Bank of Finland Research Discussion Papers
15 • 2021

Markus Sihvonen

Yield curve momentum

Bank of Finland
Research

Bank of Finland Research Discussion Papers
Editor-in-Chief Esa Jokivuolle

Bank of Finland Research Discussion Paper 15/2021
12 December 2022

Markus Sihvonen:
Yield curve momentum

ISBN 978-952-323-392-8, online
ISSN 1456-6184, online

Bank of Finland
Research Unit

PO Box 160
FIN-00101 Helsinki

Phone: +358 9 1831

Email: research@bof.fi
Website: www.suomenpankki.fi/en/research/research-unit/

Yield Curve Momentum

Markus Sihvonen*

This draft: December, 2022
First draft: October 2020

Abstract

I analyze time series momentum along the Treasury term structure. Yield curve momentum is primarily due to changes in the level factor of yields. 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, which can be explained by a specific form of bounded rationality.

Keywords: Bond risk premia, term structure models, time series momentum, spanning.

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, Mike Gallmeyer (discussant), Rubens Moura, Emanuel Mönch, George Pennacchi, Andrea Polo, Dimitri Vayanos and Michael Weber as well as participants at the Econometric Society European Winter Meeting, Fixed Income and Financial Institutions Conference, 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 the observed return autocorrelation 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.

The term structure literature features a dichotomy between variables that are *spanned* by current yields and *unspanned* variables that contain additional information useful for predicting returns (Duffee, 2011; Joslin et al., 2014). The key empirical contribution of this paper is to argue that past returns are spanned neither by current yields nor previously studied possibly unspanned variables.

While no-arbitrage term structure models can in principle allow for unspanned variables, theoretically motivated models imply full spanning. Therefore my results are problematic for nearly all models attempting to

[1]This is a growing literature, see e.g. Pitkäjärvi et al. (2020) and Zhang (2022).

[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.

explain yield curve momentum.

My investigation starts by establishing basic properties of yield curve momentum. 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 primarily due to *changes* in the first principal component of yields, also known as the level factor. This is even though the *level* of the level factor cannot explain momemtum. The associated strong factor structure in bond returns implies that most of momentum can be captured using a single return factor defined as the average past return of different maturity bonds.

Third, 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 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).

Fourth, I analyze whether past returns are spanned by current yields. Standard term structure models imply that yields are affine in a set of factors, which also determine expected bond returns. But since the yields are a simple function of the factors, controlling for sufficiently many yields is equivalent to controlling for the factors. These models can in principle generate return autocorrelation through autocorrelation in model factors. However, they imply that past returns cannot predict future returns after controlling for yields.

I find that past bond returns predict future returns also conditional on the information in the yield curve today. In fact past returns appear largely orthogonal to current yield curve factors. Hence the spanning condition is clearly violated in the data.

Several papers (e.g. Duffee, 2011; Joslin et al., 2014) have pointed that macro variables related to real activity and inflation also appear unspanned by yields. I therefore show that my results hold when controlling for macro variables including a large panel of such variables as in Ludvigson and Ng (2009). Hence past returns are also unspanned by macro variables.

As mentioned, while reduced form no-arbitrage models can be parametrized to include unspanned variables, theoretically motivated models imply full spanning. This includes macrofinance models characterized through investor preferences. While some of these models imply non-linear relationship between bond returns and past yields, they still imply full spanning after controlling for these non-linear relationships.

The majority of explanations offered for momentum feature deviations from full information rational expectations. Can such behavioral theories explain my findings? Not necessarily. The reason is that the current behavioral models tend to imply the same affine form for yields though the coefficients might be different from those in rational models.

The key modelling contribution of this paper is to i) construct term structure models consistent with my empirical findings, ii) offer a theoretical interpretation for the models. In these models agents do not understand that factors possess longer term dependencies and bond prices reflect this misunderstanding. The mispecification of true factor dynamics leads to a momentum pattern similar to that observed in the data. Moreover, it leads to a violation of the standard spanning condition so that past bond returns predict future returns conditional on standard yield curve factors. Finally, this misinterpretion also explains the forecast errors documented in interest rate surveys.

I explain that my term structure models are consistent with the form of bounded rationality discussed by Molavi (2019) and Molavi et al. (2021). Here an agent can only entertain models with at most a fixed number of factors and chooses a mispecified model that gives a best representation of the data. This constraint on model complexity leads agents to ignore longer dependencies in factors.

2 Data and Definitions

I start by describing the data sources and variable definitions.

Bond Yields and Returns I use the dataset on zero coupon US Treasury yields constructed by Liu and Wu (2021). 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. Data is available for bonds with maturities up to 10 years. 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 denote the monthly continuously compounded yield of a bond (or bill) with n months until maturity by y_t^n. The logarithmic excess monthly return of maturity n bond is then given by

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

Here the excess return is calculated as the monthly bond return deducted by the one month bill rate. The return between month t and any month $t+h$, $rx_{t,t+h}^n$, is given by the sum over the one period excess returns.

To save space, in most of the tables and figures I show results for bonds with integer annual maturities rather than all the 120 different maturities. That is these focus on bonds with maturities $12, 24, \ldots, 120$ months i.e. $1, 2, \ldots, 10$ years. Table 1 shows key descriptive statistics for these yields and excess returns.

Macro Variables and Other Data I construct a large panel of 135 other macroeconomic and financial variables. This contains all the variables in the FRED-MD database of McCracken and Ng (2016). Following Moench and Siavash (2022) I further add 8 variables. This includes the weekly hours of production and non-supervisory employees, the Philadelphia Fed

Yields (%)										
Maturity	1	2	3	4	5	6	7	8	9	10
Mean	5.06	5.31	5.50	5.69	5.82	5.96	6.06	6.14	6.22	6.27
Volatility	3.54	3.49	3.40	3.32	3.23	3.18	3.11	3.06	3.00	2.93
Skewness	0.48	0.40	0.38	0.39	0.39	0.42	0.44	0.45	0.46	0.45
Ex. kurtosis	-0.15	-0.32	-0.36	-0.37	-0.38	-0.37	-0.32	-0.31	-0.28	-0.26
Obs#	580	580	580	580	580	580	580	580	580	580

Excess Returns (%)										
Maturity	1	2	3	4	5	6	7	8	9	10
Mean	0.07	0.10	0.15	0.17	0.20	0.22	0.22	0.25	0.24	0.26
Volatility	0.44	0.83	1.18	1.51	1.80	2.08	2.33	2.61	2.86	3.13
Skewness	1.20	0.54	0.14	-0.12	-0.02	0.09	0.11	0.13	0.11	0.10
Ex. Kurtosis	17.39	13.87	8.21	5.04	4.11	4.11	3.06	2.51	2.25	2.18
Obs#	579	579	579	579	579	579	579	579	579	579

Table 1: shows descriptive statistics for bond yields and excess returns. The data is monthly but the bond yields are expressed in annual form.

leading indicator for the U.S. economy, the VXO index and the measure of realized stock market volatility of Berger et al. (2020). Moreover the data contains the Bank of America Merrill Lynch MOVE bond volatility index, a measure of financial uncertainty from Ludvigson et al. (2021), the excess bond premium from Gilchrist and Zakrajšek (2012) and the three-Month Treasury bill forecast from the Consensus Economics. The measures based on academic papers are extended to cover my sample period accordingly.

I also show results when controlling only for the Chicago Fed National Activity Index, used for example by Joslin et al. (2014), and the trend inflation measure used in Cieslak and Povala (2015). Here I apply a smoothing parameter of 0.987 in monthly updating terms to annual core inflation.[3]

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.

3 Simple Regression Evidence

I first consider a univariate regression of the form

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

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 months $t-h$ 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 2 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.

[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.

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). Moreover, in contrast to Moskowitz et al. (2012), I avert pooled regressions due to potentially biased slope estimates and issues with calculating standard errors (Huang et al., 2020; Hjalmarsson, 2010).

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 3 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 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.

[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 2: 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)

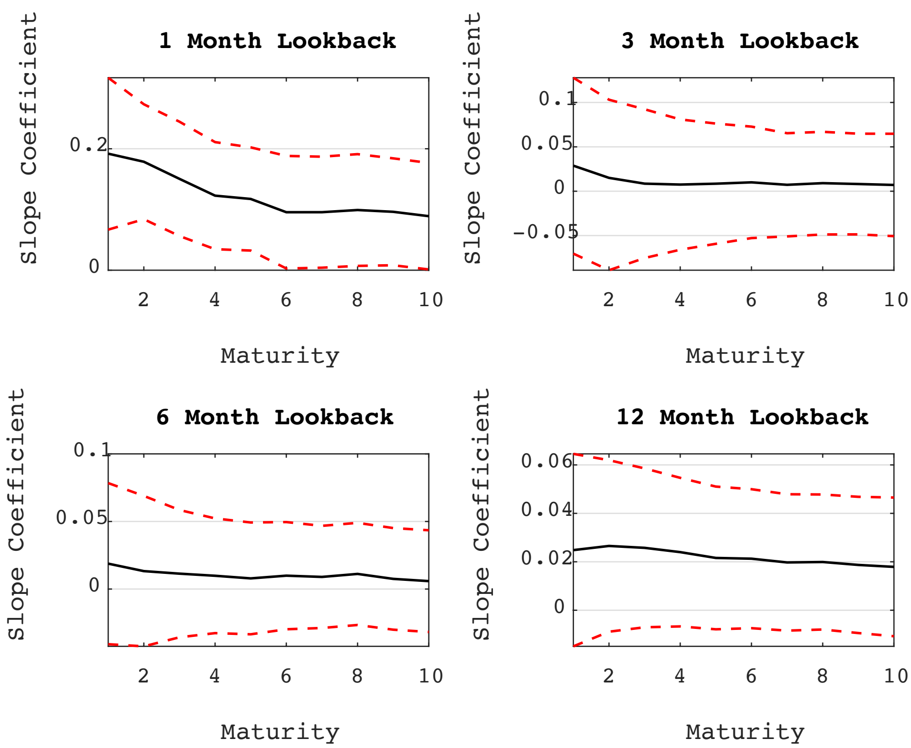

Figure 1: shows the slope coefficients and the relevant 95% confidence bounds from regressing the returns of different maturity bonds (years) on the past return for the same maturity bond for lookback horizons of 1,3,6 and 12 months.

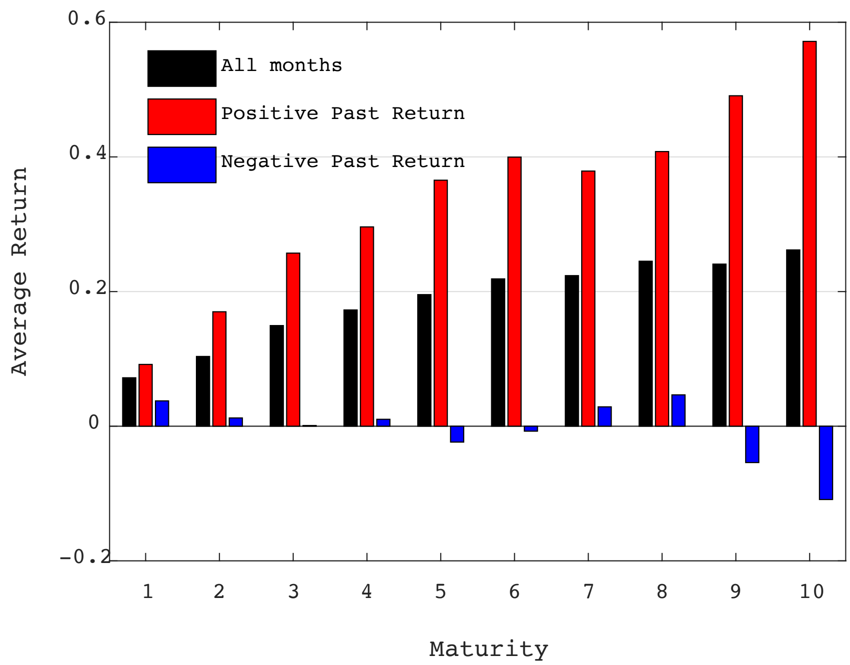

Figure 2: shows the mean returns for different maturity (years) bonds both for the full sample and in subsamples following positive and negative past month returns.

