Yield Curve Momentum

Markus Sihvonen*

April 12, 2022

Abstract

I analyze time series momentum along the Treasury term structure. Past bond returns predict future returns both due to autocorrelation in bond risk premia and because unexpected bond return shocks increase the premium. Yield curve momentum is primarily due to autocorrelation in yield changes rather than autocorrelation in bond carry and can largely be captured using a single bond return or yield change factor. Because yield changes are partly induced by changes in the federal funds rate, yield curve momentum is related to post-FOMC announcement drift. The momentum factor is unspanned by the information in the term structure today and is hence inconsistent with standard term structure, macrofinance and behavioral models. I argue that the results are consistent with a model with unpriced longer term dependencies.

Keywords: Bond risk premia, time series momentum, term structure models, post-FOMC announcement drift.

JEL classification: G12, E43, E47

*Bank of Finland, Research Unit, Snellmaninaukio, P.O. Box 160. Email: markus.sihvonen@bof.fi. I thank Michael Bauer, Ryan Chahrour, Anna Cieslak, Laura Coroneo, Rubens Moura, Emanuel Mönch, George Pennacchi, Andrea Polo, Dimitri Vayanos and Michael Weber as well as participants at the Econometric Society European Winter Meeting, Dynamic Econometrics Conference and Bank of Finland for useful comments.

1 Introduction

Past returns can predict future returns (Fama, 1965). Moskowitz et al. (2012) find evidence of medium horizon return autocorrelation among a large set of asset classes. They dub this phenomenon ”time series momentum”.[1]

Possibly due to the focus on a broad set of asset classes, the time series momentum literature has evolved largely separately from the vast literature on term-structure modelling and bond risk premia (e.g., Ang and Piazzesi, 2003; Fama and Bliss, 1987; Cochrane and Piazzesi, 2005). Because of this disconnect it is for example not clear whether time series momentum of government bonds is consistent with standard term-structure models.[2] This paper is an attempt to study the finer dynamics of time series momentum of government bonds, or yield curve momentum, and close the gap between the two literatures.

I argue that the findings of Moskowitz et al. (2012) are not necessarily inconsistent with standard models. However, I present new evidence related to yield curve momentum, which clearly is incompatible with such models.

First, I find that the term structure of momentum coefficients is downward sloping. Slope coefficients from regressing bond returns on the past return of the same maturity bond decline with the bond's maturity.

Second, I argue that yield curve momentum occurs both because of autocorrelation in bond yield changes and bond carry. However, because bond carry has small variation, most of the covariance between current and

[1] This is a growing literature, see e.g. Pitkäjärvi et al. (2020), Huang et al. (2020) and Goyal and Jegadeesh (2018).

[2] Durham (2013) analyzes the performance of a duration neutral cross-sectional momentum strategy with government bonds. He argues that some its profitability can be explained by a specific affine term structure model. However, he does not address time series momentum. Asness et al. (2013) study a cross-country momentum strategy with government bonds finding that such a strategy yields positive yet fairly small returns. Brooks and Moskowitz (2017) explain bond returns using value, momentum and carry factors. However, they do not study the sources of momentum or relate the findings to the term structure modelling literature. Osterrieder and Schotman (2017) connect bond return autocorrelations with model risk parameters but do not explicitly address momentum.

past returns is due to autocorrelation in bond yield changes.

I also decompose yield curve momentum into autocovariance in bond risk premia and covariance between bond risk premia and past unexpected news to bond returns. On average the first channel explains roughly one third of yield curve momentum while the second explains two thirds of it.

Third, I analyze the factor structure of yield curve momentum. I find that yield curve momentum can be largely captured by the change in the first principal component of yields or a single momentum factor defined as the average past return of different maturity bonds.

Fourth, I assess the relationship between monetary policy and yield curve momentum. Because changes in the Treasury yield curve are related to changes in the federal funds target rate, yield curve momentum is partly induced by monetary policy. That is, yield curve momentum is in part driven by a drift pattern following a recent, expected or unexpected, rate change by the Fed. However, because especially long maturity yields display movements unrelated to target rate changes, yield curve momentum is not identical to post-FOMC announcement drift discussed in Brooks et al. (2019).

Fifth, I analyze whether yield curve momentum is consistent with standard term-structure models. The standard models imply that yields are affine in a set of factors. This form is also implied by standard macrofinance models, at least up to first order. These models can in principle generate covariance between current and past bond returns. However, this correlation should vanish after controlling for information in the current yield curve. The intuition is that, in this class models, the current factors determine the expected bond returns and after controlling for these factors no other variable should predict bond returns. On the other hand, these current factors are priced in the yield curve today. But then controlling for sufficiently many yields today is equivalent to controlling for the factors. I explain that this intuition carries to more complicated models after controlling for the generally non-linear relationship between bond returns and past yields.

I find that past bond returns predict future returns also conditional on the information in the yield curve today. Hence the spanning condition implied by standard models is violated in the data.

This point is similar to that made by Joslin et al. (2014), who study affine term structure models with macroeconomic factors. Emprically macro variables predict returns even after controlling for current yields. They therefore argue that the data can only be explained by a model with unspanned macro factors. However, they do not address momentum or consider the predictive power of past returns.

Can behavioral theories resolve my findings? Not necessarily. The reasons is that the current behavioral models still imply the same affine form for yields though the coefficients and factors might be different from rational models. Therefore these models still cannot generate yield curve momentum conditional on all the information in the yield curve today.

I propose a model that is consistent with the above empirical findings. In this model, factors exhibit longer term dependencies. However, these longer term relations are not priced in the term structure of interest rates today. Because past returns include information about such unpriced dependencies, they predict future returns also conditional on current yields.

I discuss two possible economic explanations for longer term dependencies to be unpriced. The first is a simple behavioral narrative: agents do not understand that factor dynamics have longer dependencies and bond prices reflect this misunderstanding. Second, I sketch a model with rational arbitraugeurs and simple rule-based traders. In this model, the demand from rule-based traders affects the duration risk that must be absorbed by arbitrageurs. This effect can offset some of the effects of expected short rates on bond prices and imply a violation of the standard spanning condition.

2 Data and Definitions

I use the dataset on zero coupon US Treasury yields constructed by Liu and Wu (2020). These yields are built using a novel non-parametric method, which implies lower pricing errors compared to previous interpolation procedures. I apply a sample of end of month data between August 1971 and

December 2019 and focus on the yields and returns on 1 to 10 year bonds as well as 1 month bills. In the appendix I show that the key results are robust to using the alternative dataset constructed by Gürkaynak et al. (2007), the data concerning the German yield curve available on the Bundesbank webpage and the Bloomberg US Treasury Index.

I obtain the federal funds target rate and the relevant target ranges from FRED. For monetary policy shock identification I utilize a series of the front month federal funds futures contract listed on the CME. I apply the information on the Federal Reserve web page to create a series of the meeting dates of the Federal Open Market Committee.

I construct a trend inflation measure as in Cieslak and Povala (2015). Here I apply a smoothing parameter of 0.987 in monthly updating terms to annual core inflation.[3] My measure of real activity is the three month moving average of the Chicago Fed National Activity Index (CFNAI), also used by Joslin et al. (2014).

I denote the monthly continuously compounded yield of a bond with n months until maturity by y_t^n. The logarithmic excess monthly return of maturity n bond is then given by

$$rx_{t+1}^n = -(n-1)y_{t+1}^{n-1} + ny_t^n - y_t^1 \tag{1}$$

and the return between periods t and any $t+h$, $rx_{t,t+h}^n$, is given by the sum over the one period excess returns.

3 Regression Evidence

I start by considering a simple regression of the form

$$rx_{t+1}^n = \alpha + \beta rx_{t-h,t}^n + \epsilon_{t+1} \tag{2}$$

[3] The trend inflation τ_t^{CPI} is calculated using $\tau_t^{CPI} = (1-v)\sum_{i=0}^{t-1} v^i \pi_{t-i}$, where v is smoothing/learning parameter. Cieslak and Povala (2015) compute v using survey data.

That is I regress the excess return of an n maturity bond in month $t+1$ on the excess return of an n maturity bond between periods $t-h$ and t. When calculating excess returns I hold maturity constant by rolling over the bond each month. I focus on lookback horizons (h) of 1,3,6 and 12 months. The results are given in Table 1 and demonstrated further in Figure 1.

The results are statistically significant for the return over the past month. However, the results for longer horizon past returns are not significant. Therefore, for the rest of this paper, I focus on the one month horizon. This is in contrast to Moskowitz et al. (2012) who focus on 1 year past returns.[4] I also ignore the volatility scaling applied by Moskowitz et al. (2012) as it can induce return predictability unrelated to raw momentum in returns as discussed in Kim et al. (2016) and Huang et al. (2020).

The regression betas decline in bond maturity. Hence the term structure of momentum coefficients is downward sloping. In the theoretical section I show that this is inconsistent with one factor interest rate models.

The results for the 1 month horizon have strong economic significance as illustrated in Table 2 and Figures 2 and 3. These show the mean excess returns and annualized Sharpe ratios for different maturity bonds both for the full sample and in two subsamples with positive and negative past month excess returns for the same maturity bond. The mean returns and Sharpe ratios are substantially higher following positive rather than negative past month returns. The mean returns are increasing in bond maturity but Sharpe ratios decreasing in maturity. The Sharpe ratios of short maturity bonds are over 0.8 for months following positive excess returns in the previous month.

Figure 4 provides an alternative way to look at the above momentum patterns. It shows the share of total excess bond returns explained by excess returns in months with positive past month excess returns. For all maturities the bulk of returns comes from months with positive past month returns. For many maturities this share is more than 100 per cent because average returns in months with negative past month returns are negative. Because on average

[4]Note that here the significance of 1 year past returns is somewhat better than for 3 and 6 month past returns.

	1 month lookback					3 month lookback				
Mat.	α	t-value	β	t-value	R^2	α	t-value	β	t-value	R^2
1	0.06	3.42	0.19	3.01	3.69	0.06	2.82	0.03	0.57	0.30
2	0.08	2.67	0.18	3.70	3.20	0.10	2.40	0.02	0.31	0.08
3	0.12	2.82	0.15	3.13	2.27	0.14	2.63	0.01	0.20	0.03
4	0.15	2.57	0.12	2.73	1.51	0.17	2.45	0.01	0.20	0.02
5	0.17	2.44	0.12	2.71	1.38	0.19	2.34	0.01	0.25	0.02
6	0.20	2.38	0.10	2.02	0.91	0.21	2.30	0.01	0.31	0.03
7	0.20	2.17	0.10	2.05	0.91	0.22	2.13	0.01	0.24	0.02
8	0.22	2.11	0.10	2.11	0.98	0.24	2.08	0.01	0.31	0.03
9	0.22	1.90	0.10	2.14	0.92	0.23	1.89	0.01	0.28	0.02
10	0.24	1.90	0.09	1.99	0.79	0.26	1.89	0.01	0.24	0.02
	6 month lookback					12 month lookback				
1	0.06	2.73	0.02	0.62	0.26	0.05	1.81	0.02	1.23	0.97
2	0.10	2.28	0.01	0.47	0.12	0.07	1.56	0.03	1.47	1.05
3	0.14	2.49	0.01	0.47	0.09	0.10	1.72	0.03	1.54	0.95
4	0.16	2.34	0.01	0.45	0.06	0.12	1.69	0.02	1.53	0.79
5	0.19	2.26	0.01	0.37	0.04	0.15	1.70	0.02	1.44	0.63
6	0.21	2.21	0.01	0.49	0.06	0.17	1.69	0.02	1.45	0.61
7	0.21	2.04	0.01	0.47	0.05	0.17	1.62	0.02	1.37	0.53
8	0.23	1.99	0.01	0.58	0.08	0.19	1.61	0.02	1.40	0.54
9	0.23	1.83	0.01	0.39	0.04	0.20	1.49	0.02	1.30	0.47
10	0.26	1.84	0.01	0.31	0.02	0.22	1.51	0.02	1.23	0.42

Table 1: shows the results from regressing the excess returns of different maturity bonds (years) on the past return for the same maturity bond for lookback horizons of 1,3,6 and 12 months. The t-values are based on Newey and West (1987) standard errors and the lag selection procedure of Newey and West (1994)

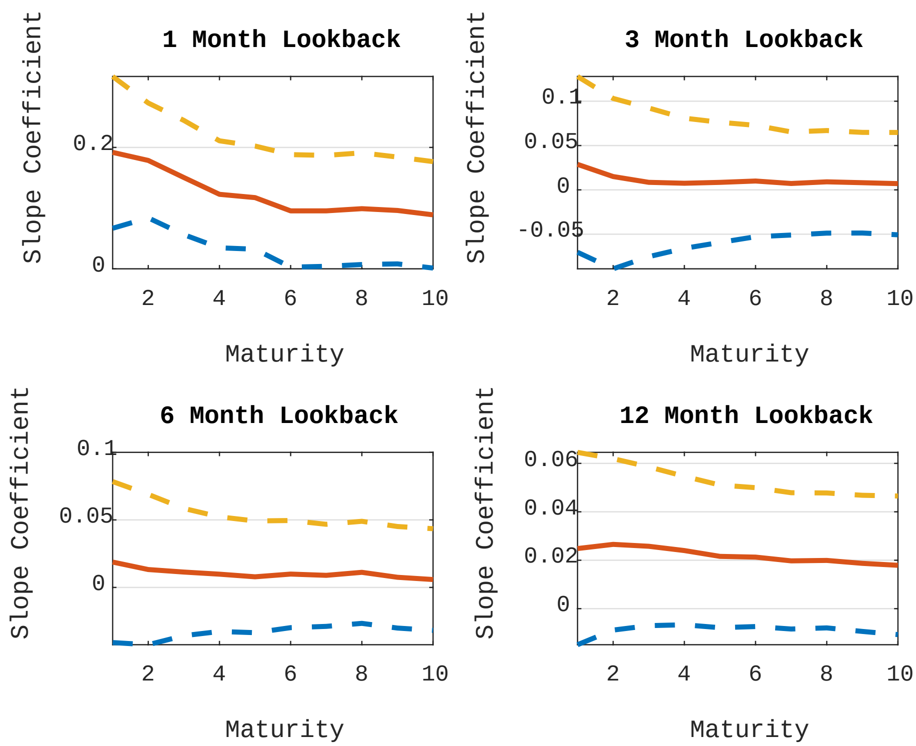

Figure 1: shows the slope coefficients and the relevant 95% confidence bounds from regressing the returns of different maturity bonds (years) on the past return for the same maturity bond for lookback horizons of 1,3,6 and 12 months.

Average excess returns (%)

Maturity	1	2	3	4	5	6	7	8	9	10
All months	0.07	0.10	0.15	0.17	0.20	0.22	0.22	0.25	0.24	0.26
Positive past month ret.	0.09	0.17	0.26	0.30	0.37	0.40	0.38	0.41	0.49	0.57
Negative past month ret.	0.04	0.01	0.00	0.01	-0.02	-0.01	0.03	0.05	-0.05	-0.11

Sharpe ratios (annual)

All months	0.57	0.43	0.44	0.40	0.38	0.36	0.33	0.33	0.29	0.29
Positive past month ret.	0.83	0.82	0.78	0.71	0.72	0.66	0.56	0.53	0.59	0.64
Negative past month ret.	0.26	0.04	0.00	0.02	-0.04	-0.01	0.04	0.06	-0.07	-0.12

Table 2: shows the mean excess returns and annualized Sharpe ratios for different maturity bonds in both the full sample and in two subsamples: following positive and negative past month excess returns.

