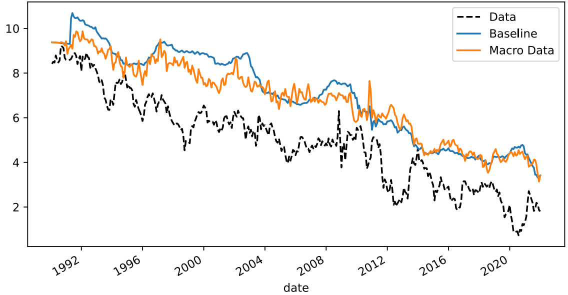
Back

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

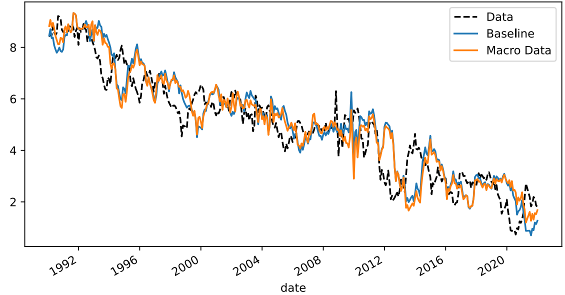
Regularization Failure for β_1

Back

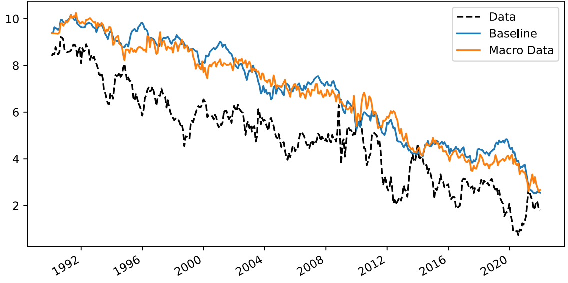
Controlling by Lagged Factors, Targeting the Level



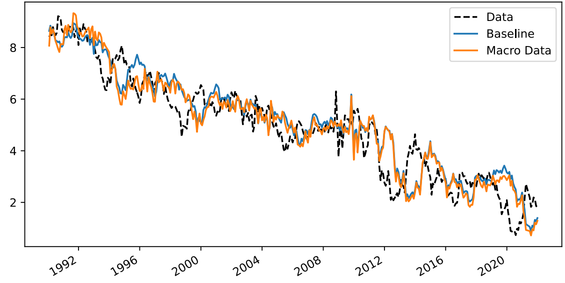
Controlling by Lagged Factors, Targeting the Innovations



Controlling by Lagged Forward Rates, Targeting the Level

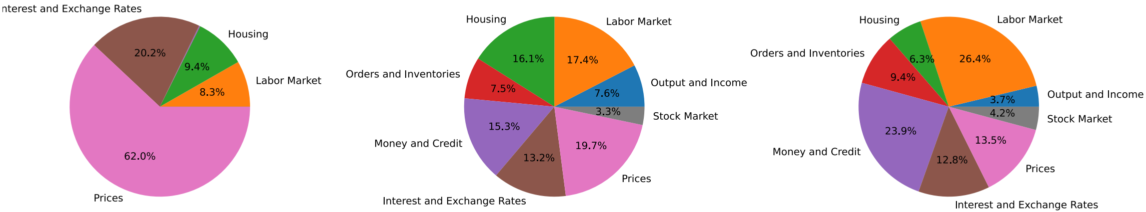


Controlling by Lagged Forward Rates, Targeting the Innovations



Model Selection - Lasso

How frequently are variables from each group chosen?

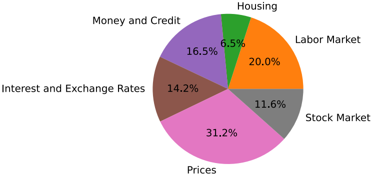


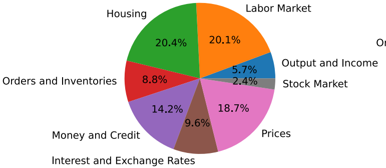
- Typical number of chosen variables is around 10-15
- Price measures are the leading predictor for β_1 - echoes Joslin et al (2014)
- Short and medium run: the "illusion of sparsity" - Giannone et al (2021)

