Forecasting Bond Risk Premia using Stationary Yield Factors*

Tobias Hoogteijling† Martin Martens Michel van der Wel
Erasmus University Rotterdam and Robeco Quantitative Investing‡

January 14, 2022

Abstract

The standard way to summarize the yield curve is to use the first three principal components of the yield curve, resulting in level, slope and curvature factors. Yields, however, are non-stationary. We analyze the first three principal components of yield changes, which correspond to changes in level, slope and curvature. The new factors based on changes in yields have strong predictive power for bond risk premia, in contrast to the factors based on yield levels. We also provide insights into the impact this has on the added value of macro data for bond risk premia predictions and the recent conclusion that machine learning provides better forecasts than linear regression.

JEL classification: G12, G17, C38, E43, C45.
Keywords: Yield curve, Bond risk premia, Forecasting, PCA, Machine learning.

---

*We thank Casper Zomerdijk for many fruitful discussions and elaborate feedback. We are grateful to Michael Bauer, Laurens Swinkels and Adam Zaremba for valuable comments on our work.

†Corresponding author. E-mail: t.hoogteijling@robeco.nl.

‡Tobias Hoogteijling and Martin Martens are at Robeco Quantitative Investing. Michel van der Wel is at Erasmus University Rotterdam. The views expressed in this paper are those of the authors and do not necessarily reflect the position of Robeco.

1 Introduction

The dominant approach to model the term structure of interest rates is to use principal components analysis (PCA) applied to yields leading to the well-known level, slope and curvature factors [see Piazzesi (2010), Gürkaynak and Wright (2012), and Duffee (2013) for surveys]. However, yields are highly persistent. Uhlig (2009) shows that unrelated persistent (macro) series may give rise to a spurious factor structure in finite samples. Onatski and Wang (2021) formalize this finding, by documenting that applying PCA to non-stationary series produces spurious common variation. Crump and Gospodinov (2019) argue that therefore yields are not good primitive processes for modeling and extracting the relevant factor space. They recommend to use returns [like Litterman and Scheinkman (1991), Garbade (1996), and, more recently, Adrian, Crump and Moench (2013) and Golinski and Spencer (2017)] or changes in forward rates rather than yield levels.

In this paper, we study the importance of using (stationary) yield changes rather than (non-stationary) yields for excess bond return prediction, by analyzing one-year changes in US treasury yields for 1971 through 2018.[1] We show that the first three principal components of yield changes explain 99.75 percent of the variation in changes in yields. The loadings on these principal components indicate that the factors can be interpreted as *changes* in level, slope and curvature. We then investigate how well these three factors can predict one-year excess bond returns for maturities ranging from two years to ten years. Using real-time regressions with the historical mean as the benchmark we find that the out-of-sample R-squareds of the three principal components of yield changes range from 14.4 percent for the excess returns of the ten-year zero-coupon bonds to 20.4 percent for the four-year bonds. This is in sharp contrast with the principal components of yield levels or forward rates as well as the Cochrane and Piazzesi (2005) linear combination of forward rates. These all give high *negative* out-of-sample R-squareds for all maturities, implying that they increase the Mean Squared Prediction Error compared to an prediction equal to the historical mean.[2,3] Applying PCA to changes in forwards

[1] The convention of using one-year periods is common in the literature and dates back to early work such as Fama and Bliss (1987).

[2] Andreasen et al. (2021) find that the regression coefficients of the yield spread, forward spread and the Cochrane and Piazzesi (2005) factor actually change over time - even switch sign - conditional on recession and expansion indicators. This at least provides a partial explanation for why it is difficult to predict bond returns out-of-sample with principal components of yields (with the slope as second factor) or a linear combination of forward rates.

[3] Another explanation provided in Thornton and Valente (2012) in the context of the Cochrane and Piazzesi (2005) factor is that bond yields are highly serially correlated and correlated across maturities. If both regressors and regressands exhibit a high serial correlation, the predictive regressions based on forward rates may suffer

gives similar results as those for changes in yields. Hence, unlike the most commonly used yield and forward factors, the new factors based on changes in yields or changes in forwards do have significant out-of-sample predictive power for excess bond returns. The conclusion that the new factors have predictive power whilst existing factors do not also holds when looking at economic value in the spirit of Thornton and Valente (2012).

Of the new factors, especially the change in the slope factor has strong predictive power. A positive change in the slope on average leads to positive excess bond returns, and vice versa. In recessions for example, we tend to see that the Fed lowers the target rate and the entire curve tends to steepen amidst strong positive bond returns.

The strong predictive power of principal components from yield changes raises the bar on non-yield curve information to have added value above and beyond these new-found yield factors. A common requirement for a new predictive variable is that it adds predictive value above and beyond level, slope and curvature. But these factors themselves have no out-of-sample predictive value nor economic value, whereas the new factors based on yield changes or changes in forward rates do. To examine whether existing non-yield variables also meet this higher bar, we re-examine Ludvigson and Ng (2009) who add principal components of a large data set of macro data as factors.[4] For final (revised) macro data we find that they only improve forecasts for the ten-year excess bond returns, and for vintage macro data we see significant improvements for seven- and ten-year excess bond returns. Interestingly, we see that in our case regressions more often select the second and third principal component of the macro data, which load more on interest rate and price variables. Several of these components are known in real time, explaining why in our case the impact of final versus vintage data is much smaller. This is in contrast to the findings of Ludvigson and Ng (2009) that it is the first principal component of the macro data that adds most value to the principal components of the yield levels, which is the 'real' factor capturing employment and production. The significance of this real factor disappears when using vintage data, as reported by Ghysels, Horan, and Moench (2014). Macro data do not add value to the shortest maturities.

Finally we look at neural networks. Bianchi, Tamoni, and Büchner (2021) and Bianchi,

from a spurious regression problem (see Ferson et al. (2003a), Ferson et al. (2003b) and the references therein). Consequently, evidence of in-sample predictability need not be a useful indicator of out-of-sample predictive performance.

[4] There is a long literature on using macro data for modeling yields and bond returns. See, e.g., Ang and Piazzesi (2003), Cooper and Priestley (2009), Hamilton and Wu (2012), Bansal and Shaliastovich (2013), Greenwood and Vayanos (2014), Joslin et al. (2014), Cieslak and Povala (2015), Bauer and Hamilton (2018), Christensen and Van der Wel (2019) and Baltussen et al. (2021).

Büchner, Hoogteijling, and Tamoni (2021) conclude neural networks - the top performer amongst a whole battery of machine learning techniques - beat linear regression both in terms of out-of-sample R-squared and economic value regardless of only using forward rates, or combining forward rates with final macro data. When replacing (the first three principal components of) forward rates with the first three principal components of changes in forward rates, the conclusion is however reversed. We show that for most maturities linear regression is significantly better than neural networks both in the case where only forward rates are used, and in the case that both forward rates and final macro data are used; and regardless of whether the neural network uses forward rates or changes in forward rates. Notably when only using forward rates the neural network forecasts also improve when providing it with changes in forward rates instead of forward rates levels. Jung and Shah (2015) and Sugiyama, Yamada, and Du Plessis (2013) already note that machine learning techniques can also run into severe problems with non-stationary data. In the yield setting, we confirm this, but our results highlight machine learning techniques are nevertheless able to cope better with non-stationary input than plain linear regressions.

Summarizing, we add to the existing literature in several ways. First and most importantly, we propose a new three-factor model based on the principal components of one-year changes in yields. Whereas the first principal component resembles the well-known trend factor, the other two principal components - changes in the slope and curvature - are new predictive variables. Papers either study explicitly a non-stationary yield setting [see, e.g., Hall, Anderson, and Granger (1992), Shea (1992) and Bowsher and Meeks (2008)], or deviations of yields around a trend. For example, Favero, Melone, and Tamoni (2021) and Berardi, Markovich, Plazzi, and Tamoni (2021) study deviations of non-stationary yields from the drift caused by monetary policy rates and document predictability of these deviations for excess bond returns. Bauer and Rudebusch (2020) show that deviations of yields from time-varying long-run trends help predicting excess bond returns. We differ by studying factors based on a simple transformation of yields, using simple PCA as commonly done for the untransformed series. Second, we show that the new factors have strong real-time predictability for bond risk premia, both in terms of out-of-sample R-squared and economic value. Our new factor model leads to three important conclusions that are different from what has been reported before: (i) We do find economic value when only using yield curve information to predict bond risk premia, whereas this is not the case when principal components of yield levels or the Cochrane and Piazzesi (2005) factor are used; (ii) vintage macro data do significantly improve forecasts for longer maturities; and

(iii) we find that linear regression still provides better forecasts than the currently best Machine Learning techniques.

2 Modeling Bond Risk Premia with Yield Changes

This section presents the result that factors based on changes in yields have strong predictive power for bond risk premia. In Section 2.1 we present our data, introduce the notation and transformations, and show factors based on both yields and yield changes. In Section 2.2 we formally test nonstationarity of the yield data, and in Section 2.3 we present the results of forecasting bond risk premia.

2.1 Yield data, transformations, and yield factors

For yield information, we use the yield data set of Liu and Wu (2021).[5] The sample period is from August 1971, when the first ten-year US government bond was issued, through December 2018. The out-of-sample period starts in January 1990. We adopt the same time windows throughout this paper. We consider maturities of one through ten years. The left panel of Figure 1 provides a plot of the data, for five of the maturities. After an increase in the first 15 years of the sample, yields peaked in the early 1980's, in a period often referred to as the Volcker regime, and have drifted down strongly ever since. After the financial crisis in 2008, particularly short rates have been close to 0%, and they have increased somewhat during the final years of the 2010's. There is a strong co-movement in the series. Over time, there is variation in the spread between the short and long yields. Particularly during the mid-1990's, mid 2000's and after the financial crisis, long yields have been considerably higher than short yields.

In our analysis, we also use changes in yields. Consistent with our analysis of one-year bond risk premia, for changes in yields we also consider a one-year period. In terms of notation, at time t we consider a zero-coupon bond with time-to-maturity n which provides a payoff of one dollar at maturity. Following the literature, t is at the observation frequency, which is monthly, and n is in years as we consider continuously-compounded annual yields. We denote the log price and continuously compounded yield of the zero-coupon bond by $p_t^{(n)}$ and $y_t^{(n)} = -\frac{1}{n}p_t^{(n)}$, respectively. With a slight notational abuse of the Δ operator, we then define the one-year

[5] Available at `https://sites.google.com/view/jingcynthiawu/`.

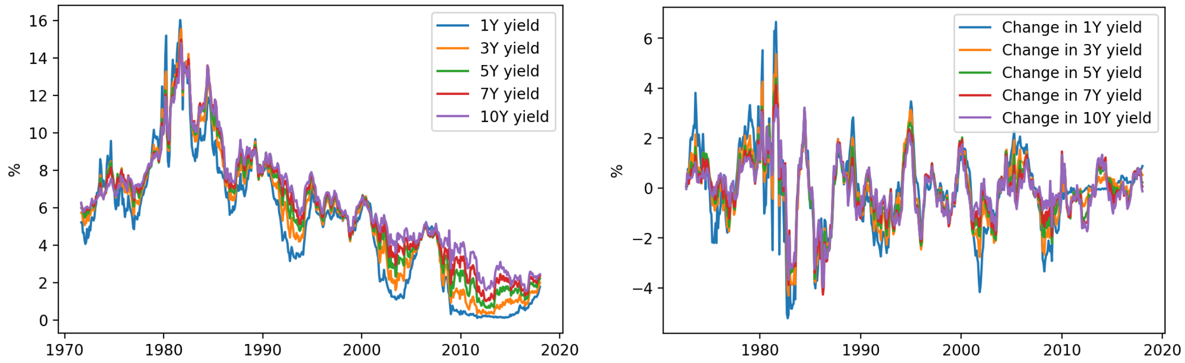

Figure 1: Plots of yields (left panel) and changes in yields (right panel) for the period August 1971 through December 2018.

change of yield as

$$\Delta y_t^{(n)} = y_t^{(n)} - y_{t-12}^{(n)}. \tag{1}$$

The right panel of Figure 1 provides a plot of the one-year yield change for five maturities. The panel presents also strong co-movement in the changes of yields, with a cyclical behavior that has decreased a bit over time.

Our analysis centers around excess bond returns. These are the returns on holding a bond for a particular period, in excess of a certain bond investment over that same period. As Cochrane and Piazzesi (2005) point out, by looking at excess returns inflation and the interest rate level is netted out and the excess returns thus represent real bond risk premia. Hence, we use the terms excess bond returns and bond risk premia interchangeable. The log excess return of holding an n-year bond from month t to $t+12$, when its remaining maturity is $n-1$ as 12 months and thus one year has progressed, can then be expressed as

$$xr_{t:t+12}^{(n)} = -(n-1)\left(y_{t+12}^{(n-1)} - y_t^{(n)}\right) + \left(y_t^{(n)} - y_t^{(1)}\right), \tag{2}$$

which is the accounting identity in Campbell and Shiller (1991). The excess returns for the two- and ten-year maturities are shown in Figure 2. Also in excess returns a co-movement is present. There is greater variation on the ten-year excess returns than on the two-year excess returns. Similar to the yield changes presented in the right panel of Figure 1, a cyclicality is present in the excess return series. The lack of a strong drift, as was present for yields, and the cyclicality, motivates our approach for using changes in yields for predicting excess returns.

In our analysis, we consider both the usual principal components of yields as well as principal components of yield changes. To get a feeling as to what the principal components look like

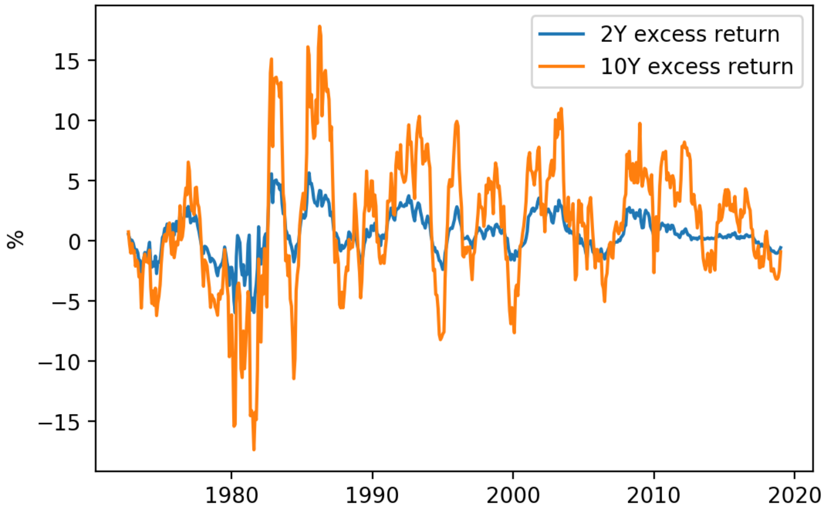

Figure 2: A plot of two- and ten-year excess returns for the period August 1971 through December 2018.

we show the full-sample factor loadings for the principal component analysis of yield changes in Figure 3, along with the corresponding loadings for yield levels. There is a striking resemblance in the loadings for the principal components of levels and changes in yields. For yield levels the lines for principal component 1, 2 and 3 are commonly referred to as level, slope and curvature due to the shape of the loadings. Hence it makes sense to refer to principal 1, 2 and 3 for yield changes as *changes* in level, slope and curvature. These three principal components explain on average 99.75 percent of the variation in the yield changes.[6] Dynamics over time are shown in Figure 4. For both yield levels and changes in yields the first principal component is by far the most import factor to explain the dynamics of respectively yields and changes in yields. In the top panel, the strong drifting behavior of the yields is also present in the first principal component. This drift is absent from the principal components based on yield changes.

2.2 Testing for stationarity

A key reason for considering (principal components of) yield changes rather than levels is that yields are non-stationary. We formally test for non-stationarity using an Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979). As the mean of $y_t^{(n)}$ is not zero, we include a

[6]In some of our analyses we also consider forward rates and changes in forward rates. For the loadings of the principal components of changes in forward rates see Appendix A, Table A.1.

