Modelling the Term Structure with Trends in Yields and Cycles in Excess Returns*

Carlo A. Favero† Rubén Fernández-Fuertes‡

First Version: November 3, 2023
This Version: November 8, 2023

Abstract

This paper proposes an Affine Macro Term Structure model in which yields are drifting, sharing a common stochastic trend driven by the drift in short-term (monetary policy) rates and excess returns are stationary as the compensation for risk is driven by the cycles in yields. We apply the approach to US data and compare the empirical results from the new specification with those obtained from standard Affine Term Structure models. The cycle-trend decomposition-based Affine Term Structure model produces much better forecasts of the dynamics of yields and, consequently, different and stationary dynamics for the term premia.

---

*We are grateful to Daniel Gros for stimulating discussions and comments and to Refet Gurkaynak for comments and suggestions. Financial Support by the European Union-Next Generation EU, project GRINS–Growing Resilient, INclusive and Sustainable PE00000018 (CUP B43C22000760006) is gratefully acknowledged. This paper is part of the research activities of the Baffi Sentre unit on Macroeconomic Trends, Cycles, and Asset Prices.

†* Institute for European Policymaking Bocconi University & Innocenzo Gasparini Institute for Economic Research (IGIER) Centre& Centre for Economic Policy Research (CEPR), Bocconi University, Department of Economics, via Roentgen 1, 20136 Milano, Italy, carlo.favero@unibocconi.it.

‡Bocconi University, Department of Finance, via Roentgen 1, 20136 Milano, Italy, ruben.fernandez@phd.unibocconi.it.

JEL codes: E43, E52, G12.
Keywords: Affine Term Structure Models, Trends and Cycles, Term Premia.

1 Introduction

This paper proposes a new Affine Macro Term Structure model in which yields are drifting, sharing a common stochastic trend driven by the drift in short-term (monetary policy) rates and excess returns are stationary as the compensation for risk depends on the cycle in yields. This approach is strongly motivated by the data and addresses a gap in the existing literature that adopts a common factor structure for yields and excess returns.

1.1 A First Look at the Data

The quarterly US data from the last fourty years on the term structure of Government bonds show the presence of a common drift in yields to maturity which disappears when 1-period excess holding returns for bonds at all maturities are considered. Figure 1 reports the quarterly time-series observations on the Treasury yield curve estimates of the Federal Reserve Board made available by Gürkaynak et al. (2007) over the period 1980-2023. Yields at maturities from 1 year to 15 year show the presence of a common drift shared with that of interest rates on 1-period bond (the three-month rate). Figure 2 reports the observed 1-quarter returns of holding bonds at all maturities from 5 to 15 years in excess over the return on three-month Treasury Bills. No trend is evident from the data. The stationarity of one-period excess holding returns has two immediate implications. First, term premia at all maturities, being average of expected one-period excess holding returns over the residual maturity of bonds, are also stationary. Second, the common drift component in the term structure is driven by the trend in the one-period bonds and it is removed when