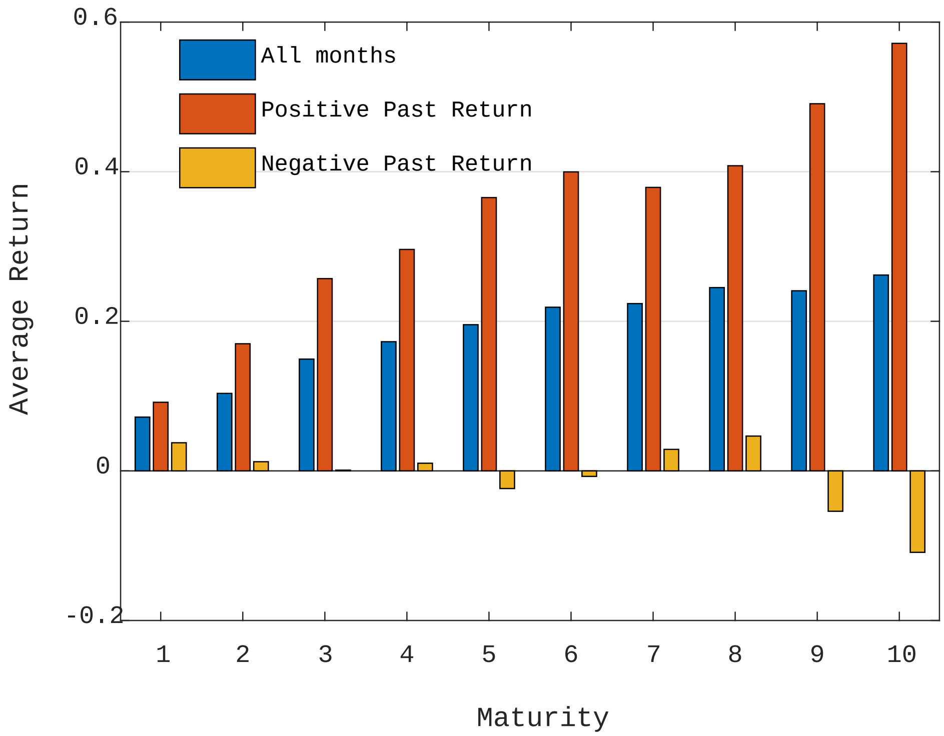

Figure 2: shows the mean returns for different maturity bonds both for the full sample and in subsamples following positive and negative past month returns.

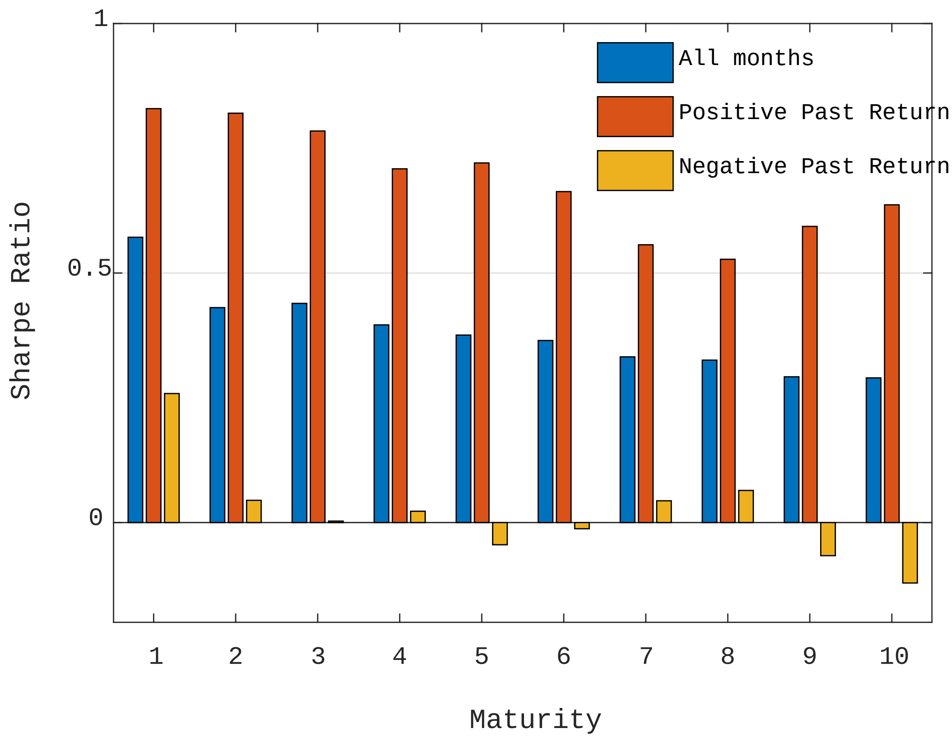

Figure 3: shows the annualized Sharpe ratios for different maturity bonds both for the full sample and in subsamples following positive and negative past month returns.

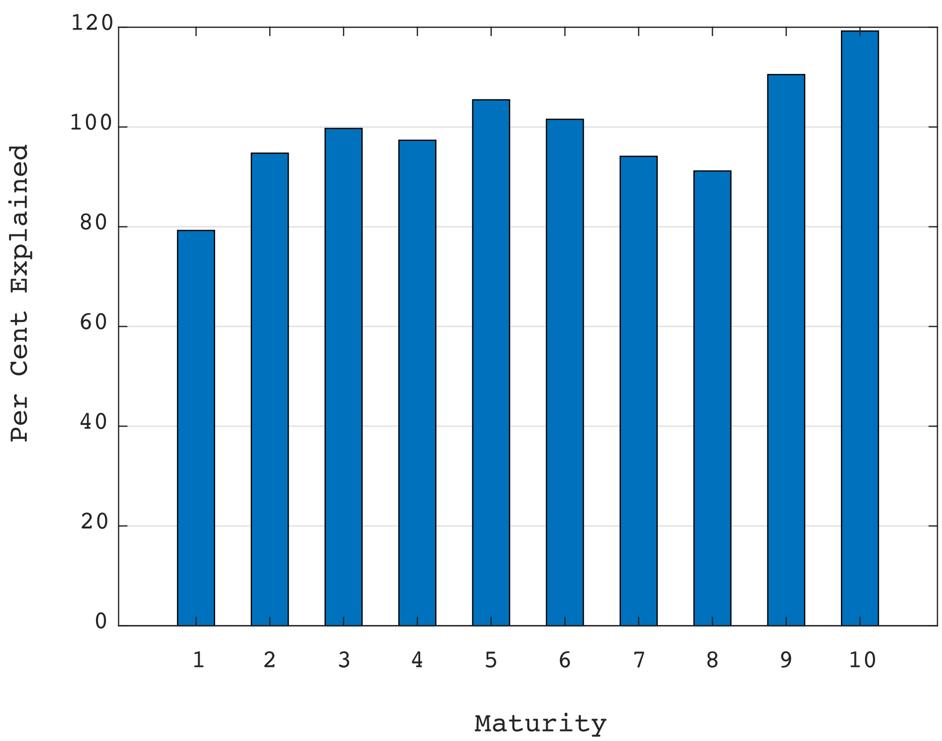

Figure 4: shows the share of total excess returns of different maturity bonds earned in months with positive past month excess returns.

only 56 % of months show positive excess returns, these relationships are not mechanical. The appendix contains additional results concerning the invesment performance of a simple momentum strategy.

Factor momentum Yields and bond returns are often found to exhibit strong factor structures (e.g., Litterman and Scheinkman, 1991). Hence yield curve momentum might also be captured well using a simple factor. I next demonstrate that most of this momentum can indeed be represented by a single factor.

Let us create a simple average of the different maturity bond returns as

$$\bar{rx}_t = \frac{1}{10} \sum_{n \in N} rx_t^n, \tag{3}$$

where $N = \{12, 24, 36, 48, 60, 72, 84, 96, 108, 120\}$. I then run a regression

$$rx_{t+1}^n = \alpha + \beta \bar{rx}_t + \epsilon_{t+1} \tag{4}$$

The results are given in table 3. Using the average of excess returns across different maturity bonds leads to only a minor loss in predictive power relative to using the past return of a bond with the corresponding maturity. For longest maturity bonds the R^2 actually increases but this improvement is small. I confirm this overall result in the next section by showing that yield curve momentum is driven by a change in the first principal component of yields. Note that the loadings for the momentum factor are still different for returns based on different maturity bonds.

4 Decompositions

What is driving the results obtained in the previous section? I next analyze the sources of yield curve momentum using three decompositions. The first is based on decomposing bond returns into a carry and yield change component. The second decomposes returns into a risk premium and news

Mat.	α	t-value	β	t-value	R^2
1	0.06	3.58	0.04	2.37	3.16
2	0.09	2.68	0.08	2.75	2.73
3	0.13	2.84	0.09	2.57	2.04
4	0.15	2.56	0.09	2.43	1.31
5	0.17	2.43	0.11	2.45	1.22
6	0.19	2.37	0.14	2.65	1.46
7	0.19	2.14	0.15	2.59	1.42
8	0.21	2.08	0.17	2.55	1.33
9	0.21	1.85	0.17	2.45	1.17
10	0.23	1.85	0.18	2.27	1.05

Table 3: shows the results from regressing the returns of different maturity bonds on the previous month average return of different maturity bonds. The t-values are based on Newey and West (1987) standard errors.

component. The third divides returns to a part that is spanned by yields and to an unspanned residual component.

Carry-yield change decomposition To begin note that we can decompose the excess return on a bond as

$$\begin{gathered} rx_{t+1}^{n} = -(n-1)y_{t+1}^{n-1} + ny_t^n - y_t^1 = \\ \underbrace{-(n-1)y_t^{n-1} + ny_t^n - y_t^1}_{\text{excess carry}} - \underbrace{(n-1)(y_{t+1}^{n-1} - y_t^{n-1})}_{\text{yield change}} \equiv c_t^n - yc_{t+1}^n. \end{gathered} \tag{5}$$

Here carry (c_t^n) describes the excess return on a bond assuming the yield curve would remain unchanged. This part of the return between t and $t+1$ is observable already at time t. On the other hand, yield change (yc_{t+1}^n) represents the effect of a change in the yield curve on the bond excess return. Therefore for the covariance between current returns and past returns we have

$$\begin{gathered} Cov(rx_{t+1}^n, rx_t^n) = Cov(rx_{t+1}^n, c_{t-1}^n - yc_t^n) = \\ Cov(c_t^n, c_{t-1}^n) + Cov(-yc_{t+1}^n, c_{t-1}^n) + Cov(c_t^n, -yc_t^n) + Cov(-yc_{t+1}^n, -yc_t^n) \end{gathered} \tag{6}$$

Maturity	$Cov(c_t^n, c_{t-1}^n)$	$Cov(c_t^n, yc_t^n)$	$Cov(yc_{t+1}^n, c_{t-1}^n)$	$Cov(yc_{t+1}^n, yc_t^n)$
1	5.3	-0.9	5.6	90.0
2	7.2	1.4	2.7	88.7
3	5.7	2.4	3.0	88.9
4	5.4	2.2	4.8	87.5
5	4.7	4.1	2.3	88.9
6	4.9	5.3	4.9	85.0
7	4.1	1.1	3.4	91.5
8	3.7	4.8	6.1	85.3
9	3.3	3.7	6.3	86.7
10	3.6	0.2	2.8	93.4

Table 4: shows the share of covariance between bond return and past month bond return in per cent accounted by the four channels

This implies that past bond returns can predict future bond returns either because (i) past carry predicts current carry, (ii) past carry predicts future yield changes, (iii) past yield changes predict current carry or (iv) past yield change predicts future yield change.

Because current carry is observable one might argue that (i) and (iii) do not constitute an economically interesting form of predictability. Moreover, it is not clear that such covariance should be called "momentum". However, an investor would certainly benefit from being able to predict future yield changes. Moreover, especially covariance between future yield changes and past yield changes would aptly be called momentum. Such separations are not clear from standard treatments of time series momentum such as Moskowitz et al. (2012).

Table 4 gives the covariance decomposition above. One can see that covariance between future and past bond returns is mainly due to covariance between future and past yield changes. I also test these dependencies using the following regressions:

$$c_t^n = \alpha + \beta c_{t-1}^n + \epsilon_{t+1} \tag{7}$$

$$c_t^n = \alpha + \beta y c_t^{n-1} + \epsilon_{t+1} \tag{8}$$

$$y c_{t+1}^n = \alpha + \beta c_t^{n-1} + \epsilon_{t+1} \tag{9}$$

$$y c_{t+1}^n = \alpha + \beta y c_t^{n-1} + \epsilon_{t+1} \tag{10}$$

The results are given in table 5.The coefficient for the past carry in the carry prediction regression and the coefficient for past yield change in the yield change prediction regression are statistically significant. On the other hand, I do not find evidence of significant cross carry-yield change predictability. Note that even though there is a statistically robust relationship between past carry and future carry, because carry does that vary much its contribution to the covariance between future and past returns is small. Autocorrelation between yields appears to be strongest for shorter maturity bonds, which explains why the relationship between past and future returns is also strongest for these maturities.

Given these findings I now revisit the question about whether yield curve momentum can be captured using a single factor. In particular I explore this further using principal component analysis. I extract the first three principal components using all the 120 maturities between 1 month and 10 years. I then consider the following regressions:

$$r x_{t+1}^n = \alpha + \beta(pc_t^1 - pc_{t-1}^1) + \epsilon_{t+1} \tag{11}$$

$$r x_{t+1}^n = \alpha + \beta(pc_t^2 - pc_{t-1}^2) + \epsilon_{t+1} \tag{12}$$

$$r x_{t+1}^n = \alpha + \beta(pc_t^3 - pc_{t-1}^3) + \epsilon_{t+1} \tag{13}$$

$$r x_{t+1}^n = \alpha + \beta_1(pc_t^1 - pc_{t-1}^1) + \beta_2(pc_t^2 - pc_{t-1}^2) + \beta_3(pc_t^3 - pc_{t-1}^3) + \epsilon_{t+1} \tag{14}$$

			c_t^n on c_{t-1}^n					c_t^n on yc_t^n		
Mat.	α	t-value	β	t-value	R^2	α	t-value	β	t-value	R^2
1	0.01	3.76	0.85	28.79	72.59	0.07	10.38	-0.00	-0.14	0.01
2	0.01	2.59	0.92	42.82	83.68	0.09	9.77	0.00	0.33	0.04
3	0.01	2.43	0.90	27.50	81.06	0.13	12.01	0.00	0.61	0.13
4	0.01	2.57	0.93	44.80	85.66	0.14	11.89	0.00	0.58	0.11
5	0.01	2.42	0.93	48.49	87.18	0.16	12.05	0.01	1.21	0.41
6	0.01	2.48	0.94	52.78	88.38	0.18	12.52	0.01	1.45	0.53
7	0.01	2.57	0.93	48.16	87.18	0.17	11.92	0.00	0.34	0.03
8	0.01	2.40	0.94	55.93	89.14	0.19	12.02	0.00	1.44	0.59
9	0.01	2.37	0.94	51.06	88.21	0.18	11.19	0.00	1.38	0.39
10	0.01	2.28	0.95	57.19	89.75	0.19	10.97	0.00	0.08	0.00
			yc_{t+1}^n on c_{t-1}^n					yc_{t+1}^n on yc_t^n		
1	0.01	0.23	-0.07	-0.03	0.00	0.00	0.23	0.16	2.38	2.72
2	-0.02	-0.25	0.55	0.32	0.15	0.01	0.30	0.16	3.13	2.64
3	-0.00	-0.04	0.23	0.20	0.04	0.02	0.35	0.14	2.70	1.84
4	-0.06	-0.45	0.80	0.59	0.25	0.02	0.40	0.11	2.28	1.19
5	-0.10	-0.69	1.15	0.82	0.40	0.03	0.44	0.11	2.38	1.12
6	-0.16	-0.97	1.49	1.13	0.64	0.04	0.47	0.08	1.71	0.68
7	-0.05	-0.30	0.74	0.59	0.15	0.04	0.47	0.09	1.84	0.78
8	-0.22	-1.15	1.73	1.44	0.82	0.05	0.48	0.09	1.82	0.74
9	-0.26	-1.32	2.00	1.78	1.09	0.05	0.49	0.08	1.85	0.72
10	-0.06	-0.25	0.74	0.66	0.15	0.06	0.48	0.08	1.82	0.70

Table 5: shows the results of regressing carry c_t^n and yield change yc_{t+1}^n on their past values. The t-values are based on Newey and West (1987) standard errors.

That is I explain returns using the change in the first three principal components of yields, first individually and then including them all in one regression. The principal components appear standard. The first principal component explains roughly 98.5% of the variation in yields. While this component is often called a level factor, the yield loadings decline slightly in bond maturity. That is they drop from around 0.098 for 1 month yields to 0.082 for 10 year bonds. The average contemporaneous correlation between the change in this factor and excess bond returns is -0.95. That is an increase in this factor is related to an upward shift in the yield curve but also to negative excess returns on long-term bonds.