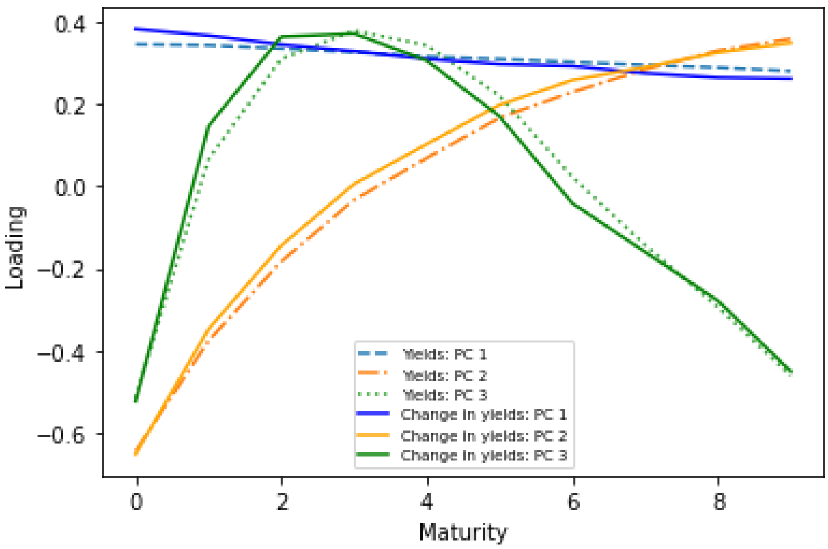

Figure 3: Plots of the loadings of the first three principal components ('PC 1', 'PC 2' and 'PC 3') of the yields (dotted lines) and the changes in yields (solid lines) for the period August 1971 through December 2018.

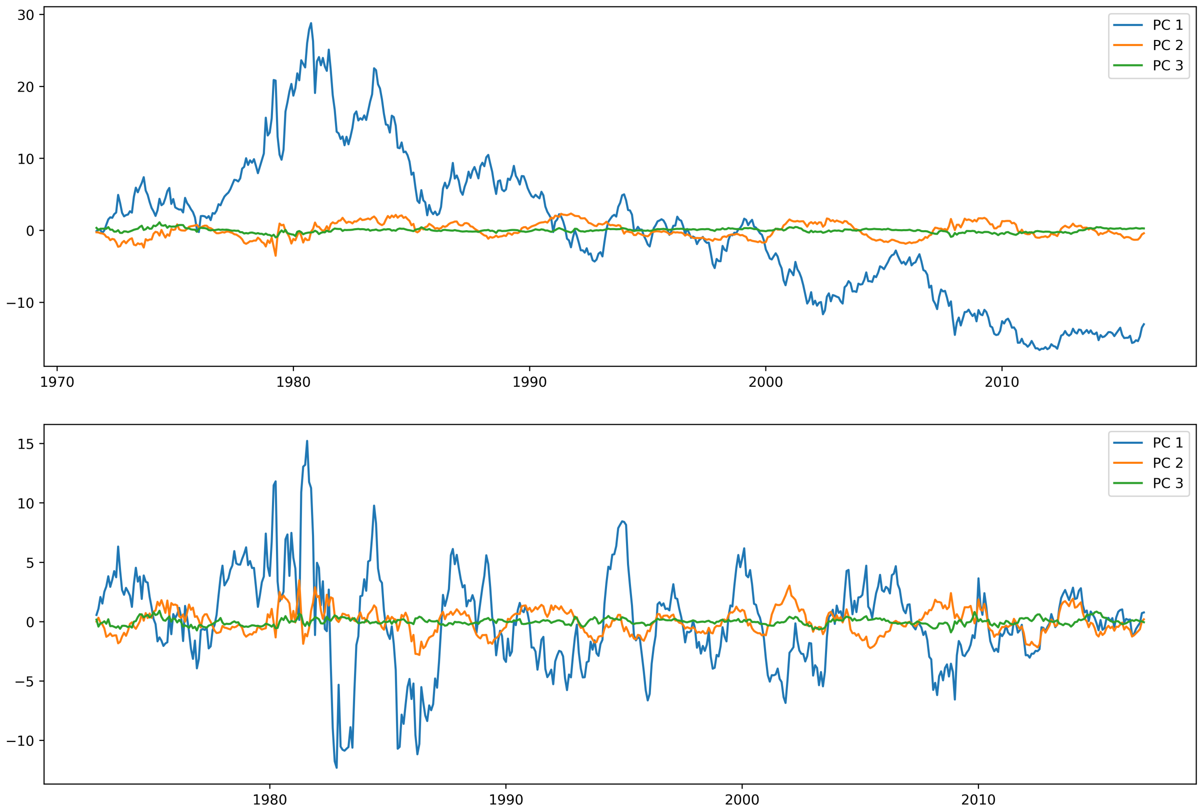

Figure 4: Plots of the first three principal components ('PC 1', 'PC 2' and 'PC 3') of the yields (top panel) and the changes in yields (bottom panel) for the period August 1971 through December 2018.

constant in the ADF regression,

$$y_t^{(n)} = \alpha + \beta t + \rho y_{t-1}^{(n)} + \sum_{j=1}^{k-1} \delta_j \left(y_{t-j}^{(n)} - y_{t-j-1}^{(n)} \right) + v_t, \tag{3}$$

where α is the drift term, β is the slope of the trend and v_t an error term. The null hypothesis of a unit root corresponds to $\rho = 1$ and the alternative hypothesis of stationarity to $\rho < 1$. We also consider the case without trend, corresponding to $\beta = 0$. The number of lags is chosen to minimize the Akaike Information Criterion. We run the test separately for each maturity n.

The results in Table 1 show that without a trend we cannot reject the null hypothesis of non-stationarity at any conventional significance level for any maturity. This is not surprising when looking at the yields over time on the left-hand side of Figure 1. The lowest p-value is 0.395, for the one-year maturity. This is expected, as the one-year yield is generally least persistent. The conclusion changes somewhat when considering the specification with trend. Now, for maturities of one and two years, the null of non-stationarity is rejected. However, for the other eight maturities the null is not rejected. Next, we transform the data and consider differences of yields instead. Specifically, we replace $y_t^{(n)}$ on the left-hand side of Equation (3) with $\Delta y_t^{(n)}$ from Equation (1), and naturally similarly do so on the right-hand side for the $y_{t-1}^{(n)}$ and $y_{t-j}^{(n)} - y_{t-j-1}^{(n)}$ terms. The results of the tests for the transformed variables are also in Table 1. Now, for all maturities the null of non-stationarity is rejected at a 1% significance level, irrespective of using the specification without or with trend. The conclusion for this analysis is that for (at least the majority) of yields the null of non-stationarity cannot be rejected, while this null is rejected for all maturities when looking at the transformed series.[7]

Augmented Dickey-Fuller (ADF) tests								
	Constant only				Constant and trend			
	Yields		Yield changes		Yields		Yield changes	
n	ADF stat	p-value	ADF stat	p-value	ADF stat	p-value	ADF stat	p-value
1	-1.77	0.395	-4.53	0.000	-3.81	0.016	-4.51	0.001
2	-1.56	0.503	-4.64	0.000	-3.55	0.034	-4.66	0.001
3	-1.40	0.580	-4.84	0.000	-3.38	0.054	-4.88	0.000
4	-1.30	0.630	-4.60	0.000	-3.27	0.072	-4.65	0.001
5	-1.11	0.712	-4.48	0.000	-3.24	0.077	-4.55	0.001
6	-1.21	0.667	-4.51	0.000	-3.24	0.077	-4.59	0.001
7	-1.01	0.750	-4.66	0.000	-2.88	0.168	-4.76	0.001
8	-0.96	0.769	-4.50	0.000	-2.84	0.181	-4.61	0.001
9	-0.91	0.786	-4.54	0.000	-2.81	0.194	-4.67	0.001
10	-0.91	0.785	-4.64	0.000	-2.83	0.186	-4.80	0.000

Table 1: This table reports the results of an Augmented Dickey-Fuller (ADF) test for a unit root in n-year yields and changes in n-year yields, over the period August 1971 through December 2018.

[7]For the corresponding results for forward rates see Appendix A, Tables A.2 and A.3

2.3 Predicting excess bond returns

We now turn to one of our key analyses, to predict excess bond returns with yield factors. To do so, we consider regressions of excess returns on a constant and a set of yield factors. These regressions are done at the maturity level. The yield factors typically used are principal components of the panel of yields $y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(10)}$. We denote the yield factors based on principal components of yield levels with Y_t. With this notation, the typical excess bond return regression takes the form

$$xr_{t:t+12}^{(n)} = a_0^{(n)} + b_0^{(n)} Y_t + e_t^{(n)}, \tag{4}$$

where $a_0^{(n)}$ is the average excess return for maturity n and $b_0^{(n)}$ measures the sensitivity of maturity n to the yield factors. We consider returns on bonds with n = 2, 3, 4, 5, 7 and 10 to cover a broad range of maturities.[8]

Our suggestion is to use factors based on changes of yields, as defined in Equation (1). We denote principal components based on changes in yields with ΔY_t. Keeping the remainder of the approach unchanged, in this new approach the excess bond return regression becomes

$$xr_{t:t+12}^{(n)} = a_1^{(n)} + b_1^{(n)} \Delta Y_t + e_t^{(n)}, \tag{5}$$

where the subscripts on $a_1^{(n)}$ and $b_1^{(n)}$ distinguish these parameters from the usual approach from Equation (4). Below we analyse the importance of considering factors based on yield changes rather than factors based on yield levels.

2.3.1 In-sample predictive regressions

Before turning to an out-of-sample analysis, we first look at the full-sample regression to learn about the usefulness of the factors to predict bond returns. The results of the regressions from Equations (4) and (5) are presented in Table 2. Focusing first on the specification using factors based on yields in Panel A, we see that the level factor (PC1) is not significant for any maturity. The slope factor (PC2) is highly significant for all maturities, which is also expected looking at Fama and French (1989) and Campbell and Shiller (1991). Finally, the curvature factor (PC3) is also significant for maturities up to five years.

[8]Excess returns on six, eight and nine year bonds are over 99% correlated with the excess returns on seven or ten year bonds and therefore show similar results. To keep down the size of the tables we do not report the results on these maturities.

Full-sample bond risk premia regressions										
	Panel A: Using Y					Panel B: Using ΔY				
n	const	PC1	PC2	PC3	R^2	const	PC1	PC2	PC3	R^2
2	0.62***	0.03	0.46***	1.22**	0.142	0.62***	−0.06	0.53***	0.72*	0.132
	(3.08)	(1.23)	(2.83)	(2.25)		(3.18)	(-1.21)	(2.77)	(1.77)	
3	1.08***	0.03	0.94***	2.22**	0.144	1.08***	−0.11	1.00***	1.34*	0.139
	(2.95)	(0.65)	(3.08)	(2.22)		(3.03)	(-1.10)	(2.93)	(1.72)	
4	1.52***	0.02	1.49***	3.10**	0.167	1.52***	-0.14	1.40***	1.93*	0.136
	(3.02)	(0.38)	(3.49)	(2.20)		(3.00)	(-0.98)	(2.94)	(1.71)	
5	1.74***	0.00	1.95***	3.70**	0.177	1.74***	-0.15	1.69***	2.23	0.125
	(2.84)	(0.05)	(3.74)	(2.09)		(2.75)	(-0.82)	(2.89)	(1.54)	
7	2.22***	0.00	2.98***	3.65	0.192	2.22**	-0.14	2.34***	2.49	0.114
	(2.67)	(-0.02)	(4.04)	(1.50)		(2.49)	(-0.56)	(3.02)	(1.20)	
10	2.62***	-0.05	4.32***	3.26	0.203	2.62**	-0.14	2.97***	2.13	0.090
	(2.33)	(-0.39)	(4.27)	(0.95)		(2.09)	(-0.37)	(2.86)	(0.70)	

Table 2: This table reports output from the full-sample bond risk premia regressions. Panel A shows results for the standard level, slope and curvature regression $xr^{(n)}_{t:t+12} = a^{(n)}_0 + b^{(n)}_0 Y_t + e^{(n)}_t$, thus using principal components based on yield levels (these factors are denoted with Y). Panel B shows the results for the same regression for changes in level, slope and curvature, $xr^{(n)}_{t:t+12} = a^{(n)}_0 + b^{(n)}_0 \Delta Y_t + e^{(n)}_t$, using principal components based on yield changes (denoted with ΔY). In parentheses are t-statistics, which are are obtained from Newey-West standard errors computed with 11 lags. The estimation is carried out on monthly data from August 1971 through December 2018 using Liu and Wu (2021) bond yields. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

Second, we look at the predictive ability of changes in level, slope and curvature in Panel B of Table 2. The first principal component has no significant predictive power. The one-year change in yields resembles a time-series trend, but with some important differences with for example Moskowitz and Pedersen (2012). We highlight three. First, whereas the lookback period of one year is the same, we predict 12-month ahead bond returns instead of next month's bond returns. Second, we use the same predictive factor for all maturities, instead of a maturity specific trend. Finally, we use past yield changes, not past bond returns. All three features reduce the predictive ability. Of course applying principal components analyses to yields is aimed at formulating a parsimonious factor model, not specifically trying to find the best predictor for excess bond returns.

The second principal component of yield changes, the change-in-slope factor, significantly predicts all bond maturities. As far as we know this is a new variable capable of predicting bond risk premia. That this variable is able to predict bond risk premia actually makes intuitive sense from an economic perspective. For example, in times of crises the central bank will lower the target rate, often resulting in a steeper slope of the curve. Such periods also often coincide with declining yields across the curve. Hence the positive sign makes sense. The third principal component, the change in the curvature, is only marginally significant for maturities up to four years. Bonds with greater curvature gain more in price when yields fall than they lose when yields rise. Hence an increase in curvature makes bonds more attractive, all else equal. Indeed we see a positive sign for changes in curvature.

Comparing Panel A to Panel B, we observe that for all maturities, the R^2s are larger when using yield levels than yield changes, with an especially sizeable difference for longer maturities. This suggests that in in-sample (!) predictive regressions of bond risk premia, principal components of yield levels have more predictive power than principal components of yield changes.

2.3.2 Out-of-sample analysis

We now move to out-of-sample forecasting where we follow the procedure also used in Ludvigson and Ng (2009). This means we estimate Equations (4) and (5) first for the in-sample period August 1971 to December 1989 and use it to produce a 12-month forecast for December 1990. Then we expand the estimation period by one month, re-estimate the equations and make another 12-month forecast, and so on.[9] The predictive performance of model $\mathcal{M}$ for maturity n is evaluated using the out-of-sample R^2 as proposed by Campbell and Thompson (2008):

$$R^{2(n)}_{oos}(\mathcal{M}) = 1 - \frac{\sum_{t=T_0}^{T-12}(xr^{(n)}_{t:t+12} - \hat{xr}^{(n)}_{t:t+12}(\mathcal{M}))^2}{\sum_{t=T_0}^{T-12}(xr^{(n)}_{t:t+12} - \bar{xr}^{(n)}_{t:t+12})^2}, \tag{6}$$

where T_0 corresponds to the first out-of-sample observation, T is the length of the data set, $\hat{xr}^{(n)}_{t+12}(\mathcal{M})$ is the prediction of model $\mathcal{M}$ for time to maturity n and $\bar{xr}^{(n)}_{t:t+12}$ is the in-sample average return, which serves as the benchmark prediction. A positive R^2_{oos} implies the method predicts better than the benchmark. The out-of-sample R^2 can be interpreted as the percentage reduction in Mean Squared Prediction Error compared to the benchmark. Significance of the out-of-sample R^2 is assessed using the CW (Clark and West, 2007) statistic. Following Clark and West (2007), we define

$$\hat{\sigma}_1^{2(n)} = \frac{1}{T-12}\sum_{t=T_0}^{T-12}(xr^{(n)}_{t:t+12} - \bar{xr}^{(n)}_{t:t+12})^2,$$

$$\hat{\sigma}_2^{2(n)}(\mathcal{M}) = \frac{1}{T-12}\sum_{t=T_0}^{T-12}(xr^{(n)}_{t:t+12} - xr^{(n)}_{t:t+12}(\mathcal{M}))^2$$

[9] Results using a rolling window are presented in Table A.10 in Appendix G.

and the adjusted measure

$$\hat{\sigma}_{2,adj}^{2(n)}(\mathcal{M}) = \hat{\sigma}_2^{2(n)} - \frac{1}{T-12}\sum_{t=T_0}^{T-12}(\bar{xr}_{t:t+12}^{(n)} - xr_{t:t+12}^{(n)}(\mathcal{M}))^2.$$

This adjusted squared error can be interpreted as the part of the squared forecast error not present in the more parsimonious model. The null hypothesis of equal R_{oos}^2's is rejected when $\hat{\sigma}_1^{2(n)}$ sufficiently exceeds $\hat{\sigma}_{2,adj}^{2(n)}(\mathcal{M})$. We can test this by regressing

$$(xr_{t:t+12}^{(n)} - \bar{xr}_{t:t+12}^{(n)})^2 - (xr_{t:t+12}^{(n)} - xr_{t:t+12}^{(n)}(\mathcal{M}))^2 + (\bar{xr}_{t:t+12}^{(n)} - xr_{t:t+12}^{(n)}(\mathcal{M}))^2$$

on a constant and considering the t-statistic.[10] The autocorrelation resulting from the overlapping excess returns is taken into account by using HAC standard errors. We also use the Clark and West (2007) statistic to assess if adding macro variables to the models significantly improves predictions.