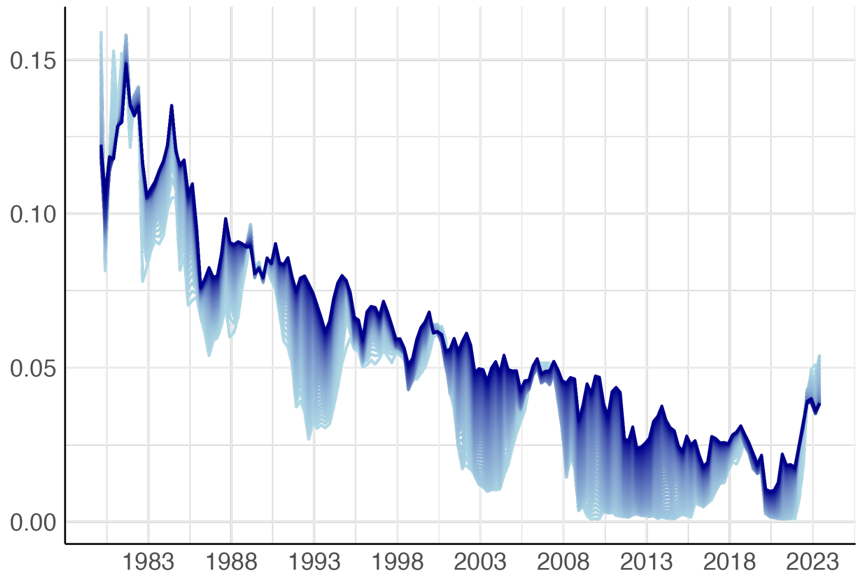


FIGURE 1. Quarterly observations on the time-series of (annualised) yields from the 3-month to the 15-year maturity. We use the same colour palette for all maturities (blue). Darkest blue indicates the highest maturity, i.e., 15 years.

spreads of bonds at all maturities on the one period bond are considered.[1]

[1]One period excess holding returns for a bond with maturity of n periods at time t are defined as $rx_{t+1}^{(n-1)} = R_t^{(n)} - r_t - (n-1)\left[R_{t+1}^{(n-1)} - R_t^{(n)}\right]$

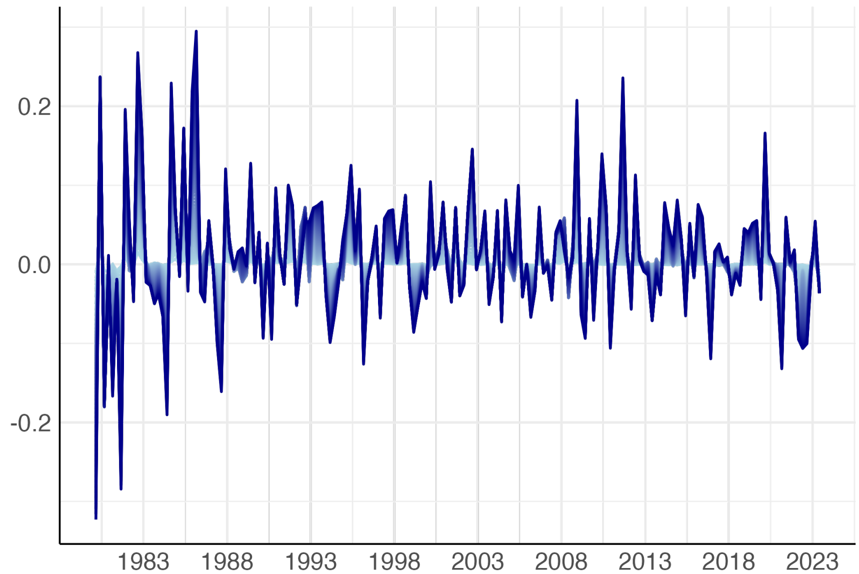


FIGURE 2. Quarterly observations on the time series of 1-quarter holding period returns for bonds at maturities between 5 and 15 years in excess of the return on three-month Treasury Bills

1.2 The Literature

Macro-finance models of the term structure mostly belong to the class of Affine Term Structure Models (Diebold et al., 2005). These models are originally designed for stationary processes in yields, as the yield dynamics is modelled as a vector autoregression (VAR) of a set of factors extracted from the term structure partially, like Ang and Piazzesi (2003), or totally, like Kim and Wright (2005) and Adrian et al. (2015); and VAR models are used for forecasting stationary processes. Importantly, The factor dynamics also drives the price of risk and holding period returns. The presence of a stochastic trend in yields has several negative consequences for this approach, in view of the stationary nature of excess returns of buy-and-hold strategies. VAR models are inappropriate for long-run forecasting of non-stationary data, biased forecast of the dynamics of short-term rates[2] do affect the measurement of term premia. The non-stationarity of factors might results in non-stationarity of term premia, which is counterfactual with respect to the empirical evidence of stationarity of holding period (excess) returns.

