Asymmetric Violations of the Spanning Hypothesis

Gustavo Freire, **Raul Riva**
Northwestern University
Kellogg School of Business
`raul.riva@kellogg.northwestern.edu`

Northwestern | Kellogg

Abstract

Many dynamic term-structure models imply that the current yield curve is *the* sufficient statistic to predict future bond yields and bond risk premia. This is the so-called *Spanning Hypothesis*. We develop an out-of-sample forecasting exercise leveraging big data and methods from the Machine Learning literature, coupled with a dynamic Nelson-Siegel model, and show that **all predictability steams from predictability of shorter rates**. Violations of the Spanning Hypothesis are *asymmetric* across the yield curve. We also show that not using macro data might cost 0.2-0.4 in terms of annual Sharpe ratio for a mean-variance utility consumer.

Introduction

- Understanding the dynamics of the yield curve is key for portfolio choice, monetary policy conduction, risk management and fiscal policy evaluation;
- Dynamic term structure models are useful for separating term premia from expectation of short rates;
- Common implication of these models: the **Spanning Hypothesis**
  - Bond prices (or yields) should reflect the underlying state of the macroeconomy;
  - Macroeconomic variables **should not** improve the forecast of bond yields and holding excess returns;
- Past literature focuses on in-sample inference exercises. Here: **out-of-sample** forecasting;

This Paper

1. **Can macroeconomic variables help predict excess bond returns out-of-sample?**
   - Yes, but asymmetrically: macro data is only useful when trying to predict shorter rates ($\approx 2-5$ years). Longer rates (≈ 10 years) behave as theory predicts.
   - This is true across many different forecasting methods.
2. **Why does it happen?**
   - We deploy a dynamic Nelson-Siegel and provide a decomposition of excess bond returns into innovations of these factors;
   - **All predictability comes from the short-run factor**, which gets distributed along the yield curve;
3. **Is this economically meaningful?**
   - A mean-variance consumer that uses macro data information to improve her forecasts and trade has a non-trivial increase in realized Sharpe ratios (from 0.3 to 0.5-0.6);
   - This improvement is available when trading bonds of shorter maturities, but is much smaller when trading longer-maturity bonds;

Methodology and Empirical Results

Let $y_t^{(n)}$ be the n-year zero-coupon yield at month t. The 1-year excess return on an n-year bond, realized between t and $t+12$, is given by:

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

First exercise: predict $xr_{t+12}^{(n)}$ using a linear model where C_t are forward rates and PC_t is a vector of principals components from the FRED-MD dataset:

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

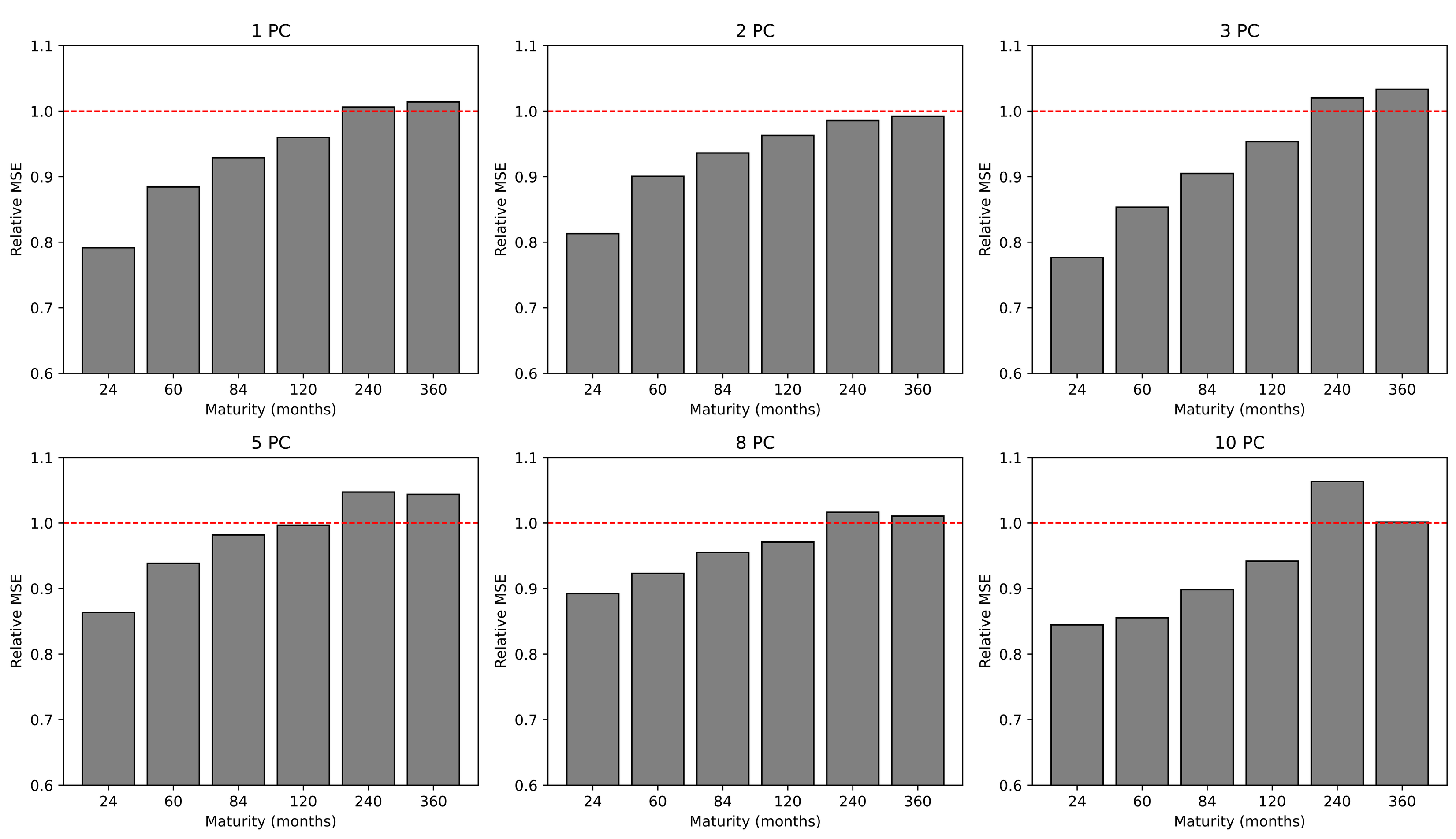

Figure 1: Ratio between out-of-sample mean-squared error without/with macro data. The spanning hypothesis implies that **bars should be around the red line - centered on 1**. The out-of-sample period is 1990-2021. Different panels run the exercise with different numbers of principal components extracted from the FRED-MD dataset.

A common statistical model for the yield curve is the **Nelson-Siegel** one:

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

We fit the model by OLS, month by month. The time-varying betas have interpretations:

- $\beta_{1,t} = \lim_{\tau\to\infty} y_t^{(\tau)} \implies$ long-run factor;
- $\beta_{2,t} = -\left(\lim_{\tau\to\infty} y_t^{(\tau)} - \lim_{\tau\to 0} y_t^{(\tau)}\right) \implies$ short-run factor, since it moves short rates disproportionately more;
- $\beta_{3,t} \implies$ medium-run factor since its coefficients starts at 0, converges to zero as τ diverges but it has a hump;

Proposition 1. *Suppose the yield curve follows a 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*

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

- Terms in brackets depend on t but not on the maturity. These are innovations in factors;
- Terms in parentheses *only* depend on maturity;

Now forecast the factors (or their innovations) with different methods, with and without a large panel of macroeconomic data X_t in addition to the forward rates C_t:

$$\beta_{i,t+12} = g(C_t, X_t) + u_{i,t+12}$$

Methods: 1) PCA + Linear regression; 2) Regularized linear models; 3) Random Forest;
Forecasting performance is measured by the out-of-sample R^2 against a random walk:

$$R^2_{i,OOS} = 1 - \left(\sum_t \left(\beta_{i,t+12} - \widehat{\beta}_{i,t+12}\right)^2\right) \Big/ \left(\sum_t \left(\beta_{i,t+12} - \beta_{i,t}\right)^2\right)$$

		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
$\Delta\beta_1$	-0.19	-0.15	-0.17	-0.14	-0.10	-0.08	0.19	0.32	0.17	0.12	0.10
$\Delta\beta_2$	-0.11	-0.12	**0.14**	**0.18**	**0.17**	**0.19**	0.52	0.00	0.02	0.02	0.02
$\Delta\beta_3$	-0.10	-0.12	-0.06	-0.05	-0.05	-0.06	0.93	0.17	0.25	0.26	0.31

Table 1: R^2_{OOS} with and without macro data achieved by a linear regression that control for forward rates and different numbers of principal components of the macro data. The p-values refer to a one-sided test in which, **under the null, macro data should not be helpful.**.