The second component has positive loadings on short term yields and negative loadings on long term yields and could be called a slope factor. The third component has positive loadings on short and long maturity yields and negative loadings on mid-maturity yields. This component represents curvature. The first three components together account for 99.97% of the variation in yields.

The results for individual regressions are given in table 6. Here only the change in the first principal component of yields is clearly significant, though in some regressions changes in the curvature factor are significant at a 10%-confidence level. The results for the regressions with all three included at the same time are given in table 7. Again only the first principal component is significant. This suggests that yield curve momentum is driven by changes in a single factor.[5]

Risk premium-news decomposition I next study the anatomy of yield curve momentum using a variant of a well-known decomposition of returns into a risk premium and news component. In particular we have

[5]These findings are related to those in Hoogteijling et al. (2021), who in contemporaneous work find evidence that yield changes can predict bond returns. Using an annual rather than monthly horizon, they also find evidence that changes in the slope and curvature factors can forecast returns.

	pc^1					pc^2				
Mat.	α	t-value	β	t-stat	R^2	α	t-value	β	t-value	R^2
1	0.07	4.00	-0.02	-2.47	3.41	0.07	3.63	-0.02	-0.85	0.27
2	0.10	3.05	-0.03	-2.86	2.59	0.10	2.75	-0.01	-0.23	0.02
3	0.14	3.14	-0.04	-2.57	1.85	0.15	2.85	-0.01	-0.11	0.00
4	0.17	2.79	-0.04	-2.29	1.07	0.17	2.60	0.01	0.10	0.00
5	0.19	2.65	-0.05	-2.22	0.94	0.19	2.48	0.02	0.31	0.02
6	0.21	2.60	-0.06	-2.39	1.17	0.22	2.41	0.02	0.27	0.02
7	0.22	2.37	-0.07	-2.37	1.22	0.22	2.21	-0.01	-0.15	0.01
8	0.24	2.30	-0.07	-2.36	1.21	0.24	2.15	-0.04	-0.40	0.04
9	0.23	2.06	-0.07	-2.23	1.02	0.24	1.94	-0.02	-0.21	0.01
10	0.25	2.05	-0.08	-2.09	0.92	0.26	1.93	-0.03	-0.23	0.01
	pc^3									
1	0.07	3.55	0.06	0.99	0.74					
2	0.10	2.76	0.12	1.17	0.71					
3	0.15	2.88	0.16	1.32	0.70					
4	0.17	2.63	0.17	1.21	0.49					
5	0.19	2.51	0.23	1.34	0.59					
6	0.22	2.45	0.37	1.85	1.20					
7	0.22	2.23	0.38	1.71	0.98					
8	0.24	2.17	0.40	1.61	0.87					
9	0.24	1.96	0.45	1.68	0.93					
10	0.26	1.95	0.45	1.55	0.76					

Table 6: shows the results of predicting returns of different maturity bonds on the change in the first three principal components separately. The t-values are based on Newey and West (1987) standard errors.

Mat.	α	t-value	β_1 (pc^1)	t-value	β_2 (pc^2)	t-value	β_3 (pc^3)	t-value	R^2
1	0.07	3.99	-0.02	-2.25	-0.01	-0.60	0.03	0.53	3.65
2	0.10	3.02	-0.03	-2.39	0.00	0.02	0.06	0.59	2.76
3	0.14	3.10	-0.04	-2.04	0.00	0.06	0.10	0.73	2.08
4	0.17	2.77	-0.04	-1.70	0.01	0.20	0.10	0.62	1.27
5	0.19	2.62	-0.04	-1.61	0.03	0.35	0.15	0.75	1.25
6	0.21	2.58	-0.05	-1.60	0.02	0.21	0.28	1.22	1.87
7	0.22	2.35	-0.05	-1.63	-0.01	-0.15	0.29	1.15	1.73
8	0.24	2.28	-0.06	-1.58	-0.04	-0.39	0.31	1.08	1.68
9	0.23	2.04	-0.06	-1.46	-0.03	-0.24	0.36	1.16	1.54
10	0.26	2.02	-0.06	-1.38	-0.03	-0.25	0.35	1.06	1.33

Table 7: shows the results of predicting returns of different maturity bonds on the change in the first three principal components together. The t-values are based on Newey and West (1987) standard errors.

$$rx_{t+1}^n = \underbrace{\mathbb{E}_t[rx_{t+1}^n]}_{\text{Expectation}} + \underbrace{rx_{t+1}^n - \mathbb{E}_t[rx_{t+1}^n]}_{\text{News}} \tag{15}$$

Therefore also

$$Cov(rx_{t+1}^n, rx_t^n) = Cov(\mathbb{E}_t[rx_{t+1}^n], \mathbb{E}_{t-1}[rx_t^n]) + Cov(\mathbb{E}_t[rx_{t+1}^n], rx_t^n - \mathbb{E}_{t-1}[rx_t^n]) \tag{16}$$

Here I used the fact that the news component should be uncorrelated with information known when forming the expectation. The decomposition implies that bond returns are correlated with past bond returns either due to autocorrelation in bond risk premia or because the bond risk premium is correlated with past unexpected shocks to the premium.

One benefit of the carry-yield change decomposition is that both components can be easily measured in the data. However, risk premia are not directly observable and must be approximated using a model. Here I consider a simple linear predictive regression:

$$rx_{t+1}^{n} = A'X_t + \epsilon_{t+1} \tag{17}$$

Note that such a form is implied by standard term structure models. I focus on yield curve factors as predictors. In particular I include the first five principal components as well as their lagged values. Now we have

$$\begin{aligned} Cov(rx_{t+1}^{n}, rx_{t}^{n}) = Cov(A'X_t, A'X_{t-1}) + Cov(A'X_t, \epsilon_t) + \\ Cov(\epsilon_{t+1}, A'X_{t-1}) + Cov(\epsilon_{t+1}, \epsilon_t) \end{aligned} \tag{18}$$

The last two terms can be seen as the effect of approximation error of the model, which arises if the model is not exactly correct or if principal components are measured with error.

The results are given in table 8. We can see that for short maturity bonds, momentum is mainly due to unexpected past bond returns increasing the next period bond risk premium. However, for longer maturity bonds, the two channels are roughly equally important. The approximation error components are fairly small, with perhaps the exception of 5 and 6 year bonds. This suggests that the model provides a reasonable approximation to bond risk premia. The appendix redoes this decomposition when the predictive regression includes macro variables, here the results are similar but the share accounted by risk premia is somewhat higher.

Spanning decomposition Past bond returns can predict future bond returns either because i) past bond returns contain information about current yield curve factors that predict future bond returns or ii) past returns contain additional information relevant for future returns. Formally the first explanation implies that past returns are *spanned* by current yields whereas the second implies that they are not. As explained later standard term structure models imply that the spanning condition holds so that yield curve momentum should be explained by the first channel.

Maturity	Risk premia	News	Error
1	21.4 %	78.6 %	0.0 %
2	22.0 %	78.4 %	-0.4 %
3	25.5 %	80.3 %	-5.9 %
4	29.2 %	71.1 %	-0.2 %
5	27.2 %	66.1 %	6.7 %
6	30.3 %	77.9 %	-8.3 %
7	34.7 %	68.1 %	-2.7 %
8	40.8 %	60.2 %	-1.1 %
9	44.3 %	55.5 %	0.2 %
10	51.6 %	49.3 %	-0.9 %

Table 8: shows the decomposition of covariance between the return of different maturity bonds and their past value into the autocovariance of risk premia, covariance between risk premia and past unexpected bond returns and covariance between past returns and an approximation error component.

To test the relevant importance of the two channels consider two linear projections of returns on the principal components of yields

$$rx^n_{t+1} = A'PC_t + \epsilon_{t+1} \tag{19}$$

$$rx^n_t = B'PC_t + \varepsilon_t \tag{20}$$

The autocovariance in bond returns can then be decomposed to spanned and unspanned parts:[6]

$$Cov(rx^n_{t+1}, rx^n_t) = \underbrace{A'Var(PC_t)B}_{\text{Spanned}} + \underbrace{Cov(rx^n_{t+1}, rx^n_t) - A'Var(PC_t)B}_{\text{Unspanned}} \tag{21}$$

[6]In a standard spanned term structure model: $\mathbb{E}_t[rx^n_{t+1}] = A'PC_t$, where the expectation is computed conditional on all information available at time t. Then $Cov(rx^n_{t+1}, rx^n_t) = Cov(A'PC_t, rx^n_t) = Cov(A'PC_t, B'PC_t) = A'Var(PC_t)B$.

Maturity	Spanned	Unspanned
1	12.1 %	87.9 %
2	7.5 %	92.5 %
3	8.0 %	92.0 %
4	4.4 %	85.6 %
5	2.6 %	97.4 %
6	5.9 %	94.1 %
7	6.6 %	93.4 %
8	11.6 %	88.4 %
9	10.3 %	89.7 %
10	7.3 %	92.7 %

Table 9: shows the decomposition of covariance between the return of different maturity bonds and their past value into a part spanned by yields and an unspanned part.

I apply seven principal components of yields as including further components has minor effects on the results. The results are given in Table 9. On average only about 10% of the covariance between current and past returns is spanned by yields.

Testing Spanning Results from the spanning decomposition above suggest that unspanned variation in returns is important to explaining yield curve momentum. I now test this result more formally by including the first three principal components into the predictive regression shown in Table 1. The results are given by in 10. The table suggests that the past return is still significant. However, for higher maturity bonds this significance is obtained only at the 10% level.

Higher principal components of yields can contain information useful for predicting bond returns. Therefore in the appendix I extend this regression by controlling for more information in the yield curve as well as potential non-linearities. Here the past return is significant for shorter but not for longer maturities. These results further confirm that, at least for short maturities, the unspanned components of returns are important for explaining

Mat.	β_1 (rx_{t-1})	t-stat	β_2 (pc^1)	t-stat	β_3 (pc^2)	t-stat	β_4 (pc^3)	t-stat	R^2
1	0.19	2.61	0.0012	1.63	0.00	-0.39	-0.03	-1.28	5.12
2	0.18	3.27	0.0014	1.15	-0.01	-0.82	-0.08	-2.22	5.02
3	0.15	2.86	0.0017	1.05	-0.02	-1.29	-0.11	-2.46	4.43
4	0.13	2.50	0.0016	0.77	-0.03	-1.64	-0.15	-2.58	3.93
5	0.12	2.51	0.0016	0.65	-0.04	-1.81	-0.17	-2.32	3.70
6	0.09	1.81	0.0013	0.47	-0.06	-2.12	-0.15	-1.80	3.00
7	0.09	1.81	0.0010	0.32	-0.07	-2.44	-0.15	-1.52	3.06
8	0.09	1.87	0.0013	0.34	-0.08	-2.67	-0.15	-1.29	3.24
9	0.09	1.90	0.0004	0.09	-0.09	-2.64	-0.13	-1.01	2.92
10	0.08	1.77	0.00	0.00	-0.10	-2.71	-0.13	-0.87	2.74

Table 10: shows the results of predicting returns of different maturity bonds on the past return of the bond and the first three principal components of yields. The t-values are based on Newey and West (1987) standard errors.

yield curve momentum.

Adding Macro Variables Macroeconomic variables are often found to forecast bond returns on top of yields (Duffee, 2011; Joslin et al., 2014; Cieslak and Povala, 2015; Coroneo et al., 2016; Moench and Siavash, 2021). As discussed later, this suggests these variables are unspanned by current yields. In theory past returns might be correlated with such unspanned macro variables, which could explain why past returns themselves are unspanned. We therefore include trend inflation and the national activity index to our previous spanning regression. Here we include five principal components of yields.

The results are given in table 11. The slope coefficients for past returns are clearly significant and of similar magnitude than before. Trend inflation is also significant, especially for longer maturities, but the activity index is mainly not significant. We conclude that accounting for these macro variables does not alter the main results of this paper though these variables possibly represent additional unspanned information useful for predicting returns.

Maturity	β_1 (rx_{t-1})	t-value	β_2 (Inflation)	t-value	β_3 (Activity)	t-value
1	0.17	2.24	-0.057	-1.35	-0.10	-2.12
2	0.18	3.21	-0.18	-2.23	-0.16	-1.87
3	0.16	2.94	-0.30	-2.67	-0.22	-1.84
4	0.14	2.65	-0.42	-2.86	-0.24	-1.57
5	0.13	2.79	-0.53	-2.92	-0.25	-1.38
6	0.12	2.26	-0.59	-2.87	-0.26	-1.23
7	0.12	2.37	-0.73	-3.19	-0.27	-1.18
8	0.12	2.47	-0.87	-3.35	-0.28	-1.12
9	0.12	2.51	-1.014	-3.54	-0.29	-1.09
10	0.12	2.45	-1.17	-3.73	-0.32	-1.11

Table 11: shows the results of predicting returns of different maturity bonds on the past return of the bond, five principal components of yields, trend inflation and the national activity index. The slope cofficients for the principal components are excluded. The t-values are based on Newey and West (1987) standard errors.

5 Momentum and Post-FOMC Announcement Drift

How are these findings related to monetary policy? Because especially the short end of the yield curve tends to be tightly controlled by the Fed, yield curve momentum might be induced by policy rate changes. This is also due to recent findings related to post-FOMC announcement drift. Brooks et al. (2019) find that longer term bond yields respond sluggishly to changes in the federal funds target rate.[7]

I now study this relationship using data on the federal funds target rate. I also utilize data on federal funds futures and the FOMC announcement dates to construct a series of surprise changes in the federal funds rate as in Kuttner (2001). The data period for the federal funds target rate begins in October 1982 and the data for monetary policy surprises in October 1988.

Figure 5 shows the correlation between changes in yields and changes in the federal funds target rate. It does so in two samples: the full sample

[7] There is a similar drift pattern in equity markets after rate changes, see Neuhierl and Weber (2018).

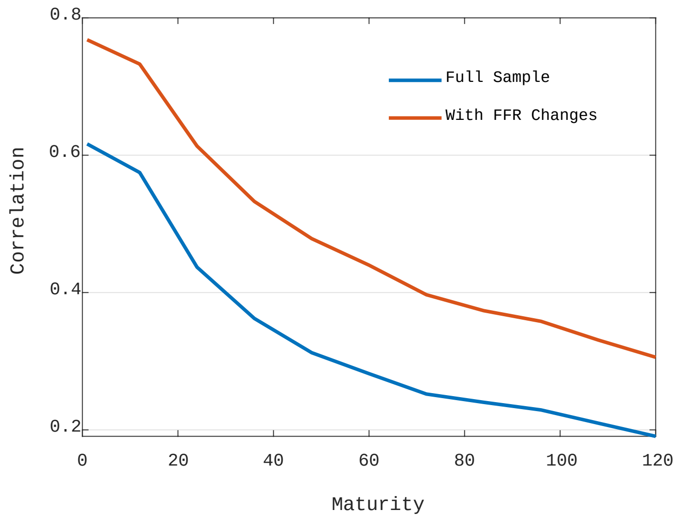

Figure 5: shows the correlation between the change in the Federal funds target rate (FFTR) and the change in the yield of different maturity (in months) bonds in two subsamples: full and months with non-zero FFTR changes.

starting in 1982 and a subsample of months with a non-zero change in this policy rate. Excluding months with no rate changes this correlation is close to 0.8 at the short end of the yield curve but only around 0.3 at the long end. The decline in correlation for longer maturity bonds is natural since the federal funds rate is an overnight rate. All of these correlations are somewhat smaller in the full sample; overall roughly 30% of months included changes in the policy rate.