Several papers have documented the existence of a slow-moving component common to the entire term structure (see, for example, Bakshi and Chen, 1994 and Fama, 2006). An important and growing literature has modeled Treasury yields using shifting endpoints (Kozicki and Tinsley, 2001), near-cointegration (Jardet et al., 2013) or long memory (Golinski and Zaffaroni, 2016), vector autoregressive models (VAR) with common trends (Negro et al., 2017), slow-moving averages of inflation (Cieslak and Povala, 2015) and consumption (Jørgensen, 2018), or an (unobserved) stochastic trend common across Treasury yields (Bauer and Rudebusch, 2020). Interestingly, (Bauer and Rudebusch, 2020), in their model that allows for a trend in yields and returns, note that the loading of returns on the unobserved common stochastic trend is an order of magnitude smaller than the loading of prices and they also report that predictive regressions of returns on de-trended yields and trend proxies lead to coefficients on the trend that are not significantly different from zero. Campbell and Shiller (1987) have proposed a stationary representation of spreads and changes in short-term rates, based on cointegration between short-term rates and yields at any maturity, but their approach has never found its way in Affine Term Structure models. Piazzesi et al. (2015) use survey data on interest rate forecasts to construct subjective bond risk premia to find that subjective premia are less volatile and not very cyclical.

[2] It has also been recognised that OLS estimates of near-unit roots are notoriously biased downward, thus overestimating the amount of mean reversion in yields.

They explain this evidence by pointing out that survey forecasts of interest rates are made as if both the level and the slope of the yield curve are more persistent than under common statistical models. Zhao (2020) proposes a structural model of trends and cycles in the term structure capable of explaining several features of the data, without relating the trend component of the yield curve to observable slow-moving variables, such as the demographic structure of the population, whose properties can be exploited for forecasting purposes.

1.3 Our Contribution

The objective of our contribution is to build an Affine No Arbitrage Term Structure model consistent with the evidence from the data that yields are non-stationary and driven by a common trend and excess returns are stationary. Following Favero et al. (2022) and Favero and Fernandez-Fuertes (2023), we decompose short term rates in a trend component and a cycle component. The trend component is driven by the very long-run forecast of the central bank for real short term rates and by its response to the very long-run forecast for inflation. The very long-run forecast for the real rates is labelled in the literature as the natural rate of interest. We model the natural rate as function of the equilibrium growth rate of output in the economy and the age structure of population, a time-varying determinant of household preferences. Given availability of long-run forecast for the growth rate of the economy, the age structure of population, and long-term expected inflation, we measure the trend component of the short term rate and its expectations and construct the current and future trend component of the short-term rates. Trends for yields at any maturity are then identified by taking the appropriate average of the future trends in the short rates over the duration of each bonds. Finally, factors are extracted from the cyclical components of bonds at all maturities. A stationary VAR for the cyclical components of yields is then estimated. Excess returns and term premia are driven by these stationary variables. Prediction for short-term rates at any future dates are then derived by combining the predictions for the trends (not based on the VAR for factors) and the predictions for the cyclical components (based on the VAR for factors). Then bonds at any maturities are priced via pricing equations that imposes no-arbitrage restrictions. Term premia are derived as the difference of bond yields obtained when the price of risk is estimated in the affine specification and when the price of risk is restricted to zero. Bond yields are non-stationary, but their trend is the average trend of short-term rates over the maturities of the bond and term-premia are driven by the stationary state variables.

Our new specification has implications for forecasting and measurement of the risk premia. We show that our approach has better forecasting performance and leads to a measurement of term premia very different from that of standard models that do not address the relevant features of the data. These differences are particularly relevant when fluctuations in the risk premia are used to evaluate the macroeconomic implications of monetary policy. (Schnabel, 2022)

2 An Affine Term Structure model with Trend and Cycle in Monetary Policy Rates

Affine models of the term structure of interest rates are a popular way of determining the term premia. The expectation of the future path of short rates can be extracted from these term structure models. The affine models typically use state variables (latent factors) to model the shocks that drive the economy. The key assumptions are: First, the pricing kernel is exponentially affine in the state variables, whose dynamics is described by a VAR. Second, market prices of risk are affine in the state variables. Finally, the innovations to state variables and one-period holding excess returns are jointly normal-distributed.

