Asymmetric Violations of the Spanning Hypothesis

Gustavo Freire[1] **Raul Riva**[2]

[1]Erasmus University
[2]Northwestern University

Fields Institute Bootcamp on Machine Learning on Finance

April 25th, 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, 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, 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;
  - The dynamics of underlying 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);
  - Inference is challenging and often in-sample: 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);
  - Inference is challenging and often in-sample: 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 at the longer end;
- Focus on **out-of-sample** prediction: closer to what a practitioner would do;
- Why should we care? Violations are economically meaningful for a mean-variance investor;
- Stronger violations when inflation is higher;

How do we do it?

1. Decompose the US zero-coupon yield curve in 3 factors related to yield maturities
   - Nelson-Siegel representation: $y_t^{(\tau)} = \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$;
   - Our estimation technique implies that β's are special linear combinations of zero-coupon yields;
   - We provide an explicit map between innovations in factors and realized excess bond returns;
   - Spanning Hypothesis $\implies$ macro data **should not** improve forecasting out-of-sample ;

How do we do it?

1. Decompose the US zero-coupon yield curve in 3 factors related to yield maturities
   - Nelson-Siegel representation: $y_t^{(\tau)} = \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$;
   - Our estimation technique implies that β's are special linear combinations of zero-coupon yields;
   - We provide an explicit map between innovations in factors and realized excess bond returns;
   - Spanning Hypothesis $\implies$ macro data **should not** improve forecasting out-of-sample ;
2. Use several different ML methods to predict bond risk premia and β's
   - We create forecasts using current β's + macro variables;
   - Macro data: FRED-MD $\implies$ monthly frequency, high-dimensional dataset, several macro signals;
   - We benchmark our forecasts against a random walk: hard to beat, and available under the SH;
   - Full sample: 1973-2021; Out-of-sample: 1990-2021; Focus on 1-year ahead forecasts;
   - Main lesson: **all** predictability of bond returns with macro data comes from β_2.

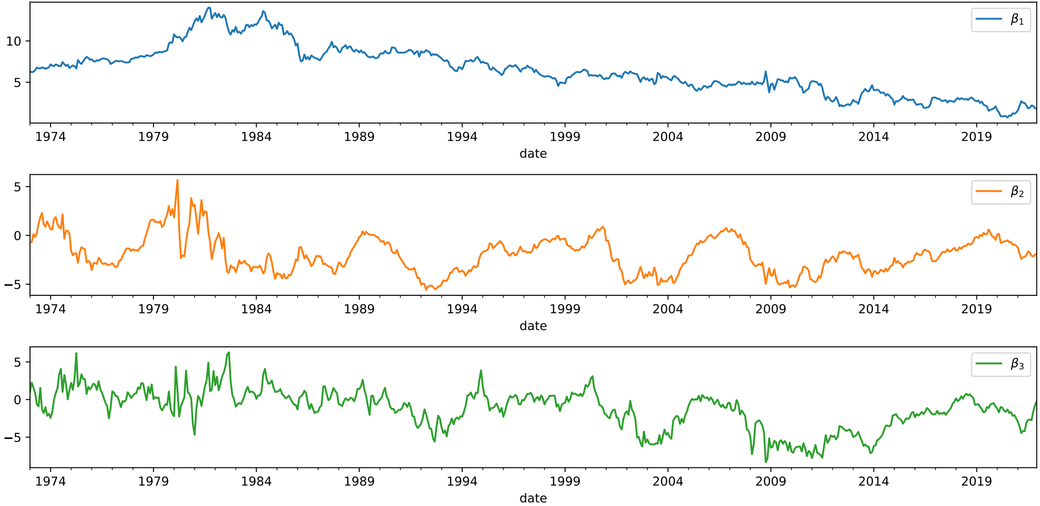
How do we do it?

1. Decompose the US zero-coupon yield curve in 3 factors related to yield maturities
   - Nelson-Siegel representation: $y_t^{(\tau)} = \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$;
   - Our estimation technique implies that β's are special linear combinations of zero-coupon yields;
   - We provide an explicit map between innovations in factors and realized excess bond returns;
   - Spanning Hypothesis $\implies$ macro data **should not** improve forecasting out-of-sample ;

2. Use several different ML methods to predict bond risk premia and β's
   - We create forecasts using current β's + macro variables;
   - Macro data: FRED-MD $\implies$ monthly frequency, high-dimensional dataset, several macro signals;
   - We benchmark our forecasts against a random walk: hard to beat, and available under the SH;
   - Full sample: 1973-2021; Out-of-sample: 1990-2021; Focus on 1-year ahead forecasts;
   - Main lesson: **all** predictability of bond returns with macro data comes from β_2.