I now consider the following regressions

$$rx_{t+1}^{n} = \alpha + \beta \Delta FFTR_t + \epsilon_{t+1} \tag{22}$$

$$rx_{t+1}^{n} = \alpha + \beta \Delta UEFFTR_t + \epsilon_{t+1}. \tag{23}$$

That is I explain the returns of different maturity bonds on the raw change of the past month federal funds target rate as well the unexpected change in this rate. These regressions are related to those considered by

	FFTR change					yc^n				
Mat.	α	t-value	β	t-value	R^2	α	t-value	β	t-value	R^2
1	0.08	6.23	-0.21	-3.08	3.80	0.08	6.33	-0.17	-3.02	5.38
2	0.13	4.58	-0.33	-2.47	1.87	0.08	6.25	-0.16	-3.36	4.99
3	0.19	4.31	-0.39	-1.92	1.02	0.08	6.07	-0.14	-3.24	3.94
4	0.23	3.97	-0.42	-1.57	0.64	0.08	5.96	-0.12	-3.16	3.15
5	0.27	3.71	-0.43	-1.28	0.43	0.08	5.90	-0.11	-3.02	2.45
6	0.31	3.60	-0.32	-0.82	0.17	0.08	5.80	-0.10	-2.61	1.98
7	0.34	3.39	-0.12	-0.27	0.02	0.08	5.75	-0.09	-2.50	1.58
8	0.38	3.34	-0.15	-0.28	0.02	0.08	5.71	-0.09	-2.30	1.37
9	0.39	3.12	-0.11	-0.19	0.01	0.08	5.69	-0.09	-2.28	1.21
10	0.43	3.09	-0.11	-0.17	0.01	0.08	5.69	-0.09	-2.21	1.07
	Unexpected FFTR change									
1	0.09	4.27	-0.26	-1.32	2.04					
2	0.14	2.49	-0.49	-1.54	1.08					
3	0.19	2.23	-0.96	-1.87	1.73					
4	0.22	1.99	-1.45	-2.08	2.20					
5	0.25	1.86	-1.84	-2.11	2.41					
6	0.28	1.83	-2.28	-2.16	2.76					
7	0.31	1.81	-2.52	-2.02	2.60					
8	0.34	1.78	-2.76	-1.95	2.54					
9	0.35	1.72	-3.02	-1.95	2.55					
10	0.39	1.78	-3.43	-2.05	2.85					

Table 12: shows the results from regressing the returns of different maturity bonds (years) on the previous change in federal funds target rate, change in the previous yield for the same maturity bond and the previous month unexpected change in the federal funds target rate (Kuttner, 2001). The t-values are based on Newey and West (1987) standard errors.

Cook and Hahn (1989) and Kuttner (2001) except that I consider the past rather than the contemporaneous change in the policy rate.[8]

The results are given in table 12. Here I also show the results from regressing bond returns on the change in the previous month change in the corresponding yield for the same period when the target rate is available. Results when using the federal funds target rate and bond yield are similar for shorter maturities, which is perhaps not surprising since these yields are highly correlated with the target rate. However, for longer maturities the target rate change is not significant while the yield change is. Therefore it seems that yield curve momentum is closely related but still separate from post-FOMC announcement drift.

Table 12 also shows the results when the independent variable is the past surprise change in the federal funds rate. Interestingly the results are not significant for 1 and 2 year bonds but become significant for longer maturities. Therefore long maturity bonds seem to have a stronger drift pattern after surprise changes in the federal funds rate. The sample period for these regressions is somewhat shorter though.

Monetary policy is related to the news part of the return decomposition analyzed in the previous section. The average correlation between target rate changes and the news component of returns is -0.3 with higher absolute values for short maturity bonds. Bond return shocks are related to but not fully driven by changes in the policy rate.

We can also analyze the contribution of target rate changes to yield curve momentum using a decomposition. I project bond returns on contemporaneous changes in the federal funds rate as follows:

$$rx_t^n = \alpha + \beta \Delta FFTR_t + \epsilon_t. \tag{24}$$

Using this projection, I can then decompose bond return autocovariance into an effect caused by changes in the federal funds target rate and a residual component:

[8] Cook and Hahn (1989) and Kuttner (2001) also look at yield changes rather than excess returns.

Maturity	FFTR effect	Other
1	47.3 %	52.7 %
2	31.0 %	69.0 %
3	20.3 %	79.8 %
4	21.3 %	78.8 %
5	17.1 %	82.9 %
6	13.6 %	86.4 %
7	5.3 %	94.7 %
8	7.4 %	92.6 %
9	4.8 %	95.2 %
10	4.8 %	95.2 %

Table 13: shows the decomposition of covariance between the return of different maturity bonds and their past value into a part explained by change in the federal funds target rate and a residual component.

$$Cov(rx_{t+1}^n, rx_t^n) = \underbrace{Cov(rx_{t+1}^n, \beta \Delta FFTR_t)}_{\text{FFR effect}} + \underbrace{Cov(rx_{t+1}^n, \epsilon_t)}_{\text{Other}} \tag{25}$$

The results are given in table 13. This simple decomposition suggests that target rate changes are an important contributor to momentum for shorter maturities but less so for longer maturities.

Overall, yield curve momentum is therefore connected with, but not identical to, post-FOMC announcement drift. Past month yield hikes predict low returns in the following month. These yield changes can be partly but not fully explained with same month movements in the policy rate. For example the momentum coefficients are still significant in the subsample of months with no policy rate changes. The appendix contains additional results concerning the post-FOMC announcement drift.

Finally, note that the above discussion is unlikely to fully capture the broad relationship between monetary policy and yield curve momentum. Yields tend to fluctuate also in periods without any formal monetary policy decisions. However, this does not imply that such changes are unrelated to monetary policy. These fluctuations might for example still reflect changes in the market participants' views about future monetary policy actions.

6 Momentum and Affine Term Structure Models

How to account for the above empirical findings in a term structure model? I start by introducing a baseline affine term structure model and discussing minimal requirements implied by the data.[9] It is seen that the violation of the spanning condition implies strong restrictions for such a model.

Assume that bond prices are a function of an $m \times 1$ dimensional factor X_t. This factor follows:

$$X_t = \mu + \phi X_{t-1} + v_t, \tag{26}$$

where v_t is multivariate Gaussian $v_t \sim N(0, V)$. The log nominal discount factor is a linear function of the factors

$$M_{t+1} = \exp\left(-\delta_0 - \delta_1' X_t - \frac{1}{2}\lambda_t' V \lambda_t - \lambda_t' v_{t+1}\right)$$

$$\lambda_t = \lambda_0 + \lambda_1 X_t.$$

We can then solve bond prices recursively using

$$p_t^1 = \log \mathbb{E}_t(M_{t+1}) \tag{27}$$

$$p_t^n = \log \mathbb{E}_t(M_{t+1} \exp(p_{t+1}^{n-1})). \tag{28}$$

In this model prices and yields take an affine form.

[9] See e.g. Ang and Piazzesi (2003) and Cochrane and Piazzesi (2009)

$$p_t^n = A_n + B_n' X_t \tag{29}$$

here

$$A_0 = 0, \ B_0 = 0$$

$$B_{n+1} = -\delta_1 + B_n' \phi^*$$

$$A_{n+1} = -\delta_0 + A_n + B_n' \mu^* + \frac{1}{2} B_n' V B_n$$

Here the risk neutral parameters are given by

$$\phi^* = \phi - V\lambda_1 \tag{30}$$

$$\mu^* = \mu - V\lambda_0 \tag{31}$$

The model implies an analytical expression for the momentum slope coefficient given in the following proposition:

Proposition 1. *The slope coefficient in the momentum regression is given by*

$$\beta^{n,m} = \frac{Cov(rx_{t+1}^n, rx_t^n)}{Var(rx_t^n)}$$

Here

$$\begin{gathered} Cov(rx_{t+1}^n, rx_t^n) = \\ (n-1)^2 B_{n-1}' \phi V B_{n-1} - (n-1) B_{n-1}' \phi^2 \Sigma (n B_n - B_1) \\ -(n-1)(n B_n' - B_1') V B_{n-1} + (n B_n' - B_1') \phi V (n B_n - B_1) \end{gathered}$$

and

$$Var(rx_t^n) =$$
$$(n-1)^2 B(n-1)' V B(n-1) - 2(n-1)B(n-1)'\phi V(nB(n) - B(1)) +$$
$$(nB(n)' - B(1)')V(nB(n) - B(1))$$

Proof: see appendix.

What type of affine term structure model can generate momentum? I first discuss the general restrictions imposed by the empirical findings. To begin note that in order to generate yield curve momentum, one needs a model with time-varying bond risk premium:

Remark 1. *The momentum slope coefficient is zero in a model with a constant (but possibly maturity specific) risk premium* $\lambda_1 = \mathbf{0}$.

The proof of the remark follows from the decomposition of bond returns into a risk premium and news component. The news component cannot be forecasted with past returns by definition. Now also the best forecast of the risk premium is a constant so the slope coefficient in the momentum regression would be zero.

Table 1 shows that the regression slope coefficient is decreasing in bond maturity. This effectively rules out models in which the coefficient is constant across maturities. In particular we have the following remark:

Remark 2. *The momentum slope coefficient is constant across maturities in a one factor model.*

Proof: see appendix.

This result is related to fact that in one factor interest rate models all bond yields are perfectly correlated (Vasicek, 1977).

In the empirical part I established that yield curve momentum is primarily driven by the change in the first principal component of yields. But does this imply that one could capture most of momentum using a single factor term structure model? This reasoning is incorrect as this finding rather suggests that the model should include information about both the first principal component and its past value rather suggesting a minimum of two factors.

Our empirical results suggest that momentum should be explained by a model in which past returns are not spanned by information in current yields. As discussed in the next section this observation poses difficulties for standard models. These models tend to imply that the same model factors that forecast bond returns also drive variation in yields. Therefore controlling for sufficiently many yields is equivalent to controlling for the factors and no other variable should contain additional information for forecasting bond returns. However, similarly to Joslin et al. (2014), we can generate a violation of this spanning condition by parametrizing the model to a knife-edge case for which an invertibility condition condition holds.

Remark 3. *Past bond returns can predict future returns conditional on the information in the term structure today only if the following condition holds:* $[B_{n(1)}, B_{n(2)}, \dots, B_{n(m)}]$ *is not invertible for* $n(i) \in \mathbb{Z}_{++}$.

Proof: see appendix.

The intuition for this result is that if an invertibility condition fails, controlling for the yields is generally not equivalent to controlling for the factors. Now some factors can predict returns and yield changes but not be priced in the current term structure of yields.

To conclude remarks 1-3 put constraints on the model that can explain the key empirical findings. In particular they imply that we need a multifactor, unspanned term structure model with a time-varying risk premium.

6.1 Spanning Puzzle and Problem with Standard Models

The finding that past returns can predict future returns controlling for information in the yield curve today poses difficulties for standard models.

These models imply that bond returns and yields are both described by the same small set of factors. The models do not naturally generate a violation of the invertibility condition described in Remark 3. I next discuss some of these models:

Macrofinance Models I first consider the three main macrofinance models used to explain asset returns: the long run risk model, the habit model and the disasters model. In the long-run risk model (see e.g. Bansal and Shaliastovich, 2012) bond yields take an affine form in the economic state variables. Therefore this model is of the form discussed in the previous section and for standard parametrizations cannot generate momentum conditional on information in the term structure today.

In the habit model, bond yields are a generally non-linear function of habit (Wachter, 2006). Therefore the argument of the previous section is strictly valid only up to a first order approximation of the underlying model. However, as discussed in the appendix one can generalize Remark 3 to any well-defined function $y_t = g(X_t)$ so that there is no conditional momentum after controlling for the generally non-linear relationship between past yields and returns. The results obtained in the appendix also suggest that controlling for non-linearities also does not alter the key conclusions.

Also the disasters model of Gabaix (2012) implies that yields are of the form $y_t = g(X_t)$ for state variables X_t. This is also true for any Markovian model such as standard DSGE models. For example Rudebusch and Swanson (2012) offer a macroeconomic interpretation of term premia using a DSGE model with Epstein-Zin preferences. Therefore the general results apply to this model subject to excluding knife-edge cases in which an invertibility condition fails.

Models with Financial Frictions Vayanos and Woolley (2013) posit that momentum might be explained by frictions in delegated asset management. Because the equilibrium is linear in state variables, the model can only generate unconditional momentum. Similarly the preferred habitat term structure model of Vayanos and Vila (2020) takes a standard affine form and hence is unable to generate conditional momentum.

Behavioral Models I now turn to behavioral models and models with heterogenous beliefs. Granziera and Sihvonen (2020) assume that agents have sticky rather than perfectly rational expectations concerning short rates. This slow updating creates a drift pattern in bond returns following short rates changes.[10] Hence the model naturally generates unconditional momentum. In this model biased beliefs enter as new state variables but again the model takes a standard affine form, which is inconsistent with conditional momentum.

In Xiong and Yan (2010) yields are a generally non-linear function of the beliefs of different types of investors. Again this model cannot generate conditional momentum controlling for non-linear dependencies between returns and past yields.

The classic momentum model of Hong and Stein (1999) features only one asset. The authors solve for an linear equilibrium. It is not obvious how to extend the model to multiple assets but assuming such an extended model were still linear the problems discussed above apply.

6.2 Accounting for Momentum in a Term Structure Model

I next discuss how to account for momentum in a term structure model. For intuition I start with a simplified example and then move to a more realistic estimated term structure model.

Simple Example

Consider a one factor model as in for example Vasicek (1977). However, make the following twist. First, instead of the standard AR(1) dynamics assume the factor follows an AR(2)-process. In such a model bond prices

[10] Brooks et al. (2019) also argue that a similar model can explain the post FOMC announcement drift.

generally depend on both the current value of the state variable x_t and its lag x_{t-1}. However, assume the second lag is not priced that is under the risk neutral pricing measure the factor follows an AR(1) process. This implies that bond prices depend only on the current value of the factor x_t.

I estimate the true factor dynamics from 1 month rates and find significant persistence parameters of 1.077 for the first lag and −0.088 for the second. For comparison fitting an AR(1) process would result in a persistence parameter of 0.98. The predictability results hinge on a single parameter, the risk neutral persistence of the factor. I calibrate this to match the relative volatility between 10 year and 1 month rates. The corresponding market price of risk parameters could be solved from equations (31) and (30) but are not relevant for the exercise.

Now consider regressing the past excess return of a 5-year bond on the previous month return of a 5 year bond. Using simulations I obtain a slope coefficient of 0.08, that is the model is able to generate momentum, though this coefficient is slightly smaller than in the data (0.12). Moreover, because this is effectively a one factor model, this coefficient is actually constant across maturities, whereas in the data it is decreasing.

But then I repeat this exercise but now explain the return using the past month return and the beginning of period yield of the bond. The coeffient on the past return is still positive at roughly 0.07. That is the model is able to generate momentum conditional on the information in the term structure today.

Why is this model able to generate conditional momentum? In the data, yields effectively follow an AR(2)-process. However, agents price bonds as if the process is AR(1). The higher lag is not priced. Still this second lag is useful for predicting future yields and returns. Because past returns incorporate information about this second lag, including them into the regression increases the model's predictive power. Note that if the second lag were priced, one could effectively back it out from the current yield curve for example using principal component analysis.

Numerical Model

I next consider a more realistic estimated term structure model. I consider a five factor model with the state variable $[pc_t^1, pc_t^2, pc_t^3, pc_t^4, pc_{t-1}^1]$. That is the state variables consists of the first four principal components of yields and the lag of the first component of yields. The demeaned factor dynamics are given by VAR(1) model in companion form with a coefficient matrix

$$\phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & 0 \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & 0 \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{32}$$

These dynamics are otherwise standard but the first principal component depends also on its second lag.[11] I estimate the coefficient matrix directly from the data using least squares. I estimate the risk neutral factor dynamics and the short rate sensitivity paratemeters δ_1 to minimize the following loss criterion:

$$\Theta \frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\sum_{t=1}^{T}(y_t^{n,m} - y_t^n)^2 + (\beta^m - \beta)'(\beta^m - \beta) + (\beta^{m,c} - \beta^c)'(\beta^m - \beta^c)$$

The first term is essentially identical to the penalty function in Ang et al. (2006) and Cochrane and Piazzesi (2009), the sum of squared deviations between model implied and actually observed yields.[12] The second term is new: the sum of squared deviations between model implied and observed momentum betas. The third term is also new: the squared deviations between model implied and observed momentum betas conditional on information in the yield curve. Finally Θ is a scaling parameter between the

[11] This is motivated by: i) the statistically significant relationship between the first principal component and its second lag, ii) the finding that yield curve momentum is primarily driven by the change in the first principal component.