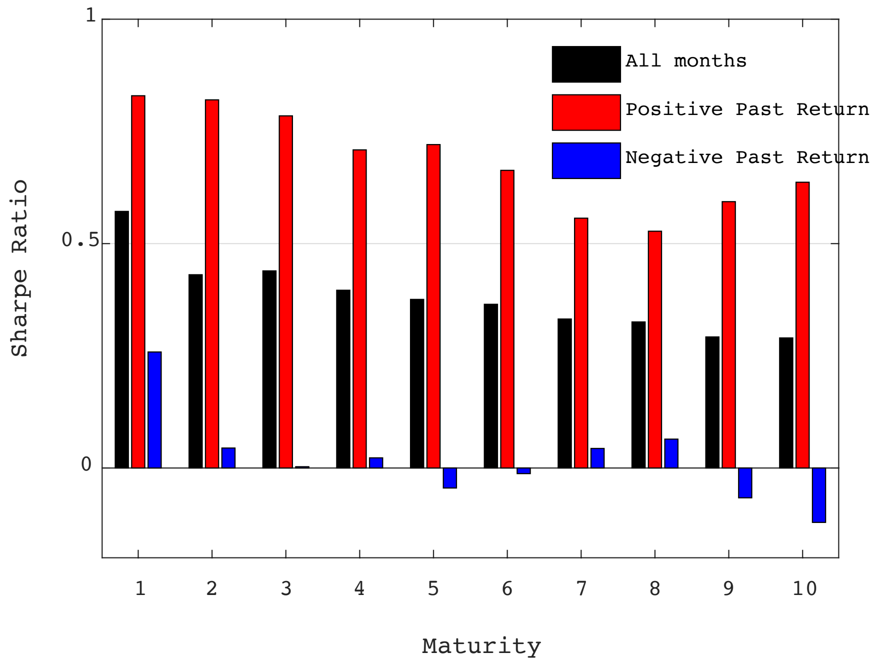

Figure 3: shows the annualized Sharpe ratios for different maturity (years) bonds both for the full sample and in subsamples following positive and negative past month returns.

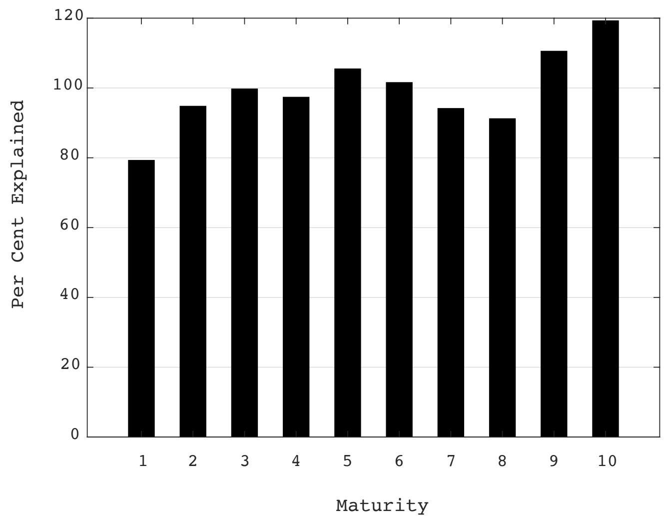

Figure 4: shows the share of total excess returns of different maturity (years) bonds earned in months with positive past month excess returns.

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 (annualized)

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

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.e. I apply the integer annual maturities between 1 and 10 years. 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 4. 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

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

returns based on different maturity bonds.

4 Sources of Momentum

What is driving the results obtained in the previous section? This section derives four key results. First, yield curve momentum is due to a change in the level factor of yields. Second, these level factor changes are not spanned by current yields. Third, results are similar when controlling for macroeconomic variables. Fourth, the change in the level factor of yields also predicts survey based forecast errors concerning interest rates.

4.1 The Effect of Level Changes

To begin note that we can decompose the excess return on a bond as

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

Maturity	$\mathbb{C}ov(c_t^n, c_{t-1}^n)$	$\mathbb{C}ov(c_t^n, yc_t^n)$	$\mathbb{C}ov(yc_{t+1}^n, c_{t-1}^n)$	$\mathbb{C}ov(yc_{t+1}^n, yc_t^n)$
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 5: shows the share of covariance between bond return and past month bond return in per cent accounted by the four channels. Bond maturities are in years.

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

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

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.

Table 5 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. In the appendix I also verify these dependencies using regressions. For Treasuries time series momentum is primarily *yield momentum*.

Given this finding I now revisit the question about whether yield curve momentum can be captured using a single factor. I extract the first principal components using all the 120 maturities between 1 month and 10 years. I then consider the following regression:

$$rx_{t+1}^{n} = \alpha + \beta \Delta pc_t^1 + \epsilon_{t+1} \tag{7}$$

I also decompose return autocovariance to a level change effect and a residual component. In particular consider the contemporaneous projection:

$$rx_t^n = a + b\Delta pc_t^1 + e_t \tag{8}$$

Now we have

$$\mathbb{C}ov(rx_{t+1}^n, rx_t^n) =$$

$$\mathbb{C}ov(\beta\Delta pc_t^1 + \epsilon_{t+1}, b\Delta pc_t^1 + e_t) =$$

$$\mathbb{C}ov(\beta\Delta pc_t^1 + \epsilon_{t+1}, b\Delta pc_t^1) + Cov(\beta\Delta pc_t^1 + \epsilon_{t+1}, e_t) =$$

$$\underbrace{\beta\mathbb{V}ar(\Delta pc_t^1)b}_{\text{Level change effect}} + \underbrace{Cov(\beta\Delta pc_t^1 + \epsilon_{t+1}, e_t)}_{\text{Residual effect}}$$

Here the 3rd line uses the fact that ϵ_{t+1} must be orthogonal to Δpc_t^1. In a standard one factor model the first component would account for 100% per cent of return autocovariance.

The first principal component explains roughly 98.5% of the variation in yields. This component is often called a level factor since it loads fairly evenly on all maturities. 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 bond returns. This high correlation between bond excess returns and changes in the level factor indicates that these level factor changes will also explain a large fraction of yield curve momentum.

Results from the predictability regression and decomposition are given in Table 6. The share of return autocovariance explained by level factor changes is on average 94% and ranges between 81% and 110%. A share of more than 100% implies that the residual component from a contemporaneous projection of returns on the change in the level factor is negatively associated with next month returns. The R^2 statistics in the regressions are of similar magnitude than in the plain momentum regressions, where returns are explained by past returns. Overall, I conclude that the bulk of yield curve momentum is explained by changes in the level factor.

The appendix further considers the predictive power of changes in higher order principal components, finding weak results for most principal components. Also note that while level changes are important predictors of bond returns, as seen from the later tables, the level of the level factor contains only minor predictive information for returns.

4.2 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.

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

$$rx_{t+1}^{n} = A'PC_t + \epsilon_{t+1} \tag{9}$$

		Regression				Decomposition	
Mat.	α	t-value	β	t-value	R^2 (%)	Δpc_t^1 change	Other
1	0.07	4.00	-0.02	-2.47	3.41	86.32 %	13.68 %
2	0.10	3.05	-0.03	-2.86	2.59	86.37 %	13.63 %
3	0.14	3.14	-0.04	-2.57	1.85	88.24 %	11.76 %
4	0.17	2.79	-0.04	-2.29	1.07	82.38 %	17.62 %
5	0.19	2.65	-0.05	-2.22	0.94	80.78 %	19.22 %
6	0.21	2.60	-0.06	-2.39	1.17	109.91 %	-9.91 %
7	0.22	2.37	-0.07	-2.37	1.22	110.25 %	-10.25 %
8	0.24	2.30	-0.07	-2.36	1.21	103.88 %	-3.88 %
9	0.23	2.06	-0.07	-2.23	1.02	96.38 %	3.62 %
10	0.25	2.05	-0.08	-2.09	0.92	96.09 %	3.91 %

Table 6: shows the results of predicting returns of different maturity (years) bonds on the change in the first principal components of yields. It also shows a decomposition of return autocovariance into an effect due to a change in this principal component and a residual component. The t-values are based on Newey and West (1987) standard errors.

Maturity	Spanned	Unspanned
1	7.4 %	92.6 %
2	1.8 %	98.2 %
3	1.3 %	98.7 %
4	0.7 %	99.3 %
5	0.2 %	99.8 %
6	-2.9 %	102.9 %
7	-2.0 %	102.0 %
8	1.2 %	98.8 %
9	1.7 %	98.3 %
10	0.7 %	99.3 %

Table 7: shows the decomposition of covariance between the return of different maturity (years) bonds and their past value into a part spanned by yields and an unspanned part.

$$rx_t^n = B'PC_t + \varepsilon_t \tag{10}$$

The autocovariance in bond returns can then be decomposed to spanned and unspanned parts:[5]

$$\mathbb{C}ov(rx_{t+1}^n, rx_t^n) = \underbrace{A'\mathbb{V}ar(PC_t)B}_{\text{Spanned}} + \underbrace{\mathbb{C}ov(rx_{t+1}^n, rx_t^n) - A'\mathbb{V}ar(PC_t)B}_{\text{Unspanned}} \tag{11}$$

I apply five principal components of yields, including further components has minor effects on the results. The results are given in Table 7. On average only about 1% of the covariance between current and past returns is spanned by yields.

Results from the spanning decomposition above suggest that unspanned variation in returns is important to explaining yield curve momentum. I now

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

mat	1	2	3	4	5	6	7	8	9	10
const.	-0.03	-0.02	-0.03	-0.03	-0.08	-0.19	-0.34	-0.53	-0.61	-0.71
t-value	-0.63	-0.20	-0.21	-0.17	-0.33	-0.69	-1.13	-1.52	-1.58	-1.66
β_1 (rx_t)	0.19	0.18	0.16	0.13	0.12	0.10	0.10	0.10	0.10	0.09
t-value	2.77	3.50	3.01	2.59	2.65	2.08	2.10	2.12	2.09	1.97
pc^1	0.001	0.001	0.002	0.002	0.002	0.001	0.001	0.001	0.0004	0.001
t-value	1.82	1.29	1.15	0.84	0.71	0.52	0.38	0.40	0.12	0.02
pc^2	0.00	-0.01	-0.02	-0.03	-0.04	-0.06	-0.07	-0.08	-0.09	-0.10
t-value	-0.40	-0.82	-1.30	-1.68	-1.85	-2.14	-2.48	-2.74	-2.72	-2.80
pc^3	-0.03	-0.08	-0.11	-0.15	-0.17	-0.15	-0.15	-0.15	-0.13	-0.13
t-value	-1.39	-2.43	-2.69	-2.83	-2.57	-2.00	-1.74	-1.50	-1.19	-1.04
pc^4	-0.12	-0.20	-0.25	-0.29	-0.33	-0.38	-0.53	-0.71	-0.86	-1.03
t-value	-2.11	-1.88	-1.73	-1.58	-1.54	-1.63	-2.08	-2.46	-2.74	-2.96
pc^5	-0.26	-0.33	-0.42	-0.60	-0.91	-1.26	-1.44	-1.41	-1.15	-0.91
t-value	-2.24	-1.60	-1.53	-1.76	-2.12	-2.21	-2.32	-2.21	-1.78	-1.38
R^2 (%)	8.7	7.2	6.2	5.7	5.9	5.8	6.4	6.6	6.0	5.9

Table 8: shows the results of predicting returns of different maturity (years) bonds on the past return of the bond and the first five principal components of yields. The t-values are based on Newey and West (1987) standard errors.

test this result more formally by including the first five principal components into the predictive regression shown in Table 1. The results are given by in 8. The table suggests that the past return is still significant and numerical values of the corresponding slope coefficient similar to before. These results confirm that past returns are largely unspanned by current yields. The appendix provides additional support that this holds when including more yields to the predictability regression as well as controlling for potential non-linearities.