The results for the out-of-sample bond risk premia analysis are shown in Table 3. The first two rows provide the R_{oos}^2 for the model with principal components based on yield levels, as well as those for components based on yield changes. The results are very poor for principal components based on yields. For none of the maturities, the out-of-sample R-squared is positive. Thus, the good full-sample regression results for factors based on yield levels do not translate into real-time predictive ability on top of the in-sample mean. In contrast, we find strong results for principal component factors based on yield changes. In this case, for all maturities the out-of-sample R-squared is positive. It ranges from 14.4% for ten-year bonds to 20.4% for four-year bonds. These results are highly significant. The Clark and West (2007) significance levels all fall below 1%. Hence, the in-sample predictive power of (stationary) yield changes translates to out-of-sample forecasting power, in strong contrast to (non-stationary) yield levels.[11] This is also indicated by Root Mean Squared Prediction Errors, presented in Table A.7 in Appendix B. Results for non-overlapping bond returns are also reported in Appendix D, Table A.9. We see qualitatively similar results as in the first two rows of Table 3.

To dive deeper into the challenge of providing good out-of-sample results for factors based

[10] This term is simply $\hat{\sigma}_1^{2(n)} - \hat{\sigma}_{2,adj}^{2(n)}$ for individual observations.

[11] Another implication of our predictive results is that there should be correlation between the first principal component of yield changes of one year and the the second principal component of yield changes of the preceding year. This is indeed what we find: a correlation of -0.25. The negative sign means that a steepening (flattening) of the curve last year is on average followed by declining (rising) yields in this year.

on levels, we also re-examine the performance of the Cochrane and Piazzesi (2005) factor. This factor is based on a linear combination of forward rates. Results for the out-of-sample analysis using this factor, denoted with 'CP', are also provided in Table 3. The results for the Cochrane and Piazzesi (2005) factor are in line with the results based on the model based on yield level factors. For the CP model, the out-of-sample R-squared for all maturities is in the order of magnitude of -40%. A possible explanation for this finding is unstable regression coefficients. Andreasen et al. (2021) show that the sign of the yield spread (the second principal component of yield levels) is positive in expansions but negative in recessions. This is also the case for the Cochrane and Piazzesi (2005) factor. Of course a switching sign over time makes real-time forecasting difficult.

Out-of-sample R-squared						
	n					
Factors	2	3	4	5	7	10
Y	-0.853	-0.653	-0.544	-0.420	-0.317	-0.147
ΔY	0.177***	0.202***	0.204***	0.188***	0.172***	0.144***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
CP	-0.411	-0.407	-0.413	-0.398	-0.434	-0.449

Table 3: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns for bonds with maturities (n) of 2 to 10 years with the first three principal components of yields (Y) or with the first three principal components of changes in yields (ΔY) or with the Cochrane and Piazzesi (2005) ('CP') factor based on a linear combination of forward rates. We only show the p-values inside parentheses for ΔY, as it is a one-sided test only applicable when doing better than the benchmark (the historical mean). Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

Cumulative squared prediction errors for the 2Y and 10Y maturities are shown in Figure 5. The poor performance of predictions based on yield levels is persistent over time. When returns are very high or very low, such as in 1992, 2002 and 2009, all predictions have large errors.

2.3.3 Economic value

Thornton and Valente (2012) and Sarno, Schneider and Wagner (2016) show that statistical predictability of bond returns does not necessary translate to economic value for investors. To analyze whether the factor model based on yield changes also provides economic value we consider a mean-variance utility investor who has as benchmark a model based on the expectations hypothesis that there is no predictability beyond just looking at the in-sample average. This investor allocates wealth across a bond of maturity n and the risk free one-year bond yield. At

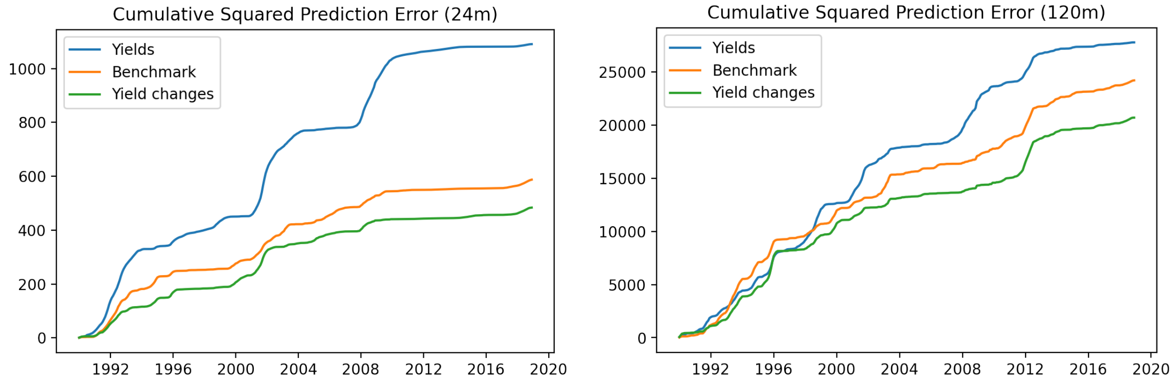

Figure 5: Plots of cumulative squared forecast errors for the benchmark model, the model based on yields and the model based on yield changes. Time to maturity is 24 months (left) or 120 months (right). The full sample period is August 1971 through December 2018, with the in-sample period being until December 1989.

time t the risky bond gets allocated weight

$$w_t^{(n)}(\mathcal{M}) = \frac{1}{\gamma}\frac{\widehat{xr}_{t:t+12}^{(n)}(\mathcal{M})}{(\hat{\sigma}_{t:t+12}^{(n)})^2}, \tag{7}$$

where $\widehat{xr}_{t:t+12}^{(n)}(\mathcal{M})$ is the excess bond return forecast from model $\mathcal{M}$, $\hat{\sigma}_{t:t+12}^{(n)}$ is a volatility estimator[12] and γ is the coefficient of relative-risk aversion. This investment results in realized portfolio returns

$$r_{p,t:t+12}^{(n)}(\mathcal{M}) = w_t^{(n)}(\mathcal{M}) \cdot xr_{t:t+12}^{(n)} + xr_{t:t+12}^{(1)}. \tag{8}$$

The realized average utility of this investment is

$$U_{t:t+12}(r_{p,t:t+12}^{(n)};\mathcal{M}) = \frac{1}{T-T_0}\sum_{t=T_0}^{T}\left(r_{p,t:t+12}^{(n)}(\mathcal{M}) - \frac{1}{2}\gamma(r_{p,t:t+12}^{(n)}(\mathcal{M}) - \hat{\mu}_p^{(n)}(\mathcal{M}))^2\right) \tag{9}$$

where T_0 is the size of the estimation window used to form the first portfolio weights and $\hat{\mu}_p^{(n)}(\mathcal{M})$ is the average portfolio return. The economic gain from a given forecasting model is evaluated using certainty equivalent returns (CER), where we compute the average utility based on the in-sample average as the benchmark prediction, denoted with $CER_0^{(n)}$, and for a given forecasting model $CER^{(n)}(\mathcal{M})$.[13] We then report $\Delta CER^{(n)}(\mathcal{M}) = CER^{(n)}(\mathcal{M}) - CER_0^{(n)}$. This can be interpreted as the fee a mean-variance investor would be willing to pay to get access to the forecasting model. As in Gargano, Pettenuzzo and Timmermann (2019), we use a Diebold-Mariano test to evaluate whether $\Delta CER^n(\mathcal{M})$ is different from zero.

[12] For this we take the historical volatility.

[13] Using the in-sample average as the benchmark prediction is consistent with Thornton and Valente (2012),

Utility gains						
	n					
Factors	2	3	4	5	7	10
Y	-3.16%	-2.83%	-2.41%	-1.90%	-1.22%	-0.08%
ΔY	0.58%*	0.64%*	0.67%**	0.74%**	0.85%**	0.83%**
	(0.054)	(0.060)	(0.055)	(0.041)	(0.035)	(0.046)
CP	-1.06%	-1.02%	-1.10%	-1.10%	-1.22%	-1.20%

Table 4: This table reports utility gains computed as the annualized difference between the Certainty Equivalent of Return from trading based on predictive regressions and from trading on the predictions of the historical average. A mean-variance investor with a relative risk aversion of $\gamma = 5$ is considered. Excess bond returns are forecast using the principal components of yields (Y) or changes in yields (ΔY) or with the Cochrane and Piazzesi (2005) ('CP') factor based on a linear combination of forward rates. The p-values, which are in parentheses, are obtained from Newey-West standard errors computed with 11 lags, and calculated only in case there are utility gains relative to the benchmark. The estimation is carried out on 12-month monthly overlapping data from August 1971 through December 2018 using Liu and Wu (2021) bond yields. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

Table 4 shows the results for economic value when considering a standard value of risk aversion with $\gamma = 5$. All CER changes are negative for the model using yields factors based on yield levels. The CER changes shrink to zero monotonically as maturity increases, and ranges from -3.16% for the two-year maturity to -0.08% for the ten-year maturity. The picture is different for the portfolio based on yield factors based on yield changes. For all maturities, the utility gain is positive. Also here, the utility gain increases monotonically with maturity. It rises from 0.58% for the two-year maturity, to 0.83% for the ten-year maturity. All changes are significant, with most at the 5% level. Thus, compared to the benchmark of the historical mean an investor would not want to have the forecasts based on the principal components of yield levels, but the same investor would like to have the forecasts based on the principal components of changes in yields. The results in Table 3 document statistical significant out-of-sample predictive power. The new results in Table 4 show that this generates also significance from an economic viewpoint. For completeness, results using the CP factor are shown in the last line of Table 4. Forecasts based on the CP factor do not offer utility gains for any maturity, in line with the negative out-of-sample R-squared found earlier. For its lack of predictive power, we disregard the CP factor in the remainder of this paper.

Sarno et al. (2016) and Gargano et al. (2019).

3 Adding macro data

Abundant literature investigates the link of the yield curve and the macroeconomy (see the references in the introduction). Macro factors are considered on top of the typical yield factors. Our finding in Section 2 that factors based on yield changes should be preferred over factors based on yield levels at least for out-of-sample analysis, begs for a re-investigation of several earlier findings. In this section we extract factors from a large set of macro data and add those to yield factors. This way our results can be directly compared with those of the macro spanning test in Ludvigson and Ng (2009), the consequences of using vintage instead of final macro data in Ghysels et al. (2018), and the competition between forecasts from linear regression and machine learning in Bianchi et al. (2021b) when both methods have access to yield data and a large set of macro data. Section 3.1 presents the macro data we use and our method for this investigation. Section 3.2 presents the results of the analysis.

3.1 Macro data and methodology

The macro data set is provided by McCracken and Ng (2016).[14] It consists of 128 variables, divided over the categories (i) output, (ii) labor market, (iii) housing sector, (iv) orders and inventories, (v) money and credit, (vi) exchange and interest rates, (vii) prices or price indices and (viii) stock market. The authors provide an extensive description of the variables and the transformations that have been used to make these data stationary. This data set closely resembles the data set used by Ludvigson and Ng (2009). We make use of the final data set of January 2019 as well as vintage data sets starting in August 1998. A full list of the macro variables and their full-sample factor loadings on the first eight principal components can be found in Table A.13 in Appendix A. Following Ludvigson and Ng (2009) we focus on the first eight principal components of the economic series of the macro data set of McCracken and Ng (2016), which we denote with Z_t.

The methodology closely follows our analysis without macro variables of Section 2 and Ludvigson and Ng (2009). Specifically, we extend Equations (4) and (5) by adding the principal components of the macro series, collected in Z_t. For the case with the yield factors based on

[14] Available at `https://research.stlouisfed.org/econ/mccracken/fred-databases/`.

yield levels, the regression thus becomes,

$$xr_{t:t+12}^{(n)} = a_0^{(n)} + b_{0,1}^{(n)} Y_t + b_{0,2}^{(n)} Z_t + e_t^{(n)}, \tag{10}$$

where the elements in $b_{0,2}^{(n)}$ capture the sensitivity of the excess returns with respect to the macro factors. For the case with yield factors based on yield changes, the regression becomes

$$xr_{t:t+12}^{(n)} = a_1^{(n)} + b_{1,1}^{(n)} \Delta Y_t + b_{1,2}^{(n)} Z_t + e_t^{(n)}. \tag{11}$$

We estimate Equations (10) and (11) in an out-of-sample fashion similar to Section 2.3.2. Thus, at each point in time we only use historically available data to estimate the parameters and make a forecast for each bond maturity. Then we expand the sample by a new observation, and so on.

In the case of macro data even with principal components the dimension is large. We therefore follow the two-stage approach outlined in Ludvigson and Ng (2009). First, we regress excess returns on individual principal components and a constant. We retain the components that are significant at a one percent level. Second, we consider regressions of excess returns on the principal components of yields or yield changes, and a subset of components that were significant in the first stage. We finally select the model with the lowest Bayesian Information Criterion (BIC; see Schwarz, 1978). This procedure is repeated after every observation, and the number of components included in the regression can thus vary over time.

3.2 Adding macro data

We start by looking in-sample at final macro data, in order to determine which macro factors are most important and how they interact with the yield and changes-in-yields factors. Panel A of Table 5 shows the results for combining yield factors with macro factors. Ludvigson and Ng (2009) look at maturities up to five years. They find that the first principal component, dubbed the real factor for its high correlation with measures of real output and employment, is the most important. In our case, we only find the second and third principal component of macro variables to be significant, although the first principal component has p-values just above 1%.[15] Additional results show that the second principal component loads heavily on several

[15] Ranging from 1.06% for $n = 3$ to 5.00% for $n = 10$.

interest rate spreads, and its significance for two-year bonds comes at the expense of the second principal component of the yields, which is the slope factor. The correlation is 35%, explaining why they are not both significant at the same time. The third principal component of the macro data is more like an inflation factor, and it is significant for all maturities except the ten-year maturity. Plots of the second and third macro principal components are shown in Figure 6.