3. Mean-variance trading strategy: a more complex model helps trading shorter maturities
   - Sharpe ratios improve when we trade based on macro signals, but *asymmetrically*

Factor Realizations (1973-2021)



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

Decomposing Excess Bond Returns

The 1-year excess bond returns for a maturity of n years are given by:

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

Proposition 1

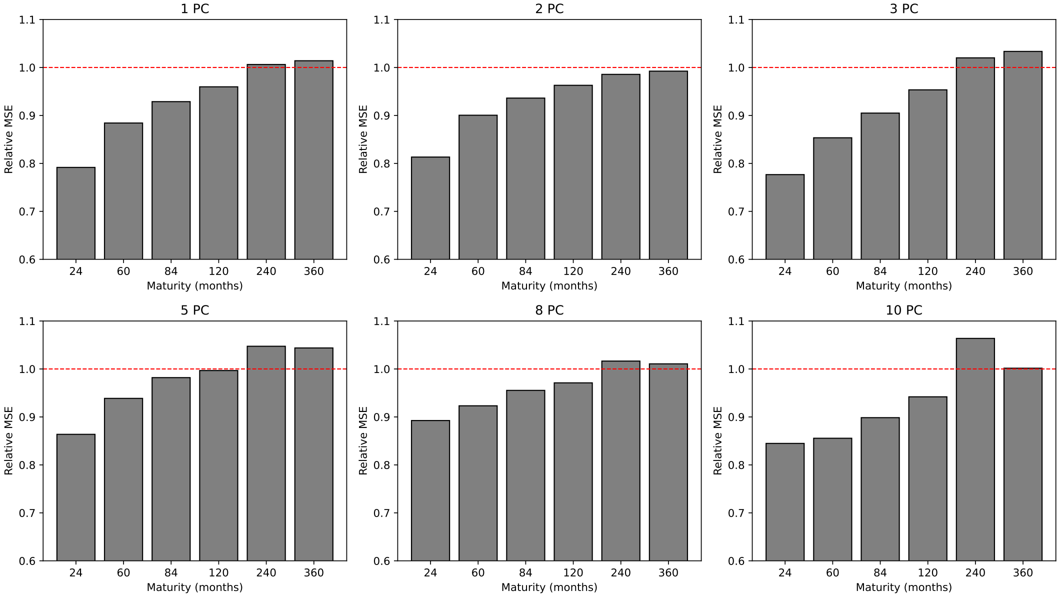
Suppose the yield curve follows the Nelson-Siegel representation and assume that the decay parameter is a positive constant $\lambda_t = \lambda > 0$. Define $\theta \equiv 12\lambda$. The following holds:

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

MSE Ratios With and Without Macro Data

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

$$xr_{t+12}(n) = \alpha_n + \theta_n' C_t + \gamma_n' PC_t + \epsilon_{t+12,n} \quad (3)$$



Regularized Linear Models

We use a supervised technique to forecast β's and compare predictions with a random walk:

$$R^2_{oos} = 1 - \left[\sum_{t=t_0}^{T} \left(\beta_{i,t} - \widehat{\beta}_{i,t} \right)^2 \Big/ \sum_{t=t_0}^{T} \left(\beta_{i,t} - \overline{\beta}_{i,t} \right)^2 \right] \tag{4}$$