4.3 Controlling for 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, 2022). 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. I next argue that my results are valid also when controlling for information in macroeconomic data.

To control for a large set of macro variables, I first follow an approach similar to that in Ludvigson and Ng (2009). Consider predicting bond returns using a factor model of the form:

$$m_{it} = \lambda' f_t + e_{it}$$
$$rx^n_{t+1} = \alpha' M_t + \beta' Z_t + \epsilon_{t+1},$$

where $M_t \subset f_t$. Here I posit that each macroeconomic variable m_{it} is driven by a smaller set of common factors f_t. These common macroeconomic factors then predict bond returns along with other variables Z_t.

The factors are extracted using principal component analysis.[6]The number of macro factors is determined using the IC2 criterion of Bai and Ng (2002). The criterion suggests that the macro data is well described by 7 common factors, which explain about 44% of the variation in the data.

The optimal combination of estimated factors $\hat{M}_t$ is determined using the BIC criterion. Following Ludvigson and Ng (2009), I also consider squares and cubes of the factors. Note that the optimal factors are generally different for different maturity bonds. I control for past bond returns and the first five principal components of yields, that is $Z_t = [rx^n_t, pc^1_t, pc^2_t, pc^3_t, pc^4_t, pc^5_t]$. This is in contrast to Ludvigson and Ng (2009) who only control for the Cochrane-Piazzesi factor. Note that as in Moench and Siavash (2022) my macro panel also includes additional macroeconomic and financial variables.

The results are given in Table 9. The set of selected factors includes the first, third and seventh principal components of the macro data. It also containts the squares of the first and seventh principal components as well as the cube of the fourth principal component. The chosen factors

[6]Missing observations are handled using the expectation maximization algorithm suggested by Stock and Watson (2002).

vary somewhat among the differerent maturities as can be seen from the table. The past returns remain significant for maturities between 1 and 5 years though less so for the longer maturity bonds. Interestingly for short maturities, the momentum slope coefficients are also larger than before.

I also examined a version of the algorithm where the set of possible macro factors includes the lags of the first seven principal components of the macro variables. For one year bonds, the algorithm includes the lags of the first and second principal components but the coefficient and t-value on the past return is similar to before. The lags are not selected for the longer maturities so the results are exactly as before.

Principal components of macroeconomic variables lack an obvious economic interpretation. Therefore I now also show results when instead controlling for trend inflation and the activity index. I also include lags of these macro variables and as well as the five principal components of yields.

The results are given in Table 10. The slope coefficients for past returns are clearly significant and of similar magnitude than before. Trend inflation and its lag is also significant. Moreover, the activity index is also significant though its lag only weakly so.

I conclude that accounting for macro variables does not alter the main results of this paper though these variables possibly represent additional unspanned information useful for predicting returns. However, in some specifications the predictive content of past returns of bonds with maturities greater than five years appears weaker than before, while the predictive of past returns of short maturity bonds can be higher.

4.4 Short Rate Forecast Errors

Forecast errors should be unpredictable under rational expectations. A large literature has instead documented that survey forecasts exhibit systematic biases. For example Coibion and Gorodnichenko (2015) associate these biases to forecast revisions. Cieslak (2017), Schmeling et al. (2022) and Granziera and Sihvonen (2021) further relate survey based short rate forecast

Mat	1	2	3	4	5	6	7	8	9	10
const	0.06	0.13	0.20	0.26	0.26	0.17	0.04	-0.10	-0.14	-0.21
t-value	1.00	1.21	1.35	1.33	1.10	0.62	0.13	-0.30	-0.38	-0.49
rx_t	0.23	0.26	0.18	0.12	0.11	0.09	0.08	0.07	0.07	0.07
t-value	2.34	3.14	2.58	2.09	2.09	1.61	1.52	1.44	1.44	1.40
pc^1	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001
t-value	1.49	1.39	1.58	1.09	0.99	0.86	0.77	0.80	0.43	0.26
pc^2	-0.01	-0.02	-0.03	-0.05	-0.06	-0.07	-0.09	-0.10	-0.11	-0.12
t-value	-1.96	-2.56	-2.65	-3.09	-3.18	-3.28	-3.43	-3.62	-3.58	-3.62
pc^3	-0.07	-0.18	-0.25	-0.31	-0.34	-0.33	-0.33	-0.34	-0.34	-0.35
t-value	-3.45	-4.22	-4.24	-4.19	-3.99	-3.43	-3.11	-2.82	-2.53	-2.33
pc^4	-0.08	-0.16	-0.17	-0.19	-0.22	-0.27	-0.43	-0.60	-0.72	-0.88
t-value	-1.70	-1.92	-1.45	-1.23	-1.23	-1.32	-1.87	-2.31	-2.63	-2.85
pc^5	-0.23	-0.32	-0.40	-0.55	-0.85	-1.21	-1.39	-1.36	-1.08	-0.83
t-value	-2.05	-1.47	-1.36	-1.53	-1.94	-2.09	-2.23	-2.10	-1.67	-1.26
$\hat{M}_1$	-0.02	-0.05	-0.07	-0.09	-0.10	-0.11	-0.13	-0.14	-0.14	-0.15
t-value	-3.95	-5.74	-6.18	-6.08	-5.67	-5.27	-5.17	-5.21	-5.07	-4.77
$\hat{M}_3$	-0.02			-0.05	-0.06	-0.07	-0.09	-0.11	-0.13	-0.14
t-value	-3.18			-3.11	-3.52	-3.54	-3.60	-3.77	-3.82	-3.77
$\hat{M}_7$	-0.03		-0.06	-0.09	-0.12	-0.14	-0.15	-0.18	-0.21	-0.22
t-value	-2.50		-2.47	-2.98	-3.31	-3.41	-3.54	-3.79	-3.91	-3.68
$\hat{M}_1^2$		0.00		0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
t-value		-2.85		-3.45	-3.77	-3.83	-3.89	-3.86	-3.60	-3.25
$\hat{M}_7^2$	-0.01		0.00							
t-value	-2.58		-3.08							
$\hat{M}_4^3$	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003
t-value	3.20	3.20	3.28	3.29	3.12	2.91	2.67	2.53	2.57	2.74
R^2 (%)	25.7	21.5	19.7	18.7	18.1	16.2	15.7	15.8	15.4	15.0

Table 9: shows the results of predicting returns of different maturity (years) bonds on the past return of a same maturity bond, the first five principal components of yields as well as the optimal set of macroecomic predictors $\hat{M}_t$. The t-values are based on asymptotic standard errors (Bai and Ng, 2006) with the Newey and West (1987) correction.

Mat	1	2	3	4	5	6	7	8	9	10
cons	0.04	0.13	0.21	0.28	0.30	0.25	0.17	0.07	0.04	0.01
t-value	0.63	1.25	1.38	1.41	1.25	0.89	0.55	0.20	0.10	0.03
rx_t	0.13	0.15	0.13	0.11	0.11	0.09	0.10	0.10	0.10	0.10
t-value	2.04	2.97	2.63	2.28	2.42	1.92	2.10	2.16	2.18	2.14
Inf	-3.79	-6.44	-8.64	-10.73	-12.88	-15.66	-17.03	-18.51	-18.23	-18.41
t-value	-2.97	-2.95	-2.94	-2.83	-2.88	-2.96	-2.94	-2.98	-2.74	-2.56
Inf lag	3.73	6.26	8.34	10.31	12.37	15.07	16.29	17.63	17.22	17.24
t-value	2.95	2.88	2.85	2.72	2.77	2.85	2.81	2.83	2.59	2.39
Act	-0.38	-0.59	-0.78	-0.93	-1.04	-1.16	-1.23	-1.27	-1.37	-1.32
t-value	-3.25	-3.38	-3.31	-3.13	-3.01	-2.63	-2.51	-2.36	-2.30	-1.94
Act lag	0.24	0.36	0.47	0.58	0.66	0.74	0.79	0.80	0.90	0.81
t-value	1.93	1.88	1.79	1.76	1.71	1.55	1.48	1.38	1.42	1.14
pc^1	0.00	0.01	0.02	0.02	0.03	0.03	0.03	0.04	0.05	0.05
t-value	2.33	2.92	3.29	3.39	3.42	3.40	3.69	3.85	3.94	4.14
pc^2	0.00	-0.01	-0.03	-0.05	-0.06	-0.07	-0.09	-0.11	-0.13	-0.15
t-value	-0.02	-0.85	-1.50	-1.95	-2.08	-2.22	-2.69	-3.04	-3.27	-3.51
pc^3	-0.03	-0.08	-0.11	-0.14	-0.15	-0.13	-0.12	-0.11	-0.09	-0.07
t-value	-1.85	-2.51	-2.62	-2.67	-2.35	-1.76	-1.43	-1.13	-0.78	-0.58
pc^4	-0.10	-0.17	-0.21	-0.25	-0.29	-0.33	-0.48	-0.66	-0.81	-0.99
t-value	-1.83	-1.61	-1.51	-1.39	-1.38	-1.44	-1.91	-2.33	-2.66	-2.94
pc^5	-0.32	-0.40	-0.49	-0.66	-0.97	-1.36	-1.51	-1.46	-1.12	-0.81
t-value	-2.61	-1.77	-1.60	-1.72	-2.03	-2.16	-2.23	-2.11	-1.62	-1.15
R^2 (%)	14.0	11.7	10.7	10.0	10.1	10.0	10.7	11.0	10.4	10.1

Table 10: shows the results of predicting the returns of different maturity (years) bonds on the past return of the bond, trend inflation, the national activity index, the past month values of these two macroeconomic variables and the first five principal components of yields. The t-values are based on Newey and West (1987) standard errors.

errors to bond and money market predictability patterns. However, these papers do not associate the results to those in the momentum literature nor measure biases when empirically controlling for all the current information in the term structure of interest rates.[7]

I measure short rate expectations using the three month Treasury-bill forecast from Consensus economics. The forecast horizon is 3 months and the data begins in October 1989. I calculate realized forecast errors as $y^3_{t+3} - \mathbb{E}^S_t[y^3_{t+3}]$, where $\mathbb{E}^S_t[y^3_{t+3}]$ is the survey forecast. I explain these forecast errors with the (time t) first 5 principal components of yields as well as the lag of the first principal component of yields. The results are given in Table 11.

Both the first principal component of yields and its lag are significant predictors of forecast errors. The sign of the level factor is positive but that of its lag is negative. Since these coefficients are of similar magnitude, the *change* in the level factor is an important predictor of forecast errors. That is a recent positive change in the level of interest rates is associated with forecaster underpredicting future short rates relative to rational expectations.

These results also suggest that a variable unspanned by the information in the current term structure of interest rates, namely the lag of the level factor, is important to predicting short rate forecast errors. Later I show that these observations are consistent with my theoretical pricing model.

5 Momentum and Post-FOMC Announcement Drift

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

[7]While Cieslak (2017) does not control for yield curve principal components, her theoretical model in Cieslak (2017) does account for unspanned macroeconomic variables. Note that Cieslak (2017) concentrates on explaining predictability on a quarterly and annual horizon, unlike this paper which focuses on more short horizon monthly predictability.

cons	0.28
t-value	2.46
pc^1	0.03
t-value	2.84
pc^2	-0.002
t-value	-0.24
pc^3	-0.16
t-value	-3.75
pc^4	-0.37
t-value	-4.14
pc^5	-0.24
t-value	-1.56
pc^1 lag	-0.03
t-value	-3.01
R^2 (%)	35.1

Table 11: shows the results of explaining forecast errors concerning 3 month rates, 3 month ahead using the current first five principal components of yields as well as the lag of the first principal component of yields. The t-values are based on Newey and West (1987) standard errors.

sluggishly to changes in the federal funds target rate.[8]

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 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

$$rx_{t+1}^{n} = \alpha + \beta \Delta FFTR_{t} + \epsilon_{t+1} \tag{12}$$

$$rx_{t+1}^{n} = \alpha + \beta \Delta UEFFTR_{t} + \epsilon_{t+1}. \tag{13}$$

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 Cook and Hahn (1989) and Kuttner (2001) except that I consider the past rather than the contemporaneous change in the policy rate.[9]

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.