[12] Here for the model implied yield we have $y_t^{n,m} = -\frac{A_n + B_n X_t}{n}$.

first and the two other moment conditions. Overall, we can view this as a GMM-type estimator with a weighting matrix

$$\begin{bmatrix} \frac{\Theta}{N} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \tag{33}$$

To generate a violation of the spanning condition, I assume that the lag of the first principal component is not priced that is under the risk neutral measure the factor follows a standard VAR(1) model. Hence the yield factor coefficients are of the form $[B_1(n), B_2(n), B_3(n), B_4(n), 0]$. This implies that as in Remark 3, this lag cannot be inferred from the current yield curve, which results in non-zero values for the conditional momentum betas.

Figure 6 shows the resulting population momentum betas and conditional momentum betas along with the values measured from the data. Overall one can see that the model is able to replicate yield curve momentum in the data quite accurately. The root mean squared error between model implied and actual yields is 0.2%.[13] Fit could be further improved by including additional factors to the model.

The model presented above is different from previous non-Markovian term structure models. In particular Feunou and Fontaine (2014) construct a model in which only the expected future values of factors are spanned by yields but their actual values are not. On the other hand, the above model features a factor, whose current value is spanned but its expectation is not.[14]

6.3 Spanning Puzzle and Measurement Error

Cochrane and Piazzesi (2005) find that taking lags of a factor computed from forward rates can help forecast returns. They suggest that this could be explained in a model where yields are observed with error. However, in the appendix I show that their results are largely unrelated to mine.

[13] I set $\Theta = 2000$.

[14] Apparent failures of the Markov property appear also in models with slow updating as in Granziera and Sihvonen (2020). However, these models can be recast in Markov form.

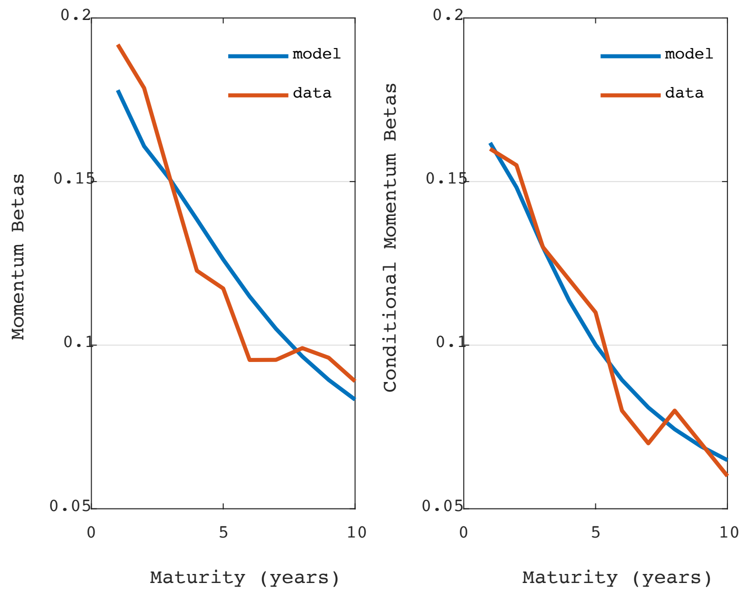

Figure 6: shows the plain and conditional momentum coefficients observed in the data and those implied by the estimated term structure model.

Duffee (2011), Joslin et al. (2014) and Cieslak and Povala (2015) find evidence that measures of inflation and real activity can help forecast bond returns on top of yields. That is some macro variables appear to be unspanned by yields. However, Cieslak and Povala (2015) argue that the evidence is rather consistent with measurement error in yields and inflation. Similarly, Bauer and Rudebusch (2017) argue that the results of Joslin et al. (2014) are due to measurement error. Feunou and Fontaine (2014) postulate that measurement error can explain why expected inflation is not spanned by yields but not why current inflation is unspanned.

I next argue that while measurement error can possibly explain why macro variables are unspanned by yields, it does not explain why past returns are unspanned. I estimate a spanned version of the 4 factor term structure model discussed in the previous section to match average yield errors and momentum betas. Here I impose a VAR(1)-structure on the yields and assume all factors are generally priced. I then simulated the regression slope coefficients controlling for all yield curve information. Similarly to Duffee (2011), Cieslak (2017) and Bauer and Rudebusch (2017), I introduce a normally distributed noise term to yields with the standard deviation based on the yield measurement error found by Liu and Wu (2020).

The simulated 5 per cent critical value for the momentum betas when controlling for all yield curve information is roughly 0.09.[15] Because for short maturities the empirically observed values are above this threshold, measurement error does not appear to explain the violations of the spanning hypothesis documented in this paper.

Of course I cannot fully rule out that there exists some spanned term structure model that is consistent with my empirical results after accounting for the effects of measurement error. However, I have not found much support for the measurement error explanation. In the appendix I further show that introducing noise to the term structure model of Cieslak and Povala (2015) does not explain my findings.

[15] The critical values in previous version of the paper were higher due to a typo in the code.

6.4 Economic Interpretations

I have argued that the empirical results of this paper are problematic for standard theories that do not naturally generate a violation of the spanning condition. But what is the economic reason that the spanning condition is not satisfied? Why are past returns important for predicting future returns but not be priced in the term structure of interest rates today?

Answering this question is challenging because unspanned models still lack a full structural interpretation. Moreover, as in Duffee (2011) and Joslin et al. (2014), these models require knife-edge restrictions on model parameters. However, I now discuss two possible explanations.

A Behavioral Narrative

The first possible explanation is behavioral. The unspanned model presented above is consistent with a situation where the true expected value of the first principal component of yields depends also on its second lag yet agents price bonds as if it does not. That is agents ignore longer term dependencies in the state variable process. Note that assuming AR(1) dynamics is also fairly common in the term structure literature. The results of this paper show that relaxing this assumption has important implications for bond return dynamics.

This narrative is similar but not indentical to that discussed by Cieslak (2017). She argues that that the spanning condition might be violated because some variables available to the econometrician are missing from the agents' information set. Here the narrative should be modified so that the agents' information set is missing higher lags of variables important for determining bond yields.

Rational Arbitrageurs and Rule-Based Traders

Another possible explanation is that demand from rule-based traders offsets some of the effects of short rates on bond prices. I next sketch such a model that is loosely motivated by the term structure model of Vayanos and Vila (2020) and its modification in Hamilton and Wu (2012). This exercise also gives an economic interpretation to the simple unspanned one factor model discussed before.

Assume there are two types of investors: rational arbitrageurs and rule-based traders. Arbitrageurs maximize a mean variance objective over the return of their portfolio r_{t+1}:

$$\mathbb{E}_t[r_{t+1}] - \frac{1}{2}\gamma \mathbb{V}ar_t[r_{t+1}]$$

The portfolio return is given by

$$r_{t+1} = \sum_{n=1}^{N} z_n r^n_{t+1}$$

Here z_n is the number of n maturity bonds held by the arbitrageur and r^n_{t+1} is the return of the corresponding bond. Somewhat similarly to Hong and Stein (1999), the model also features rule-based traders. Assume their demand for each bond ($n \geq 2$) is given by χy^1_{t-1}, where χ is a constant.[16] By market clearing

$$z_n = -\chi y^1_{t-1}$$

Finally assume, as in the simple one factor example, that short rate dynamics are given by:

$$y^1_{t+1} = c + \rho_1 y^1_t + \rho_2 y^1_{t-1} + \epsilon_{t+1}$$

We obtain the following result:

Proposition 2. *There is a $\chi > 0$ such that $p^n_t = A_n + B_n y^1_t$. Here $B_n = B_{n-1}\rho_1 - 1$.*

Proof: see appendix.

The interest rate sensitivity parameters B_n are identical to our previous one factor model that was able to generate both unconditional and conditional

[16] We could also add a maturity specific constant to this demand.

momentum. However, here we do not have a free risk parameter to calibrate the persistence separately from its objective counterpart. Using the already estimated AR(2)-process for short rates, we can now simulate a plain momentum slope coefficient of 0.26 and a conditional coefficient of 0.15. While this model generates a sligthly stronger momentum pattern than in the data, given its simplicity it does surprisingly well.

Given our estimated process, the model implies that high interest rates are associated with high bond returns. An interpretation of the rule χy_{t-1}^1 is then that these traders demand more bonds during times of high interest rates because they associate this with high returns. We could justify the use of past month rather than current month interest rate, if they have some sluggishness in executing trades and must decide their period t holdings already at $t-1$. Interest rate levels could also be related to financial institutions hedging demands.

Why do the arbitrageurs price bonds as if the process is AR(1)? Conditional on the current month rate y_t^1, a high past month rate y_{t-1}^1 predicts a lower value for next month rate y_{t+1}^1. This force pushes bond prices up already this month. However, a high past month rate y_{t-1}^1 implies high demand from rule based traders. This implies that the arbitraugers must absorb more duration risk. Because these two forces cancel out bond prices do not depend on y_{t-1}^1.

7 Conclusion

I analyze time series momentum along the Treasury term structure. I find that past returns predict future bond returns largely because of autocorrelation in yield changes. This autocorrelation is further due to both autocorrelation in bond risk premia and correlation between bond risk premia and past unexpected bond returns. Because Treasury yields are correlated with the federal funds rate, yield curve momentum is partly driven by post-FOMC announcement drift. Finally, past returns are not spanned by information in the current term structure of interest rates.

The last finding is particularly problematic for standard theory models, which predict that yield curve momentum should vanish after controlling for sufficiently many yields. However, I show that the results are consistent with a term structure model with unpriced longer term dependencies.

8 Appendix

8.1 Controlling for More Yield Curve Information

The main text shows the result from predicting bond returns using past bond returns and the first three principal components of yields. I now extend these results using the following regression:

$$rx_{t+1}^{n} = \alpha + \beta_1 rx_t^n + \sum_{i \in S} \beta_i y_t^i + \epsilon_{t+1}, \tag{34}$$

where the selected yields are the 1 month and 1 to 10 year rates. Note that this is equivalent to controlling for the 1 month rate and the corresponding 10 forward rates and spans the tent-shaped factor discussed by Cochrane and Piazzesi (2005). The results are shown in table 14. The coefficient on the past return is statistically significant for shorter maturities though less so for longer maturities. This suggests that at least for shorter maturities yield curve momentum exists after controlling for the information in the yield curve today.

In some models, for example in the habit model of Wachter (2006), yields affect future returns non-linearly. We now test this possibility by considering the more general partially linear regression

$$rx_{t+1}^{n} = \beta_1 rx_t^n + f(\mathbf{y}_t) + \epsilon_{t+1}. \tag{35}$$

As explained later, assuming an invertibility condition, any Markovian model of yields implies that

$$rx_{t+1}^{n} = f(\mathbf{y}_t) + \epsilon_{t+1}. \tag{36}$$

Mat. (y)	1	2	3	4	5	6	7	8	9	10
α	-0.02	-0.01	-0.05	-0.10	-0.22	-0.27	-0.39	-0.59	-0.77	-0.99
t-value	-0.42	-0.10	-0.32	-0.49	-0.86	-0.89	-1.15	-1.49	-1.70	-2.01
β_1 (rx_t^1)	0.16	0.16	0.13	0.12	0.11	0.08	0.07	0.08	0.07	0.06
t-value	2.41	2.97	2.64	2.30	2.24	1.54	1.56	1.60	1.52	1.42
β_2 (y_t^1)	-0.23	-0.41	-0.63	-0.92	-1.11	-1.19	-1.38	-1.61	-1.72	-1.91
t-value	-4.42	-3.94	-4.27	-4.52	-4.66	-4.51	-4.70	-4.70	-4.62	-4.70
β_3 (y_t^{12})	0.51	0.64	1.32	1.95	2.33	2.53	2.88	3.35	3.60	3.95
t-value	2.68	1.74	2.59	2.85	2.96	2.84	3.06	3.22	3.30	3.37
β_4 (y_t^{24})	-0.44	-0.18	-1.40	-1.33	-1.25	-1.46	-1.76	-2.14	-2.56	-3.08
t-value	-0.92	-0.22	-1.24	-0.94	-0.75	-0.77	-0.86	-0.99	-1.16	-1.33
β_5 (y_t^{36})	0.46	0.01	0.53	-1.74	-1.67	-0.95	-0.57	-0.50	0.38	1.15
t-value	0.51	0.01	0.28	-0.83	-0.69	-0.34	-0.19	-0.16	0.11	0.32
β_6 (y_t^{48})	-0.22	0.49	1.51	3.90	1.56	1.10	1.30	1.79	0.31	-0.00
t-value	-0.18	0.23	0.57	1.28	0.43	0.28	0.30	0.36	0.05	-0.00
β_7 (y_t^{60})	-1.19	-2.39	-3.71	-4.14	-2.37	-5.23	-6.47	-7.15	-5.49	-5.67
t-value	-1.64	-1.95	-2.17	-1.86	-0.89	-1.43	-1.45	-1.39	-0.94	-0.91
β_8 (y_t^{72})	1.39	2.84	4.04	4.59	5.71	9.15	9.01	9.79	9.93	11.31
t-value	1.83	2.00	2.05	1.80	1.88	2.31	1.93	1.89	1.78	1.94
β_9 (y_t^{84})	-0.05	-0.83	-1.57	-2.23	-2.34	-2.88	-1.71	-3.35	-4.43	-6.18
t-value	-0.11	-1.02	-1.29	-1.28	-1.06	-1.06	-0.46	-0.78	-0.95	-1.23
β_{10} (y_t^{96})	0.01	0.22	0.41	0.33	-0.80	-0.04	0.02	0.89	-0.97	-1.36
t-value	0.03	0.26	0.32	0.18	-0.36	-0.01	0.00	0.20	-0.20	-0.26
β_{11} (y_t^{108})	-0.28	-0.81	-1.26	-1.32	-0.89	-2.21	-2.52	-2.56	-1.04	-1.80
t-value	-0.52	-0.89	-0.96	-0.79	-0.43	-0.76	-0.66	-0.57	-0.21	-0.35
β_{12} (y_t^{120})	0.02	0.41	0.76	0.87	0.81	1.15	1.17	1.47	1.97	3.60
t-value	0.08	0.83	1.11	1.02	0.77	0.86	0.65	0.69	0.85	1.44
R^2	0.11	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Table 14: shows the results of predicting returns of different maturity bonds on the past return of the bond and the yields of 1 month bill and 1 to 10 year bonds. The t-values are based on Newey and West (1987) standard errors.

	Semipar.		Squares	
Mat	β_1	t-value	β_1	t-value
1	0.22	4.51	0.20	2.23
2	0.25	5.09	0.22	2.62
3	0.23	4.63	0.18	2.40
4	0.20	4.12	0.16	2.02
5	0.19	3.98	0.14	1.90
6	0.16	3.27	0.10	1.42
7	0.14	2.87	0.09	1.42
8	0.12	2.65	0.08	1.25
9	0.11	2.42	0.06	1.02
10	0.11	2.40	0.06	0.93

Table 15: shows the slope coefficient on past return when explaining excess bond returns on past excess bond returns on an arbitrary non-linear function of yields, estimated using a semiparametric method, as well as a linear regression with yields and squared yields. The t-values for the first regression are obtained using quasi-maximum likelihood (Wood, 2011). The t-values for the second regression are based on Newey and West (1987) standard errors.

Therefore these models imply that $\beta_1 = 0$. However, the challenge is that f is generally unknown. I tackle this using two approaches. The first approach is to estimate the model using the semiparametric approach described by Wood (2011). Here the standard errors are calculated using quasi-maximum likelihood.[17] The second approach is to simply add the squared yields, on top of the yields, to the regression. The results are given in table 15, which shows the results for the β_1 parameter. For the first approach β_1 is always significant. However, the model produces a high in sample fit and might achieve low standard errors by overfitting. For the second approach, the slope coefficient is significant for shorter but not for longer maturity bonds. These exercises suggest that accounting for non-linearities does not strongly alter the main conclusions of this paper.

[17] To avoid problems with overfitting I only include yields of every second year.