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

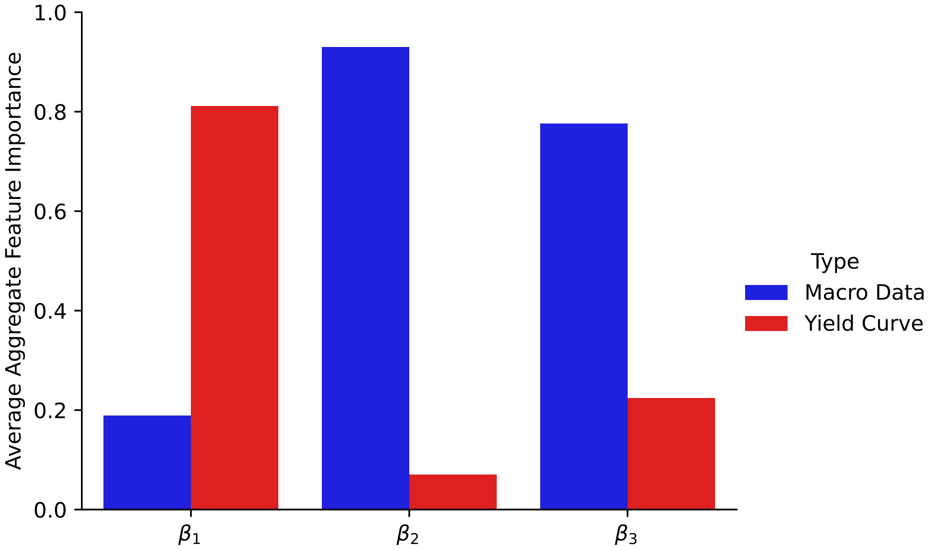
▸ Other Controls ▸ Stationarity Matters ▸ Chosen variables

Random Forrest

- Why to constrain ourselves to linear methods? We deploy a standard RF methodology;
- Medeiros et al. (2021), Goulet-Coulombe (2023) show how RF is well suited for macro data;
- We grow 500 trees at each step and forecast one year ahead;

	Control: Lagged Factors			Control: 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

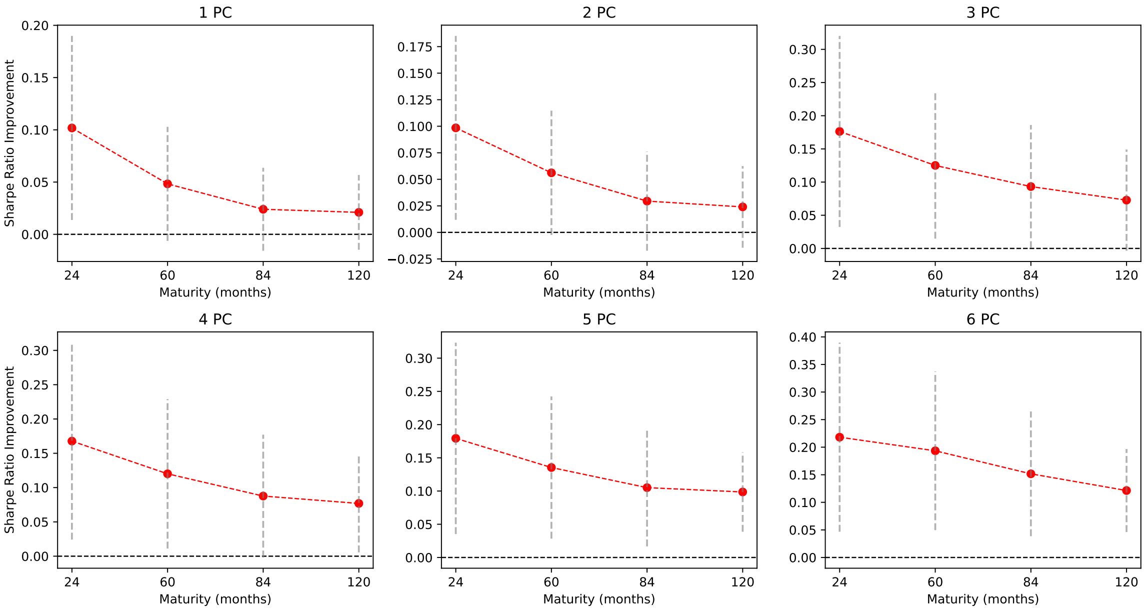
Average Feature Importance (Macro Variables vs Yield Curve)



▸ Individual Feature Importance

Baseline Sharpe Ratio ≈ 0.2 (Constrained Case)

▸ Unconstrained Case



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 $\approx 0.1 - 0.2$;
- (Not shown today) 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 $\approx 0.1 - 0.2$;
- (Not shown today) 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 DSTM 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 $\approx 0.1 - 0.2$;
- (Not shown today) 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 DSTM with an unknown reaction function look like? Future work!