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

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

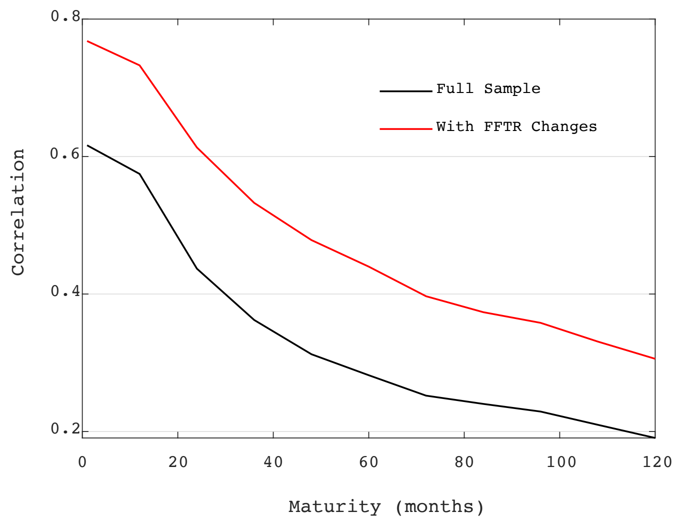

Figure 5: shows the correlation between the change in the federal funds target rate (FFTR) and the change in the yield of different maturity (in months) bonds in two subsamples: full and months with non-zero FFTR changes.

	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 (years) bonds 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.

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.

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:

$$rx_t^n = a + b\Delta FFTR_t + e_t. \tag{14}$$

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:

$$\mathbb{C}ov(rx_{t+1}^n, rx_t^n) = \underbrace{\beta \mathbb{V}ar(\Delta FFTR_t) b}_{\text{FFR effect}} + \underbrace{\mathbb{C}ov(rx_{t+1}^n, e_t)}_{\text{Other}}, \tag{15}$$

where β is the slope coefficient from a regression of bond returns on past month change in the target rate. 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

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 (years) bonds and their past value into a part explained by change in the federal funds target rate and a residual component.

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. It is seen that the violation of the spanning condition implies strong restrictions for such a model.

For generality, and similarly to Piazzesi et al. (2015), consider three probability measures. $\mathbb{P}$ represents objective probabilities as viewed by a rational econometrician. For simplicity I omit this $\mathbb{P}$ symbol from expectations taken under rational beliefs. $\mathbb{S}$ expresses subjective beliefs of a representative agent. Finally, $\mathbb{Q}$ is a pricing measure defined below.

Assume that the state perceived important for determining bond prices is an $m \times 1$ dimensional factor X_t^s. This may generally be different from the $m^e \times 1$ true state vector in the economy X_t, which can in particular include additional factors $X_t^s \subset X_t$, $m \leq m^e$. Under the subjective measure the factor X_t^s follows:

$$X_t^s = \mu^s + \phi^s X_{t-1}^s + v_t, \tag{16}$$

where v_t is multivariate Gaussian $v_t \sim N(0, V)$. I assume that under the objective measure, the true state of the economy also follows a Gaussian VAR-model with coefficients μ and ϕ.

The log nominal discount factor, expressed under the subjective measure, is a linear function of the subjective factors

$$\begin{aligned} \mathcal{M}_{t+1} &= \exp\left(-\delta_0 - \delta_1' X_t^s - \frac{1}{2}\lambda_t' V \lambda_t - \lambda_t' v_{t+1}\right) \\ \lambda_t &= \lambda_0 + \lambda_1 X_t^s \end{aligned} \tag{17}$$

I can then solve bond prices recursively using

$$p_t^1 = \log \mathbb{E}_t^S(\mathcal{M}_{t+1}) \tag{18}$$

$$p_t^n = \log \mathbb{E}_t^S(\mathcal{M}_{t+1} \exp(p_{t+1}^{n-1})). \tag{19}$$

In this model prices and yields take a standard affine form.

$$p_t^n = A_n + B_n' X_t^s \tag{20}$$

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

$$\phi^* = \phi^s - V\lambda_1 \tag{21}$$

$$\mu^* = \mu^s - V\lambda_0 \tag{22}$$

These risk neutral parameters define the pricing measure $\mathbb{Q}$ under which the factor X_t^s follows a VAR-process with modified parameters. This pricing measure is equivalent to the subjective measure $\mathbb{S}$.

General Model Restrictions 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 either needs a model with a time-varying subjective risk premium, a model with non-rational beliefs $\mathbb{S} \neq \mathbb{P}$ or a model with both channels.

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

If $\lambda_1 = \mathbf{0}$, all subjective risk premia are constant. When beliefs are rational, so is the objective risk premium i.e. there is no return predictability. Now also the best forecast of future excess returns is a constant so the slope coefficient in the momentum regression would be zero.

However, this argument does not hold once we relax full information rational expectations, $\mathbb{S} \neq \mathbb{P}$. Now subjective risk premia can be zero yet variables such as past returns predict bond returns under the true measure due to expectational errors. My theoretical pricing models feature both channels. Here expectational errors is the primary cause of momentum consistently with forecast biases documented in surveys. However, risk premia explain the average slope of the yield curve as well as the predictive performance of the levels of standard yield curve factors.

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 $m = 1$.*

Proof: see appendix.

This result is related to the 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.

To build intuition, in the theoretical part, I first consider a single factor model, which explains momentum but does not generate the downward sloping pattern for momentum betas. However, I then consider a more general multi-factor model which also accounts for this pattern.

My empirical results suggest that momentum should be explained by a model in which past returns are not spanned by information in current yields. This observation implies strong restrictions for term structure 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, this spanning condition can be violated in cases in which an invertibility condition fails. Denote by $B_{n,e}$ the yield loadings on the true state variable X_t, that is extend B_n by zeros for all state variables not considered by the agents, $B_{n,e} = [B_n, \mathbf{0}']$. Formally, we have the following result

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

Proof: see appendix.

Intuitively if we pick any combination of yields, we cannot back out the true state variable X_t. If such an invertibility condition fails, controlling for the yields is generally not equivalent to controlling for the factors. In standard unspanned models invertibilty fails because some state variable relevant for forecasting takes a zero yield loading. Now some factors can predict returns and yield changes but not be priced in the current term structure of yields. In the theoretical pricing models of this paper $\mathbb{P} \neq \mathbb{S}$ and the spanning condition fails because agents ignore longer lags of a state variable important for determining bond prices.

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. While reduced form no-arbitrage models can in principle be parametrized to knife-edge cases in which the invertibility condition in Remark 3 fails, standard theoretically motivated models tend to assume full spanning.[10] I next discuss some of these models:

[10]While the theoretical models are generically spanned, one could also parametrize them to imply a violation of the spanning condition. However, I am not aware of any such

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 true economic state variable. 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.[11]

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

exercise. Here one would have to give an explanation for such knife-edge restrictions on model parameters.

[11]There are also some non-linear reduced form term structure models such as Feldhütter et al. (2018). Here my argument remains valid when controlling for the non-linear relationships between future returns and yields.

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 (2021) 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.[12] Hence the model naturally generates unconditional momentum. In this model biased beliefs enter as new state variables but again bond prices are affine in the true state variable, 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. I initially consider a reduced form no-arbitrage setting but later offer a theoretical interpretation for the approach in Section 6.4. For intuition I start with a simplified example and then move to a more realistic estimated multi-factor term structure model.

Simple Example

Consider a one factor model as in for example Vasicek (1977). However,

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

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 generally depend on both the current value of the state variable x_t and its lag x_{t-1} that is the true state variable is $X_t = [x_t, x_{t-1}]$.

However, assume the second lag is not priced that is under the risk neutral pricing measure the factor follows an AR(1) process. This occurs if the process is considered to be AR(1) under the subjective measure $\mathbb{S}$ so that $X_t^S = x_t$. However, it can also happen when $\mathbb{P} = \mathbb{S}$ under knife-edge restrictions on model parameters, though I later argue in favor of the first interpretation.

A finite factor term structure model typically cannot price all yields perfectly. A common approach to estimating such models is to assume that only some of the yields are priced or observed without error (Hamilton and Wu, 2012b). I estimate the risk parameters by assuming that 5 year yields are priced perfectly. Moreover, I estimate (δ_0, δ_1) directly using OLS. The key predictability results hinge on a single parameter, the risk neutral persistence of the factor. I obtain $\rho^Q \approx 0.9999$. The corresponding market price of risk parameters could be solved from equations (22) and (21) but are not relevant for the exercise.

I estimate the true factor dynamics and find significant persistence parameters of 1.077 for the first lag and −0.088 for the second. For comparison fitting AR(1) factor dynamics would result in a persistence parameter of 0.989 under the real measure.

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 0.11, that is the model is able to generate momentum and this coefficient is of similar magnitude to that in the data (0.12). However, 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 0.10. 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.

A Modified ACM Model

I next consider a more realistic term structure model with five principal component factors similar to that in Adrian et al. (2013). That is let $X_t^s = [pc_t^1, pc_t^2, pc_t^3, pc_t^4, pc_t^5]$. Since the model features a large set of parameters I estimate it using linear regressions as proposed by Adrian et al. (2013).

Now consider the following twist. Assume that under $\mathbb{P}$, and as suggested by the data[13], the first principal component of yields pc^1 depends also on its second lag. That is the true state variable is $X_t = [pc_t^1, pc_t^2, pc_t^3, pc_t^4, pc_t^5, pc_{t-1}^1]$. These factor dynamics can be represented in a VAR(1) model in companion form with a coefficient matrix

$$\phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} & \phi_{16} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} & 0 \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & \phi_{25} & 0 \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & \phi_{25} & 0 \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & \phi_{55} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{23}$$

I solve for the momentum betas and conditional momentum betas when controlling for the first five principal components of yields by simulating

[13]See the appendix for the he estimation results when controlling for all the five principal components.

the model under the above $\mathbb{P}$ dynamics, that is when the first principal component depends also on its second lag. Figure 6 shows the resulting momentum betas along with those measured from the data. Overall one can see that the model is able to replicate yield curve momentum in the data fairly accurately. The model also matches variation in yields reasonably well. Here it implies a root mean squared error of 1.8bps in annual terms. As documented in the appendix, the model additionally captures the predictive ability of the levels of the yield curve factors.

Figure 6 also shows the momentum betas implied by the standard ACM model, where the true factor dynamics are simulated assuming the level factor does not depend on its second lag. This model implies only mild autocorrelation in returns. Moreover, the conditional momentum betas are exactly zero since the model satisfies full spanning.

Adrian et al. (2013) also generalize their estimation approach to cover unspanned variables though do not apply the method to unspanned lags. As in Joslin et al. (2014), here the violation of full spanning occurs due to knife-edge restrictions on model parameters. Note that I instead estimate my model as if the standard ACM model is correct but simulate the model under different, more general and realistic dynamics for the state variables. This is consistent with my bounded rationality interpretation of the model $\mathbb{P} \neq \mathbb{S}$, discussed later, whereas in the approach of Adrian et al. (2013) the agent effectively understands the true dynamics, $\mathbb{P} = \mathbb{S}$.

Alternatively, in the standard approach to unspanned variables, the change to the risk neutral measure $\mathbb{Q}$ is done from the true dynamics $\mathbb{P}$ whereas I instead make the change from the subjective measure $\mathbb{S} \neq \mathbb{P}$. These two approaches lead to a different covariance matrix of shocks. Accounting for the lag in the level factor leads to a smaller estimate of level factor residual variance.

The two approaches, however, lead to the same yield loadings B_n and therefore have identical predictions for momentum betas. The constant terms of bond prices A_n are instead generally different though in my case this difference is numerically small.