Using these assumptions together with no-arbitrage restrictions delivers generating processes for continuously compounded excess returns and continuously compounded yields at any maturity that are a function of the state variables. Yields can be decomposed into a term premium, a convexity correction, and a part reflecting expectations for the one-period rate over the residual life of the bond. In the light of the evidence reported in the introduction, this specification strategy suffers from a clear shortcoming: the state variables have to capture the drift in the data, and a VAR model is not the most appropriate specification for long-run projections of the relevant variables. Indeed, long-run projections are needed because pricing a long-dated bond with quarterly data will require to project of the three-month rates over an horizon equal to the maturity of the bond.

To deal with this problem, we propose to specify an Affine Term Structure model with two sets of states variables: the trending ones and the stationary ones. The trending variables will be related to slow moving components in the structure of the economy and will not be predicted by a VAR, the VAR specification will then be limited to the stationary state variables.

2.1 Detrending the Term Structure to model excess returns

The identification of the two sets of state variables is implemented starting from the specification of the one-period nominal risk free rate $r_t^{(1)}$. The risk free rate can be decomposed in a trend and a cyclee. The trend, i.e. long-run risk free rate, is made of two components: the natural rate of interest, r_t^*, and a component that reflects long-term inflation expectations.

Laubach and Williams (2003) show that in the standard Ramsey model households intertemporal optimization delivers a positive relationship between the natural rate of interest and both the growth rate of output in the economy and household preferences. This motivates the inclusion of (log) growth rate of potential output, Δy_t^{pot}, as a variable explaining the trend. However, Jordà and Taylor (2019) and Mian et al. (2021) illustrate that fluctuations in output growth (*per capita*) of the economy cannot fully explain the drift in natural rate, therefore, other time-varying determinants of the rate of time preference of the agents in the economy should be considered. On the one hand, we follow Favero et al. (2016), Lunsford and West (2019), and Favero et al. (2022), and consider the age structure of the population as the driver of changing preferences, in particular MY_t, the ratio of middle-aged (40-49) to young (20-29) population. On the other hand, Gürkaynak et al. (2005) convincingly argue that private agents views of long-run infations are subject to fluctuations. In line with this evidence we use the survey-based measure of long-run inflation expectations, π_t^{LR}, also considered in the Fed's FRB/US model[3] as the proxy for long-run inflation expectations. This is a reasonable proxy under the assumption that the central bank is credible. The cyclical part of the yield can be thus identified with the residual after

[3] Available at https://www.federalreserve.gov/econres/us-models-package.htm.

regressing the short rate on those three variables, Δy_t^{pot}, MY_t, and π_t^{LR}.

Once the trend and the cycle in the one-period rate are identified, the trend and the cycle for yields at all maturities can be constructed by taking the appropriate average of the expected trends in the one-period rate:

$$r_t^{(1)} = r_t^{*,(1)} + u_t^{(1)} \tag{1}$$

$$r_t^{*,(1)} = \gamma_1 MY_t + \gamma_2 \Delta y_t^{pot} + \gamma_3 \pi_t^{LR} \tag{2}$$

$$r_t^{(n)} = r_t^{*,(n)} + u_t^{(n)} \tag{3}$$

$$r_t^{*,(n)} = \frac{1}{n}\sum_{i=0}^{n-1} r_{t+i}^{*,(1)} \tag{4}$$

The model is naturally interpreted within a cointegration approach (Engle and Granger, 1987) to model the stochastic drift in rates: if demographics, productivity and the inflation target of the central bank successfully capture the trend in nominal rates, then $u_t^{(1)}$ should be stationary. Stationarity of $u_t^{(1)}$, paired with stationarity of the term premia, implies that $u_t^{(n)}$ are stationary. Note also that, in this framework the stochastic trends in yields at all maturities are all driven by the trend in one period rates.