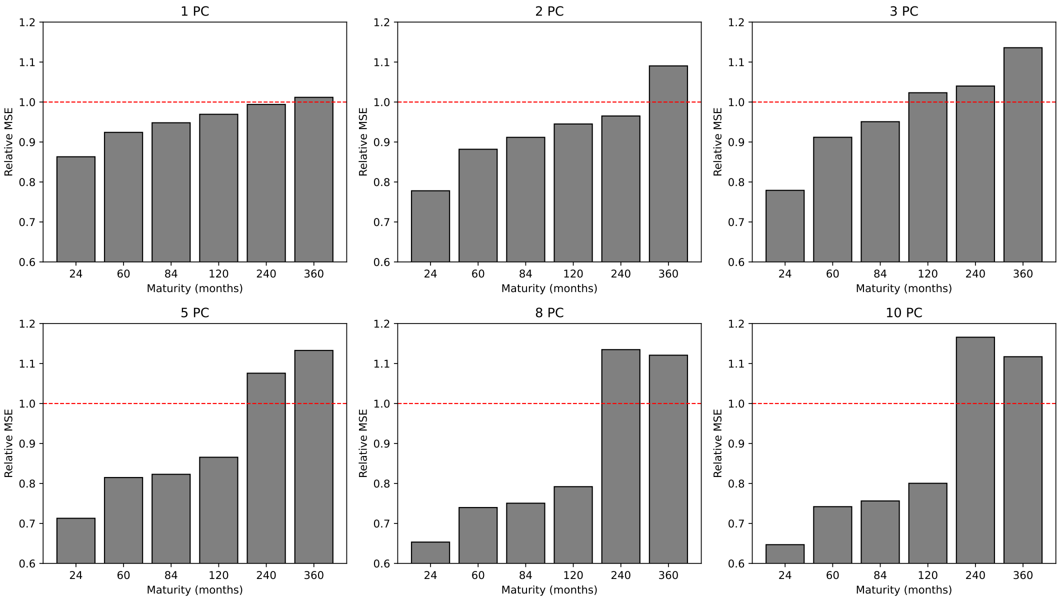
Thank you!

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

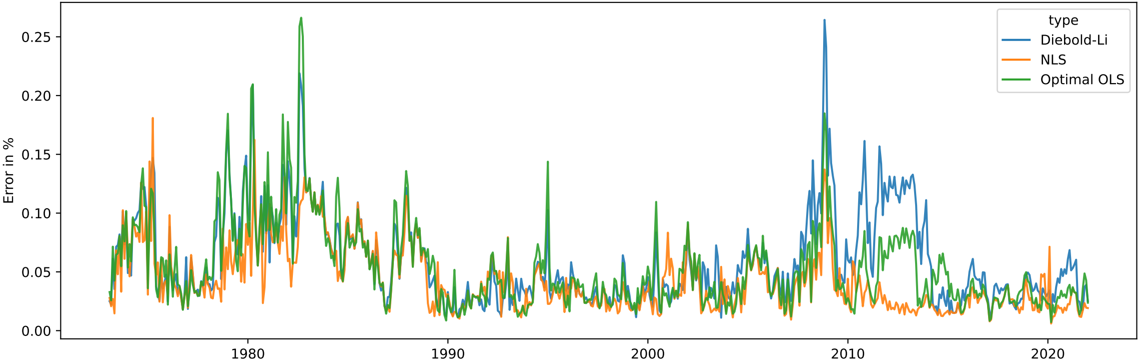
In-Sample Evidence Forecasting Returns

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



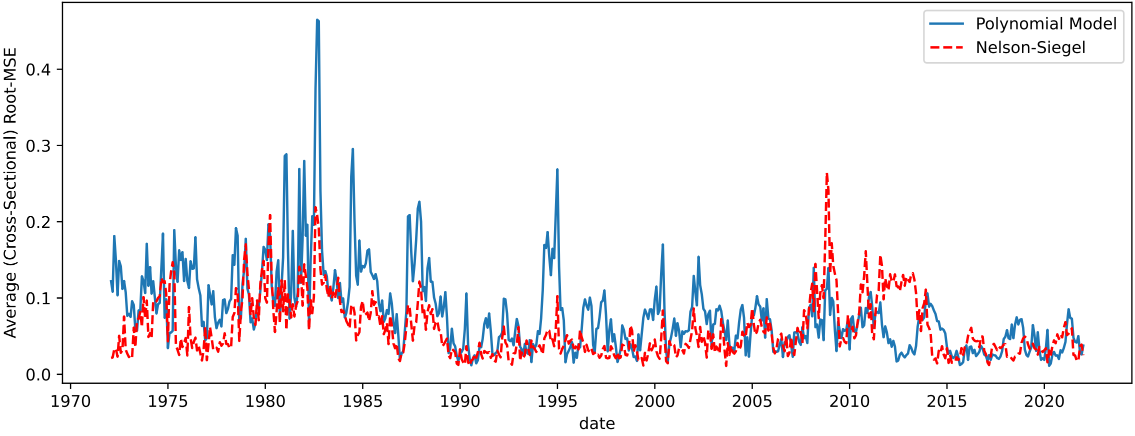
- NLS stands for Non-Linear Least Squares Date by Date
- Optimal OLS is the in-sample best OLS-implied decay fit