Full-sample bond risk premia regressions – Including macro factors							
		Yield factors			Macro factors		
n	const.	PC 1	PC 2	PC 3	PC 2	PC 3	R^2
Panel A: Using Y as yield factors							
2	0.61***	0.07	0.06***	1.42**			0.142
	(3.08)	(1.23)	(2.80)	(2.25)			
3	1.07***	0.04	0.79***	2.14**		0.13***	0.157
	(3.02)	(0.95)	(2.58)	(2.21)		(2.73)	
4	1.51***	0.08	0.90**	2.72**	0.27***	-0.26**	0.198
	(3.12)	(1.17)	(1.77)	(1.95)	(2.75)	(-2.02)	
5	1.72***	0.07	1.22**	3.23*	0.33***	-0.31	0.208
	(2.95)	(0.90)	(1.98)	(1.86)	(2.79)	(-1.96)	
7	2.19***	0.10	2.00**	3.03	0.47***	-0.41**	0.221
	(2.75)	(0.84)	(2.27)	(1.26)	(2.85)	(-1.91)	
10	2.57**	-0.01	3.90***	3.03	0.40**		0.214
	(2.38)	(-0.04)	(3.76)	(0.90)	(2.59)		
Panel B: Using ΔY as yield factors							
2	0.62***	-0.06	0.53***	0.72*			0.132
	(3.16)	(-1.21)	(2.76)	(1.78)			
3	1.09***	-0.09	0.95***	1.20	0.13*		0.154
	(3.08)	(-0.94)	(2.69)	(1.59)	(1.93)		
4	1.53***	-0.07	1.22**	1.27	0.23**	-0.21**	0.176
	(3.12)	(-0.49)	(2.44)	(1.21)	(2.48)	(-2.09)	
5	1.76***	-0.04	1.42**	1.24	0.34***	-0.32**	0.184
	(2.91)	(-0.24)	(2.29)	(0.93)	(2.90)	(-2.44)	
7	2.25***	0.03	1.91**	0.92	0.55***	-0.51***	0.191
	(2.69)	(0.11)	(2.36)	(0.50)	(3.48)	(-2.89)	
10	2.68**	0.12	2.32**	-0.24	0.82***	-0.76***	0.182
	(2.31)	(0.34)	(2.16)	(-0.09)	(3.64)	(-3.03)	

Table 5: This table reports output from the full-sample bond risk premia regressions including the macro factors. Panel A shows results for the standard level, slope and curvature regression with macro factors $xr^{(n)}_{t:t+12} = a^{(n)}_0 + b^{(n)}_1 Y_t + b^{(n)}_2 Z_t + e^{(n)}_t$, where Y_t are the first three principal components of yields and Z_t are the most significant principal components of the macro data. Panel B shows the results for the same regression for changes in level, slope and curvature, $xr^{(n)}_{t:t+12} = a^{(n)}_0 + b^{(n)}_1 \Delta Y_t + b^{(n)}_2 Z_t + e^{(n)}_t$, where ΔY_t are the first three principal components of yield changes. In parentheses are t-statistics, which are are obtained from Newey-West standard errors computed with 11 lags. The estimation is carried out on monthly data from August 1971 through December 2018 using Liu and Wu (2021) bond yields. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

Next, we look at the results in Panel B of Table 5 for combining factors from changes in yields with macro data. As with yield levels, the second and third principal components are

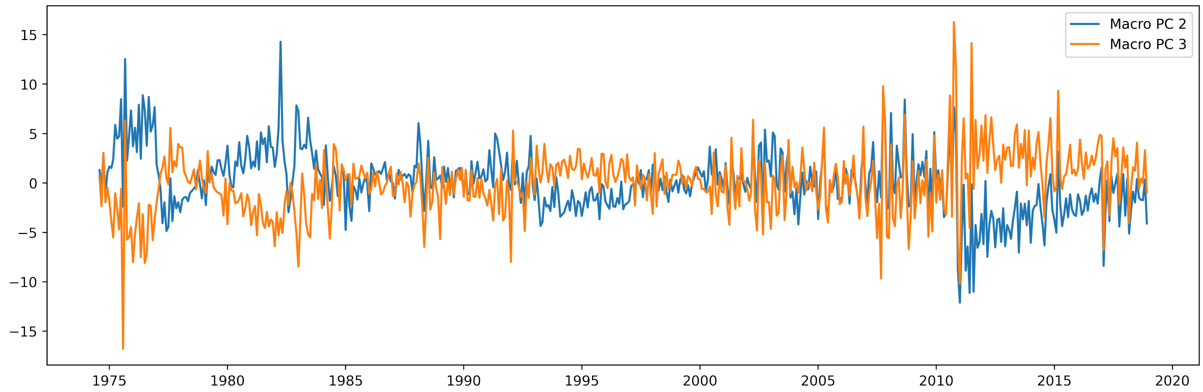

Figure 6: Plots of the second and third principal component ('Macro PC 2' and 'Macro PC 3') of the macro data for the period August 1971 through December 2018.

included, a noticeable difference being that the second principal component is now also selected for three- and ten-year bonds. Hence both interest spread information - which we no longer have as we do not have the second principal component of yield levels - and inflation are in-sample important additions to the principal components of the yield changes. It is also important to highlight that in terms of R^2, a model based on yield levels again offers a better in-sample fit than a model based on yield changes.

Of course the key question is whether macro data add to the principal components of the yield changes when we move from in-sample to out-of-sample forecasting. Thus, we repeat the analysis of Sections 2.3.2 and 2.3.3, with the macroeconomic variables included. Table A.6 shows the results for the out-of-sample analysis. For comparison purposes, the results from Table 3 using yield levels or yield changes as yield factors are added in the first row of Panel A (yield levels) and Panel B (yield changes), respectively. The second row of each panel presents the results for the out-of-sample analysis when using the final data, corresponding to the macro data used in Table 5. The yield level factors in combination with macro data still lead to negative out-of-sample R-squareds. For the principal components of yield changes as yield factors we see that adding final macro data generally reduces the out-of-sample R-squared. This is strongest for the shorter maturities. The difference increases for longer maturities, and for the ten-year maturity the R-squared is actually higher when using final macro data when compared to using no macro data and only factors based on yield changes. The R-squared for ten-year bonds improves from 14.4% to 23.0%, with the difference having a Clark and West t-statistic of 2.67.

Ghysels et al. (2018) show the importance of using real-time macroeconomic variables rather than final data. We follow their approach and repeat the out-of-sample analysis using vintage data. Unfortunately, vintage data are only available for a part of the full sample period, from August 1998 onward. For the part of the sample where vintage data is not available, we use lagged final data, to allow for a one-month publication delay. The results using vintage data are also available in Table A.6.[16] For short maturities, the conclusion is similar to using final data: The out-of-sample R-squared decreases relative to using only yield data. They are however a bit larger than using final macro data, indeed highlighting the importance of real-time data. For maturities of five years and longer, the vintage macro data do however add value even when compared to yield changes. The largest out-of-sample R-squared is obtained for ten-years, and reaches 26.0%.

Out-of-sample R-squared – Including macro factors						
	n					
Factors	2	3	4	5	7	10
Panel A: Using Y as yield factors						
Y	-0.853	-0.653	-0.544	-0.420	-0.317	-0.147
Y and Z_f	-0.593	-0.444	-0.394	-0.297	-0.226	-0.079
Y and Z_v	-0.745	-0.559	-0.470	-0.365	-0.285	-0.141
Panel B: Using ΔY as yield factors						
ΔY	0.177***	0.202***	0.204***	0.188***	0.172***	0.144***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
ΔY and Z_f	0.057***	0.093***	0.140***	0.184***	0.166***	0.230***
	(0.008)	(0.006)	(0.001)	(0.000)	(0.000)	(0.000)
ΔY and Z_v	0.127***	0.161***	0.186***	0.197***	0.220***	0.260***
	(0.011)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)

Table 6: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns with the first three principal components of yield levels (Y) or yield changes (ΔY) alone and with adding significant (at the 1% level) macro factors. The macro factors are either based on vintage data (Z_v) or final data (Z_f). The p-values are provided in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively. As vintage data are only available from August 1998 onward, lagged final data are used before this period.

Table 7 shows the Certainty Equivalent of Return (CER) gains a mean-variance investor would obtain using forecasts of the model with factors based on yield levels (first row Panel A) or yield changes (first row Panel B) or the forecasts from combining the these factors with final (second row) or vintage (third row) macro data, compared to the benchmark (the historical mean). As before, whereas the prior analysis focuses on statistical significance, this analysis investigates economic significance. Economic value is negative for yield levels, also when combining them with macro data. For factors based on yield changes we see that for final macro data only for intermediate maturities it pays to switch from the naive benchmark to the factor model. However, for all but the two-year maturity, vintage macro data significantly increase the

[16] Results from August 1999 onwards using strictly only vintage data are reported in Table A.8 in the Appendix. Also here, the out-of-sample R2 is negative when using yields/forwards rates, and positive when using changes in yields/forward rates.

Certainty Equivalent Returns. This is in sharp contrast to Ghysels et al. (2018) who find that the results of Ludvigson and Ng (2009) become insignificant when moving from final to vintage macro data. We think this is mainly caused by the fact that with yield changes we more often select the interest rate and inflation factors, which contain a lot of market data and hence are less affected by using vintage data. In the case of principal components of yield levels, Ludvigson and Ng (2009) find that the real factor has most added value. The real factor is dominated by pure economic data that are revised over time. The results for vintage data are even stronger than for final data, which only significantly increase the Certainty Equivalent Returns for the four- through ten-year maturities.

Utility gains – Including macro factors						
	n					
Factors	2	3	4	5	7	10
Panel A: Using Y as yield factors						
Y	-3.16%	-2.83%	-2.41%	-1.90%	-1.22%	-0.08%
Y and Z_f	-2.12%	-1.84%	-1.66%	-1.22%	-0.69%	-0.20%
Y and Z_v	-2.82%	-2.45%	-2.11%	-1.67%	-1.09%	-0.12%
Panel B: Using ΔY as yield factors						
ΔY	0.58%*	0.64%*	0.67%**	0.74%**	0.85%**	0.83%**
	(0.054)	(0.060)	(0.055)	(0.041)	(0.035)	(0.046)
ΔY and Z_f	0.55%	0.59%	0.88%**	1.09%***	1.08%**	0.99%
	(0.197)	(0.233)	(0.025)	(0.007)	(0.020)	(0.114)
ΔY and Z_v	0.36%	0.76%**	1.04%***	1.14%***	1.37%***	1.64%***
	(0.272)	(0.030)	(0.002)	(0.000)	(0.000)	(0.000)

Table 7: This table reports utility gains computed as the annualized difference between the Certainty Equivalent of Return from trading based on predictive regressions and from trading on the predictions of the historical average. A mean-variance investor with a relative risk aversion of $\gamma = 5$ is considered. Excess bond returns are forecast using the principal components of yield levels (Panel A) or changes in yields (ΔY) (Panel B) alone and with adding significant (at the 1% level) macro factors. The macro factors are either based on vintage data (Z_v) or final data (Z_f). The p-values are in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively. As vintage data are only available from August 1998 onward, lagged final data are used before this period.

4 Machine learning

Machine learning has been introduced to financial forecasting only recently. Heaton et al. (2017) study portfolio construction using deep neural networks. Gu et al. (2020) compare machine learning techniques for asset pricing. Even more recently, Bianchi et al. (2021b), forecast excess bond returns using a multitude of machine learning techniques. They consider both a yields-only setting, as in Cochrane and Piazzesi (2005), and a setting with unspanned macro variables, as in Ludvigson and Ng (2009).

The potential benefit of Machine Learning is that it is not bound by a linear relationship between future bond returns and current forward rates and macro data. As such it can contribute to the academic debate on how macroeconomic variable relate to the yield curve, a discussion typically centered around the spanning hypothesis (see, e.g., Bauer and Rudebusch (2016)). Also, unlike linear regression, it does not require principal components analysis to reduce the dimension of the problem. It simply searches for the best mapping of all information available at time t, x_t, using the function $g(\cdot)$,

$$xr_{t:t+12} = g(x_t). \tag{12}$$

Bianchi et al. (2021b) and Bianchi et al. (2021a) conclude Neural Networks - the top performer amongst a whole battery of machine learning techniques - beat linear regression both in terms of out-of-sample R-squared and economic value when predicting bond risk premia, regardless of only using forward rates or combining forward rates with final macro data.

The results in these two papers are, however, based on using principal components analysis applied to the level of forward rates. We have seen in Section 2 that both for out-of-sample R-squared and economic value principal components of yield levels have no predictive nor economic value, whereas forecasts from linear regressions making use of principal components of changes in yields do have predictive and economic value. Hence the top performing machine learning technique has been pitted against non-performing forecasts from linear regression. It is therefore of interest to examine how forecasts from neural networks compare to better real-time forecasts based on linear regressions applied to the stationary principal components of yield changes.

We therefore repeat the Bianchi et al. (2021b) and Bianchi et al. (2021a) analysis, but now using changes rather than levels. To stay as close as possible to these two papers, we show results for forward rates and changes in forward rates. In our implementation we closely follow the implementation of the two aforementioned papers and thus omit a full detailed description of the implementation.[17] First, we turn to out-of-sample forecasting. Table 8 presents the results of four analyses. We consider both regressions, as in Section 2, as well as neural networks. For both, we use a variant using forward levels and a variant using changes in forwards. For the neural network approach this means the input data x_t from Equation (12) are either forwards

[17] For more information on the estimation of the Neural Network see Appendix B. The results we present are for NNs with 1 hidden layer of 3 nodes (only forwards) or 32 nodes (forwards + macro), which were among the best performing ML techniques in Bianchi et al. (2021a). We have also considered alternative configurations and obtain similar results (results available upon request).

(which we denote with f) or changes in forwards (Δf).[18] For the regression approach this means that we adjust Equation (4) to allow for principal components based on forward levels (which we denote with F), and adjust Equation (5) to allow for principal components based on forward changes (ΔF). The first row of the table, for the regression approach using principal components of forward rates, shows a decrease in the out-of-sample R-squared, much like the results in Table 3 for yield levels. An exception is the ten-year maturity. When using the level of forwards in a neural network, presented in the second row, one of the key results of Bianchi et al. (2021b) and Bianchi et al. (2021a) is replicated: The neural network increases the out-of-sample R-squared, with the exception of the shortest (two-year) maturity, although not by much.

Interestingly, rows 3 and 4 in Table 8, however, show that when using changes in forward rates linear regression provides much higher reductions in mean squared prediction errors. The out-of-sample R-squared is large and statistically significant. It improves not only on the regression using factors based on forward levels, but also on the neural network! The out-of-sample R-squared decreases with maturity, and ranges from 12.5% for the ten-year maturity to 19.9% for the two-year maturity. Row 4 of the table shows that the results for neural networks also improve when using changes in forward rates. Jung and Shah (2015) and Sugiyama et al. (2013) already note that machine learning techniques can also run into severe problems with non-stationary data. However, it is also clear that neural networks cope much better with non-stationary data than linear regressions. The deterioration in results is modest for neural networks when moving from changes in forward rates to levels of forward rates. In contract, the differences for linear regression are enormous. Nevertheless, our results show that regressions perform better than neural networks, when in both cases data are suitably transformed, and that the best performing model is obtained when just simply using principal component analysis based on one-year changes of the series.

It is also interesting to compare the regression results for changes in forward rates in Table 8 with those for changes in yields in Table 3. Both results are highly significant, showing robustness to the choice for yields or forward rates. The out-of-sample R-squared is a bit higher for changes in forwards for the two-year maturity, and a bit higher for changes in yields for all other maturities. The qualitative result is identical for forwards and yields: One should use the one-year change in the series rather than the level.

[18] As neural networks are a data compression technique themselves (and can even be considered a flexible generalization of PCA), we do not separately consider the case of taking principal components as input data.

Out-of-sample R-squared – With neural networks							
		n					
Model	Factors	2	3	4	5	7	10
Regression	F	-0.315	-0.320	-0.165	-0.084	-0.056	0.033^{**} (0.003)
Neural Network	f	-0.021	0.009 (0.239)	0.016 (0.189)	0.024 (0.133)	0.031 (0.108)	0.036^{*} (0.003)
Regression	ΔF	0.199^{***} (0.002)	0.197^{***} (0.002)	0.180^{***} (0.003)	0.157^{***} (0.004)	0.139^{***} (0.003)	0.125^{**} (0.004)
Neural Network	Δf	-0.015	0.031^{**} (0.047)	0.046^{**} (0.029)	0.054^{**} (0.024)	0.053^{***} (0.021)	0.066^{**} (0.013)

Table 8: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns using both regressions and neural networks. For regressions factors based on both forwards (denoted with F) and changes in forwards (ΔF) are used, while for neural networks the input data are either forwards themselves (f) or changes in forwards (Δf). The p-values are provided in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

Next, we turn to economic value of the neural network. Table 9 shows the results. The Certainty Equivalent Returns clearly improve when using regression applied to the principal components of changes in forward rates (row 3) compared to using regression applied to the principal components of forward rate levels (row 1). Also the neural network results improve when providing it with changes in forward rates instead of forward rate levels. In terms of economic value, the forecasts from linear regression based on stationary data are more valuable to a utility investor than the forecasts from neural networks, regardless of whether these neural networks use forward rate levels or changes in forward rates. This is consistent with the findings on statistical significance.