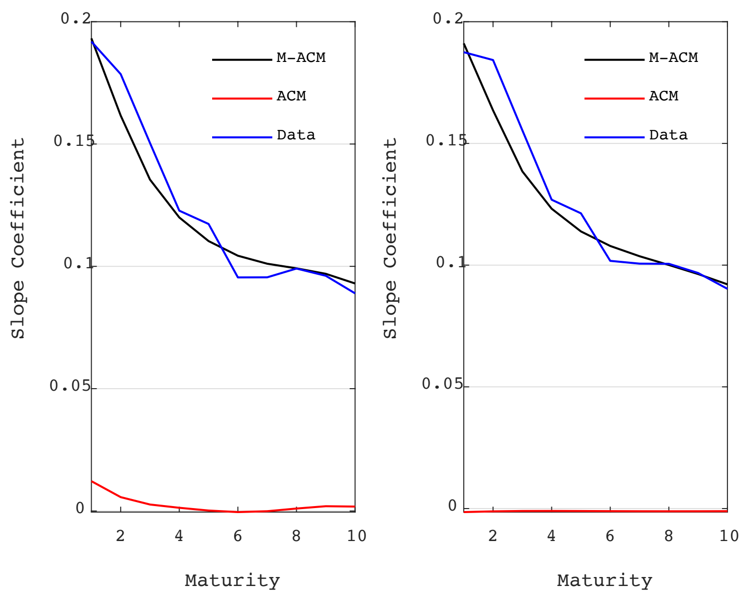

Figure 6: shows the plain (left) and conditional (right) momentum coefficients observed in the data and those implied by a modified ACM model (M-ACM), where the model is simulated under a true model in which the level factor also depends on its second lag. It also shows the coefficients implied by the standard ACM simulated ignoring this second lag. Maturity is measured in years.

Note that 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.

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 the standard 5-factor ACM model and simulate it under the assumption that the postulated dynamics are correct. Similarly to Duffee (2011), Cieslak (2017) and Bauer and Rudebusch (2017), I introduce a normally distributed noise term to yields that is independent across maturities. I set the volatility of the error to a conservative value of 10bps annually. This is higher than the value employed by Bauer and

[14] Apparent failures of the Markov property appear also in models with slow updating as in Granziera and Sihvonen (2021). However, these models can be recast in Markov form.

Rudebusch (2017) (5.8bps) and also higher than the yield measurement error found by Liu and Wu (2021).

The results are given in Figure 7, which shows the simulated 5% (two-sided) critical values for the conditional betas. While the model implies full spanning and hence zero conditional betas, measurement error can in principle explain positive observed betas. However, the betas measured from the data are clearly above the critical values so that the momentum in the data is larger than can reasonably be accounted by measurement error. In the appendix I further show that introducing noise to the term structure model of Cieslak and Povala (2015) does not explain my findings.

6.4 A Bounded Rationality Interpretation

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? I next argue that my results are consistent with the form of bounded rationality discussed by Molavi (2019) and Molavi et al. (2021)[15].

Molavi (2019) considers a form of model misspecification in which an agent can only entertain factor models with at most d factors, where d represents the agent's sophistication. On the other hand, the agent can consider any cross-sectional relationship between model variables. Therefore the approach captures the difficulty in dealing with time-series complexity.

More formally, the agent can only hold beliefs over the set of d-factor models:

[15] Molavi et al. (2021) study the asset pricing implications of this mechanism and also discuss an application to equity, but not yield curve momentum.

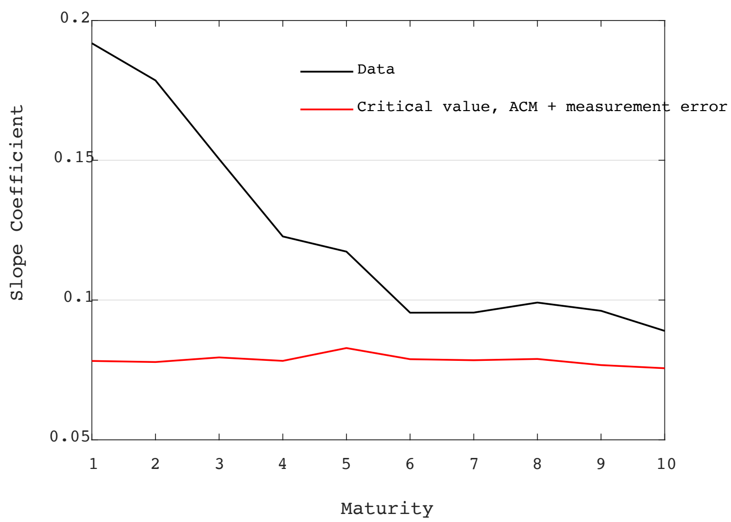

Figure 7: shows the 5% (two-sided) critical values for the conditional momentum betas obtained by introducing measurement error to the yields obtained by simulating a standard 5-factor ACM model. It also shows betas measured from the data. Maturity is expressed in years.

$$F_t = \bar{\mu} + \bar{\phi} F_{t-1} + \epsilon_t$$
$$X_t^s = C' F_t$$

Here $\bar{\mu} \in \mathbb{R}^{d,1}, C \in \mathbb{R}^{d,m}$ and $\bar{\phi} \in \mathbb{R}^{d,d}$.[16] The factor might represent for example a deep macroeconomic state that is not directly observed by the econometrician. How does the agent choose the parameters $\theta = (\bar{\mu}, \bar{\phi})$? Define the Kullback–Leibler (KL) divergence of model as:

$$KL(\theta) = \mathbb{E}[-\log f^{\theta}(X_{t+1}^s | X_t^s, \ldots)] - \mathbb{E}[-\log f(X_{t+1}^s | X_t^s, \ldots)] \tag{24}$$

Here f is the true density, f^{θ} is the agent's misspecified density and the expectations are taken under the true measure. Molavi et al. (2021) show that, among the d-factor models with a positive prior probability, the agent's posterior beliefs concentrate on the subset of models that have minimum KL divergence relative to the true data-generating process. This result is in line with the literature on learning under model mispecification (Berk, 1966).

Similarly to Molavi et al. (2021), I focus on the case where the agent can only entertain single factor models, $d = 1$. I next argue that this model gives an economic justification to our previous 1-factor term structure model.

Assume the agent can observe the factor without error and has monitored a long history.[17] Minimization of the KL-divergence amounts to a population maximum likelihood estimate of the factor dynamics, which further coincides with a population OLS estimate. Effectively, our agent learns the factor dynamics that best represent the true dynamics under a mispecified model.

Assume the representative agent's risk preference is represented by stochastic discount factor of the form 17. The bond prices are then solved using the standard price recursions. Here the state variable is one dimensional since the agent puts zero probability on models of higher dimension.

[16]Molavi (2019) defines the model without a constant and using a scaling matrix on the error term but this formulation is equivalent.

[17]Without loss of generality $c = I$.

However, in the appendix I explain that the key results do not rest on the form of the stochastic discount, and could also be derived for example in the case when the agent possesses mean-variance preferences.

This framework and the assumption that the true factor dynamics are AR(2) gives an exact justification for the one factor unspanned term structure model discussed in Section 6.2. Effectively the agent cannot consider the second lag of the factor as a state variable and views the process as an AR(1). Estimation of the model can now proceed as before.

The argument could be generalized to higher dimensions for example to justify the multifactor unspanned term structure model in this paper. Assume the agent is constrained to entertain models with at most d factors. On the other hand assume an econometrician observes that yield levels must be described by at least d factors. Since the representative agent correctly understands all cross-sectional relationships between variables, these factors must also be taken into account in pricing. Now as the agent has exhausted its available sophistication, it cannot consider lagged values of any of these factors as a state variable, that is $m = d$ and these lags must be ignored in pricing.

This explanation, when applied to yield dynamics, has similarities with the narrative discussed by Cieslak (2017). She argues that full spanning 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.

As explained by Duffee (2011) unspanned macroeconomic variables can in some cases be consistent with rational expectations $\mathbb{S} = \mathbb{P}$. This can occur for example if some macroeconomic variable has opposing effects for short rate expectations and risk premia and these two effects happen to net out exactly. However, it is unclear why such netting would occur for the lag of interest rate levels but not for their contemporaneous value. Such rational expectations explanations of unspanned variables are also inconsistent with the expectational errors documented in interest rate surveys. I next argue that my interpretation is instead coherent with the errors observed in the data.

Matching Short Rate Forecast Errors

Section 4.4 showed that forecast errors concerning short rates are predictable. This is inconsistent with rational expectations explanations of yield dynamics $\mathbb{S} = \mathbb{P}$. Moreover, the lag of the level factor predicts these errors conditional on the current information in the yield curve. This is incoherent with standard behavioral models (see e.g, Granziera and Sihvonen, 2021), which predict that the forecast error predictability could be captured using current information in the yield curve.

I next show that the observed errors are instead consistent with the unspanned term structure models presented in this paper and the interpretation that $\mathbb{S} \neq \mathbb{P}$. First, consider the simple modified Vasicek model. Assume the agents perceive the single factor dynamics to be correct that is they use a distorted process to forecast the state variable. I simulate the implied forecast errors by assuming that under the true measure $\mathbb{P}$ the factor depends also on its second lag.

I replicate the predictability regression in Table 11. Here the simple model predicts a value of 0.012 for the first principal component and -0.13 for its lag. The emprical values are 0.030 and -0.032. The model captures the fact that these coefficients are of similar magnitude but opposite in sign. However, the coefficients are of smaller absolute size than those in the data. However since this is effectively a one factor model, we cannot control for higher principal components so the exercise is not directly comparable to that in Table 11.

Now repeat the exercise but using the general modified ACM model. Now I can also control for all the five principal components of yields. The model implies a value of 0.029 for the first principal component and -0.030 for its lag. Comparing these to the empirical values of 0.030 and -0.032, this model replicates the short rate forecasts errors quite accurately. Note that this survey data has not been in any way targeted in model estimation.

7 Conclusion

Yield curve momentum cannot be explained by yield curve factors, as predicted by standard models. Moreover, it cannot be captured by unspanned macroeconomic variables and therefore represents an independent source of predictability. However, I show that the data is consistent with a term structure model in which agents ignore longer term dependencies in model factors.

This paper bears important implications to three strands of literatures. First, by showing that past returns are an economically important yet unspanned source of predictability, it contributes to the literature on bond return predictability and risk premia. Second, it shows how momentum can be incorporated to a standard no-arbitrage setting, a useful addition to the term structure modelling literature.

Finally, the paper contributes to the literature attempting to provide a theoretical explanation for momentum. When applied to government bonds, the standard theories tend to make predictions clearly violated in the data. My results indicate that momentum is best explained by the form of bounded rationality discussed by Molavi (2019) and Molavi et al. (2021).

8 Internet Appendix

8.1 Controlling for More Yield Curve Information

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

$$rx_{t+1}^{n} = \alpha + \beta_1 rx_t^n + \sum_{i \in S} \beta_i y_t^i + \epsilon_{t+1}, \tag{25}$$

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. I now test this possibility by considering the more general partially linear regression

$$rx_{t+1}^{n} = \beta_1 rx_{t}^{n} + f(\mathbf{y}_t) + \epsilon_{t+1}. \tag{26}$$

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

$$rx_{t+1}^{n} = f(\mathbf{y}_t) + \epsilon_{t+1}. \tag{27}$$

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 method is to estimate the model using the semiparametric approach described by Wood (2011). Here the standard errors are calculated using quasi-maximum likelihood.[18] 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.

8.2 Predicting Bonds Returns with Carry and Yield Change

Section 4 concluded that autocovariance in bond returns is primarily due to autocovariance in yield changes. I now test these relationships using

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

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 (%)	11.2	9.4	9.2	8.8	8.5	9.1	9.1	9.3	9.0	9.3

Table 14: shows the results of predicting returns of different maturity (years) 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. Maturity is expressed in years.

regressions. Consider the following specifications.

$$c_t^n = \alpha + \beta c_{t-1}^n + \epsilon_{t+1} \tag{28}$$

$$c_t^n = \alpha + \beta y c_t^{n-1} + \epsilon_{t+1} \tag{29}$$

$$y c_{t+1}^n = \alpha + \beta c_t^{n-1} + \epsilon_{t+1} \tag{30}$$

$$y c_{t+1}^n = \alpha + \beta y c_t^{n-1} + \epsilon_{t+1} \tag{31}$$

The results are given in Table 16. 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.