Alternative Estimation Procedures

▸ Back

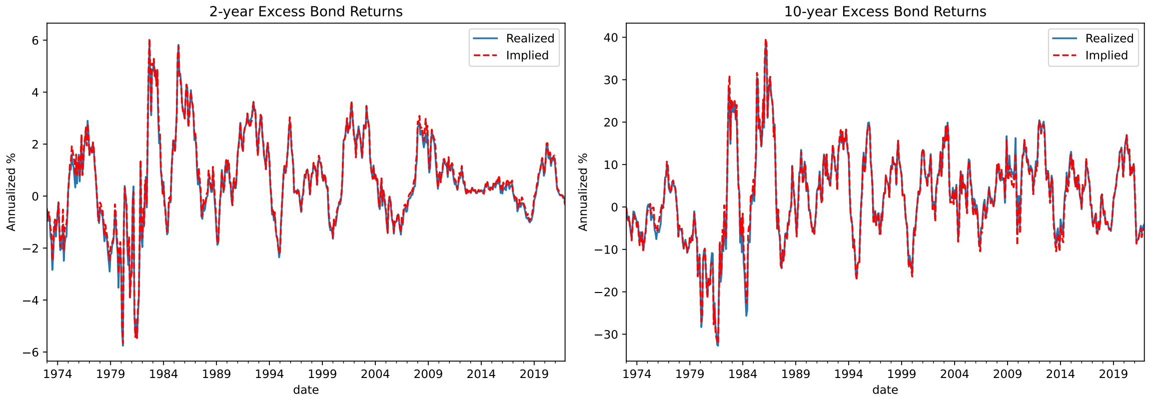
A quadratic polynomial model:

$$y_t^{(\tau)} = c_{1,t} + c_{2,t} \cdot \tau + c_{3,t} \cdot \tau^2 \tag{5}$$



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:

$$X \equiv \begin{bmatrix} 1 & \left(\frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1}\right) & \left(\frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1}\right) \\ \vdots & \vdots & \vdots \\ 1 & \left(\frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N}\right) & \left(\frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N}\right) \end{bmatrix}, \quad Y_t = \begin{bmatrix} y_t^{(\tau_1)} \\ \vdots \\ y_t^{(\tau_N)} \end{bmatrix} \tag{6}$$

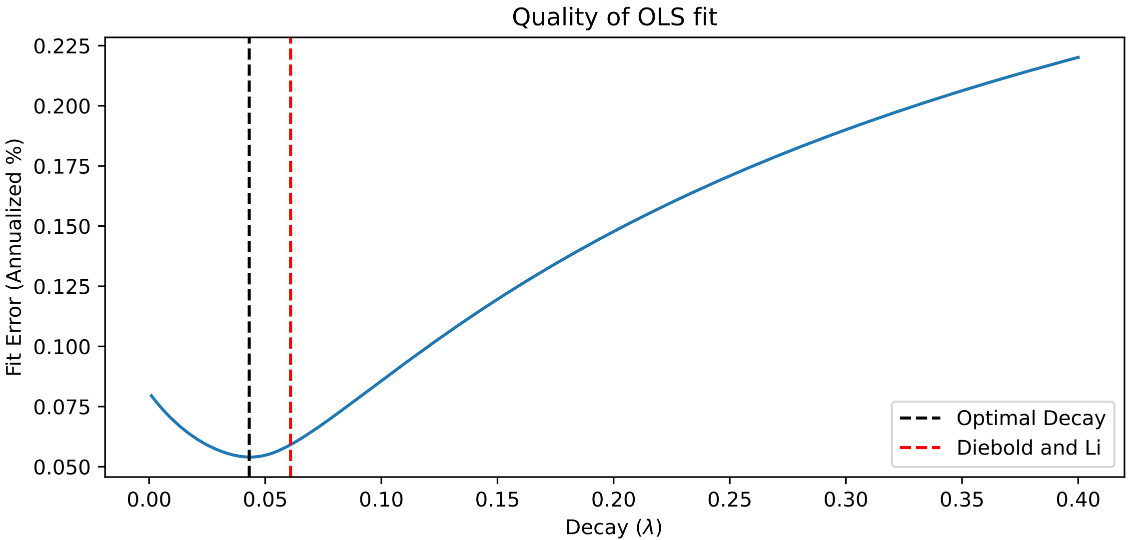
Now estimate betas using OLS:

$$\begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{bmatrix} = \left(X'X\right)^{-1} X'Y_t \tag{7}$$