Utility gains – With neural networks							
		n					
Model	Factors	2	3	4	5	7	10
Regression	F	-0.72%	-0.61%	-0.31%	0.08% (0.448)	0.26% (0.357)	0.73% (0.149)
Neural Network	f	-0.01%	0.13% (0.124)	$0.19\%^{*}$ (0.064)	$0.20\%^{*}$ (0.053)	$0.25\%^{**}$ (0.049)	$0.25\%^{*}$ (0.054)
Regression	ΔF	0.04% (0.427)	0.39% (0.126)	$0.60\%^{*}$ (0.070)	$0.76\%^{**}$ (0.047)	$0.89\%^{**}$ (0.038)	$0.94\%^{**}$ (0.032)
Neural Network	Δf	-0.09%	0.09% (0.261)	$0.29\%^{*}$ (0.062)	$0.40\%^{**}$ (0.042)	$0.47\%^{**}$ (0.039)	$0.53\%^{**}$ (0.028)

Table 9: This table reports utility gains computed as the annualized difference between the Certainty Equivalent of Return from trading based on predictive regressions or neural networks, and from trading on the predictions of the historical average. A mean-variance investor with a relative risk aversion of $\gamma = 5$ is considered. For regressions factors based on both forwards (denoted with F) and changes in forwards (ΔF) are used, while for neural networks the input data are either forwards themselves (f) or changes in forwards (Δf). The p-values are in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

The added benefit of Machine Learning techniques should lie in combining different sources of information, as this increases the opportunities to find non-linear relationships in the data. We therefore also look at combining forward rates with macro data. Whereas Bianchi et al. (2021b) use final macro data in forecasts, Feng et al. (2020) use real-time macro data to forecast bond returns with machine learning. They report some forecasting power for both linear and machine learning techniques with vintage macro data. This does not translate to positive economic value for investors in a mean-variance setting. However, the differences between the results of Bianchi et al. (2021b) and Feng et al. (2020) cannot solely be attributed to the use of final versus vintage macro data. Firstly, Feng et al. (2020) do not include yield information in their forecasts, which was responsible for a substantial share of the predictability in the work of Bianchi et al. (2021b). This also implies that the paper does not offer evidence pro or contra the aforementioned spanning hypothesis. As yield information was not included in the forecasts, we do not know if any of the macro variables' forecasting power was also contained in ('spanned by') the yield curve. Secondly, they only consider 56 'true' macro variables, such as employment and production, but disregard financial variables, such as price, stock market, bond market and exchange rate information. These variables were part of the 128 variable data set of Bianchi et al. (2021b) and were also important for forecasting. Finally, they consider a different time period, different maturities[19] and a different approach to some of the machine learning techniques. Thus, it is not yet clear what the impact is of the use of vintage macro data on machine learning.

Table 10 presents the out-of-sample R-squared for both regression models and neural networks, using levels and changes in the variables, and considering final and vintage macro data. Panel A of the table presents the results using final macro data. Focusing first on the neural network, making use of all individual forward rates and final macro data, cf. Bianchi et al. (2021b), we see that the neural network provides sizeable improvements compared to the historical mean as the benchmark. The out-of-sample R-squared ranges from 2.0% for the two-year maturity to 20.9% for the ten-year maturity. It is also interesting to see that in contrast to the case of using forward rates only, when combining forward rates with macro data the neural network does not benefit from using changes in forward rates (final row of Panel A). Most important however, is the comparison with linear regressions based on the principal components of changes in forward rates (third row). For all maturities but the seven-year maturity, regression does better than the neural network. Hence by applying regression to changes in forward rates instead of forward rate levels, the conclusion of Bianchi et al. (2021b), that neural networks outperform linear

[19] They only forecast bonds with maturities of up to five years, whereas the strongest predictability was found for seven and ten year bonds in Bianchi et al. (2021b).

regressions, is no longer valid.

Next, we look at Panel B, using vintage macro data. We now see that whereas the out-of-sample R-squared of linear regression improves with the use of vintage data, those of neural networks deteriorate substantially. When using forward levels and neural networks, there now is positive performance for bonds with maturities of 4 years and higher. The level of the R-squared is furthermore reduced, for example going from 12.4% to 2.3% for the four-year maturity, and from 20.9% to 17.1% for the ten-year maturity. For neural networks using changes in forwards, results are similar when using final compared to vintage macro data. Regressions however only do better when using vintage macro data when compared to final macro data, just as was the case for using yield changes rather than forward changes. Thus, the outperformance of neural networks over linear regressions also does not hold for vintage macro data.

Out-of-sample R-squared – With neural networks and including macro data							
		n					
Model	Factors	2	3	4	5	7	10
Panel A: Final macro data							
Regression	F	−0.185	−0.108	−0.090	−0.019	−0.061	−0.003
Neural Network	f	0.020** (0.016)	0.099*** (0.006)	0.124*** (0.004)	0.154*** (0.003)	0.173*** (0.001)	0.209*** (0.001)
Regression	ΔF	0.108*** (0.004)	0.128*** (0.002)	0.160*** (0.001)	0.171*** (0.001)	0.171*** (0.000)	0.237*** (0.000)
Neural Network	Δf	−0.358	−0.150	−0.099	−0.024	0.017** (0.003)	0.116*** (0.001)
Panel B: Vintage macro data							
Regression	F	−0.243	−0.200	−0.113	−0.036	−0.003	0.032*** (0.005)
Neural Network	f	−0.132	−0.015	0.023** (0.042)	0.067** (0.025)	0.106** (0.013)	0.171*** (0.003)
Regression	ΔF	0.195** (0.005)	0.175*** (0.001)	0.208*** (0.001)	0.225*** (0.000)	0.232*** (0.000)	0.270*** (0.000)
Neural Network	Δf	−0.345	−0.144	−0.084	−0.022	0.016** (0.012)	0.105*** (0.004)

Table 10: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns using both regressions and neural networks, both with final macro data (Panel A) and vintage macro data (Panel B). For regressions factors based on both forwards (denoted with F) and changes in forwards (ΔF) are used, while for neural networks the input data are either forwards themselves (f) or changes in forwards (Δf). The p-values are provided in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively. As vintage data are only available from August 1998 onward, lagged final data are used before this period.

We finally turn to economic significance for all variants. The results on economic value in Table 11 are slightly different, in the sense that also for the neural networks investors would be willing to pay a premium. For final macro data we see that a mean-variance investor would be willing to pay the most for the neural network forecasts for all maturities. However, for vintage macro data it is the other way around again, and the investor would be willing to pay most for the linear regression forecasts for all maturities. Comparing all output, the investor gains most when using regressions based on one-year changes in forward rates, for the majority of the maturities.

Utility gains – With neural networks and including macro data							
		n					
Model	Factors	2	3	4	5	7	10
Panel A: Final macro data							
Regression	F	-0.66%	-0.30%	-0.32%	0.05%	-0.10%	-0.06%
					(0.469)		
Neural Network	f	0.90%***	1.10%***	1.10%***	1.17%***	1.21%***	1.38%***
		(0.007)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
Regression	ΔF	0.77%**	0.86%**	1.07%***	1.13%***	1.16%***	1.02%***
		(0.026)	(0.020)	(0.003)	(0.003)	(0.009)	(0.063)
Neural Network	Δf	0.74%*	0.76%	0.80%	0.81%	0.91%*	1.30%**
		(0.079)	(0.111)	(0.114)	(0.126)	(0.098)	(0.027)
Panel B: Vintage macro data							
Regression	F	-1.07%	-0.71%	-0.32%	0.02%	0.20%	0.53%
					(0.490)	(0.388)	(0.209)
Neural Network	f	0.45%	0.66%*	0.64%*	0.71%*	0.84%**	1.21%**
		(0.150)	(0.075)	(0.083)	(0.079)	(0.044)	(0.010)
Regression	ΔF	0.50%*	0.92%**	1.20%***	1.34%***	1.50%***	1.75%***
		(0.085)	(0.010)	(0.001)	(0.000)	(0.000)	(0.000)
Neural Network	Δf	0.54%	0.53%	0.51%*	0.48%*	0.63%*	1.10%**
		(0.181)	(0.144)	(0.059)	(0.061)	(0.068)	(0.037)

Table 11: This table reports utility gains computed as the annualized difference between the Certainty Equivalent of Return from trading based on predictive regressions or neural networks, and from trading on the predictions of the historical average, both with final macro data (Panel A) and vintage macro data (Panel B). A mean-variance investor with a relative risk aversion of $\gamma = 5$ is considered. For regressions factors based on both forwards (denoted with F) and changes in forwards (ΔF) are used, while for neural networks the input data are either forwards themselves (f) or changes in forwards (Δf). The p-values are in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively. As vintage data are only available from August 1998 onward, lagged final data are used before this period.

5 Conclusion

The most popular way to model the yield curve is the apply principal components analysis to yields, resulting in the well-known level, slope and curvature factors. Yield levels, however, are non-stationary. Moreover, recent literature documented that real-time forecasts for bond risk premia based on these three factors are very poor.

We propose to look at the same methodology, but now based on changes in yields to overcome the non-stationarity problem. Applying principal components analysis to changes in yields results in changes in level, slope and curvature as the three most important factors. We show that these new factors are more than capable of providing good real-time forecasts for one-year excess bond returns. Especially changes in the slope are a strong predictor of the variation in bond risk premia. We use these new factors to revisit the spanning test for macro data. We find that macro data do provide added value for predicting the longer bond maturities, both for final and vintage macro data. This result for vintage data contrasts earlier findings in the academic literature failing to reject the spanning test.

We also use our new factors to check the recent finding that machine learning techniques provide better forecasts for bond risk premia than linear regression forecasts. Whereas neural networks do provide remarkably good forecasts without providing structure to the data - not using principal components and working with non-stationary forward rates - we do find that with our new factors linear regressions provide better forecasts than neural networks.

These results pose a new challenge to the Efficient Market Hypothesis. The new factor model also provides a higher bar to beat in terms of real-time forecasting of bond risk premia.

Bibliography

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

Andreasen, M. M., Engsted, T., Moller, S. V., and Sander, M. (2021). The yield spread and bond return predictability in expansions and recessions. *Review of Financial Studies*, 34(6):2773–2812.

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(4):745–787.

Baltussen, G., Martens, M., and Penninga, O. (2021). Predicting international bond returns: 70 years of evidence. *Financial Analysts Journal*, 77(3):1–23.

Bansal, R. and Shaliastovich, I. (2013). A long-run risks explanation of predictability puzzles in bond and currency markets. *The Review of Financial Studies*, 26(1):1–33.

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

Bauer, M. D. and Rudebusch, G. D. (2016). Resolving the spanning puzzle in macro-finance term structure models. *Review of Finance*, 21(2):511–553.

Bauer, M. D. and Rudebusch, G. D. (2020). Interest rates under falling stars. *American Economic Review*, 110(5):1316–1354.

Berardi, A., Markovich, M., Plazzi, A., and Tamoni, A. (2021). Mind the (convergence) gap: bond predictability strikes back! *Management Science*, 67(12):7888–7911.

Bianchi, D., Büchner, M., Hoogteijling, T., and Tamoni, A. (2021a). Corrigendum: Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1090–1103.

Bianchi, D., Büchner, M., and Tamoni, A. (2021b). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

Bowsher, C. G. and Meeks, R. (2008). The dynamics of economic functions: modeling and forecasting the yield curve. *Journal of the American Statistical Association*, 103:1419–1437.

Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Christensen, B. J. and Van der Wel, M. (2019). An asset pricing approach to testing general term structure models. *Journal of Financial Economics*, 134:165–191.

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

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

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

Cooper, I. and Priestley, R. (2009). Time-varying risk premiums and the output gap. *The Review of Financial Studies*, 22(7):2801–2833.

Crump, R. K. and Gospodinov, N. (2019). Deconstructing the yield curve. *Working Paper*.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431.

Duffee, G. R. (2013). Forecasting interest rates. *Handbook of Economic Forecasting*, 2:385–426.

Fama, E. and French, K. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 77(4):680–692.

Favero, C. A., Melone, A., and Tamoni, A. (2021). Monetary policy and bond prices with drifting equilibrium rates and diagnostic expectations. *Working Paper*.

Feng, G. G., Fulop, A., and Li, J. (2020). Real-time macro information and bond return predictability: Does deep learning help? *Available at SSRN 3517081*.

Ferson, W. E., Sarkissian, S., and Simin, T. T. (2003a). Is stock return predictability spurious? *Journal of Investment Management*, 1:1–10.

Ferson, W. E., Sarkissian, S., and Simin, T. T. (2003b). Spurious regressions in financial economics. *Journal of Finance*, 58:1393–1413.

Garbade, K. D. (1996). *Fixed income analytics*. Mit Press.

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

Ghysels, E., Horan, C., and Moench, E. (2018). Forecasting through the rearview mirror: Data revisions and bond return predictability. *The Review of Financial Studies*, 31(2):678–714.

Golinski, A. and Spencer, P. (2017). The advantages of using excess returns to model the term structure. *Journal of Financial Economics*, 125:163–181.

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

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 3(5):2223—-2273.

Gürkaynak, R. S. and Wright, J. H. (2012). Macroeconomics and the term structure. *Journal of Economic Literature*, 50(2):331–367.

Hall, A. D., Anderson, H. M., and Granger, C. W. J. (1992). A cointegration analysis of treasury bill yields. *Review of Economics and Statistics*, 74(1):116–126.

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

Heaton, J., Polson, N., and Witte, J. H. (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12.

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

Jung, K. and Shah, N. H. (2015). Implications of non-stationarity on predictive modeling using ehrs. *Journal of Biomedical Informatics*, 58:168–174.

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

Liu, Y. and Wu, J. C. (2021). Reconstructing the yield curve. *Journal of Financial Economics*, 142(3):1395–1425.

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

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

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

Onatski, A. and Wang, C. (2021). Spurious factor analysis. *Econometrica*, 89(2):591–614.

Piazzesi, M. (2010). Affine term structure models. *Handbook of Financial Econometrics*, pages 691–766.

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

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shea, G. S. (1992). Benchmarking the expectations hypothesis of the interest-rate term structure: an analysis of cointegration vectors. *Journal of Business and Economic Statistics*, 10:347–366.

Sugiyama, M., Yamada, M., and du Plessis, M. C. (2013). Learning under nonstationarity: covariate shift and class-balance change. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6):465–477.

Thornton, D. L. and Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *The Review of Financial Studies*, 25(10):3141–3168.

Uhlig, H. (2009). Comment on "How has the euro changed the monetary transmission mechanism?". In *NBER Macroeconomics Annual 2008, Volume 23*, pages 141–152. University of Chicago Press.