These results 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.

$$r x_{t+1}^n = \alpha + \beta_1 c_t^n + \beta_2 y c_t^n + \epsilon_{t+1} \tag{32}$$

Note that because period t carry is observable I include this rather than the previous period carry into the regression. The results are given in Table 17. 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.

	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 16: 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. Maturity is expressed in years.

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

8.3 Changes in Higher Order Principal Components

Table 6 shows the results from predicting bond returns using changes in the first principal component of yields. I now present additional results for the predictive power of changes in higher order principal components. Table 18 considers the predictive power of the changes in 2nd, 3rd, 4th and 5th principal component separately. Table 19 jointly includes the first five principal components to a predictability regression.

On top of changes in the level factor, changes in the fifth principal component contain additional predictive information for bond returns. This component is also known to be correlated with the predictability factor proposed by Cochrane and Piazzesi (2005). However, since it is only weakly correlated with contemporaneous bond returns, it cannot explain much of yield curve momentum. The results also suggest that changes in the 3rd principal components have some weak additional predictive content for bond returns.

These results are loosely related to those in Hoogteijling et al. (2021), which is contemporaneous work to this paper. Motivated by Crump and Gospodinov (2021), they consider the predictive content of changes in the

	Δpc^2					Δpc^3				
Mat	α	t-value	β	t-value	R^2 (%)	α	t-value	β	t-value	R^2 (%)
1	0.07	3.63	-0.02	-0.85	0.27	0.07	3.55	0.06	0.99	0.74
2	0.10	2.75	-0.01	-0.23	0.019	0.10	2.76	0.12	1.17	0.71
3	0.15	2.85	-0.01	-0.11	0.004	0.15	2.88	0.16	1.32	0.70
4	0.17	2.60	0.01	0.10	0.003	0.17	2.63	0.17	1.21	0.49
5	0.19	2.48	0.02	0.31	0.02	0.19	2.51	0.23	1.34	0.59
6	0.22	2.41	0.02	0.27	0.02	0.22	2.45	0.37	1.85	1.2
7	0.22	2.21	-0.01	-0.15	0.01	0.22	2.23	0.38	1.71	0.98
8	0.24	2.15	-0.04	-0.40	0.04	0.24	2.17	0.40	1.61	0.87
9	0.24	1.94	-0.02	-0.21	0.01	0.24	1.96	0.45	1.68	0.93
10	0.26	1.93	-0.03	-0.23	0.01	0.26	1.95	0.45	1.55	0.76

	Δpc^4					Δpc^5				
Mat	α	t-value	β	t-value	R^2 (%)	α	t-value	β	t-value	R^2 (%)
1	0.07	3.56	0.08	0.86	0.39	0.07	3.54	-0.42	-2.37	3.7
2	0.10	2.76	0.21	1.41	0.69	0.10	2.75	-0.65	-2.09	2.4
3	0.15	2.86	0.29	1.52	0.66	0.15	2.86	-0.84	-2.14	2.0
4	0.17	2.61	0.35	1.42	0.58	0.17	2.61	-1.12	-2.56	2.2
5	0.19	2.49	0.42	1.42	0.59	0.19	2.48	-1.54	-2.82	2.9
6	0.22	2.43	0.51	1.26	0.65	0.22	2.42	-2.00	-2.85	3.6
7	0.22	2.21	0.35	0.73	0.24	0.22	2.20	-2.24	-3.05	3.6
8	0.24	2.15	0.15	0.29	0.03	0.24	2.15	-2.20	-2.82	2.8
9	0.24	1.94	0.10	0.21	0.01	0.24	1.94	-1.98	-2.44	1.9
10	0.26	1.94	-0.01	-0.03	0.0001	0.26	1.94	-1.80	-2.03	1.3

Table 18: shows the results of predicting returns of different maturity (years) bonds on changes in the 2nd, 3rd, 4th and 5th principal components separately. The t-values are based on Newey and West (1987) standard errors.

Mat	1	2	3	4	5	6	7	8	9	10
const.	0.07	0.10	0.14	0.17	0.19	0.21	0.22	0.24	0.23	0.25
t-value	4.05	3.07	3.16	2.82	2.67	2.62	2.39	2.32	2.08	2.06
Δpc^1	-0.02	-0.04	-0.05	-0.05	-0.06	-0.07	-0.08	-0.08	-0.08	-0.08
t-value	-2.90	-3.00	-2.54	-2.26	-2.32	-2.29	-2.31	-2.23	-2.07	-1.94
Δpc^2	0.00	0.01	0.02	0.03	0.05	0.05	0.03	0.00	0.01	0.01
t-value	-0.20	0.24	0.25	0.39	0.58	0.52	0.26	0.03	0.11	0.08
Δpc^3	0.01	0.02	0.05	0.04	0.06	0.18	0.20	0.24	0.30	0.31
t-value	0.25	0.25	0.39	0.24	0.35	0.90	0.89	0.92	1.03	0.97
Δpc^4	0.05	0.14	0.21	0.27	0.32	0.36	0.18	-0.04	-0.12	-0.25
t-value	0.72	1.18	1.20	1.10	1.12	1.05	0.46	-0.10	-0.27	-0.53
Δpc^5	-0.50	-0.79	-1.01	-1.31	-1.76	-2.25	-2.50	-2.45	-2.21	-2.03
t-value	-2.82	-2.63	-2.66	-3.13	-3.46	-3.43	-3.51	-3.21	-2.84	-2.38
R^2 (%)	8.7	6.4	5.1	4.4	5.1	6.5	6.1	5.0	3.8	3.0

Table 19: shows the results of predicting returns of different maturity (years) bonds on changes in the first five principal components jointly. The t-values are based on Newey and West (1987) standard errors.

first three principal components of yields. However, since they focus on an annual rather than monthly horizon, they exclude most of the short horizon autocorrelation in bond returns, which is the focus of this paper. Moreover, they do not relate their results to the term structure modelling literature. However, using an annual horizon, they find evidence that changes in the second principal component of yields contain additional information for predicting returns, which does not occur at the monthly horizon.

8.4 Post Announcement Drift: Further Analysis

This section provides some further results related to the post-FOMC announcement drift. Figure 8 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.

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 the regression results are stronger for long maturity bonds when using unexpected rather than plain changes in the target rate.

Figure 9 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. For example theoretically longer maturity yields should reflect expectations about the long run path of future short rates and also anticipate target rate changes.[19]

8.5 ACM Model: Further Results

Table 20, Panel A shows the estimated ϕ^s matrix. Panel B shows the first row of the estimated ϕ matrix, that is the results for the first principal component of yields. The rest of the coefficients are as in Panel A. We can see that the second lag of the first principal component is clearly significant.

Table 21 shows the full estimation results for the conditional beta regression depicted in Figure 6. Comparing the results to those in Table 8, the model also captures the predictive performance of the yield curve factors.

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

Liu and Wu (2021) 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?

Table 22 replicates the results in Table 2 for the 1 month lookup using the Gürkaynak et al. (2007) data updated on the Federal reserve webpage.

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

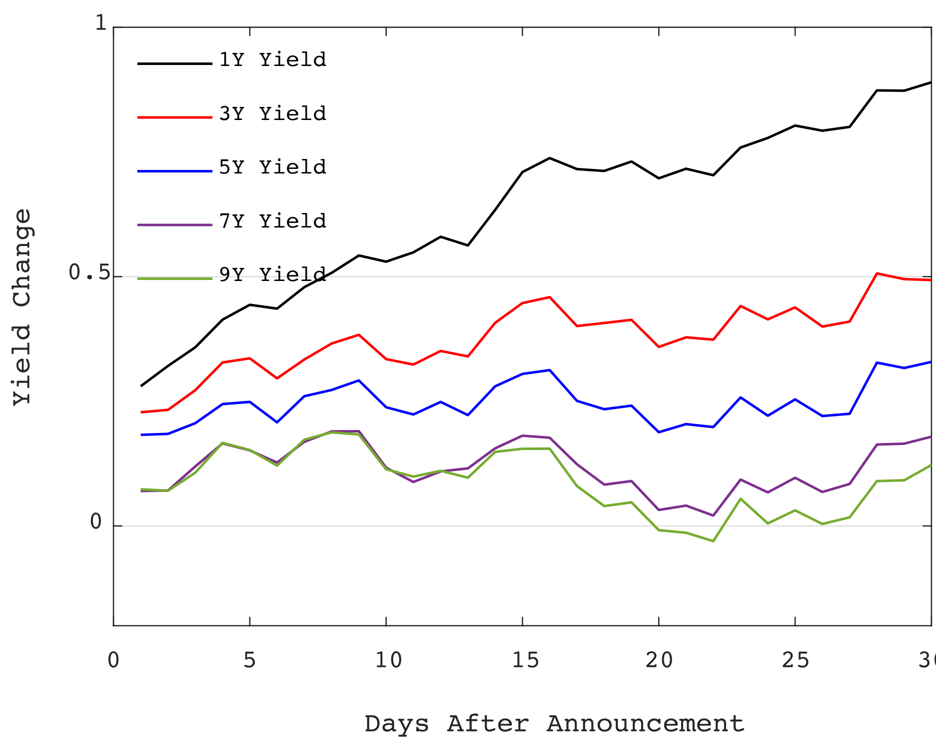

Figure 8: shows the change in different maturity yields after a change in the federal funds target rate (FFTR). Changes are measured per one basis point change in the FFTR. Days after announcement are measured using trading days.

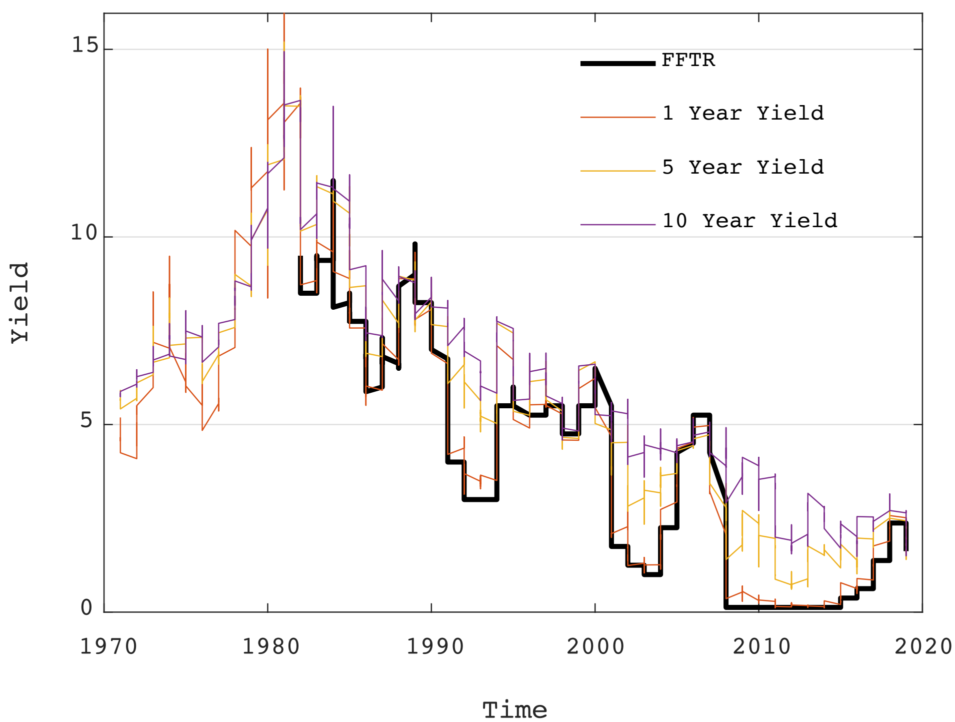

Figure 9: shows the historical development of 1,5 and 10 year yields along with the federal funds target rate.