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

		Number of Macro PCs						p-values					
Target	No Macro	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

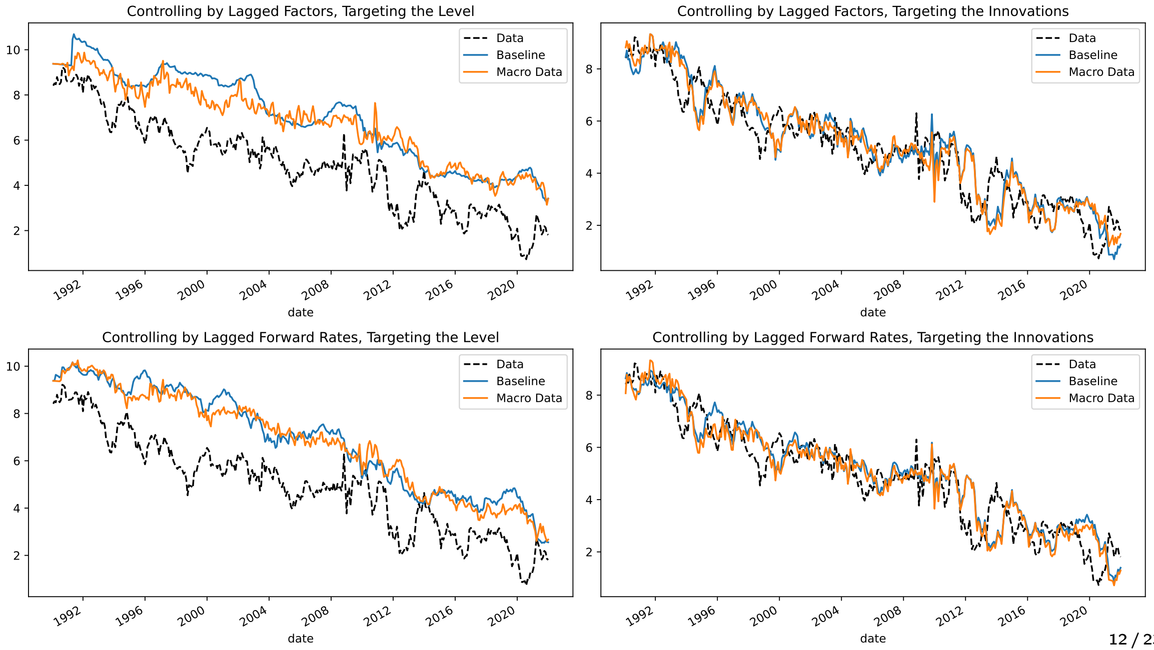
Regularization Methods - Controlling by Lagged Betas

▸ 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

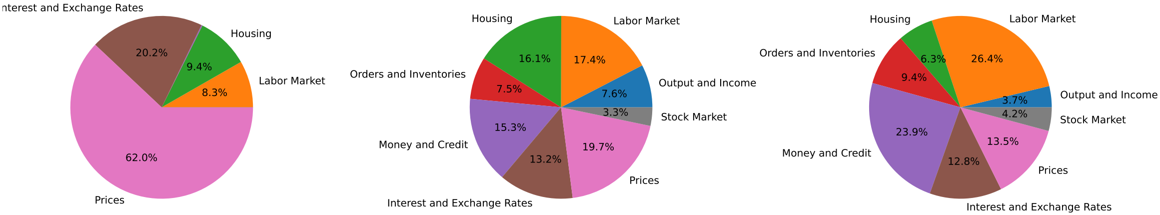
Regularization Failure for β_1

▸ Back



Model Selection - Lasso

How frequently are variables from each group chosen?