Appendix

A Results using forward rates and changes in forward rates

For robustness, we here repeat our analysis using forward rates and changes in forward rates instead of yields and changes in yields. Results are comparable: predictions based on forward rate levels are worse than the benchmark and predictions based on changes in forward rates outperform the benchmark, with positive and significant out-of-sample R^2s.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10
n=12	0.35	0.69	0.10	0.27	0.34	0.16	0.25	0.29	0.10	0.15
n=24	0.36	0.36	-0.04	-0.02	-0.13	-0.12	-0.23	-0.43	-0.33	-0.60
n=36	0.34	0.14	-0.09	-0.17	-0.22	-0.19	-0.22	-0.37	-0.04	0.75
n=48	0.33	-0.01	-0.10	-0.23	-0.24	-0.19	-0.08	0.17	0.80	-0.23
n=60	0.29	-0.09	0.08	-0.40	-0.09	-0.18	-0.13	0.69	-0.46	0.00
n=72	0.31	-0.21	-0.43	0.02	-0.22	0.01	0.77	-0.08	-0.15	-0.02
n=84	0.35	-0.24	0.17	0.02	-0.20	0.85	-0.18	-0.02	0.01	0.01
n=96	0.21	-0.22	0.02	-0.51	0.76	0.05	0.08	-0.25	0.06	-0.04
n=108	0.25	-0.30	-0.54	0.52	0.31	-0.10	-0.40	0.15	-0.02	0.01
n=120	0.34	-0.36	0.67	0.39	0.02	-0.35	0.14	-0.10	0.01	0.00
λ	12.70	2.26	1.59	0.74	0.41	0.29	0.21	0.11	0.07	0.04

Table A.1: This Table shows the loadings of the 10 principal components of the 10 changes in forward rates $\Delta f_t^{(m)} = f_t^{(n)} - f_{t-12}^{(n)}$ and their corresponding eigenvalues λ. Times to maturity are n and the period is 1971:08-2018:12.

	No trend			Trend		
n	DF	p-value	lags	DF	p-value	lags
12	-1.77	0.395	11	-3.81	0.016	17
24	-1.35	0.605	11	-3.29	0.068	11
36	-1.23	0.659	0	-2.97	0.141	0
48	-1.02	0.746	5	-2.80	0.196	5
60	-0.97	0.765	5	-2.93	0.154	5
72	-1.00	0.755	14	-2.96	0.144	6
84	-0.92	0.781	15	-2.85	0.179	15
96	-1.15	0.695	1	-2.79	0.200	1
108	-1.36	0.601	3	-3.01	0.131	3
120	-0.95	0.772	15	-2.71	0.232	15

Table A.2: The results of an Augmented Dickey-Fuller Test for a unit root in the n month forward rates in the period 1971:08-2018:12. The ADF test statistic is computed as $\frac{\hat{\rho}}{\text{se}(\hat{\rho})}$, where $\hat{\rho}$ is the estimate of ρ and se$(\hat{\rho})$ its standard error. The null hypothesis is the existence of a unit root.

n	DF	p-value	lags
12	-4.53	0.000	19
24	-4.23	0.001	16
36	-5.13	0.000	12
48	-5.32	0.000	12
60	-4.91	0.000	15
72	-5.19	0.000	15
84	-6.20	0.000	12
96	-3.88	0.002	13
108	-4.73	0.000	17
120	-5.80	0.000	16

Table A.3: The results of an Augmented Dickey-Fuller Test for a unit root in the first differences of the forward rates $\Delta f_t^{(n)}$ in the period 1971:08-2018:12. The ADF test statistic is computed as $\frac{\hat{\rho}}{\mathrm{se}(\hat{\rho})}$, where $\hat{\rho}$ is the estimate of ρ and $\mathrm{se}(\hat{\rho})$ its standard error. The null hypothesis is the existence of a unit root. The tests are performed with intercept but without trend.

Full-sample bond risk premia regressions										
	Panel A: Using F					Panel B: Using ΔF				
n	const	PC1	PC2	PC3	R^2	const	PC1	PC2	PC3	R^2
2	0.54***	0.04	0.22**	0.19	0.094	0.58***	-0.03	0.38***	0.13	0.117
	(2.69)	(1.60)	(2.19)	(0.87)		(2.96)	(-0.40)	(3.53)	(1.05)	
3	0.95***	0.05	0.48***	0.36	0.092	1.01***	-0.04	0.70***	0.30	0.126
	(2.60)	(1.10)	(2.67)	(0.81)		(2.82)	(-0.31)	(3.71)	(1.43)	
4	1.34***	0.06	0.79***	0.54	0.109	1.42***	-0.04	0.97***	0.45	0.123
	(2.67)	(0.88)	(3.17)	(0.80)		(2.82)	(-0.22)	(3.73)	(1.46)	
5	1.52**	0.05	1.07***	0.56	0.115	1.63**	-0.03	1.16**	0.49	0.113
	(2.49)	(0.60)	(3.50)	(0.63)		(2.59)	(-0.12)	(3.73)	(1.33)	
7	1.93**	0.06	1.73***	0.76	0.150	2.08**	0.03	1.57***	0.93	0.118
	(2.36)	(0.57)	(4.17)	(0.60)		(2.38)	(0.10)	(3.79)	(1.63)	
10	2.22**	0.04	2.69***	0.71	0.179	2.46**	0.10	2.04***	1.30*	0.110
	(2.03)	(0.27)	(4.84)	(0.41)		(2.02)	(0.23)	(3.75)	(1.70)	

Table A.4: This table reports output from the full-sample bond risk premia regressions. Panel A shows results for the first three principal components of the forward rates (F), panel B for the first three principal components of the changes in forward rates (ΔF). In parentheses are t-statistics, which are are obtained from Newey-West standard errors computed with 11 lags. The estimation is carried out on monthly data from August 1971 through December 2018 using Liu and Wu (2021) bond yields. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

Full-sample bond risk premia regressions – Including macro factors							
		Yield factors			Macro factors		
n	const.	PC 1	PC 2	PC 3	PC 2	PC 3	R^2
Panel A: Using F as yield factors							
2	0.61***	0.04	0.21**	0.21			0.084
	(2.93)	(1.40)	(2.00)	(0.91)			
3	1.06***	0.06	0.35*	0.31	0.17***		0.104
	(2.89)	(1.23)	(1.86)	(0.70)	(2.89)		
4	1.50***	0.12*	0.26	0.19	0.35***	-0.36***	0.159
	(3.00)	(1.71)	(0.91)	(0.28)	(3.69)	(-3.15)	
5	1.71***	0.12	0.41	0.11	0.44***	-0.45***	0.168
	(2.82)	(1.49)	(1.15)	(0.12)	(3.79)	(-3.12)	
7	2.17***	0.16	0.89*	0.15	0.59***	-0.56***	0.196
	(2.65)	(1.41)	(1.77)	(0.12)	(3.88)	(-3.03)	
10	2.55**	0.14	1.76**	0.04	0.69***	-0.59**	0.207
	(2.30)	(0.91)	(2.52)	(0.02)	(3.09)	(-2.20)	
Panel B: Using ΔF as yield factors							
2	0.62***	-0.02	0.38***	0.13			0.118
	(3.09)	(-0.31)	(3.53)	(1.07)			
3	1.09***	-0.01	0.65***	0.31	0.15**		0.147
	(3.05)	(-0.07)	(3.33)	(1.48)	(2.15)		
4	1.53***	0.05	0.80***	0.45	0.26**	-0.25**	0.181
	(3.12)	(0.28)	(2.85)	(1.47)	(2.82)	(-2.52)	
5	1.76***	0.09	0.93***	0.49	0.37***	-0.36***	0.191
	(2.93)	(0.44)	(2.71)	(1.35)	(3.22)	(-2.85)	
7	2.25***	0.21	1.20***	0.93	0.57***	-0.55***	0.214
	(2.74)	(0.71)	(2.67)	(1.65)	(3.80)	(-3.25)	
10	2.68**	0.36	1.51**	1.30*	0.84***	-0.78***	0.214
	(2.36)	(0.83)	(2.55)	(1.71)	(3.86)	(-3.25)	

Table A.5: This table reports output from the full-sample bond risk premia regressions including the macro factors. Panel A shows results for the first three principal components of the forward rates (F), panel B for the first three principal components of the changes in forward rates (ΔF). In parentheses are t-statistics, which are are obtained from Newey-West standard errors computed with 11 lags. The estimation is carried out on monthly data from August 1971 through December 2018 using Liu and Wu (2021) bond yields. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

Out-of-sample R-squared – Including macro factors						
	n					
Factors	2	3	4	5	7	10
Panel A: Using F as yield factors						
F	-0.319	-0.233	-0.167	-0.085	-0.056	0.034
F and Z_f	-0.185	-0.108	-0.090	-0.019	-0.061	-0.003
F and Z_v	-0.238	-0.189	-0.115	-0.037	-0.003	0.043
Panel B: Using ΔF as yield factors						
ΔF	0.199***	0.197***	0.180***	0.157***	0.139***	0.125***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
ΔF and Z_f	0.108***	0.128***	0.160***	0.171***	0.171***	0.237***
	(0.008)	(0.006)	(0.001)	(0.000)	(0.000)	(0.000)
ΔF and Z_v	0.190***	0.175***	0.208***	0.225***	0.232***	0.270***
	(0.011)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)

Table A.6: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns with the first three principal components of forward rates (F) or forward rate changes (ΔF) alone and with adding significant (at the 1% level) macro factors. The macro factors are either based on vintage data (Z_v) or final data (Z_f). The p-values are provided in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively. As vintage data are only available from August 1998 onward, lagged final data are used before this period.

B Root Mean Squared Prediction Errors (RMSPE)

Root Mean Squared Prediction Errors of predictions using yields, yield changes, forwards rates and forward rates changes are shown in Table A.7. Predictions using yields and forward rates generally produce higher prediction errors than the benchmark, and predictions using changes in yields/forward rates lead to lower predictions errors, indicating that the latter approach provides the most accurate predictions.

RMSPE						
	n					
Factors	2	3	4	5	7	10
Benchmark	1.30	2.50	3.52	4.44	6.11	8.34
Y	1.77	3.22	4.38	5.30	7.01	8.93
ΔY	1.18	2.23	3.14	4.00	5.55	7.71
Y and Z_f	1.64	3.01	4.16	5.06	6.76	8.66
ΔY and Z_f	1.26	2.38	3.27	4.01	5.58	7.32
Y and Z_v	1.72	3.13	4.31	5.21	6.90	8.96
ΔY and Z_v	1.21	2.29	3.18	3.98	5.39	7.17

Table A.7: This table shows the Root Mean Squared Prediction Error (RMSPE) when predicting out-of-sample one-year excess bond returns for bonds with maturities (n) of 2 to 10 years with the first three principal components of yields/forwards (Y) or with the first three principal components of changes in yields (ΔY).

C Results using strictly only vintage macro data

Vintage data is only available from August 1999 onwards. In our main analysis with vintage data, we have used lagged revised data for the period 1990-1999. To confirm that this does not change conclusions, we repeat the analysis using strictly only vintage macro data (from August 1999 onwards). Again, using factors based on changes in yields/forward rates outperforms factors based on yield/forward rate levels, and generally produces positive and significant out-of-sample R^2s.

Out-of-sample R-squared – Including macro factors (1999:08-2018:12)						
	n					
Factors	2	3	4	5	7	10
Y and Z_v	-1.003	-0.831	-0.799	-0.639	-0.496	-0.285
ΔY and Z_v	0.089**	0.130***	0.145***	0.177***	0.196***	0.281***
	(0.033)	(0.027)	(0.018)	(0.011)	(0.005)	(0.001)
F and Z_v	-0.452	-0.376	-0.322	-0.203	-0.142	-0.002
ΔF and Z_v	0.202***	0.178***	0.208***	0.245***	0.238***	0.309***
	(0.004)	(0.005)	(0.005)	(0.003)	(0.002)	(0.001)

Table A.8: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns with the first three principal components of yield levels (Y), yield changes (ΔY), forward rates (F) and forward rate changes (ΔF), in combination with vintage macro data (Z_v). The p-values are provided in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively. The out-of-sample period starts in August 1999, the first period from which vintage data is available.

D Results using nonoverlapping bond returns

For robustness, we also consider an out-of-sample analysis using only strictly nonoverlapping bond returns. I.e., for every year we only use the excess return of January. Also in this case, using factors based on changes in yields/forward rates outperforms factors based on yield/forward rate levels, and generally produces positive and significant out-of-sample R^2s.

Out-of-sample R-squared (non-overlapping returns)						
	n					
Factors	2	3	4	5	7	10
Y	-0.591	-0.400	-0.305	-0.210	-0.109	0.041*** (0.163)
ΔY	0.249*** (0.000)	0.262*** (0.000)	0.279*** (0.000)	0.277*** (0.000)	0.224*** (0.000)	0.163*** (0.000)
F	-0.418	-0.257	-0.156	-0.066	0.006** (0.024)	0.109*** (0.008)
ΔF	0.283*** (0.000)	0.275*** (0.000)	0.283*** (0.000)	0.276*** (0.000)	0.238*** (0.000)	0.196*** (0.000)

Table A.9: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns for bonds with maturities (n) of 2 to 10 years with the first three principal components of yields (Y) or with the first three principal components of changes in yields (ΔY). We consider only nonoverlapping returns, i.e. excess returns solely from January. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

E Computational details for the Neural Network

For neural networks, we have made use of the Keras[20] package, which is built on TensorFlow[21]. We have generally followed Bianchi et al. (2021b) in the neural network design, as tuning many settings was beyond our computing power. For an in-depth discussion of neural network settings, we refer to Bianchi et al. (2021b). We here simply state our approach for replication purposes. The numerical optimization is done using Stochastic Gradient Descent, with the learning rate set to 0.01 and Nesterov Momentum to 0.9. An Early Stopping procedure is used that stops optimization if the validation loss does not improve for 20 consecutive periods. In this case, the optimal weights are restored. Explanatory variables are scaled to be between -1 and 1. Kernel weights are initialized with the normal distribution. To combat overfitting, we make use of four techniques. Firstly, we fit each neural network 20 times, and take the average prediction of the 10 networks with the lowest validation errors.[22] Secondly, each batch is standardized before it is passed on to the final layer. Thirdly, after each layer, a fraction of the nodes is dropped out. Finally, L1L2 Regularization is applied to kernel weights. Every 60 observations, the values of the dropout fraction (0.1, 0.3 or 0.5) and L1L2 regularization parameter (0.001, 0.01, 0.1 or 1) are chosen to minimize the validation loss.

[20] `https://keras.io/`

[21] `https://www.tensorflow.org/`

[22] Bianchi et al. (2021b) select 10 from 100 networks, but this is beyond our computing power. Regardless, our research indicated that selecting 10 works out of 20 or out of 100 does not appear to have much impact on the results.

F Rolling window results

In our main analysis, we have used an expanding estimation window, starting with 209 in-sample observations. For robustness, we repeat our analysis using a rolling window of 209 observations. Again, using factors based on changes in yields/forward rates outperforms factors based on yield/forward rate levels, and generally produces positive and significant out-of-sample R^2s.

Out-of-sample R-squared – Including macro factors						
	n					
Factors	2	3	4	5	7	10
Panel A: Using Y as yield factors						
Y	-0.483	-0.434	-0.384	-0.323	-0.291	-0.189
Y and Z_f	-0.445	-0.355	-0.321	-0.307	-0.242	-0.171
Y and Z_v	-0.397	-0.378	-0.387	-0.312	-0.277	-0.191
Panel B: Using ΔY as yield factors						
ΔY	0.157***	0.191***	0.186***	0.182***	0.167***	0.169***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)
ΔY and Z_f	-0.227	-0.058	-0.008	0.034***	0.154***	0.189***
				(0.000)	(0.001)	(0.000)
ΔY and Z_v	-0.074	0.035***	0.066***	0.147***	0.148***	0.170***
		(0.002)	(0.003)	(0.002)	(0.001)	(0.000)
Panel C: Using F as yield factors						
F	-0.477	-0.509	-0.500	-0.463	-0.430	-0.359
F and Z_f	-0.367	-0.323	-0.343	-0.374	-0.388	-0.322
F and Z_v	-0.354	-0.393	-0.455	-0.457	-0.415	-0.356
Panel D: Using ΔF as yield factors						
ΔF	0.145***	0.170***	0.160***	0.156***	0.141***	0.153***
	(0.003)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)
ΔF and Z_f	-0.159	-0.015	0.036***	0.072***	0.175***	0.205***
			(0.000)	(0.000)	(0.000)	(0.000)
ΔF and Z_v	-0.011	0.076***	0.093***	0.168***	0.160***	0.180***
		(0.001)	(0.001)	(0.001)	(0.000)	(0.000)

Table A.10: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns with the first three principal components of yield/forward rate levels (Y/F) or yield/forward rate changes ($\Delta Y/\Delta F$) alone and with adding significant (at the 1% level) macro factors. The macro factors are either based on vintage data (Z_v) or final data (Z_f). The p-values are provided in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively. The rolling window consists of 209 periods.