Panel A: ϕ^s

	Ind: pc^1	Ind: pc^2	Ind: pc^3	Ind: pc^4	Ind: pc^5
Dep: pc^1	0.995	0.030	0.155	0.761	1.854
t-value	221.22	0.80	1.13	2.37	2.69
Dep: pc^2^2	-0.001	0.955	-0.144	-0.184	0.051
t-value	-0.94	77.04	-3.22	-1.75	0.22
Dep: pc^3^3	0.0004	0.003	0.862	-0.098	-0.001
t-value	0.59	0.43	41.28	-1.99	-0.01
Dep: pc^4	0.000	0.000	-0.012	0.784	-0.040
t-value	0.04	0.03	-1.11	30.50	-0.72
Dep: pc^5	0.00002	-0.001	0.001	-0.003	0.631
t-value	0.11	-0.58	0.19	-0.19	19.65

Panel B: ϕ_1

	Ind: pc^1	Ind: pc^2	Ind: pc^3	Ind: pc^4	Ind: pc^5	Ind: lag pc^1
Dep: pc^1	1.14	0.02	0.20	0.79	1.89	-0.15
t-value	27.78	0.52	1.45	2.46	2.75	-3.64

Table 20: Panel A shows the estimated ϕ^s matrix from the ACM model. Panel B shows the first row of the estimated ϕ matrix. The rest of the coefficients are as in Panel A.

Mat	1	2	3	4	5	6	7	8	9	10
const	-0.04	-0.04	-0.03	-0.05	-0.11	-0.22	-0.36	-0.51	-0.67	-0.82
rx_t	0.19	0.16	0.14	0.12	0.11	0.11	0.10	0.10	0.10	0.09
pc^1	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001
pc^2	-0.003	-0.01	-0.02	-0.03	-0.04	-0.05	-0.07	-0.08	-0.09	-0.09
pc^3	-0.03	-0.08	-0.12	-0.15	-0.17	-0.17	-0.16	-0.15	-0.14	-0.13
pc^4	-0.12	-0.23	-0.28	-0.33	-0.40	-0.50	-0.63	-0.80	-1.00	-1.22
pc^5	-0.29	-0.33	-0.45	-0.70	-0.99	-1.24	-1.38	-1.40	-1.27	-1.02

Table 21: shows the results from simulated predictability regressions when returns of different maturity bonds are explained by the past return of a same maturity bond as well as the first five principal component of yields. The results are based on the estimated ACM model simulated under the assumption that the first principal component of yields depends also on its second lag. Maturity is expressed in years.

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.

8.7 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, 2021), 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 23 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 short-term yields that are highly autocorrelated. This suggests that the results are robust to the German yield curve though this curve might be

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 22: shows the results from regressing the excess returns of different maturity (years) 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.

measured with larger pricing errors.

8.8 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:

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

	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 23: shows the results from regressing returns of different maturity (years) bonds on their past returns in both Germany and US. The regressions concern plain bond returns rather than excess returns. The t-values are based on Newey and West (1987) standard errors.

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

The results from this decomposition are given in Table 24. 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.[20]

[20] However, Asness et al. (2013) and Goyal and Jegadeesh (2018) provide evidence that cross-sectional strategies between sovereign bonds issued by different countries can generate reasonably high profits. This is because the term structures between different countries have lower correlations.

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

Table 24: shows a decomposition of the mean return from a cross sectional momentum strategy (%)

8.9 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 10 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. 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%.

This momentum stategy also enjoys a positive skewness of 1.16 compared to 0.08 for the buy and hold strategy. It has a higher excess kurtosis of 8.8 compared to 3.4 for buy and hold strategy. However, due to positive skewness and higher mean returns the momentum strategy exhibits lower tail risk. In particular the 1% sample quantile is -3.3% for the momentum strategy compared to -5.0% for the buy and hold strategy.

Figure 3 conveys an interesting additional point. The mean excess re-

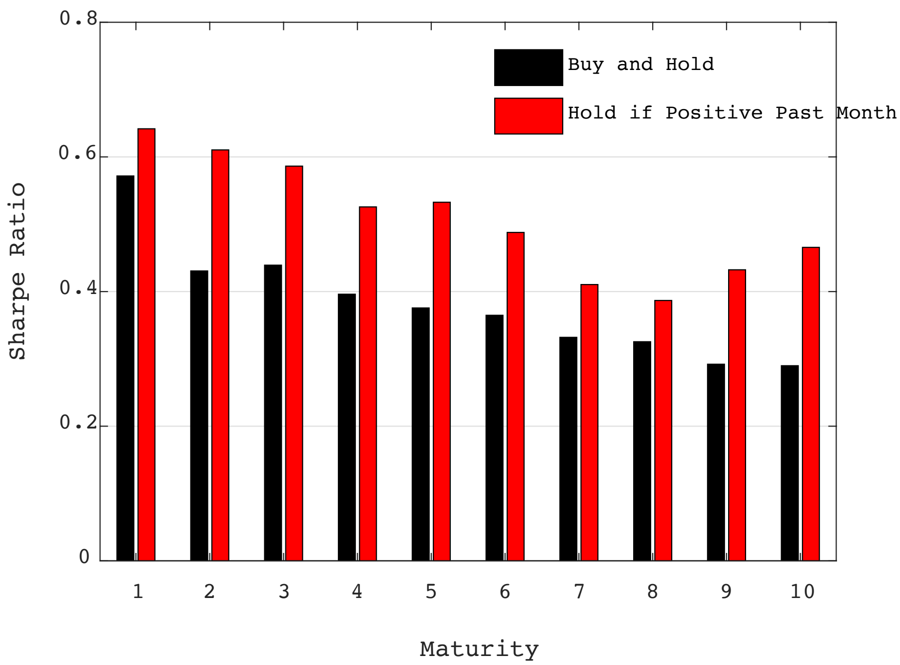

Figure 10: shows the annualized Sharpe ratios for different maturity (years) bonds for a simple momentum strategy and a buy and hold strategy.

turns 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.[21]

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.

8.10 Out-of-Sample Significance

The investment strategy results above constituted an out-of-sample evaluation for the performance of yield curve momentum strategies. However, since the focus was on economic rather than statistical significance, I now show show out-of-sample counterparts to the baseline regression results in Table 2.

As in Huang et al. (2020) I choose a 15 year in-sample training period used for model estimation. The rest of the sample is used for out-of-sample forecast evaluation. I compare two predictive models. In the first case, the forecast is given by a constant estimated as the sample mean return. In the second model, the forecast is obtained using our baseline regression in which future returns are explained by the past return and a constant. Note that the second model nests the first model as a special case.

[21]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.

Mat	$MSPE_{avg}$	Adj. $MSPE_{mom}$	$\mathbb{P}(\mathbb{E}(MSPE_{avg}) = \mathbb{E}(MSPE_{mom}))$
1	0.044	0.040	0.001
2	0.26	0.24	0.001
3	0.67	0.64	0.002
4	1.26	1.22	0.006
5	1.95	1.90	0.008
6	2.70	2.62	0.014
7	3.63	3.55	0.034
8	4.63	4.54	0.062
9	5.79	5.67	0.103
10	7.02	6.90	0.139

Table 25: shows the mean square prediction errors when forecasting bond excess returns (in per cent) using the past mean excess return and the baseline momentum regression. It also shows the results from a test of equal predictive accuracy against the alternative that the momentum regression shows more accurate results (Clark and West, 2007).

Table 25 shows the mean square prediction errors (MSPE) for the two models. For the momentum regression the sample prediction error is adjusted to correct for estimation error as suggested by Clark and West (2007). One can see that the momentum regression yields more accurate results for all maturities.

The table also shows results from a test of equal predictive accuracy against the alternative that the momentum regression produces more accurate results. The testing procedure is described in Clark and West (2007). For maturities between 1 and 5 years, we can reject the null of equal accuracy at the 1% level. The significance is below 5% for maturities of 6 and 7 years and below 10% for the 8 year maturity. For the 9 and 10 year maturities the results are not significant at the 10% level. We can conclude that, with the possible exception of the very longest maturities, the key results of this paper also hold out-of-sample. Note that here some reduction in statistical significance is mechanically expected since we have 15 years less data for the significance test.

Using Treasury futures data Huang et al. (2020) find significant out-of-sample results only for the 2 year maturity. However, they apply a lookback horizon of 12 months. For this horizon momentum is weaker and, as can be seen from Table 2, the results are mainly not significant even in-sample.

8.11 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.[22] 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

[22] 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.

because the strategy is effectively implemented for a portfolio of bonds, it cannot benefit from any individual time series predictability for different maturity bonds.

8.12 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 first consider a simple 10 year rolling regression. Figure 11 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 seem to fall somewhat after the financial crisis.

I test for structural breaks in the slope coefficients using the supremum test of Andrews (1993). The test does not identify significant break points after the financial crisis.

For shorter maturities we cannot reject the null of no structural breaks. Interestingly for bonds with maturities greater than 7 years, the test suggests a breakpoint in the fall of 1981, when for example 10 year bonds saw two consecutive months with excess returns over 10 per cent.

This break point, however, does not appear robust. First, the test does not indentify statistically significant breakpoints if we cut the sample to two parts just before or after the suggested break point month. Second, the break point is significant under a standard Chow (1960) test but not if I modify this with Newey and West (1987) standard errors. Similarly a standard Chow (1960) test gives insignificant results if I allow for the constant term to be different during the two sample periods.

Hanson et al. (2018) argue that the sensitivity of long rates to short rate changes increased since 2000. However, note that I find no evidence of structural breaks in momentum betas around that year.

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 uncondi-

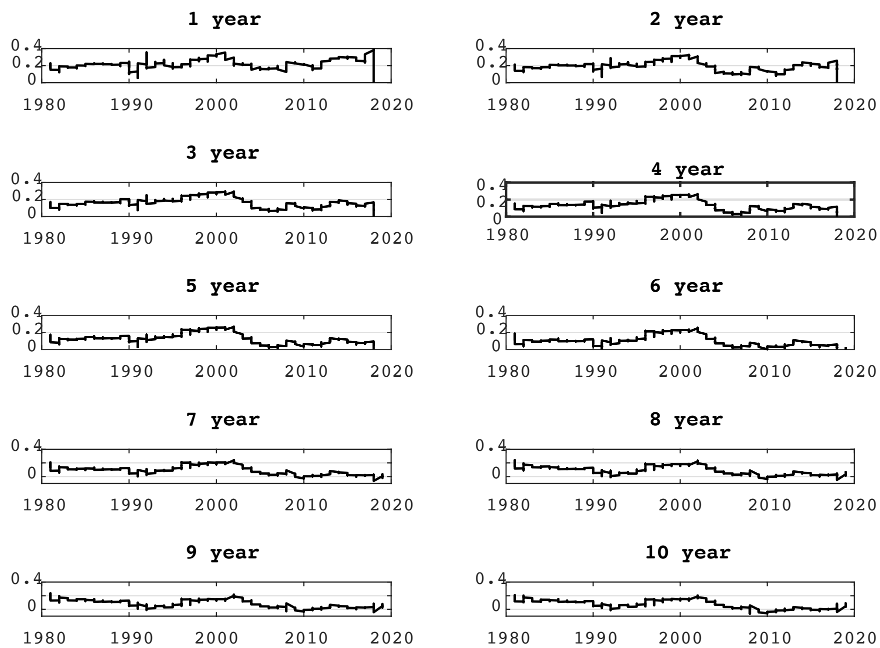

Figure 11: Momentum slope coefficient in a rolling 10 year sample for different maturity (years) bonds.

tional 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. This is true even absent any structural breaks in the data generating process.

8.13 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

$$\Delta y^n_{t+h} = \alpha + \beta \Delta y^n_t + \epsilon_{t+h} \tag{33}$$

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 $t-1$. 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 12. 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 $(t+1)$.