Long-run forecast for MY_{t+i}, Δy_{t+1}^{pot}, π_{t+i}^{LR} are readily available in that demographics and potential output long-term forecast can be respectively downloaded from the Bureau of Census and the Fred database, while credibility of the central bank implies that long forecast for inflation cannot diverge from the CB target. Therefore, no VAR is needed to obtain $R_{t+i}^{(1),*}$, as these forecasts can be derived directly by using (1) with the appropriate scenario for the exogenous variables MY_{t+i}, Δy_{t+1}^{pot}, π_{t+i}^{LR}. After this, K factors can now be extracted by obtaining the principal components to the

N cyclical components of the yield curve $u_t^{(j)}$, for $j = 1, \ldots, n$, which we stack into a $T \times N$ matrix, $\mathbf{U}$. We denote these K factors as $X_t \in \mathbb{R}^K$, and they are the first K principal components of $\mathbf{U}$. This procedure ensures the stationarity of X_t to specoify a VAR, i.e.,

$$X_{t+1} = \mu + \Phi X_t + v_{t+1} \tag{5}$$

$$v_{t+1} | \left(X_s\right)_{s=0}^{t} \sim \mathcal{N}(0, \Sigma), \tag{6}$$

where $\mu \in \mathbb{R}^K$, $\Phi \in \mathbb{R}^{K \times K}$ and $\Sigma \in \mathbb{R}^{K \times K}$. On the other hand, the variables in X_t determine the market price of risk, λ_t, in the following affine form:

$$\lambda_t = \Sigma^{-1/2}(\lambda_0 + \lambda_1 X_t), \tag{7}$$

The assumption of no-arbitrage implies that there exists a pricing kernel, M_t, such that:

$$P_t^{(n)} = \mathbb{E}_t \left(M_{t+1} P_{t+1}^{(n-1)} \right), \tag{8}$$

for every $n > 0$ and $t \geq 0$, and where $P_t^{(n)} = \exp\left[-nr_t^{(n)}\right]$ is the price of a zero coupon bond with maturity n. We are strictly following Adrian et al. (2015). Hence we assume that the pricing kernel is exponentially affine, i.e.,

$$m_{t+1} = -r_t^{(1)} - \frac{1}{2}\lambda_t^{\mathrm{T}}\lambda_t - \lambda_t^{\mathrm{T}}\Sigma^{-1/2}v_{t+1}, \tag{9}$$

where $r_t^1 = -\log(P_t^{(1)}) = -p_t^{(1)}$ is the continuously compounded risk-free rate, and

$m_t = \log M_t$. The excess log returns are given by:

$$rx_{t+1}^{(n-1)} = p_{t+1}^{(n-1)} - p_t^{(n)} - r_t, , \tag{10}$$

where $p_t^{(n)} = \log P_t^{(n)}$. After some derivations using (9) and (8) (see Appendix A.1), we arrive to

$$\mathbb{E}_t \left(rx_{t+1}^{(n-1)} \right) = \text{cov}_t \left[rx_{t+1}^{(n-1)}, v'_{t+1} \Sigma^{-1/2} \lambda_t \right] - \frac{1}{2} \mathbb{V}_t \left(rx_{t+1}^{(n-1)} \right), \tag{11}$$

In the same fashion as Adrian et al. (2013), we can define $\beta_t^{(n-1)'}$ as

$$\beta_t^{(n-1)} \coloneqq \Sigma^{-1} \text{cov}_t \left(rx_{t+1}^{(n-1)}, v_{t+1} \right) \in \mathbb{R}^K. \tag{12}$$

By substituting from (12) into (11) and using (7), we have:

$$\mathbb{E}_t \left[rx_{t+1}^{(n-1} \right] = \lambda_t \cdot \beta_t^{(n-1)} - \frac{1}{2} \mathbb{V}_t \left[rx_{t+1}^{(n)} \right], \tag{13}$$