G Different lookback & holding period

In this section, we vary the holding period and lookback period in our analysis. That is, we predict excess returns of an n month bond over an m month bond for $n = 24$, 36, 48, 60, 84 or 120 and $m = 1$, 3, 6, 12, 24. These are computed as

$$xr_{t:t+m}^{(n)} = -\left(\frac{n-m}{12}\right)\left(y_{t+m}^{(n-m)} - y_t^{(n)}\right) + \left(y_t^{(n)} - y_t^{(m)}\right), \tag{13}$$

where we now denote time in months. We also consider various lookback periods k for the yield factors, denoted as

$$\Delta_k y_t^{(n)} = y_t^{(n)} - y_{t-k}^{(n)}, \tag{14}$$

for $k = 1$, 3, 6, 12 and 24. The goal of this exercise is not to obtain equally strong results across holding periods and lookback periods. The excess returns on a one month bond are more noisy than the returns on a one year bond, making it harder to accurately fit and predict. When using a smaller lookback period, we include less information in our model. Movements in the yield curve arguably not only respond to developments of last month, but also of the months before. Rather, we perform this analysis to confirm that the predictability of bond returns does not entirely break down with alternative time horizons.

Results are shown in Tables A.11 and A.12. As expected, predictions for the one month holding period are not very accurate. However, for holding period of 3 months and higher, using changes in yields/forward rates clearly outperforms using yield/forward rate levels, with out-of-sample R^2s that are generally positive and significant.

Results are also less strong with a smaller lookback period. However, even with only a one month lookback period, using changes in yield factors produces much better results than using yield levels. For lookback periods of 6, 12 or 24 months, predictive performance remains strong.

Out-of-sample R-squared							
		n					
Holding period	Factors	2	3	4	5	7	10
	Y	-0.099	-0.085	-0.065	-0.057	-0.016	0.002*** (0.042)
	Y and Z_f	-0.016	-0.046	-0.052	-0.048	0.054*** (0.006)	0.037** (0.011)
	Y and Z_v	-0.159	-0.116	-0.091	-0.085	-0.067	-0.030
	ΔY	-0.037	-0.027	-0.027	-0.025	-0.013	-0.015
	ΔY and Z_f	-0.135	-0.128	-0.153	-0.162	-0.007	-0.020
1	ΔY and Z_v	-0.094	-0.049	-0.058	-0.051	-0.054	-0.064
	F	-0.053	-0.037	-0.022	-0.018	-0.006	0.004** (0.029)
	F and Z_f	-0.022	-0.055	-0.049	-0.026	0.052*** (0.004)	0.028*** (0.001)
	F and Z_v	-0.124	-0.082	-0.058	-0.055	-0.066	-0.033
	ΔF	-0.021	-0.011	-0.012	-0.012	-0.007	-0.011
	ΔF and Z_f	-0.076	-0.104	-0.136	-0.123	0.017*** (0.004)	0.006*** (0.008)
	ΔF and Z_v	-0.075	-0.035	-0.042	-0.045	-0.054	-0.061

Holding period	Factors	n 2	3	4	5	7	10
	Y	-0.280	-0.214	-0.149	-0.121	-0.070	-0.016
	Y and Z_f	-0.151	-0.099	-0.054	-0.033	0.023	0.008
	Y and Z_v	-0.155	-0.132	-0.099	-0.091	-0.055	-0.004
	ΔY	0.021* (0.063)	0.018* (0.070)	0.008	0.002	0.002* (0.050)	-0.008
	ΔY and Z_f	-0.064	-0.011	-0.028	-0.007	0.007	-0.004
3	ΔY and Z_v	-0.032	0.001	-0.001	-0.013	-0.020	-0.027
	F	-0.123	-0.085	-0.040	-0.026	-0.009	0.016** (0.018)
	F and Z_f	-0.056	-0.009	0.016	0.015	0.043	0.020
	F and Z_v	-0.023	-0.016	-0.016	-0.012	-0.011	0.022
	ΔF	0.044*** (0.004)	0.037** (0.010)	0.029** (0.023)	0.023** (0.036)	0.018*** (0.001)	0.010
	ΔF and Z_f	-0.002	0.023*** (0.001)	0.013*** (0.002)	0.029*** (0.002)	0.034*** (0.003)	0.020*** (0.001)
	ΔF and Z_v	0.015*** (0.008)	0.030*** (0.006)	0.021** (0.013)	0.006** (0.034)	-0.006	-0.013
	Y	-0.488	-0.357	-0.250	-0.189	-0.106	-0.019
	Y and Z_f	-0.222	-0.153	-0.131	-0.080	-0.039	0.035*** (0.003)
	Y and Z_v	-0.258	-0.220	-0.156	-0.111	-0.058	-0.018
	ΔY	0.108*** (0.003)	0.089*** (0.002)	0.063*** (0.002)	0.047*** (0.002)	0.039*** (0.002)	0.022*** (0.003)
	ΔY and Z_f	0.005*** (0.002)	0.030*** (0.004)	-0.003	0.000 ** (0.0017)	0.007** (0.015)	0.039*** (0.006)
6	ΔY and Z_v	0.061*** (0.008)	0.059** (0.010)	0.031** (0.019)	0.011** (0.038)	0.005** (0.045)	0.024** (0.040)
	F	-0.161	-0.102	-0.033	-0.003	0.021** (0.031)	0.058*** (0.008)
	F and Z_f	-0.015	0.007*** (0.003)	0.024*** (0.004)	0.046*** (0.003)	0.043*** (0.004)	0.075*** (0.001)
	F and Z_v	0.009** (0.017)	0.006** (0.023)	0.031** (0.016)	0.046** (0.012)	0.046** (0.011)	0.041** (0.010)
	ΔF	0.156*** (0.000)	0.127*** (0.001)	0.103*** (0.002)	0.083*** (0.005)	0.070** (0.011)	0.056** (0.021)
	ΔF and Z_f	0.088*** (0.000)	0.099*** (0.001)	0.071*** (0.002)	0.066*** (0.003)	0.064*** (0.003)	0.084*** (0.001)
	ΔF and Z_v	0.123*** (0.001)	0.106*** (0.002)	0.082*** (0.005)	0.051** (0.014)	0.047** (0.015)	0.059** (0.015)
	Y	-0.857	-0.656	-0.547	-0.422	-0.319	-0.148
	Y and Z_f	-0.593	-0.444	-0.394	-0.297	-0.226	-0.079
	Y and Z_v	-0.751	-0.568	-0.493	-0.378	-0.278	-0.155
	ΔY	0.177*** (0.004)	0.202*** (0.003)	0.204*** (0.002)	0.188*** (0.002)	0.172*** (0.002)	0.144*** (0.002)
	ΔY and Z_f	0.057*** (0.008)	0.093*** (0.006)	0.140*** (0.001)	0.184*** (0.000)	0.166*** (0.000)	0.230*** (0.000)
12	ΔY and Z_v	0.127*** (0.011)	0.161*** (0.002)	0.186*** (0.001)	0.197*** (0.001)	0.220*** (0.000)	0.260*** (0.000)
	F	-0.319	-0.233	-0.167	-0.085	-0.056	0.034** (0.014)
	F and Z_f	-0.185	-0.108	-0.090	-0.019	-0.061	-0.003
	F and Z_v	-0.238	-0.189	-0.115	-0.037	-0.003	0.043
	ΔF	0.199*** (0.004)	0.197*** (0.003)	0.180*** (0.002)	0.157*** (0.002)	0.139*** (0.002)	0.125*** (0.002)
	ΔF and Z_f	0.108*** (0.008)	0.128*** (0.006)	0.160*** (0.001)	0.171*** (0.000)	0.171*** (0.000)	0.237*** (0.000)
	ΔF and Z_v	0.190*** (0.011)	0.175*** (0.002)	0.208*** (0.001)	0.225*** (0.001)	0.232*** (0.000)	0.270*** (0.000)

Holding period	Factors	n					
		2	3	4	5	7	10
24	Y		-1.391	-1.313	-1.214	-1.173	-0.871
	Y and Z_f		-1.238	-1.204	-1.101	-1.064	-0.792
	Y and Z_v		-1.234	-1.182	-1.085	-1.050	-0.784
	ΔY		0.065*** (0.002)	0.087*** (0.001)	0.103*** (0.001)	0.131*** (0.000)	0.131*** (0.000)
	ΔY and Z_f		0.059*** (0.002)	0.080*** (0.001)	0.089*** (0.001)	0.093*** (0.001)	0.119*** (0.000)
	ΔY and Z_v		0.035*** (0.002)	0.064*** (0.002)	0.085*** (0.001)	0.103*** (0.001)	0.084*** (0.002)
	F		-0.683	-0.599	-0.490	-0.483	-0.291
	F and Z_f		-0.561	-0.490	-0.406	-0.419	-0.264
	F and Z_v		-0.589	-0.525	-0.405	-0.405	-0.247
	ΔF		0.152*** (0.005)	0.145*** (0.007)	0.136*** (0.009)	0.142*** (0.005)	0.102** (0.012)
	ΔF and Z_f		0.122*** (0.005)	0.134*** (0.006)	0.128*** (0.008)	0.097*** (0.008)	0.090*** (0.008)
	ΔF and Z_v		0.133*** (0.005)	0.128*** (0.009)	0.114** (0.012)	0.126*** (0.006)	0.063** (0.021)

Table A.11: This table shows the out-of-sample R-squared when predicting out-of-sample m-month excess bond returns with the first three principal components of yield/forward rate levels (Y/F) or yield/forward rate changes ($\Delta Y/\Delta F$) alone and with adding significant (at the 1% level) macro factors. The macro factors are either based on vintage data (Z_v) or final data (Z_f). The p-values are provided in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

Out-of-sample R-squared							
		n					
Lookback	Factors	2	3	4	5	7	10
Using Y as yield factors							
	Y	-0.857	-0.656	-0.547	-0.422	-0.319	-0.148
	Y and Z_f	-0.593	-0.444	-0.394	-0.297	-0.226	-0.079
	Y and Z_v	-0.751	-0.568	-0.493	-0.378	-0.278	-0.155
Using ΔY as yield factors							
1	ΔY	-0.016	-0.012	-0.012	-0.013	-0.006	-0.012
	ΔY and Z_f	-0.031	-0.069	-0.048	-0.015	-0.030	0.043** (0.001)
	ΔY and Z_v	0.058** (0.045)	-0.062	-0.098	-0.048	0.012* (0.058)	0.060** (0.022)
3	ΔY	0.015** (0.047)	0.015* (0.056)	0.010	0.004	0.007	-0.003
	ΔY and Z_f	-0.050	-0.073	-0.037	0.004*** (0.005)	-0.007	0.077*** (0.001)
	ΔY and Z_v	0.111** (0.023)	-0.033	-0.071	-0.008	0.035** (0.043)	0.068** (0.018)
6	ΔY	0.092 (0.007)	0.077 (0.012)	0.060** (0.023)	0.043** (0.044)	0.037* (0.066)	0.018
	ΔY and Z_f	0.036*** (0.007)	-0.024	0.012** (0.011)	0.052*** (0.004)	0.022*** (0.002)	0.114*** (0.001)
	ΔY and Z_v	0.127** (0.014)	-0.005	-0.001	0.046** (0.020)	0.078* (0.017)	0.065** (0.020)

Lookback	Factors	n = 2	3	4	5	7	10
	ΔY	0.177*** (0.004)	0.202*** (0.003)	0.204*** (0.002)	0.188*** (0.002)	0.172*** (0.002)	0.144*** (0.002)
12	ΔY and Z_f	0.057*** (0.008)	0.093*** (0.006)	0.140*** (0.001)	0.184*** (0.000)	0.166*** (0.000)	0.230*** (0.000)
	ΔY and Z_v	0.127*** (0.011)	0.161*** (0.002)	0.186*** (0.001)	0.197*** (0.001)	0.220*** (0.000)	0.260*** (0.000)
	ΔY	0.027*** (0.004)	0.100*** (0.000)	0.145*** (0.000)	0.150*** (0.000)	0.167*** (0.000)	0.153*** (0.001)
24	ΔY and Z_f	0.053*** (0.002)	0.137*** (0.000)	0.169*** (0.000)	0.152*** (0.000)	0.197*** (0.000)	0.201*** (0.000)
	ΔY and Z_v	0.026*** (0.004)	0.097*** (0.000)	0.126*** (0.000)	0.140*** (0.001)	0.204*** (0.000)	0.202*** (0.001)
Using F as yield factors							
	F	-0.319	-0.233	-0.167	-0.085	-0.056	0.034** (0.014)
	F and Z_f	-0.185	-0.108	-0.090	-0.019	-0.061	-0.003
	F and Z_v	-0.238	-0.189	-0.115	-0.037	-0.003	0.043
Using ΔF as yield factors							
	ΔF	-0.012	-0.010	-0.011	-0.010	-0.005	-0.010
1	ΔF and Z_f	-0.021	-0.064	-0.041	-0.002	-0.016	0.063** (0.001)
	ΔF and Z_v	0.054* (0.047)	-0.055	-0.096	-0.045	0.012* (0.056)	0.062** (0.022)
	ΔF	0.036* (0.006)	0.036** (0.010)	0.033** (0.017)	0.026** (0.032)	0.026** (0.043)	0.016
3	ΔF and Z_f	-0.025	-0.044	-0.015	0.026	0.018	0.105
	ΔF and Z_v	0.135** (0.019)	-0.007	-0.046	0.013** (0.042)	0.038** (0.030)	0.072** (0.012)
	ΔF	0.114*** (0.002)	0.101*** (0.003)	0.089*** (0.006)	0.072* (0.013)	0.067** (0.019)	0.054** (0.034)
6	ΔF and Z_f	0.080*** (0.003)	0.019** (0.011)	0.043*** (0.006)	0.087*** (0.002)	0.068*** (0.001)	0.163*** (0.000)
	ΔF and Z_v	0.175*** (0.006)	0.020** (0.041)	0.018*** (0.022)	0.068*** (0.010)	0.096*** (0.009)	0 094 *** (0.008)
	ΔF	0.199*** (0.004)	0.197*** (0.003)	0.180*** (0.002)	0.157*** (0.002)	0.139*** (0.002)	0.125*** (0.002)
12	ΔF and Z_f	0.108*** (0.008)	0.128*** (0.006)	0.160*** (0.001)	0.171*** (0.000)	0.171*** (0.000)	0.237*** (0.000)
	ΔF and Z_v	0.190*** (0.011)	0.175*** (0.002)	0.208*** (0.001)	0.225*** (0.001)	0.232*** (0.000)	0.270*** (0.000)
	ΔF	0.144*** (0.002)	0.146*** (0.002)	0.125*** (0.003)	0.111*** (0.004)	0.073*** (0.005)	0.029** (0.013)
24	ΔF and Z_f	0.040*** (0.003)	0.044*** (0.003)	0.077*** (0.001)	0.117*** (0.001)	0.062*** (0.000)	0.053*** (0.000)
	ΔF and Z_v	0.090*** (0.009)	0.110*** (0.002)	0.121*** (0.001)	0.141*** (0.001)	0.184*** (0.000)	0.151*** (0.001)

Table A.12: This table shows the out-of-sample R-squared when predicting out-of-sample one-year excess bond returns with the first three principal components of yield/forward rate levels (Y/F) or k month changes in yields/forward rates ($\Delta_k Y/\Delta_k F$) alone and with adding significant (at the 1% level) macro factors. The macro factors are either based on vintage data (Z_v) or final data (Z_f). The p-values are provided in parentheses. Three (***), two (**) and one (*) asterisk(s) denote significance at the 1, 5 and 10 percent significance level, respectively.