Maturity	α	t-value	β_1 (c_t^n)	t-value	β_2 (yc_t^n)	t-value	R^2
1	0.01	0.19	0.97	2.50	0.16	2.39	5.17
2	-0.02	-0.29	1.29	2.40	0.16	3.06	5.02
3	-0.00	-0.03	1.15	1.40	0.14	2.63	3.18
4	-0.06	-0.46	1.54	2.19	0.11	2.19	2.95
5	-0.09	-0.66	1.74	2.51	0.10	2.23	2.98
6	-0.15	-0.95	2.06	2.75	0.08	1.56	2.84
7	-0.06	-0.34	1.58	2.05	0.09	1.81	1.83
8	-0.21	-1.15	2.35	2.94	0.08	1.66	2.97
9	-0.25	-1.33	2.71	3.13	0.08	1.69	3.23
10	-0.07	-0.30	1.66	1.97	0.08	1.80	1.63

Table 16: shows the results of predicting returns of different maturity bonds on carry and past yield change. The t-values are based on Newey and West (1987) standard errors.

8.2 Predicting Bonds Returns with Carry and Yield Change

The results of the main section suggest that including information about both past carry and yield change might be beneficial to predicting bond returns. I now test this prediction by including both variables separately into the predictability regression.

$$rx_{t+1}^n = \alpha + \beta_1 c_t^n + \beta_2 yc_t^n + \epsilon_{t+1} \tag{37}$$

Note that because period t carry is observable we include this rather than the previous period carry into the regression. The results are given in table 16. For most maturities both carry and past yield change are significant. There is a small increase in R^2 relative to a regression with past return.

8.3 Post Announcement Drift: Further Analysis

This section provides some further results related to the post-FOMC announcement drift. Figure 7 shows the changes in different maturity yields per one basis point change in the federal funds target rate. Shorter maturity yields show a clear drift pattern after target rate changes.

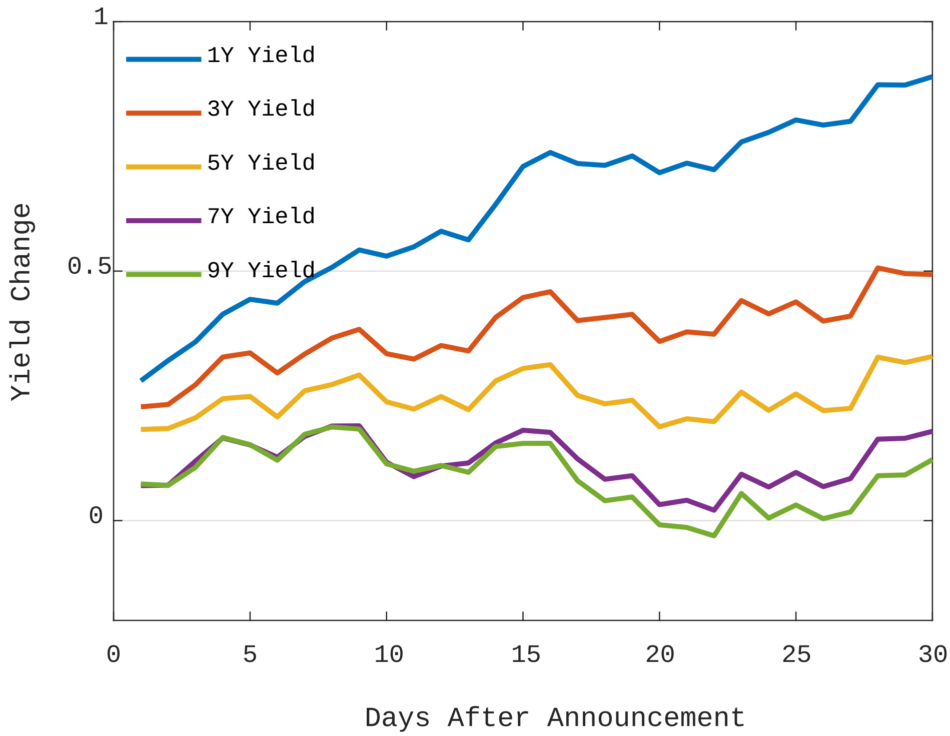

Figure 7: shows the change in different maturity yields after a change in the federal funds target rate (FFTR). Changes are measured per one basis point change in the FFTR. Days after announcement are measured using trading days.

In this particular sample long maturity yields do not exhibit similar drifts. However, as explained by Brooks et al. (2019) results for long maturities are stronger when considering unexpected target rate changes. This can explain why are regression results are stronger for long maturity bonds when using unexpected rather than plain changes in the target rate.

Figure 8 plots the historical development of different maturity yields along with that for the target rate. One can see that all the yields share the same broad developments. However, the contemporaneous correlation between yield changes and changes in the federal funds target rate is far from perfect. Post-FOMC announcement drift seems to contribute to this correlation being fairly low. However, this is likely not the only reason.

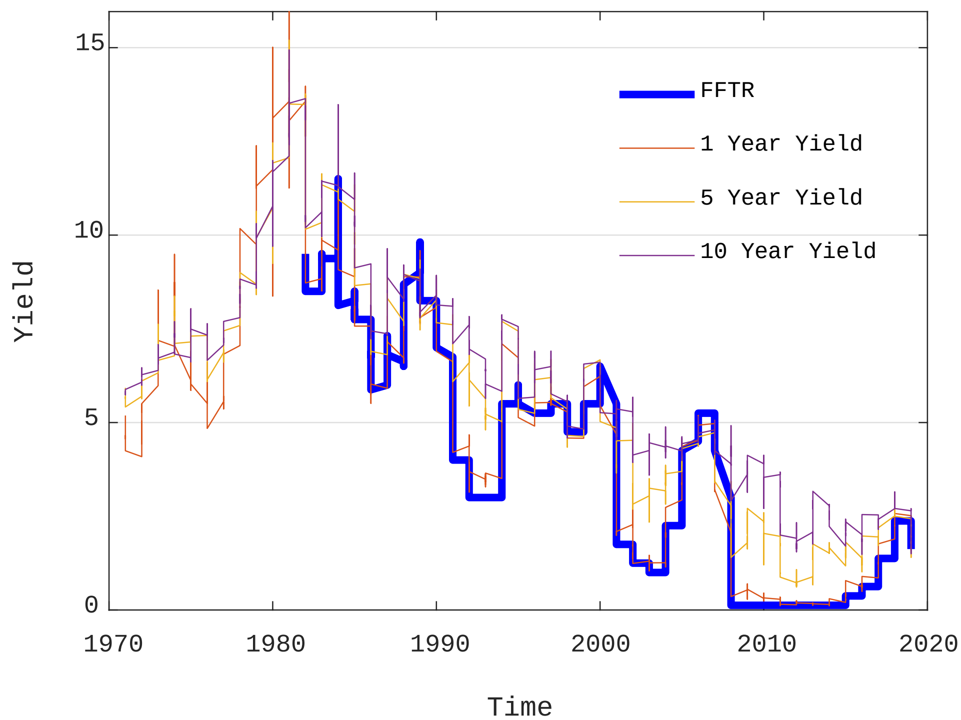

Figure 8: shows the historical development of 1,5 and 10 year yields along with the federal funds target rate.

For example theoretically longer maturity yields should reflect expectations about the long run path of future short rates and also anticipate target rate changes.[18]

8.4 Robustness with Respect to Gürkaynak et al. (2007) data

Liu and Wu (2020) construct the yield curve using a novel procedure that results in lower pricing errors compared to standard procedures such as the Svensson (1994) method applied by Gürkaynak et al. (2007). How does this affect the key results of this paper?

[18]Also yield levels reflect the cumulative effect of yield changes and hence tend to be more correlated.

Mat	α	t-value	β	t-value	R^2
1	0.06	3.31	0.19	3.12	3.77
2	0.09	2.86	0.17	3.39	2.83
3	0.12	2.70	0.14	3.12	2.02
4	0.15	2.59	0.12	2.78	1.46
5	0.18	2.49	0.10	2.47	1.08
6	0.20	2.38	0.09	2.19	0.83
7	0.22	2.28	0.08	1.94	0.65
8	0.23	2.17	0.07	1.69	0.51
9	0.25	2.08	0.06	1.46	0.39
10	0.26	1.99	0.05	1.23	0.29

Table 17: shows the results from regressing the excess returns of different maturity bonds on their past returns using the atlernative data from Gürkaynak et al. (2007). The t-values are based on Newey and West (1987) standard errors.

Table 17 replicates the results in table 1 for the 1 month lookup using the Gürkaynak et al. (2007) data updated on the Federal reserve webpage. The sample period is as before. While this alternative data yields somewhat smaller coefficients for long maturity bonds, overall the results are fairly similar across the two datasets.

8.5 Robustness with Respect to German Data

Are the results robust to data from other developed countries? Next I study this using data on the German government yield curve available on the Bundesbank webpage. These curves are constructed using the interpolation procedure of Svensson (1994). Because standard interpolation procedures often have large pricing errors for short maturity yields (Liu and Wu, 2020), I focus on actual rather than excess returns that do not require specifying a 1 month risk-free rate.

I replicate the exercise of explaining the return of different maturity bonds on their return in the prior month. The results are given in table 18 and are fairly similar for both countries. The R^2 is quite high for short maturity bonds in both countries as their returns are strongly related to

	Germany					USA				
Mat.	α	t-value	β	t-value	R^2	α	t-value	β	t-value	R^2
1	0.23	5.10	0.40	3.86	15.65	0.26	7.55	0.42	7.31	17.28
2	0.28	7.76	0.35	5.47	12.59	0.37	9.53	0.24	4.46	5.82
3	0.35	8.22	0.27	4.56	7.68	0.43	9.12	0.18	3.27	3.16
4	0.39	8.22	0.25	4.66	6.17	0.48	8.18	0.13	2.67	1.77
5	0.43	7.88	0.22	4.52	5.01	0.50	7.32	0.12	2.59	1.46
6	0.47	7.44	0.20	4.13	3.92	0.54	6.58	0.09	1.91	0.90
7	0.50	7.01	0.17	3.60	2.95	0.55	6.01	0.09	1.88	0.85
8	0.54	6.61	0.15	3.02	2.14	0.56	5.52	0.10	1.95	0.91
9	0.57	6.26	0.12	2.45	1.45	0.56	5.04	0.09	1.94	0.81
10	0.61	5.93	0.09	1.86	0.87	0.59	4.80	0.08	1.78	0.68

Table 18: shows the results from regressing returns of different maturity bonds on their past returns in both Germany and US. The t-values are based on Newey and West (1987) standard errors.

short-term yields that are highly autocorrelated. This suggests that the results are robust to the German yield curve though this curve might be measured with larger pricing errors.

8.6 Time Series vs Cross-Sectional Momentum

The literature on equity momentum (e.g. Chan et al. (1996)) has focused on a cross sectional strategy that goes long stocks with relatively high past returns and short stocks with relatively low past returns. Could a similar strategy be applied with different maturity government bonds?

The finding that time series momentum is largely associated with a single factor suggests that such a strategy is unlikely to provide high returns. I now demostrate this further by considering a simple cross-sectional momentum strategy. I consider the returns of bonds with maturities from 1 to 10 years. As in Lewellen (2002) assume the weight of each bond is given by $w_i = (r_{i,t} - r_{p,t})/10$, where $r_{i,t}$ is the return of the bond and $r_{p,t}$ is the return of an equal weighted portfolio of all the ten bonds. The mean return of this strategy can be decomposed as follows:

$\mathbb{E}[r_{s,t}]$	$\frac{1}{10}\sum_{i=1}^{10}\rho_i$	$\frac{1}{10}\sum_{i=1}^{10}\mathbb{E}[r_{i,t}]^2$	$-\rho_m$	$-\mu_m^2$
0.0003	0.0044	0.0033	-0.0041	-0.0032

Table 19: shows a decomposition of the mean return from a cross sectional momentum strategy (%)

$$\mathbb{E}[r_{s,t}] = \frac{1}{10}\sum_{i=1}^{10}(\rho_i + \mathbb{E}[r_{i,t}]^2) - (\rho_m + \mu_m^2).$$

Here ρ_i and ρ_p are the autovariances of the individual bonds and the equally weighted portfolio of bonds respectively. Moreover, μ_i and μ_p are the bonds' unconditional mean returns.

The results from this decomposition are given in table 19. The strategy yields a 0.0003 per cent monthly return with a modest annualized Sharpe ratio of 0.087. This is largely because the mean autocovariance of the bonds is close to the autocovariance of an equally weighted portfolio of the bonds. This zero net investment strategy cannot benefit from time series momentum related to shifts in a single factor that manifests itself somewhat similarly for all the different maturity bonds.

8.7 Investment Performance

The results of this paper suggest that an investor could gain using momentum strategies in Treasury bonds. But how big are these gains? Answering this question is complicated because such momentum strategies can be implemented in multiple ways. While more sophisticated strategies might provide higher returns, for transparency I focus on a particularly simple strategy. In particular assume an investor buys a bond assuming its past month excess return was positive. On the other hand, if this past return was negative, assume the investor instead chooses to hold short term bills earning her zero excess returns. Note that this simple strategy naturally also constitutes an "out of sample" evaluation for the relevant trading performance.

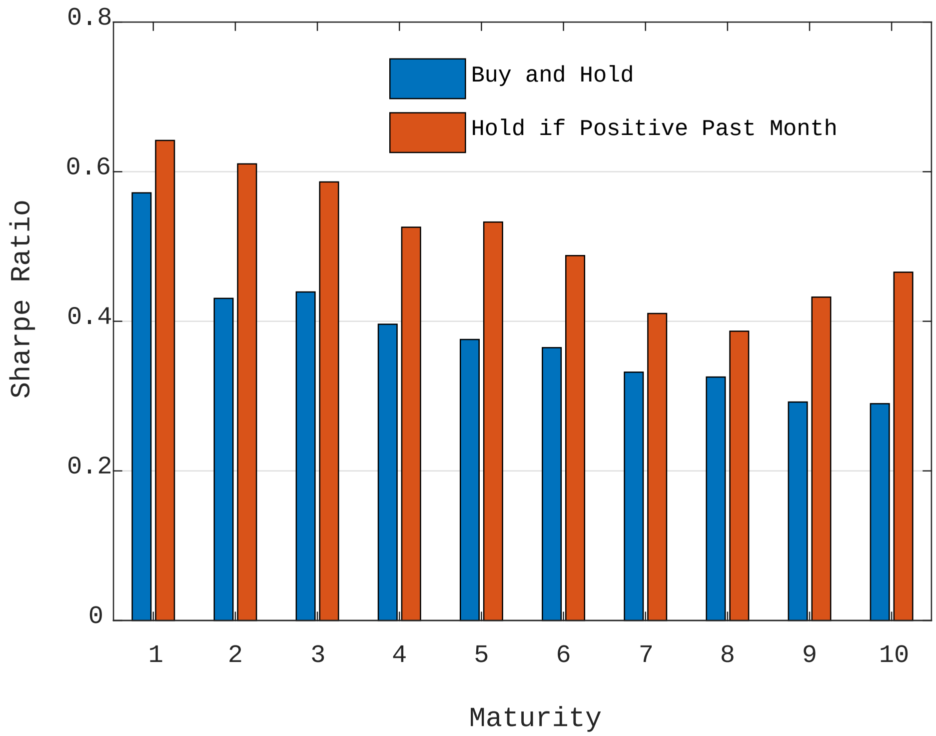

Figure 9: shows the annualized Sharpe ratios for different maturity bonds for a simple momentum strategy and a buy and hold strategy.

Figure 9 shows the Sharpe ratios from this simple momentum strategy along with those for a buy and hold strategy that passively holds given maturity bonds. One can see that the momentum strategy earns higher Sharpe ratios for all maturities. The average Sharpe ratio of the momentum strategy is 0.51 compared to 0.38 for the buy and hold strategy. This momentum stategy also enjoys a positive average skewness of 1.27 compared to 0.22 for the buy and hold strategy. The Sharpe ratios for an equally weighted portfolio of simple momentum strategies would be 0.50 compared to 0.36 for an equally weighted buy and hold strategy. Here the improvement in Sharpe ratio is therefore 39%.