The unexpected excess return can be decomposed in a component that is correlated with v_{t+1}, and whose correlation vector coincides with $\beta_t^{(n-1)}$, and another component which is conditionally orthogonal to v_t, and which can be interpreted as the return pricing error:

$$rx_{t+1}^{(n-1)} - \mathbb{E}_t \left(rx_{t+1}^{(n-1)} \right) = \beta_t^{(n-1)} \cdot v_{t+1} + e_{t+1}^{(n-1)}, \tag{14}$$

Under the assumption that the return pricing error are i.i.d. with variance σ^2 and that β_t is constant, the generating process for log excess returns becomes:

$$
\begin{aligned}
rx_{t+1}^{(n-1)} = (\lambda_0 + \lambda_1 X_t)^{\mathrm{T}} \beta^{(n-1)} - \frac{1}{2}\left(\left(\beta^{(n-1)}\right)^{\mathrm{T}} \Sigma \beta^{(n-1)} + \sigma^2\right) \\
+ v_{t+1}^{\mathrm{T}} \beta^{(n-1)} + \varepsilon_{t+1}^{(n-1)},
\end{aligned}
\tag{15}
$$

and so it's clear now that the (log) excess returns can be decomposed into the expected return (first term), a convexity correction (second term), and a return innovation. This expression also allows us to see that the time-varying component of expected excess returns is stationary and driven by the dynamics of the stationary state variables. We can thus stack (15) across N maturities and T time-periods we have the following matrix-form representation:

$$
\mathbf{rx} = \left(\lambda_0 \mathbb{1}_{T\times 1}^{\mathrm{T}} + \lambda_1 \mathbf{X}_{-}^{\mathrm{T}}\right)^{\mathrm{T}} \mathbf{B} - \frac{1}{2}\left(\mathbf{B}^{*} \mathrm{vec}\left(\Sigma\right) + \sigma^2 \mathbb{1}_{K\times 1}\right) \mathbb{1}_{T\times 1}^{\mathrm{T}} + \mathbf{V}^{\mathrm{T}}\mathbf{B} + \mathbf{E} \tag{16}
$$

where $\mathbb{1}_{l\times m}$ is a matrix of ones for each $l, m \in \mathbb{N}$, and

1. $\mathbf{rx} \in \mathbb{R}^{T\times N}$.
2. $\lambda_0 \in \mathbb{R}^{K}$, $\lambda_1 \in \mathbb{R}^{K\times K}$,
3. $\mathbf{X}_{-} = [X_1 \mid X_2 \mid \cdots \mid X_{T-1}]^{\mathrm{T}} \in \mathbb{R}^{T\times K}$,
4. $\mathbf{B} \in \mathbb{R}^{K\times N}$,
5. $\mathbf{B}^{*} = [\mathrm{vec}\left(B_1 B_1^{\mathrm{T}}\right) \mid \cdots \mid \mathrm{vec}\left(B_n B_n^{\mathrm{T}}\right)]^{\mathrm{T}} \in \mathbb{R}^{K\times N^2}$,
6. $\mathbf{V} \in \mathbb{R}^{T\times K}$ and $\mathbf{E} \in \mathbb{R}^{T\times N}$.

2.2 Parameters' Estimation

Estimation of the parameters in our model is implemented by extending the 3-step procedure proposed by Adrian et al. (2013) to a 4-step procedure. All details are found in the Appendix A.2.

1. Construct the cyclical components of yields at all maturities by estimating a (cointegrating) regression of the one-period rate as a function of predictable slow-moving variables and use the available predictions on the drivers of the trend in one-year yields to construct the trend for yields at all maturities by taking the appropriate average of the expected trends in the one-period rate as described in (4).

2. Construct the pricing factors, $\mathbf{X}$, from principal component analysis (PCA) of the cyclical components of yields derived in the first step, $\mathbf{U}$. Estimate the equation (5) using OLS, decomposing the pricing factors into predictable components and factor innovations $\hat{V}$.