	No Macro Data				All Macro Data				p-value			
Target	Ridge	Lasso	ENet	RF	Ridge	Lasso	ENet	RF	Ridge	Lasso	ENet	RF
$\Delta\beta_1$	0.12	0.12	0.09	-0.53	0.01	0.12	0.12	-0.05	0.96	0.50	0.27	0.00
$\Delta\beta_2$	0.01	-0.02	-0.01	-0.42	**0.15**	**0.22**	**0.19**	**0.32**	0.02	0.00	0.00	0.00
$\Delta\beta_3$	0.04	-0.02	-0.03	-0.34	-0.13	-0.09	-0.08	-0.26	1.00	0.95	0.95	0.23

Table 2: R^2_{OOS} with and without macro data achieved by regularized linear methods and the Random Forest. The p-values refer to a one-sided test in which, **under the null, macro data should not be helpful.**.

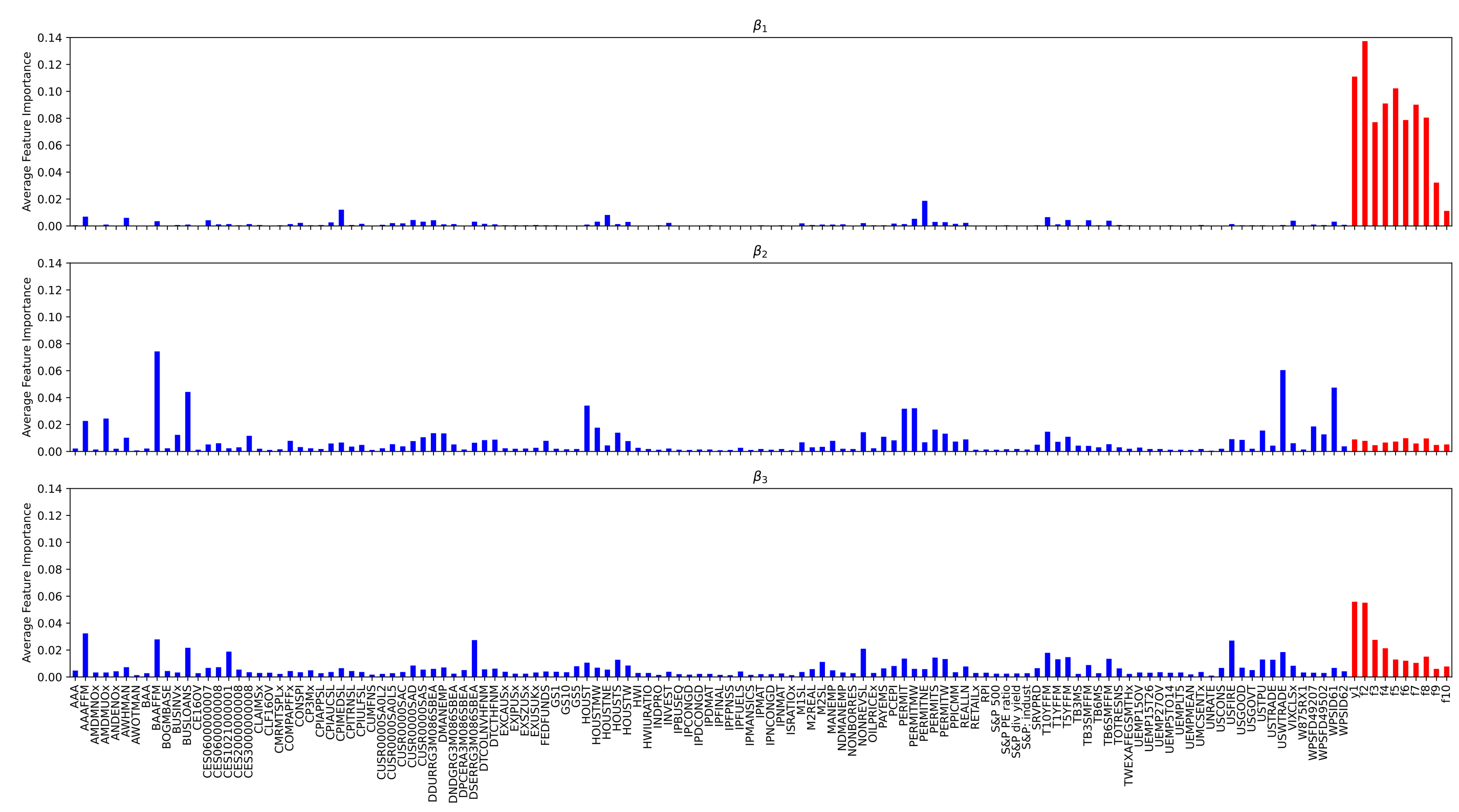

Figure 2: Feature importance computed by the Random Forest methodology. The bars represent how much splits on each variable contributed to the overall decrease in in-sample MSE reduction. Red bars are forward rates while blue bars are variables from macro data. From top to bottom, β_1, β_2, and β_3.