Figure 3 conveys an interesting additional point. The mean excess returns are fairly close to zero following months with negative past month returns. Hence it is not clear that an investor could benefit from twisting our momentum strategy by also going short bonds after such months. This long short strategy would improve mean returns for some maturity bonds but not all. Moreover, because this improvement in mean returns is fairly small but such a strategy involves higher volatility, the Sharpe ratios for this long-short strategy are lower for all maturities.[19]

Finally note that a more comprehensive analysis of the investment performance of yield curve momentum strategies should take into account the broader constitution of the investor's portfolio and other signals used. For example Hurst et al. (2017) notes that trend followers can clearly improve Sharpe ratios by diversifying exposures to momentum strategies for different asset classes. They also show that momentum returns tend to survive after controlling for reasonable estimates of transaction costs.

[19]This point is somewhat nuanced though. If the unconditional bond risk premium represents rational compensation for risk, going short following months with negative returns might hedge macroeconomic risk and is not necessarily suboptimal.

8.8 Results for a Bond Index

The key results of this paper are based on a yield curve constructed using a numerical approximation scheme. A possible concern is that these errors contribute to the key findings regarding yield curve momentum.[20] I next demonstrate that these errors are unlikely to invalidate the main regression results of this paper.

In particular, I use the excess returns on the Bloomberg Aggregate Treasury bond index, available from 1973, that is a few years before the start of our main data. This index is calculated directly using Treasury bonds and hence represents tradable returns. It serves as perhaps the most widely followed benchmark index for Treasuries. However, the results obtained with this index are not fully comparable with our main results because of two reasons. First, this index is based on coupon paying bonds, while our main results are for zero coupon bonds. Second, this index represents a broad portfolio of different maturity Treasury bonds.

I replicate the key regression of this paper by explaining one month excess return on this index by its past value. The slope coefficient is 0.11, which is close to the slope coefficients for longer maturity bonds in table 1. The corresponding t-value is 2.67 and hence the results are strongly significant.

I also replicated the investment strategy that holds bonds only in months following positive past month excess returns. The Sharpe ratio for this strategy is 0.55 compared to 0.44 for a buy and hold strategy. Note that because the strategy is effectively implemented for a portfolio of bonds, it cannot benefit from any individual time series predictability for different maturity bonds.

[20] Zero coupon Treasury bonds are traded as Treasury STRIPS introduced in 1985. Before that the Treasury issued some zero-coupon bonds. However, overall there are not enough such bonds to create long histories of zero coupon curves.

8.9 Stability Analysis

Is yield curve momentum stronger during some periods than others? I now analyze potential structural breaks in the relationship between current and past returns. I consider a simple 10 year rolling regression. Figure 10 plots the results when one month return is explained with the one month return in the past month. One can see that the slope coefficients are fairly stable overall but clearly fall after the financial crisis.

Explaining this break is beyond the scope of the paper. However, the period is characterized by extraordinarily low interest rates and unconventional monetary policies. For example an effective lower bound on yields can alter the relationship between current and past bond returns. The Fed policies during the period pushed yields down and led to high bond returns. However, if yields are close to an effective floor, these high bond returns do not predict similar elevated returns going forward.

As discussed in the section on investment performance, bond excess returns tend to be close to zero following months with negative returns. Effectively the negative momentum effect is offset by a substantial unconditional bond risk premium. On the other hand, following positive months the positive momentum effect increases expected bond returns on top of the unconditional risk premium. Because high bond returns are associated with increasing interest rates, momentum strategy returns tend to be higher during subperiods with declining rather than increasing interest rates.

8.10 Predicting Yield Changes: the Longer Run

In this section I study the longer run effects of a shock to bond yields. I consider a regression of the form

$$\Delta y_{t+h}^{n} = \alpha + \beta \Delta y_{t}^{n} + \epsilon_{t+h} \tag{38}$$

for different horizons h. That is I predict yield changes between $t+h$ and $t+h-1$ by the change in the same maturity bond yield between t and $t-1$.

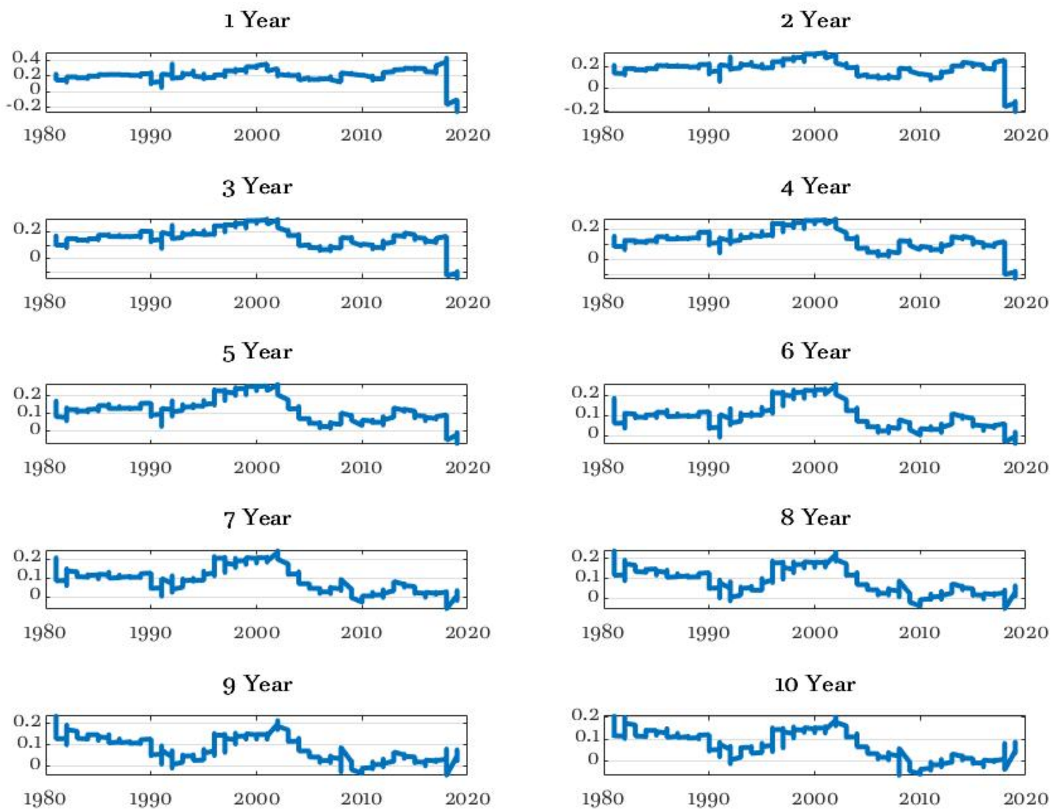

Figure 10: Momentum slope coefficient in a rolling 10 year sample for different maturity bonds

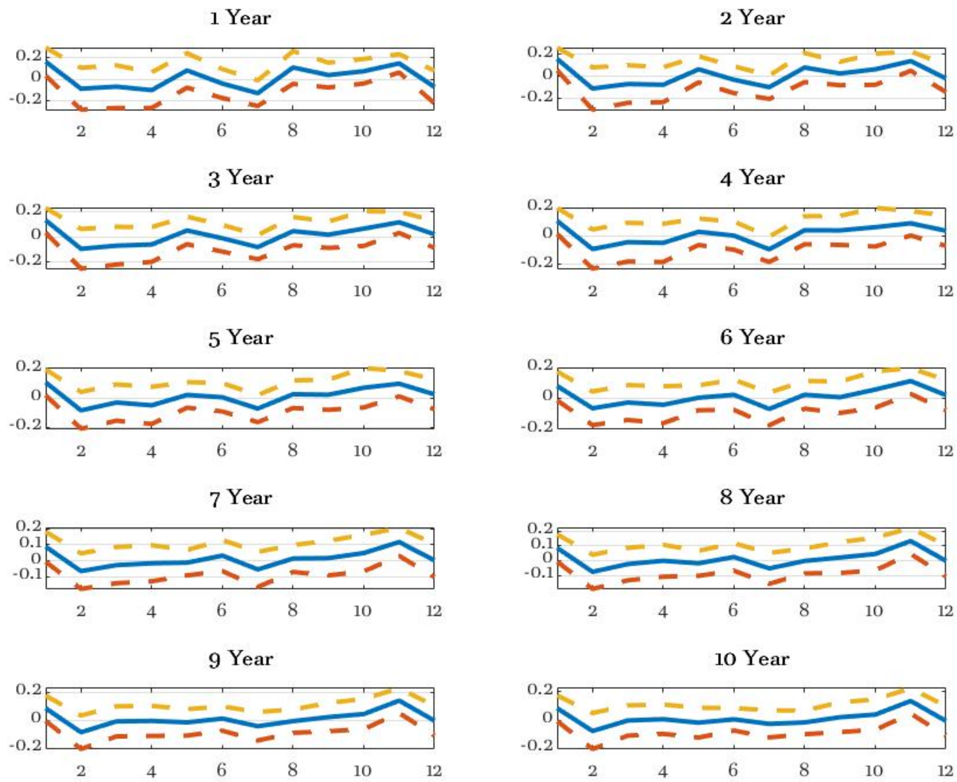

Figure 11: shows the slope coefficients on a regression of bond yield change on future bond yield changes

As in the local projection method of Jordà (2005), the slope coefficients can be interpreted as a type of impulse response function.

The resulting slope coefficients along with the 95% confidence intervals are shown in figure 11. The coefficients are high for the horizon of one month and then again high for the 11 month horizon. Many of the coefficients in between are negative though not statistically different from zero. These results can explain why the 1 month horizon works best in the regressions reported in table 1.

The slope coefficients for different horizons sum to a numbers slightly smaller than the coefficient for the first year. Therefore the total effect to yields after a year is positive but fairly small. Put alternatively, assume there is an increase in bond yields at period t. Because of short horizon autocorrelation in yields, this predicts a further increase in yields in the next month. The longer horizon autocorrelations largely offset each other so that on average yields after a year remain sligthly below but close to the level after a month following the yield change ($t+1$).

8.11 On the Cochrane-Piazzesi-Factor (CP)

Cochrane and Piazzesi (2005) find that a single tent shaped factor of forward rates can forecast annual excess returns for bonds with maturities between 2 and 5 years. This factor is weakly correlated with the standard level, slope and curvature factors and is rather connected to the fourth and fifth principal component of yields. Cochrane and Piazzesi (2005) also argue that including lags of this factor can improve forecasting performance. They postulate that this might be related to measurement error in yields, though do not investigate the issue formally.

Are the results in this paper related to those in Cochrane and Piazzesi (2005)? First note that the spanning results when controlling 11 yields are effectively also accounting for the CP-factor, which can be expressed as a function of these yields. Second, we found that momentum can be largely captured by the change in the first principal component of yields (PC1). This is simply because bond returns have a high contemporaneous correlation with changes in PC1. On the other hand the average contemporaneous correlation between changes in the CP-factor and bond excess returns is merely -0.12.[21] While this factor can forecast returns, it cannot explain a high share of return autocorrelation.

However, in theory the finding that a lagged CP-factor can help forecast returns might help explain why past bond returns can predict returns conditional on the information in the yield curve today. This is not the

[21] Here I construct the factor as in Cochrane and Piazzesi (2005).

Maturity	β_1 (rx_{t-1})	t-value	β_2 (CP_{t-1})	t-value
1	0.26	2.70	0.16	1.54
2	0.21	3.03	0.20	1.13
3	0.17	2.87	0.17	0.84
4	0.13	2.50	0.16	0.69
5	0.12	2.53	0.23	0.90
6	0.10	2.00	0.30	1.06
7	0.10	2.02	0.21	0.68
8	0.10	2.04	0.16	0.45
9	0.095	1.99	0.21	0.54
10	0.089	1.87	0.19	0.45

Table 20: shows the slope coefficients from a regression of bond excess returns on past bond excess returns and a lagged Cochrane-Piazzesi factor. The regression also controls for the first five principal components of yields. The t-values for the second regression are based on Newey and West (1987) standard errors.

case empirically. Table 20 shows the results from a spanning regression that explains bond returns on their lag, the lagged CP-factor and five principal components of yields. The results are similar to before and past returns are clearly significant. Here the lagged CP-factor does not appear to improve forecasting performance at a monthly frequency.

8.12 Risk Premium-News Decomposition with Macro Variables

Table 21 repeats the exercise in table 8 but now adding trend inflation and the activity index to the predictive regression used to calculate the risk premium. The results are similar to before though now the risk premium component accounts for a somewhat larger component of autocovariance in returns.

8.13 On Macro- vs. Yield-Based Factors

Joslin et al. (2014) assume the spanned variables possess VAR(1)-dynamics. Hence their framework does not nest the setting discussed in this paper.

Maturity	Risk premia	News	Error
1	36.1 %	64.1 %	-0.3 %
2	37.4 %	63.1 %	-0.6 %
3	44.8 %	61.5 %	-6.3 %
4	51.6 %	49.3 %	-0.9 %
5	50.2 %	43.4 %	6.4 %
6	57.4 %	50.7 %	-8.1 %
7	64.5 %	38.9 %	-3.4 %
8	70.1 %	31.1 %	-1.2 %
9	75.7 %	24.1 %	0.3 %
10	87.8 %	14.2 %	-2.0 %

Table 21: shows the decomposition of covariance between the return of different maturity bonds and their past value into the autocovariance of risk premia, covariance between risk premia and past unexpected bond returns and covariance between past returns and an approximation error component. The risk premium accounts for macro variables

Because of such differences we now discuss the relation between deep macroeconomic factors and principal component based factors.

Consider a spanned macroeconomic factor Z_t. Under the risk neutral measure it follows a VAR(2) process (demeaned and written in companion form):

$$\begin{bmatrix} Z_t \\ Z_{t-1} \end{bmatrix} = \begin{bmatrix} \phi^*_{1,Z} & \phi^*_{2,Z} \\ I & 0 \end{bmatrix} \begin{bmatrix} Z_{t-1} \\ Z_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{t,Z} \\ 0 \end{bmatrix}$$

On the other hand we assumed the principal components of yields follow.

$$\begin{bmatrix} PC_t \\ PC_{t-1} \end{bmatrix} = \begin{bmatrix} \phi^*_1 & 0 \\ I & 0 \end{bmatrix} \begin{bmatrix} PC_{t-1} \\ PC_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$$

These representations are equivalent assuming $\phi^*_{2,Z} = 0$. That is we are *de facto* assuming the longer lags of the macro factors are not priced. This and the assumption that Z_t is spanned implies (demeaned) yields of the form $y_t(n) = \hat{B}(n)Z_t$. Given the spanning assumption, there is a direct rotation between factors and yields $Z_t = R \times PC_t$. Now we also have $\phi^*_{1,Z} = R\phi^*_1$. If the deep factors are unobservable we do not have to solve for Z_t and R but can rather employ principal components as factors. The mapping between the physical law of motion for the deep and principal component based factors is similar. Here the coefficient matrices for principal components can be transformed to those for the deep factors by multiplying by R.

Above we effectively assumed the factors Z_t are spanned by yields, merely their lags are not. Unspanned factors do not affect the risk neutral process, here only spanned variables are included. However, such unspanned factors occur in the physical factor law of motion.

8.14 Relation to Crump and Gospodinov (2021)

In a recent contribution Crump and Gospodinov (2021) criticize standard practices of characterizing the factor structure of interest rates. They propose i) modelling the factor structure of bond returns rather than that of yield levels, ii) being cautious with standard goodness of fit measures. Following their advice, my spanning decomposition applies more principal components than would be required to simply produce a high fit to yield levels. However, I use principal components of yield levels rather than those of bond returns. Using principal components of past returns would make the decomposition meaningless as these are mechanically related to past returns.

The form for my decomposition is also implied by standard term structure models. Crump and Gospodinov (2021), on the other hand, do not offer an economical explanation for their findings. However, a theoretical model that could generate their results might also help in explaining yield curve momentum.

8.15 Measurement Error in the Cieslak-Povala-Model

Cieslak and Povala (2015) build a three factor macro-finance term structure model. They apply the model to argue that measurement error can explain why trend inflation appears unspanned by yields. This model is different from the model I use to study the effects of measurement error. Therefore as a robustness check I now argue that combining measurement error with the Cieslak-Povala model does not explain my findings.