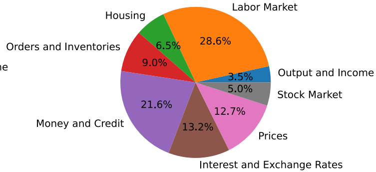
▸ Elastic Net ▸ Back

Model Selection - Elastic Net

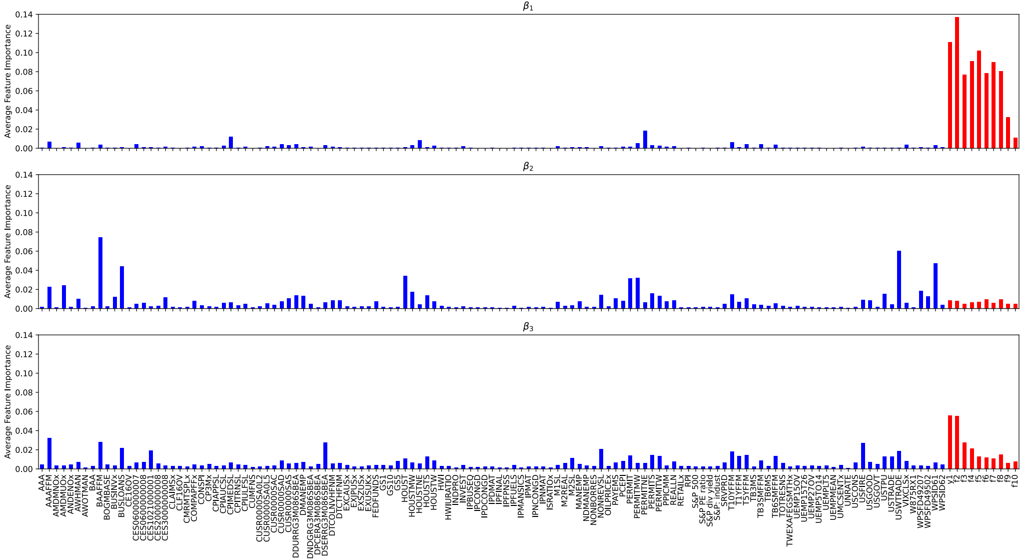
Back





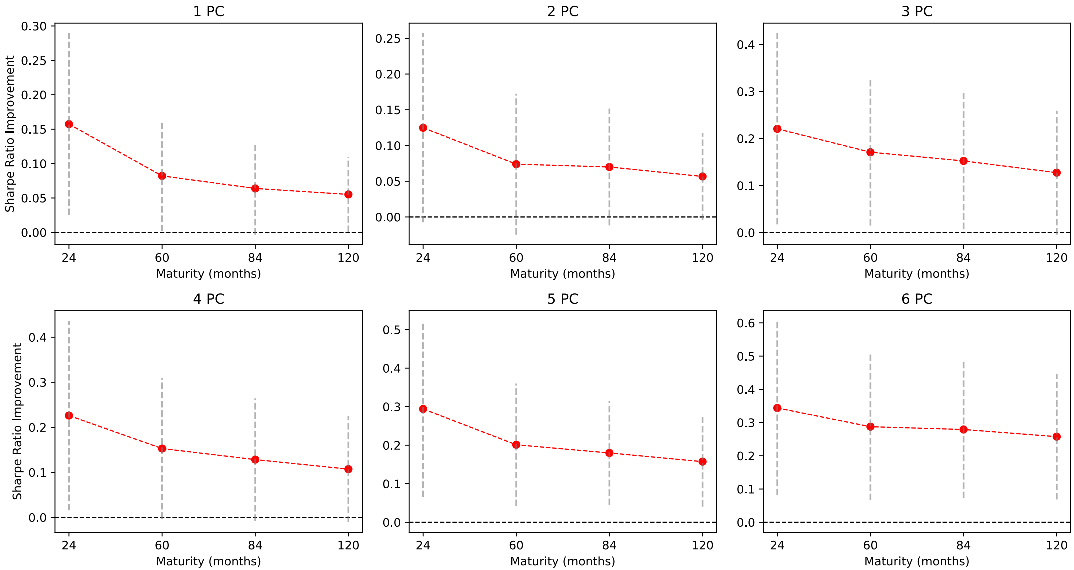


Feature Importance

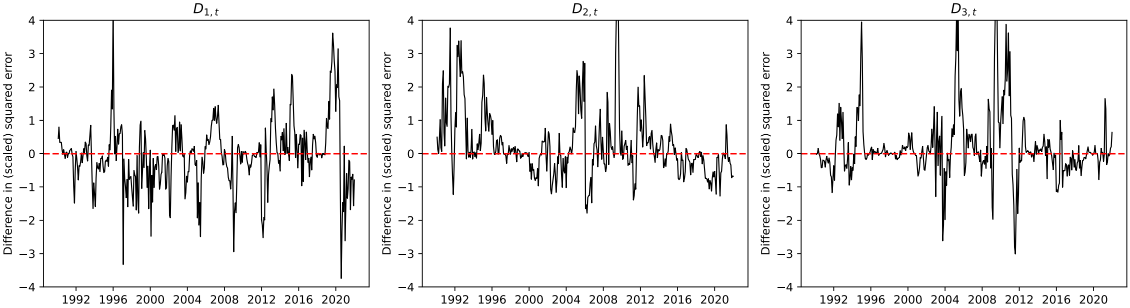


Unconstrained Sharpe Ratio Improvement

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

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

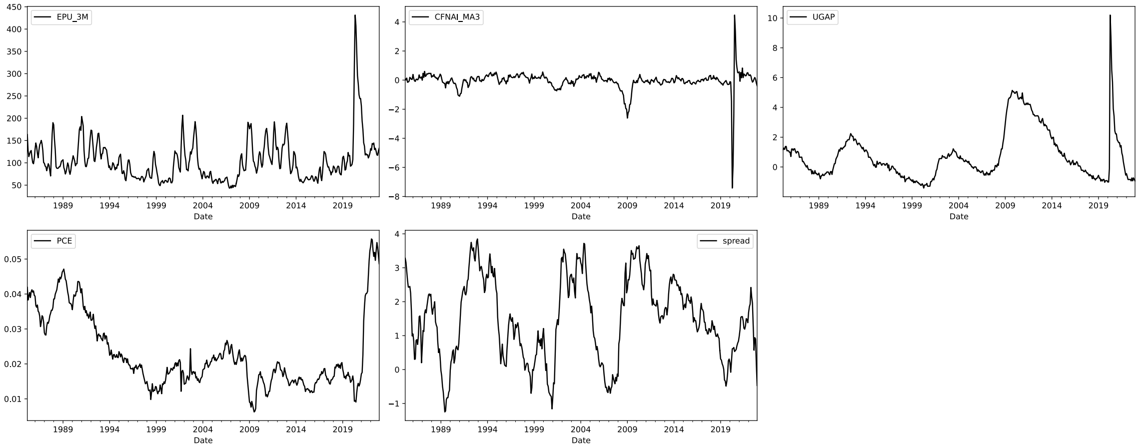
$$\widehat{\Omega}_T \equiv \frac{1}{T-h-t_0} \sum_{t=t_0}^{T-h} \mathbf{z}_{t+h}\mathbf{z}'_{t+h} + \frac{1}{T-h-t_0} \sum_{j=1}^{h-1} w_{j,T} \sum_{t=t_0+j}^{T-h} \left(\mathbf{z}_{t+h-j}\mathbf{z}'_{t+h} + \mathbf{z}_{t+h}\mathbf{z}'_{t+h-j} \right)$$

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

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

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

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

References I

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

Bianchi, D., BÃijchner, 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.

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

Dewachter, H. and Lyrio, M. (2006). Macro factors and the term structure of interest rates. *Journal of Money, Credit and Banking*, 38(1):119–140.

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.

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

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

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

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*.

References II

HÃďnnikÃďinen, 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.

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.

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

Piazzesi, M. and Schneider, M. (2007). Equilibrium Yield Curves. In *NBER Macroeconomics Annual 2006, Volume 21*, NBER Chapters, pages 389–472. National Bureau of Economic Research, Inc.

Rudebusch, G. D. and Swanson, E. T. (2012). The bond premium in a dsge model with long-run real and nominal risks. *American Economic Journal: Macroeconomics*, 4(1):105–143.

Rudebusch, G. D. and Wu, T. (2008). A macro-finance model of the term structure, monetary policy and the economy. *The Economic Journal*, 118(530):906–926.

Sarno, L., Schneider, P., and Wagner, C. (2016). The economic value of predicting bond risk premia. *Journal of Empirical Finance*, 37(C):247–267.

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.

Wachter, J. (2006). A consumption-based model of the term structure of interest rates. *Journal of Financial Economics*, 79(2):365–399.