H Principal component loadings of macro factors

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Production								
Real Personal Income	0.063	0.069	0.016	0.046	0.035	0.026	-0.070	-0.071
Real personal income ex transfer receipts	0.085	0.082	0.020	0.042	0.037	-0.004	-0.078	-0.066
Real personal consumption expenditures	0.067	0.042	-0.034	0.066	0.003	0.174	-0.026	0.106
Real Manu. and Trade Industries Sales	0.119	0.058	-0.060	0.038	0.077	0.152	-0.012	-0.014
Retail and Food Services Sales	0.069	-0.044	-0.050	0.041	0.057	0.170	-0.055	0.066
IP Index	0.180	0.077	-0.062	-0.025	0.120	0.062	0.130	0.004
IP: Final Products and Nonindustrial Supplies	0.173	0.060	-0.061	-0.007	0.133	0.090	0.152	0.023
IP: Final Products (Market Group)	0.157	0.064	-0.056	-0.009	0.145	0.095	0.175	0.027
IP: Consumer Goods	0.125	0.066	-0.080	-0.012	0.116	0.125	0.215	0.073
IP: Durable Consumer Goods	0.119	0.081	-0.078	-0.018	0.104	0.098	0.138	-0.064
IP: Nondurable Consumer Goods	0.076	0.022	-0.044	-0.003	0.084	0.095	0.191	0.167
IP: Business Equipment	0.153	0.049	-0.016	0.003	0.132	0.021	0.064	-0.051
IP: Materials	0.158	0.080	-0.054	-0.036	0.089	0.029	0.089	-0.015
IP: Durable Materials	0.172	0.054	-0.067	-0.014	0.086	0.016	0.056	-0.073
IP: Nondurable Materials	0.113	0.047	-0.047	-0.032	0.046	0.060	0.080	0.002
IP: Manufacturing (SIC)	0.185	0.064	-0.068	-0.011	0.115	0.062	0.112	-0.036
IP: Residential Utilities	0.000	0.005	-0.014	-0.032	0.003	0.059	0.093	0.162
IP: Fuels	0.021	0.017	-0.001	-0.006	0.022	0.003	0.095	-0.008
ISM Manufacturing: Production Index	0.172	0.081	-0.098	-0.020	0.113	0.056	0.118	-0.052
Capacity Utilization: Manufacturing	0.072	0.007	-0.067	0.015	-0.008	-0.024	-0.045	-0.042
Labor market								
Help-Wanted Index for United States	0.105	0.032	-0.084	0.021	-0.016	-0.042	-0.055	-0.028
Ratio of Help Wanted/No. Unemployed	0.042	-0.040	0.049	0.028	-0.014	0.008	-0.010	-0.094
Civilian Labor Force	0.130	-0.012	0.007	0.027	-0.007	-0.063	-0.050	-0.070
Civilian Employment	-0.131	-0.032	0.054	0.000	-0.005	0.097	0.064	-0.024
Civilian Unemployment Rate	-0.041	0.019	-0.047	-0.037	0.019	0.145	0.045	-0.151
Average Duration of Unemployment (Weeks)	-0.022	-0.017	0.043	0.023	0.032	-0.041	0.000	0.025
Civilians Unemployed - Less Than 5 Weeks	-0.065	-0.016	0.047	-0.007	-0.024	0.013	0.033	0.034
Civilians Unemployed for 5–14 Weeks	-0.115	-0.019	0.013	-0.052	-0.031	0.190	0.082	-0.099
Civilians Unemployed - 15 Weeks & Over	-0.075	-0.011	0.036	-0.038	-0.048	0.097	0.029	-0.041
Civilians Unemployed for 15–26 Weeks	-0.092	-0.024	-0.020	-0.036	0.000	0.173	0.090	-0.097
Civilians Unemployed for 27 Weeks and Over	-0.087	-0.066	0.078	0.006	-0.015	-0.091	0.032	0.078
Initial Claims	0.190	-0.009	0.016	0.042	0.034	-0.105	-0.077	0.046
All Employees: Total nonfarm	0.190	0.022	-0.005	0.018	0.051	-0.142	-0.087	0.015
All Employees: Goods-Producing Industries	0.040	-0.018	0.059	-0.066	0.106	-0.061	-0.057	-0.023
All Employees: Mining and Logging: Mining	0.145	0.016	-0.001	0.081	-0.011	-0.083	-0.050	0.015
All Employees: Construction	0.183	0.030	-0.015	-0.015	0.069	-0.147	-0.077	0.011
All Employees: Manufacturing	0.181	0.030	-0.006	-0.021	0.058	-0.144	-0.073	-0.011
All Employees: Durable goods	0.143	0.024	-0.035	0.006	0.082	-0.123	-0.068	0.069
All Employees: Nondurable goods	0.157	-0.039	0.047	0.055	0.025	-0.062	-0.070	0.063
All Employees: Service-Providing Industries	0.168	-0.032	0.017	0.031	0.027	-0.103	-0.103	0.028
All Employees: Trade, Transportation & Utilities	0.161	-0.051	0.050	0.011	0.028	-0.121	-0.113	0.010
All Employees: Wholesale Trade	0.143	-0.027	0.010	0.041	0.012	-0.065	-0.060	0.025
All Employees: Retail Trade	0.123	-0.054	0.098	0.062	-0.019	-0.011	-0.047	0.069
All Employees: Financial Activities	0.020	-0.028	0.072	0.036	-0.037	0.027	0.059	0.073
All Employees: Government	0.064	0.063	-0.065	-0.004	0.040	-0.206	-0.135	0.090
Avg Weekly Hours: Goods-Producing	0.069	0.029	-0.066	-0.034	0.027	0.077	0.020	-0.056
Avg Weekly Overtime Hours: Manufacturing	0.065	0.063	-0.069	-0.006	0.024	-0.209	-0.130	0.074
Avg Weekly Hours: Manufacturing	0.138	-0.086	0.150	0.120	-0.171	0.071	0.072	0.010
ISM Manufacturing: Employment Index	0.110	-0.085	0.148	0.117	-0.123	0.067	0.067	-0.007
Housing								
Housing Starts: Total New Privately Owned	0.120	-0.071	0.136	0.102	-0.150	0.078	0.083	-0.024
Housing Starts, Northeast	0.132	-0.076	0.138	0.107	-0.161	0.062	0.058	0.024
Housing Starts, Midwest	0.134	-0.087	0.135	0.121	-0.177	0.058	0.063	0.017
Housing Starts, South	0.139	-0.078	0.124	0.122	-0.193	0.057	0.061	0.023
Housing Starts, West	0.121	-0.084	0.132	0.119	-0.156	0.058	0.064	0.002
New Private Housing Permits (SAAR)	0.128	-0.066	0.108	0.116	-0.180	0.059	0.076	-0.012
New Private Housing Permits, Northeast (SAAR)	0.120	-0.060	0.093	0.099	-0.172	0.038	0.036	0.040
New Private Housing Permits, Midwest (SAAR)	0.136	-0.083	0.134	0.118	-0.183	0.061	0.060	0.025
New Private Housing Permits, South (SAAR)	0.090	-0.136	-0.041	-0.181	0.065	0.135	-0.035	-0.138
New Private Housing Permits, West (SAAR)	0.078	0.016	-0.034	0.005	0.080	0.092	-0.018	-0.104
Orders & inventories								
ISM: PMI Composite Index	0.043	0.004	-0.006	0.008	0.066	0.057	-0.015	-0.062
ISM: New Orders Index	0.099	-0.073	0.103	-0.027	0.062	-0.064	-0.078	-0.055
ISM: Supplier Deliveries Index	0.073	-0.078	0.133	-0.050	0.092	-0.143	-0.100	-0.066
ISM: Inventories Index	-0.091	0.015	0.120	-0.028	-0.067	-0.206	-0.025	0.026

New Orders for Consumer Goods	-0.024	-0.001	0.010	0.075	0.005	-0.042	0.032	-0.031
New Orders for Durable Goods	-0.014	0.005	0.006	0.088	0.043	-0.117	0.081	-0.080
New Orders for Nondefense Capital Goods	-0.022	0.122	-0.031	0.097	-0.081	-0.036	0.085	0.046
Unfilled Orders for Durable Goods	-0.006	0.037	0.017	0.004	0.005	-0.069	0.088	0.015
Total Business Inventories	0.005	0.067	0.054	-0.005	-0.010	-0.038	0.095	-0.021
Total Business: Inventories to Sales Ratio	-0.009	0.044	0.039	0.018	-0.007	-0.047	0.109	-0.004
Money & credit								
M1 Money Stock	0.024	-0.008	-0.030	-0.042	-0.038	-0.013	0.042	-0.054
M2 Money Stock	0.002	0.037	-0.002	-0.004	-0.016	-0.026	0.083	0.008
Real M2 Money Stock	0.010	-0.019	-0.030	-0.006	-0.019	0.004	0.020	0.076
St. Louis Adjusted Monetary Base	0.012	-0.015	-0.034	0.004	-0.098	-0.097	0.021	0.151
Total Reserves of Depository Institutions	0.010	-0.016	-0.086	0.133	0.035	0.216	-0.311	0.062
Reserves Of Depository Institutions	0.010	-0.018	-0.089	0.126	0.031	0.217	-0.310	0.053
Commercial and Industrial Loans	0.008	-0.004	0.091	-0.158	-0.026	-0.230	0.249	-0.049
Real Estate Loans at All Commercial Banks	-0.036	0.002	-0.083	0.114	-0.014	0.256	-0.182	0.035
Total Nonrevolving Credit	0.101	-0.033	0.014	-0.165	-0.032	0.029	-0.075	-0.037
Nonrevolving consumer credit to Personal Income	0.100	-0.018	-0.002	-0.233	-0.116	0.030	-0.046	-0.067
S&P's Common Stock Price Index: Composite	0.089	-0.027	-0.020	-0.231	-0.135	0.065	-0.096	-0.052
S&P's Common Stock Price Index: Industrials	0.092	-0.026	-0.014	-0.263	-0.138	0.057	-0.080	-0.059
S&P's Composite Common Stock: Dividend Yield	0.091	-0.021	-0.018	-0.271	-0.152	0.053	-0.078	-0.064
S&P's Composite Common Stock: Price-Earnings Ratio	0.074	-0.030	-0.026	-0.265	-0.154	0.059	-0.070	-0.053
Effective Federal Funds Rate	0.063	-0.028	-0.019	-0.254	-0.148	0.051	-0.052	-0.071
3-Month AA Financial Commercial Paper Rate	0.058	-0.021	0.011	-0.264	-0.121	0.009	0.004	-0.082
3-Month Treasury Bill	0.039	-0.003	0.065	-0.254	-0.095	-0.040	0.048	-0.051
6-Month Treasury Bill	0.005	0.122	-0.140	-0.022	-0.210	-0.080	0.050	-0.011
1-Year Treasury Rate	0.035	0.133	-0.202	0.043	-0.206	-0.064	-0.028	-0.008
5-Year Treasury Rate	0.040	0.136	-0.198	0.054	-0.225	-0.068	-0.009	-0.013
10-Year Treasury Rate	0.052	0.124	-0.177	0.055	-0.233	-0.038	0.007	-0.008
Moody's Seasoned Aaa Corporate Bond Yield	0.021	0.142	-0.206	0.088	-0.194	-0.030	0.038	-0.023
Moody's Seasoned Baa Corporate Bond Yield	0.007	0.150	-0.219	0.079	-0.169	-0.045	0.028	-0.035
3-Month Commercial Paper Minus FEDFUNDS	-0.018	0.154	-0.224	0.074	-0.155	-0.054	0.038	-0.029
3-Month Treasury C Minus FEDFUNDS	-0.041	0.152	-0.221	0.072	-0.144	-0.026	0.053	-0.030
6-Month Treasury C Minus FEDFUNDS	0.006	0.066	0.028	-0.158	-0.017	0.017	0.035	0.411
1-Year Treasury C Minus FEDFUNDS	0.014	0.041	0.001	-0.139	-0.029	0.065	-0.023	0.374
5-Year Treasury C Minus FEDFUNDS	0.004	0.010	0.012	-0.123	-0.014	0.057	-0.061	0.306
10-Year Treasury C Minus FEDFUNDS	0.007	-0.052	-0.012	0.105	0.007	-0.020	-0.048	-0.349
Moody's Aaa Corporate Bond Minus FEDFUNDS	0.007	0.081	0.048	-0.090	-0.022	-0.075	0.144	0.157
Moody's Baa Corporate Bond Minus FEDFUNDS	-0.002	-0.200	-0.129	0.017	-0.008	-0.040	0.024	0.046
Trade Weighted U.S. Dollar Index: Major Currencies	-0.003	-0.203	-0.133	0.018	-0.010	-0.039	0.025	0.049
Switzerland/U.S. Foreign Exchange Rate	0.008	-0.190	-0.138	0.031	0.014	-0.019	0.003	0.003
Japan/U.S. Foreign Exchange Rate	-0.008	-0.142	-0.094	0.024	-0.002	-0.025	0.049	0.027
U.S./U.K. Foreign Exchange Rate	-0.004	-0.106	-0.069	0.007	-0.016	-0.014	-0.003	0.031
Canada/U.S. Foreign Exchange Rate	0.001	-0.039	-0.041	0.037	0.002	0.017	-0.033	-0.090
Prices								
PPI: Finished Goods	0.010	-0.239	-0.161	-0.003	0.008	-0.055	0.066	0.028
PPI: Finished Consumer Goods	0.012	-0.029	-0.035	-0.010	0.007	0.033	0.005	0.047
PPI: Intermediate Materials	0.001	-0.223	-0.154	0.014	0.001	-0.025	0.038	0.012
PPI: Crude Materials	0.007	0.009	0.011	0.011	0.032	0.002	-0.016	0.007
Crude Oil, spliced WTI and Cushing	0.002	-0.245	-0.165	0.007	-0.003	-0.042	0.057	0.039
PPI: Metals and metal products:	0.001	-0.051	-0.034	-0.025	-0.008	0.017	0.016	0.043
ISM Manufacturing: Prices Index	0.022	-0.063	-0.042	-0.024	0.028	0.008	0.009	-0.039
CPI: All Items	0.012	-0.224	-0.154	-0.001	0.013	-0.040	0.054	0.022
CPI: Apparel	0.005	-0.243	-0.166	0.000	-0.002	-0.044	0.063	0.030
CPI: Transportation	0.012	-0.239	-0.162	-0.005	0.011	-0.052	0.070	0.021
CPI: Medical Care	0.008	-0.223	-0.154	0.006	0.005	-0.022	0.042	0.025
CPI: Commodities	0.007	-0.048	-0.033	-0.002	0.007	0.015	-0.005	0.023
CPI: Durables	0.004	-0.240	-0.164	0.009	-0.002	-0.043	0.059	0.035
CPI: Services	0.008	-0.059	-0.041	0.001	0.020	-0.018	-0.013	-0.048
CPI: All Items Less Food	-0.003	-0.003	0.004	0.024	-0.016	-0.050	-0.044	-0.150
CPI: All items less shelter	-0.022	-0.013	0.015	0.014	-0.046	-0.053	-0.051	-0.059
CPI: All items less medical care	0.007	0.005	-0.006	-0.002	-0.012	-0.034	-0.015	-0.151
Consumption								
Personal Cons. Expend.: Chain Index	-0.003	0.031	-0.056	0.015	-0.072	0.123	-0.188	0.103
Personal Cons. Exp: Durable goods	-0.030	0.005	0.015	0.132	0.067	-0.127	0.058	-0.057
Personal Cons. Exp: Nondurable goods	0.003	0.000	0.007	-0.021	-0.017	0.016	-0.001	0.067
Personal Cons. Exp: Services	0.006	0.004	0.003	-0.015	-0.013	0.008	-0.003	0.068
Avg Hourly Earnings: Goods-Producing	-0.004	-0.021	-0.024	0.004	0.011	0.024	0.033	-0.048
Avg Hourly Earnings: Construction	-0.091	0.012	0.040	-0.038	-0.004	0.007	0.226	-0.049

Table A.13: The factor loadings of the 128 macro variables on the first 8 components. The sample period is 1971:08-2018:12. Variable descriptions are taken from McCracken and Ng (2016).