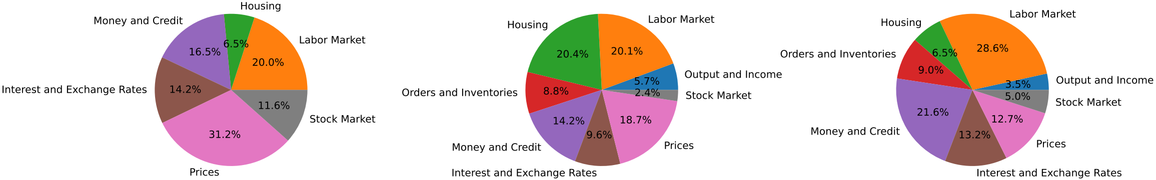


- Typical number of chosen variables is around 10-15
- Price measures are the leading predictor for β_1 - echoes Joslin et al (2014)
- Short and medium run: the “illusion of sparsity” - Giannone et al (2021)

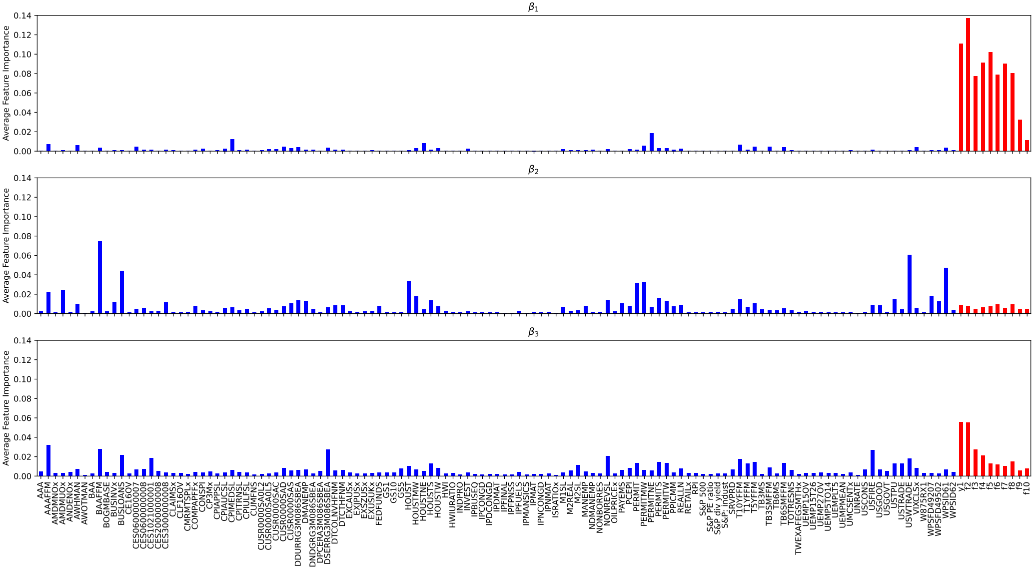
▸ Elastic Net ▸ Back

Model Selection - Elastic Net

▸ Back

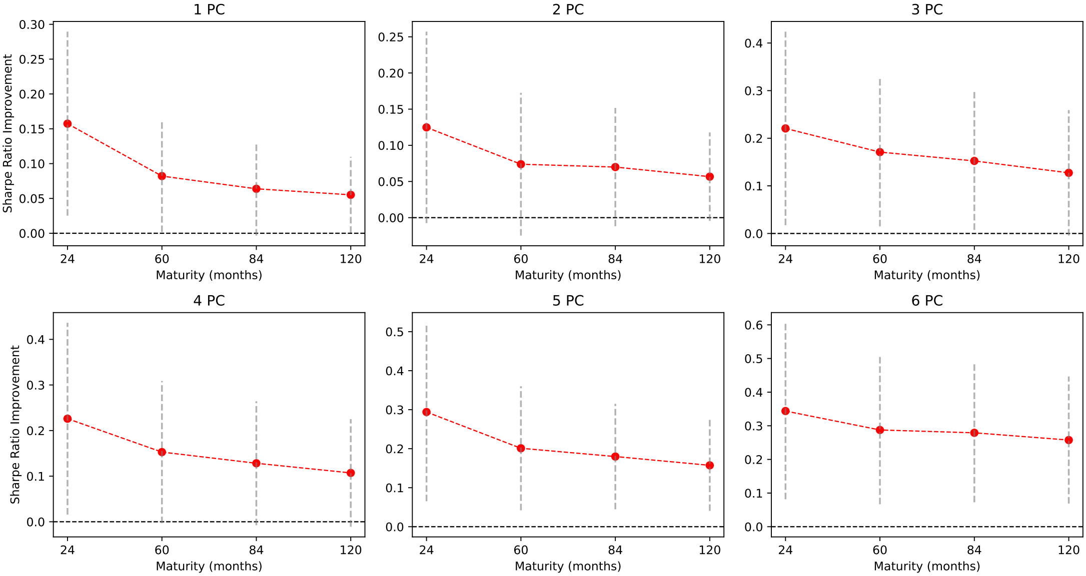


Feature Importance

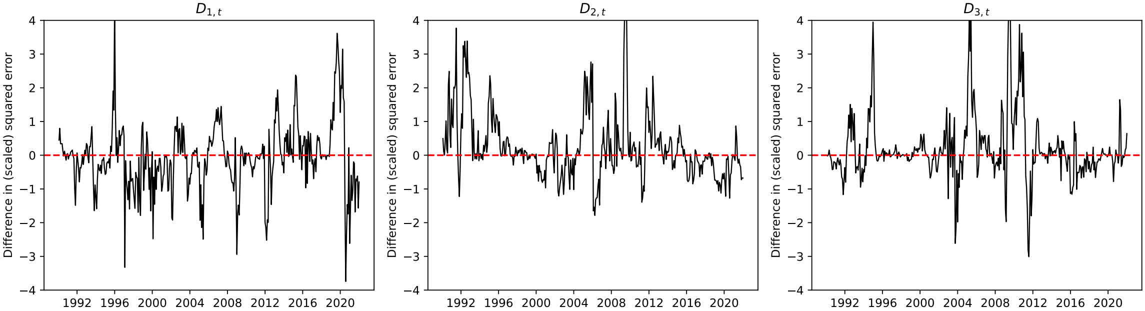


Unconstrained Sharpe Ratio Improvement

▸ 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 $z_{t+h} \equiv x_t \left(L^{m'}_{t+h} - L^{m}_{t+h} \right)$ for a given forecasting horizon h
- Define

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

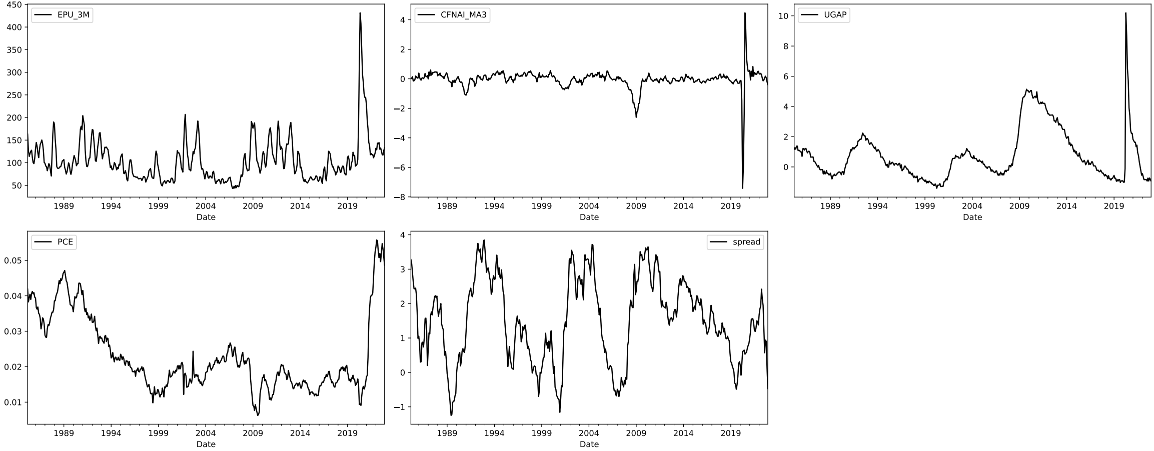
$$\widehat{\Omega}_T \equiv \frac{1}{T-h-t_0} \sum_{t=t_0}^{T-h} z_{t+h} z'_{t+h} + \frac{1}{T-h-t_0} \sum_{j=1}^{h-1} w_{j,T} \sum_{t=t_0+j}^{T-h} \left(z_{t+h-j} z'_{t+h} + z_{t+h} z'_{t+h-j} \right)$$

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

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

$$W \equiv T \cdot z'_{t+h} \widehat{\Omega}_T^{-1} z_{t+h} \xrightarrow{d} \chi^2_q \tag{8}$$

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.

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.

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

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

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

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.

References II

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

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.

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.

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