3. Regress excess returns on a constant, lagged pricing factors and contemporaneous pricing factor innovations according to

$$\mathbf{rx} = a\mathbb{1}_{T\times K}\mathbb{1}_{K\times N} + \hat{\mathbf{V}}b + \mathbf{X}_{-}c + \mathbf{E} \tag{17}$$

4. We show in the Appendix A.2 that

$$a = \left(\lambda_0 \mathbb{1}_{T\times 1}^{\mathrm{T}}\right)^{\mathrm{T}} \mathbf{B} - \frac{1}{2}\left(\mathbf{B}^{*}\mathrm{vec}\left(\Sigma\right) + \sigma^2 \mathbb{1}_{K\times 1}\right)\mathbb{1}_{T}^{\mathrm{T}} \tag{18}$$

$$c = \lambda_1^{\mathrm{T}}\mathbf{B} \tag{19}$$

From these, market price of risk's estimates are given by

$$\hat{\lambda}_0 = \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^{\mathrm{T}}\right)^{-1}\hat{\mathbf{B}}\left[\hat{a}^{\mathrm{T}} + \frac{1}{2}\mathbb{1}_{T\times 1}\left(\mathbf{B}^{*}\mathrm{vec}\left(\Sigma\right) + \sigma^2 \mathbb{1}_{N\times 1}\right)^{\mathrm{T}}\right], \tag{20}$$

$$\hat{\lambda}_1 = \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^{\mathrm{T}}\right)^{-1}\hat{\mathbf{B}}\hat{c}^{\mathrm{T}}. \tag{21}$$

2.3 Modelling Trending Yields

Bond prices at any maturity can be obtained by recursive forward substitution of prices in (**??**), keeping in mind that the (log) price of all bonds at maturity is zero, i.e., $p_{t+n}^{(0)} = 0$. The cyclical component of the one-period bond r_t^1, i.e., $u_t^{(1)} \coloneqq r_t^{(1)} - r_t^{*,(1)}$, can be expressed as a linear function of the underlying factors, i.e.,

$$
\begin{aligned}
r_t^{(1)} &= r_t^{*,(1)} + \delta_0 + \delta_1 \cdot X_t + e_t^{(1)}, \qquad (22)\\
p_t^{(1)} &= -r_t^{(1)}, \quad p_t^{1,*} = r_t^{*,(1)},
\end{aligned}
$$

Where parameters $\hat{\delta}_0$ and $\hat{\delta}_1$ can be estimated by projecting the cycle $u_t^{(1)}$ on the stationary factors X_t.

Our specification of the no-arbitrage model implies that bond prices depend linearly on a trend component and on a stationary component[4]:

$$
p_t^n = p_t^{n,*} + A_n + B_n' X_t + u_t^n, \qquad (23)
$$

where $p_t^{n,*}$ captures the trend component of bond prices. The model also implies cross-equation restrictions on the parameters A_n, B_n and on the trend $p_t^{n,*}$.

$$
A_n = A_{n-1} + (\mu - \lambda_0)^{\mathrm{T}} B_{n-1} + \frac{1}{2}\left(B_{n-1}^{\mathrm{T}} \Sigma B_{n-1} + \sigma^2\right) - \delta_0 \qquad (24)
$$

$$
B_n = (\Phi - \lambda_1)^{\mathrm{T}} B_{n-1} - \delta_1 \qquad (25)
$$

$$
p_t^{(n),*} = p_{t+1}^{(n-1),*} - r_t^* \qquad (26)
$$

[4] See Appendix A.3 for more details

In this specification, trends affect yields but excess returns are driven exclusively by stationary variables. The main innovation in our proposal is that the vector X_t is extracted from the de-trended term structure and therefore the drivers of the excess-returns are the factors extracted from the cyclical components of the yield curve. Note that our specification imposes on the dynamics of de-trended bond prices exactly the same restrictions that a standard model imposes on the dynamics of bond prices. Hence, the comparison of the output of our model with that of a comparable ATSM model is immediate. In the case of a standard ATSM, the same VAR structure that we use for factors extracted from the cyclical components of yields is adopted directly for factors extracted from yields, without de-trending them. In this specification,

$$r_t^{(1)} = \delta_0 + \delta_1 \cdot X_t^{ACM} + \epsilon_t, \tag{27}$$

$$p_t^{(n)} = C_n + D_n^{\mathrm{T}} X_t^{ACM} + u_t^{(n)}, \tag{28}$$

where the recursive restrictions apply to C_n, and D_n. Basically, everything is the same but the trendy terms are drifting prices and yields. Hence, in this specification yields (trendy) and excess returns (stationary) are driven by the same set of state variables, X_t^{ACM} (Adrian et al., 2013).