The estimation of the model follows Cieslak and Povala (2015) with three exceptions. First, my sample includes nine years of additional more recent data. Second, I apply only bonds with maturities less than 10 years. Third, I cast the model in monthly form even though I also calibrate the model to match annual coefficients.

The model features three factors: trend inflation, cycle (real rate) and a return forecasting factor. Trend inflation and real rate persistence as well as the market price of risk parameters are calibrated to match reduced form yield loadings. The rest of the parameters are estimated directly using regressions.

I first use the model to simulate momentum betas obtained by regressing monthly excess returns on their past values. The population coefficients decline in maturity and range between 0.026 and -0.017. The empirical counterparts for the momentum coefficients range between 0.19 and 0.09. Therefore the model generates momentum only for shorter maturity bonds and even there the magnitude is much smaller than in the data.

Cieslak and Povala (2015) consider the effects of different values for measurement error. I use a conservative 10 basis point independent error on all yields. I then simulate five percentage point (two sided) confidence intervals for the momentum betas when controlling for three principal components of yields. The critical values range between 0.02 at the short end of the curve to 0.096 at the long end. The empirical values for the momentum betas are below the critical values only for 10 year bonds.

Therefore again measurement error does not appear to explain my results with the exception of perhaps the very longest maturities. However, note that this model is not estimated to match momentum and hence appears to provide a less stringent test on the measurement error explanation than my numerical model.

8.16 Proof of Proposition 1

We have

$$
\begin{gathered}
rx_{t+1}^{n} = -(n-1)y_{t+1}^{n-1} + ny_t^n - y_t^1 = \\
-(n-1)(A(n-1) + B(n-1)'X_{t+1}) + n(A(n) + B(n)'X_t) - (A(1) + B(1)'X_t) = \\
-(n-1)B(n-1)'X_{t+1} + (nB(n)' - B(1)')X_t - (n-1)A(n-1) + nA(n) - A(1)
\end{gathered}
$$

$$
\begin{gathered}
Cov(rx_{t+1}^n, rx_{t-1}^n) = \\
Cov(-(n-1)B_{n-1}'X_{t+1}, -(n-1)B_{n-1}'X_t) + Cov(-(n-1)B_{n-1}'X_{t+1}, (nB_n' - B_1')X_{t-1}) + \\
Cov((nB_n' - B_1')X_t, -(n-1)B_{n-1}'X_t) + Cov(nB_n' - B_1')X_t, (nB_n' - B_1')X_{t-1}) = \\
(n-1)^2 B_{n-1}'\phi V B_{n-1} - (n-1)B_{n-1}'\phi^2 V(nB_{n-1} - B_1) - (n-1)(nB_n' - B_1')VB_{n-1} + \\
(nB_n' - B_1')\phi V(nB_n - B_1)
\end{gathered}
$$

$$
\begin{gathered}
Var(rx_t^n) = \\
(n-1)^2 B_{n-1}' V B_{n-1} - 2(n-1)B_{n-1}'\phi V(nB_n - B_1) + \\
(nB_n' - B_1')V(nB_n - B_1)
\end{gathered}
$$

The regression slope coefficient is given by the ratio of the covariance and variance terms.

8.17 Proof of Remark 2

Due to normality, the standard pricing formula applies:

$$
p_t^n = -y_t^1 + \mathbb{E}_t[p_{t+1}^{n-1}] + \frac{1}{2}Var_t(p_{t+1}^{n-1}) + Cov_t(m_{t+1}, p_{t+1}^{n-1})
$$

Hence

$$rx_{t+1}^{n} = p_{t+1}^{n-1} - p_t^n - y_t^1 = p_{t+1}^{n-1} - \mathbb{E}_t[p_{t+1}^{n-1}] - Cov_t(m_{t+1}, p_{t+1}^{n-1}) - \frac{1}{2}Var_t(p_{t+1}^{n-1})$$

$$rx_{t+1}^{n} = B_{n-1}v_{t+1} + B_{n-1}V\lambda_t - \frac{1}{2}B_{n-1}'VB_{n-1}$$

Therefore

$$Var(rx_{t+1}^{n}) = B_{n-1}^{2}Var(v_{t+1} + V\lambda_t)$$

and

$$Cov(rx_{t+1}^{n}, rx_t^n) = B_{n-1}^{2}Cov(v_{t+1} + V\lambda_t, v_t + V\lambda_{t-1})$$

and the slope coefficient in the momentum regression (this is given $n \geq 2$, if $n = 1$, excess returns are always zero and the coefficient undefined) is

$$Var(rx_{t+1}^{n}) = B_{n-1}^{2}Var(v_{t+1} + V\lambda_t)$$

and

$$\frac{Cov(rx_{t+1}^{n}, rx_t^n)}{Var(rx_{t+1}^{n})} = \frac{Cov(v_{t+1} + V\lambda_t, v_t + V\lambda_{t-1})}{Var(v_{t+1} + V\lambda_t)}$$

which is independent of bond maturity.

8.18 Proof of Remark 3

Excess return of an n maturity bond is given by

$$rx_{t,t+1}^{n} = -(n-1)y_{t+1}^{n-1} + ny_t^n - y_t^1 =$$
$$-(n-1)(A_{n-1} + B_{n-1}'X_{t+1}) + n(A_n + B_n'X_t) - (A_1 + B_1'X_t).$$

This implies the expected excess return is of the form

$$\mathbb{E}_t[rx^n_{t,t+1}] = \tilde{A}_n + \tilde{B}_n X_t,$$

where

$$\tilde{A}_n = -(n-1)A_{n-1} + nA_n - A_1$$

and

$$\tilde{B}_n = -(n-1)B_{n-1}\phi + nB_n - B_1.$$

Now consider an m dimensional collection of yields $\hat{y}_t$. Note that we have

$$\hat{y}_t = \hat{A} + \hat{B}X_t,$$

where $\hat{A}$ are $\hat{B}$ simply collect the relevant A_n and B_n for the corresponding maturities. If $\hat{B}$ is invertible:

$$X_t = \hat{B}^{-1}(\hat{y}_t - \hat{A})\hat{y}_t.$$

Therefore we have

$$\mathbb{E}_t[rx^n_{t,t+1}] = \tilde{A}_n + \tilde{B}_n\hat{B}^{-1}(\hat{y}_t - \hat{A})\hat{y}_t,$$

now we can write the conditional expectation for the excess return as a linear (affine) function of the yields $\hat{y}_t$. Therefore we can write the excess returns as

$$rx^n_{t+1} = \tilde{A}_n + \tilde{B}_n\hat{B}^{-1}(\hat{y}_t - \hat{A})\hat{y}_t + \varepsilon_{t+1},$$

where ε_{t+1} is independent white noise. Now conditional on the yields $\hat{y}_t$, no other variable like past returns or previous period returns should forecast excess returns.

However, the argument fails if $\hat{B}$ is not invertible. Then controlling for current yields is not generally equivalent to controlling for the factors. Then past bond returns can also predict future returns conditional on the information in the yield curve today.

Remark 3: The Effect of Nonlinearities Remark 3 assumes that yields are an affine function of state variables. However, it can be generalized to arbitrary functions. Now assume yields are of the form

$$y_t^n = g_n(X_t).$$

and that

$$X_{t+1} = \xi(X_t) + \epsilon_{t+1}$$

for some g_n and ξ. We can view this as a generalized Markovian model. Now pick any m yields stacked into a vector $\tilde{y}_t$. Moreover, define $\tilde{g}$ as

$$\tilde{y} = \tilde{g}(X_t),$$

where this function simply collects the relevant elements using g_n. Assuming the inverse exists, we can solve

$$X_t = \tilde{g}^{-1}(\tilde{y}).$$

Now note that we have

$$\begin{gathered} rx_{t,t+1}^n = -(n-1)y_{t+1}^{n-1} + ny_t^n - y_t^1 = \\ -(n-1)g_{n-1}(X_{t+1}) + ng_n(X_t) - g_1(X_t) = \\ -(n-1)g_{n-1}(\xi(X_t) + \epsilon_{t+1}) + ng_n(X_t) - g_1(X_t) \end{gathered}$$

By the definition of a state variable

$$\mathbb{E}_t[rx_{t,t+1}^n] = \mathbb{E}[rx_{t,t+1}^n | X_t] \equiv \Pi_n(X_t) = \Pi_n(\tilde{g}^{-1}(\tilde{y})).$$

Now no other variable should predict excess returns controlling for $\Pi_n(\tilde{g}^{-1}(\tilde{y}))$.

8.19 Proof of Proposition 2

Conjecture $p_t^n = A_n + B_n y_t^1$. Similarly to Hamilton and Wu (2012) then approximate:

$$\mathbb{E}_t[r_{t+1}] \approx -z_{1t} y_t^1 +$$
$$\sum_{n=2}^{N} z_{tn} \left[A_{n-1} + B_{n-1}(c + \rho_1 y_t^1 + \rho_2 y_{t-1}^1) - A_n - B_n y_t^1 - y_t^1 + \frac{1}{2} B(n-1)^2 \sigma_\epsilon^2 \right]$$

and

$$\mathbb{V}ar_t[r_{t+1}] \approx \left(\sum_{n=2}^{N} z_{tn} B_{n-1} \right)^2 \sigma_\epsilon^2$$

Maximizing the arbitrauger's objective for maturity n bond gives:

$$A_{n-1} - A_n + B_{n-1}[c + \rho_1 y_t^1 + \rho_2 y_{t-1}^1] - B_n y_t^1 - y_t^1 = \gamma B_{n-1} \sigma_\epsilon^2 \left(\sum_{n=2}^{N} z_{tn} B_{n-1} \right)$$

Plugging in $z_{tn} = -\chi y_{t-1}^1$

$$A_{n-1} - A_n + B_{n-1}[c + \rho_1 y_t^1 + \rho_2 y_{t-1}^1] - B_n y_t^1 - y_t^1 = -\gamma B_{n-1} \sigma_\epsilon^2 \left(\sum_{n=2}^{N} (\chi y_{t-1}^1) B_{n-1} \right)$$

Set

$$\chi = -\frac{\rho_2}{\gamma \sigma_\epsilon^2 \sum_{n=2}^{N} B_{n-1}}$$

Then we obtain:

$$-A_{n-1} + A_n + B_{n-1}[c + \rho_1 y_t^1] - B_n y_t^1 - y_t^1 = 0$$

We can solve:

$$B_n = B_{n-1}\rho_1 - 1$$

$$A_n = A_{n-1} - B_{n-1}c$$

One can see that $B_n < 0$ and hence $\chi > 0$ assuming $\rho_2 < 0$ as in the data. The time-varying part of expected bond returns is

$$B_{n-1}(\rho_1 y_t^1 + \rho_2 y_{t-1}^1) - B_n y_t^1 = y_t^1 + \rho_2 B_{n-1} y_t^1$$

This is increasing in y_t^1 assuming $\rho_2 < 0$.

References

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, 50:745–787.

Ang, A., Piazzesi, M., and Wei, M. (2006). What does the yield curve tell us about gdp growth? *Journal of Econometrics*, 131(1-2):359–403.

Asness, C., Moskowitz, T., and Pedersen, L. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985.

Bansal, R. and Shaliastovich, I. (2012). A long-run risks explanation of predictability puzzles in bond and currency markets. *Review of Financial Studies*, 26(1):1–33.

Bauer, M. and Rudebusch, G. (2017). Resolving the spanning puzzle in macro-finance term structure models. *Review of Finance*, 21(2):511–553.

Brooks, J., Katz, M., and Lustig, H. (2019). Post-FOMC announcement drift in U.S. bond markets. Working Paper.

Brooks, J. and Moskowitz, T. (2017). Yield curve premia. Working paper.

Chan, L., Jegadeesh, N., and Lakonishok, J. (1996). Momentum strategies. *The Journal of Finance*, 51(5):1681–1713.

Cieslak, A. (2017). Short-rate expectations and unexpected returns in treasury bonds. *Review of Financial Studies*, 31(9):3265–3306.

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

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

Cochrane, J. and Piazzesi, M. (2009). Decomposing the yield curve. Working Paper.

Cook, T. and Hahn, T. (1989). The effect of changes in the federal funds rate target on market interest rates in the 1970s. *Journal of Monetary Economics*, 24(3):331–351.

Coroneo, L., Giannone, D., and Modugno, M. (2016). Unspanned macroeconomic factors in the yield curve. *Journal of Business & Economic Statistics*, 34(3):472–485.

Crump, R. K. and Gospodinov, N. (2021). On the factor structure of bond returns. *Econometrica (forthcoming)*.

Duffee, G. (2011). Information in (and not in) the term structure. *The Review of Financial Studies*, 24(9):2895–2934.

Durham, J. (2013). Momentum and the term structure of interest rates. *FRB of New York Staff Report*, (657).

Fama, E. (1965). The behavior of stock-market prices. *The journal of Business*, 38(1):34–105.

Fama, E. and Bliss, R. (1987). The information in long-maturity forward rates. *American Economic Review*, 77:680–692.

Feunou, B. and Fontaine, J.-S. (2014). Non-markov gaussian term structure models: The case of inflation. *Review of Finance*, 18(5):1953–2001.

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly journal of economics*, 127(2):645–700.

Goyal, A. and Jegadeesh, N. (2018). Cross-sectional and time-series tests of return predictability: What is the difference? *The Review of Financial Studies*, 31(5):1784–1824.

Granziera, E. and Sihvonen, M. (2020). Bonds, currencies and expectational errors. *Bank of Finland Research Discussion Paper*, (7).

Gürkaynak, R., Sack, B., and Wright, J. (2007). The us treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304.

Hamilton, J. and Wu, J. (2012). The effectiveness of alternative monetary policy tools in a zero lower bound environment. *Journal of Money, Credit and Banking*, 44:3–46.

Hong, H. and Stein, J. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6):2143–2184.

Hoogteijling, T., Martens, M., and van der Well, M. (2021). Forecasting bond risk premia using stationary yield factors. Working Paper.

Huang, D., Li, J., Wang, L., and Zhou, G. (2020). Time series momentum: Is it there? *Journal of Financial Economics*, 135(3):774–794.

Hurst, B., Ooi, Y.-H., and Pedersen, L. (2017). A century of evidence on trend-following investing. *The Journal of Portfolio Management*, 44(1):15–29.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, 95(1):161–182.

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

Kim, A., Tse, Y., and Wald, J. (2016). Time series momentum and volatility scaling. *Journal of Financial Markets*, 30:103–124.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of Monetary Economics*, 47(3):523–544.

Lewellen, J. (2002). Momentum and autocorrelation in stock returns. *The Review of Financial Studies*, 15(2):533–564.

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

Liu, Y. and Wu, J. (2020). Reconstructing the yield curve. Journal of Financial Economics, Forthcoming.

Moench, E. and Siavash, S. (2021). What moves treasury yields? Working paper.

Moskowitz, T. J., Ooi, Y., and Pedersen, L. (2012). Time series momentum. *Journal of Financial Economics*, 104(2):228–250.

Neuhierl, A. and Weber, M. (2018). Monetary momentum. NBER working paper no. 24748.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703?708.

Newey, W. and West, K. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):631–653.

Osterrieder, D. and Schotman, P. (2017). The volatility of long-term bond returns: Persistent interest shocks and time-varying risk premiums. *Review of Economics and Statistics*, 99(5):884–895.

Pitkäjärvi, A., Suominen, M., and Vaittinen, L. (2020). Cross-asset signals and time series momentum. *Journal of Financial Economics*, 136(1):63–85.

Rudebusch, G. and Swanson, E. (2012). The bond premium in a dsge model with long-run real and nominal risks. *American Economic Journal: Macroeconomics*, 4(1):105–43.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. NBER Working Paper w4871.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

Vayanos, D. and Vila, J.-L. (2020). A preferred-habitat model of the term structure of interest rates. *Econometrica*.

Vayanos, D. and Woolley, P. (2013). An institutional theory of momentum and reversal. *The Review of Financial Studies*, 26(5):1087–1145.

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

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36.

Xiong, W. and Yan, H. (2010). Heterogeneous expectations and bond markets. *Review of Financial Studies*, 23(4):1433–1466.