In concurrent work Hillenbrand (2021) finds that a 3 day window around

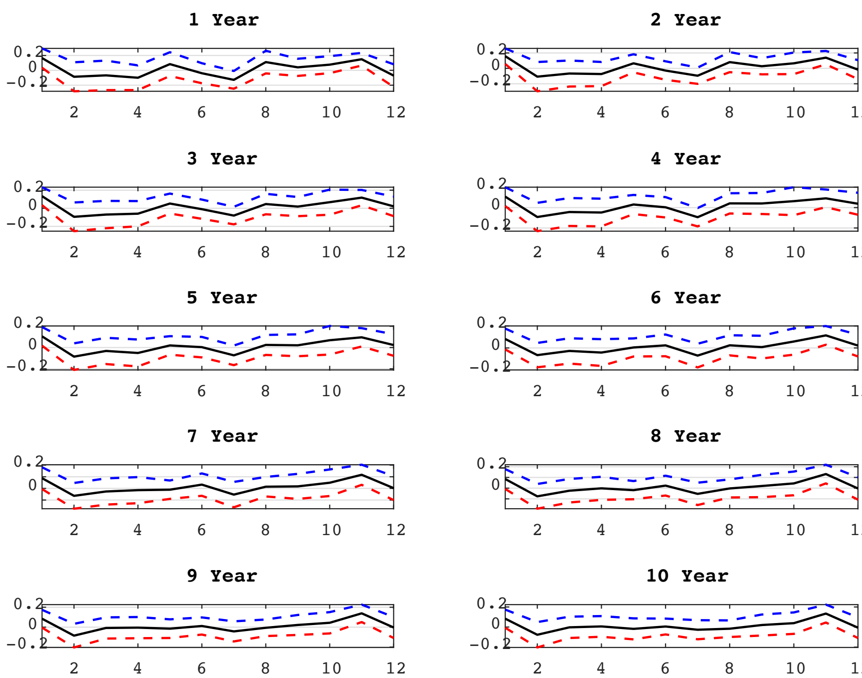

Figure 12: shows the slope coefficients on a regression of bond yield change on future bond yield changes for a horizon of up to one year.

Fed meetings fully captures the secular decline in interest rates, which started in the 80s. That is he argues that yield changes outside of this window are transitory. In this later sample period considered by Hillenbrand (2021) there is also a clearer reversion for the slope coefficients so that momentum effectively has smaller long run effects. However, full reversion does not occur for the longer sample period of this paper. This longer run view is also preferable, at least for the purposes of this paper, since it also includes a period of rising interest rates.

8.14 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.[23] 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 con-

[23] 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 26: 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. Maturity is expressed in years.

ditional on the information in the yield curve today. This is not the case empirically. Table 26 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.

8.15 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. 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 I 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 I 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.

Note that yield levels, and especially their higher frequency changes, cannot be fully explained by observable macroeconomic variables. Therefore a full macroeconomic model of the yield curve must assume that some macroeconomic information, important for explaining yields, is hidden from the econometrician.

8.16 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 a theoretical or no-arbitrage explanation for their findings. However, a theoretical model that could generate their results might also help in explaining yield curve momentum. On the other hand, since they focus on a quarterly rather than monthly horizon, their empirical results are not directly comparable to those in this paper.

8.17 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. Again 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.

8.18 Mean-Variance Prefererences

Our example concerning model misspecification assumed that the representative agent's risk preference is given by a stochastic discount factor of the form 17. Here I note that similar results can be derived under the alternative assumption that the agent possesses mean-variance preferences.

In particular assume a representative investor maximizes a mean variance objective over the return of its portfolio r_{t+1}:

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

The expectation is naturally taken under the agent's subjective measure. The portfolio return is given by

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

where z_n is the number of n maturity bonds held by the investor and r^n_{t+1} is the return of the corresponding bond. As is standard in mean-variance models, assume short rate dynamics are given exogenously. However, instead of the standard AR(1) specification, let these be given by an AR(2) process:

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

Here the natural state variable is then the short rate. As before assume our agent can only entertain single factor models, that is she puts zero probability on short rate dynamics other than AR(1). The estimation of these subjective dynamics proceeds as before. Assume the supplies of each maturity bond are given by s_n.

Conjecture $p^n_t = A_n + B_n y^1_t$. Similarly to Hamilton and Wu (2012a) then approximate:

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

and

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

Maximizing the investor's objective for an maturity n bond gives:

$$A_{n-1} - A_n + B_{n-1}[c + \rho y^1_t] - B_n y^1_t - y^1_t = \gamma B_{n-1} \sigma^2_\epsilon \left(\sum_{n=2}^{N} z_{tn} B_{n-1} \right)$$

Plugging in the market clearing condition $z_{tn} = -s_n$

$$A_{n-1} - A_n + B_{n-1}[c + \rho y^1_t] - B_n y^1_t - y^1_t = -\gamma B_{n-1} \sigma^2_\epsilon \left(\sum_{n=2}^{N} s_n B_{n-1} \right)$$

We can solve

$$B_n = -1 + B_{n-1}\rho, \quad A_n = A_{n-1} + B_{n-1}c + \gamma B_{n-1}\sigma_\epsilon^2\left(\sum_{n=2}^{N} s_n B_{n-1}\right)$$

The interest rate sensitivity parameters B_n are identical to our previous one factor model that was able to generate both unconditional and conditional momentum. However, here I do not have a free risk parameter to calibrate the persistence separately from its objective counterpart. Moreover, I naturally assume the short rate process rather than 5 year yields are observed without arror.

Similarly to before the agent estimates the AR(1) dynamics using the population OLS estimators. I approximate this with sample OLS estimation. Fitting AR(1) factor dynamics results in a persistence parameter of 0.989. In comparison estimating AR(2) dynamics results in persistence parameters of 1.077 for the first lag and −0.088 for the second.

Using the estimated AR(2)-process for short rates, I can now simulate a plain momentum slope coefficient of 0.26 and a conditional one of 0.15. While this model generates a sligthly stronger momentum pattern than in the data, given its simplicity it does surprisingly well.

My results would also hold under the assumption of risk neutrality. However, this would have the counterfactual implication that the yield curve is flat on average.

8.19 Proof of Remark 2

Due to normality, the standard pricing formula applies:

$$p_t^n = -y_t^1 + \mathbb{E}_t^S[p_{t+1}^{n-1}] + \frac{1}{2}\mathbb{V}ar_t^S(p_{t+1}^{n-1}) + \mathbb{C}ov_t^S(\log\mathcal{M}_{t+1}, p_{t+1}^{n-1})$$

Hence

$$rx_{t+1}^n = p_{t+1}^{n-1} - p_t^n - y_t^1 = p_{t+1}^{n-1} - \mathbb{E}_t^S[p_{t+1}^{n-1}] - \mathbb{C}ov_t^S(\log(\mathcal{M}_{t+1}), p_{t+1}^{n-1}) - \frac{1}{2}\mathbb{V}ar_t^S(p_{t+1}^{n-1})$$

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

Therefore

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

and

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

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

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

which is independent of bond maturity.

8.20 Proof of Remark 3

Excess return of an n maturity bond is given by

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

This implies the expected excess return is of the form

$$\mathbb{E}_t[rx_{t,t+1}^{n}] = \tilde{A}_n + \tilde{B}'_{n,e}X_t,$$

where

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

and

$$\tilde{B}_{n,e} = -(n-1)B_{n-1,e}\phi + nB_{n,e} - B_{1,e}.$$

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 $B_{n,e}$ 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 $\Pi_n(\tilde{g}^{-1}(\tilde{y}))$.

References

Adrian, T., Crump, R., and Moench, E. (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1):110–138.

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

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.

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

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

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.

Berger, D., Dew-Becker, I., and Giglio, S. (2020). Uncertainty shocks as second-moment news shocks. *The Review of Economic Studies*, 87(1):40–76.

Berk, R. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, 37(1):51–58.

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

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

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

Chow, G. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605.

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.

Clark, T. and West, K. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

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

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8).

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*, 90(1):295–314.

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.

Feldhütter, P., Heyerdahl-Larsen, C., and Illeditsch, P. (2018). Risk premia and volatilities in a nonlinear term structure model. *Review of Finance*, 22(1):337–380.

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.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720.

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. (2021). Bonds, currencies and expectational errors.

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

Hamilton, J. and Wu, J. (2012b). Identification and estimation of gaussian affine term structure models. *Journal of Econometrics*, 168(2):315–331.

Hanson, S., Lucca, D., and Wright, J. (2018). The excess sensitivity of long-term rates: A tale of two frequencies. *Federal Reserve Bank of New York Staff Report*, 810.

Hillenbrand, S. (2021). The fed and the secular decline in interest rates.

Hjalmarsson, E. (2010). Predicting global stock returns. *Journal of Financial and Quantitative Analysis*, 45(1):49–80.

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.

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

Ludvigson, S., Ma, S., and Ng, S. (2021). Uncertainty and business cycles: exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, 13(4):369–410.

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

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

Moench, E. and Siavash, S. (2022). What moves treasury yields? *Journal of Financial Economics*, 146(3):1016–1043.

Molavi, P. (2019). Macroeconomics with learning and misspecification: A general theory and applications.

Molavi, P., Tahbaz-Salehi, A., and Vedolin, A. (2021). Model complexity, expectations, and asset prices. NBER wp. 28408.

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 wp. 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.

Piazzesi, M., Salomao, J., and Schneider, M. (2015). Trend and cycle in bond premia.

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.

Schmeling, M., Schrimpf, A., and Steffensen, S. (2022). Monetary policy expectation errors. *Journal of Financial Economics*, 146(3):841–858.

Stock, J. and Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. NBER wp. 4871.

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*, 89(1):77–112.

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.

Zhang, S. (2022). Dissecting currency momentum. *Journal of Financial Economics*, 144(1):154–173.

Bank of Finland Research Discussion Papers 2021

ISSN 1456-6184, online

1/2021 Giovanni Caggiano and Efrem Castelnuovo
Global Uncertainty
ISBN 978-952-323-365-2, online

2/2021 Maija Järvenpää and Aleksi Paavola
Investor monitoring, money-likeness and stability of money market funds
ISBN 978-952-323-366-9, online

3/2021 Diego Moreno and Tuomas Takalo
Precision of Public Information Disclosures, Banks' Stability and Welfare
ISBN 978-952-323-368-3, online

4/2021 Gene Ambrocio
Euro Area Business Confidence and Covid-19
ISBN 978-952-323-370-6, online

5/2021 Markus Haavio and Olli-Matti Laine
Monetary policy rules and the effective lower bound in the Euro area
ISBN 978-952-323-371-3, online

6/2021 Seppo Honkapohja and Nigel McClung
On robustness of average inflation targeting
ISBN 978-952-323-373-7, online

7/2021 Eleonora Granziera, Pirkka Jalasjoki and Maritta Paloviita
The bias and efficiency of the ECB inflation projections: A state dependent analysis
ISBN 978-952-323-374-4, online

8/2021 Manuel M. F. Martins and Fabio Verona
Inflation dynamics and forecast: frequency matters
ISBN 978-952-323-378-2, online

9/2021 Juha Kilponen, Jouko Vilmunen and Oskari Vähämaa
Revisiting intertemporal elasticity of substitution in a sticky price model
ISBN 978-952-323-379-9, online

10/2021 Ewa Stanisławska and Maritta Paloviita
Medium- vs. short-term consumer inflation expectations: evidence from a new euro area survey
ISBN 978-952-323-380-5, online

11/2021 Stefano Eusepi, Chris Gibbs and Bruce Preston
Forward guidance with unanchored expectations
ISBN 978-952-323-384-3, online

12/2021 Francesco D'Acunto, Daniel Hoang, Maritta Paloviita and Michael Weber
Human frictions in the transmission of economic policy
ISBN 978-952-323-385-0, online

13/2021 Martin M. Andreasen, Giovanni Caggiano, Efrem Castelnuovo and Giovanni Pellegrino
Why does risk matter more in recessions than in expansions?
ISBN 978-952-323-388-1, online

14/2021 Yunjong Eo and Nigel McClung
Determinacy and E-stability with interest rate rules at the zero lower bound
ISBN 978-952-323-391-1, online

15/2021 Markus Sihvonen
Yield curve momentum
ISBN 978-952-323-392-8, online