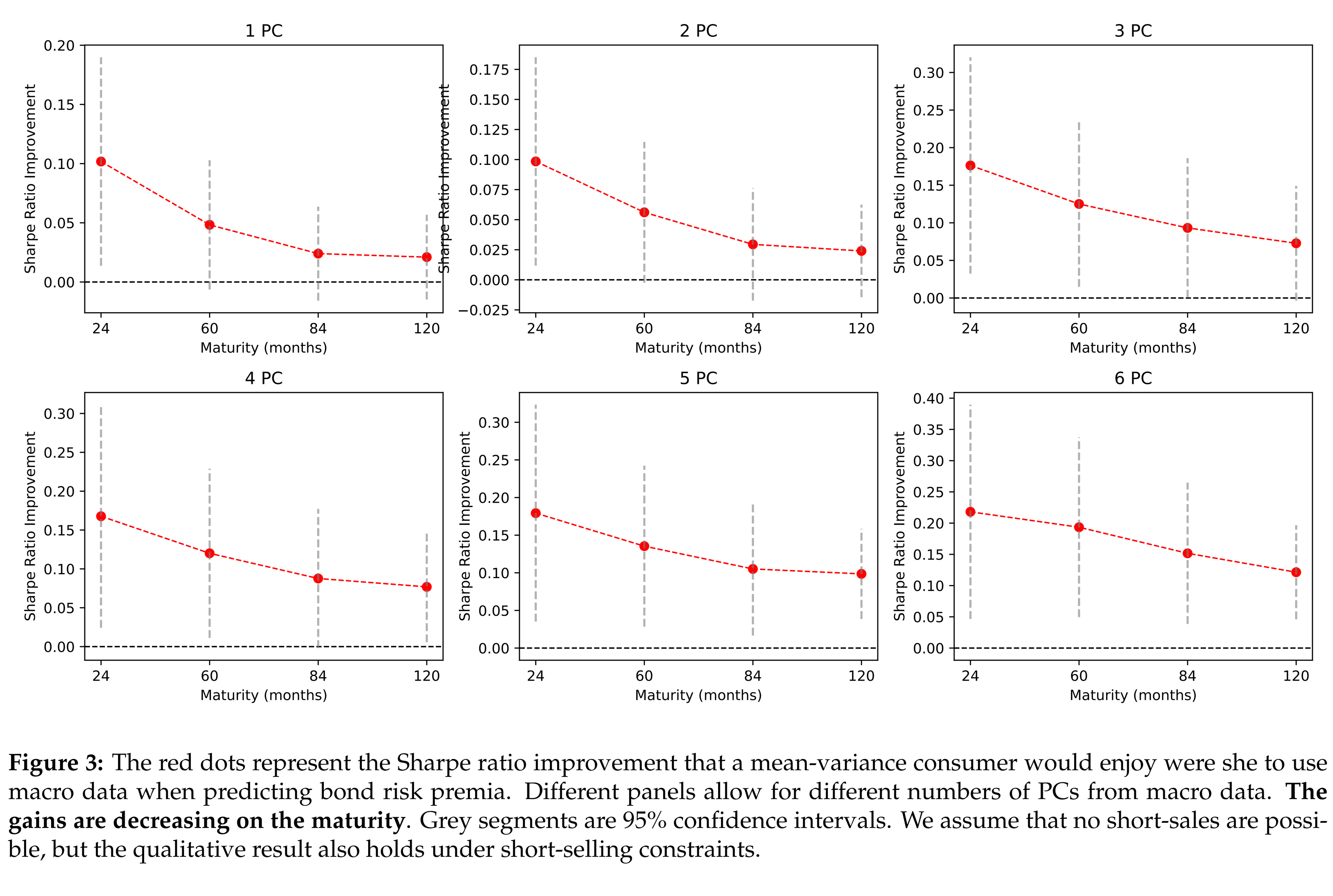

Figure 3: The red dots represent the Sharpe ratio improvement that a mean-variance consumer would enjoy were she to use macro data when predicting bond risk premia. Different panels allow for different numbers of PCs from macro data. **The gains are decreasing on the maturity**. Grey segments are 95% confidence intervals. We assume that no short-sales are possible, but the qualitative result also holds under short-selling constraints.