2.4 Model Simulation, Forecasting and Term Premia

After the estimation is completed, we have the following model:

$$r_t^{(1)} = r_t^{*,(1)} + u_t^{(1)} \tag{29}$$

$$r_t^{*,(1)} = -\gamma_1 MY_t - \gamma_2 \Delta y_t^{pot} - \gamma_3 \pi_t^{LR} \tag{30}$$

$$r_t^{(n)} = r_t^{*,(n)} + u_t^{(n)} \tag{31}$$

$$r_t^{*,(n)} = \sum_{i=0}^{n-1} r_{t+i}^{*,(1)} \tag{32}$$

$$p_t^{(n)} = p_t^{*,(n)} + A_n + B_n^{\mathrm{T}} X_t + \varepsilon_t^{(n)}, \tag{33}$$

$$X_t = \mu + \Phi X_{t-1} + v_{t+1} \tag{34}$$

in which the factors X_t are extracted from the cyclical components of yields, u_t^n, after the completion of the first stage of estimation. The model fit can be readily assessed, by comparing actual data with fitted data from the model, model forecast are also naturally constructed using the factor structure. Finally, model simulation in two scenarios, a baseline with all parameters set are their fitted values and an alternative one in which the market price of risk is set to zero, i.e. $\lambda_0 = \lambda_1 = 0$, allows to compute term premia as the differences between the model implied yields and the risk neutral yields.

The performance of our model in terms of fit, forecast and the properties of the derived term premia can be compared with that of a standard ATSM model:

$$p_t^{(n)} = C_n + D_n^{\mathrm{T}} X_t^{ACM} + \varepsilon_t^{(n)}, \tag{35}$$

$$X_t^{ACM} = \mu + \Phi X_{t-1}^{ACM} + v_{t+1}, \tag{36}$$

in which estimation is implemented in three steps and the factors X_t^{ACM} are extracted directly from the yield curve (i.e., not detrended).

3 Empirical Results

Estimation and simulation[5] is performed by using the zero coupon yields provided by the FED[6] (Gürkaynak et al., 2007), data on MY_t, the ratio of middle-aged (40-49) to young (20-29) obtained from the Bureau of Census, the survey-based measure of long-run inflation expectations, used in the Fed's FRB/US model[7] and the measure of potential Gross Domestic Product available from the FRED database.[8] Quarterly data over the period 1980:1-2023:2 are considered. In this section, we shall report evidence based on the comparison between the simulation of our model estimated in four steps and a standard ATS model estimated in three steps.

3.1 Detrending Yields

The trend in the one-period (three-month) rate is captured by projecting it on the proxy for the age structure of the population, potential output growth and the survey-based measure of long-run inflation expectations. The results, reported in Table 1, show that the estimated model produces stationary residuals with estimated coefficients on the drivers of the drift on short-term rates in line with previous studies Favero and Fernandez-Fuertes (2023), Bauer and Rudebusch (2020), with a negative

[5]a full replication package in R is available from the authors' website

[6]https://www.federalreserve.gov/econres/feds/the-us-treasury-yield-curve-1961-to-the-present.htm.

[7]https://www.federalreserve.gov/econres/us-models-package.htm.

[8]https://fred.stlouisfed.org/series/GDPPOT.

coefficients on MY capturing the effects of the age structure of the population on the supply of savings, and positive and slightly larger than one coefficients on potential output growth and long-run inflation expectations.

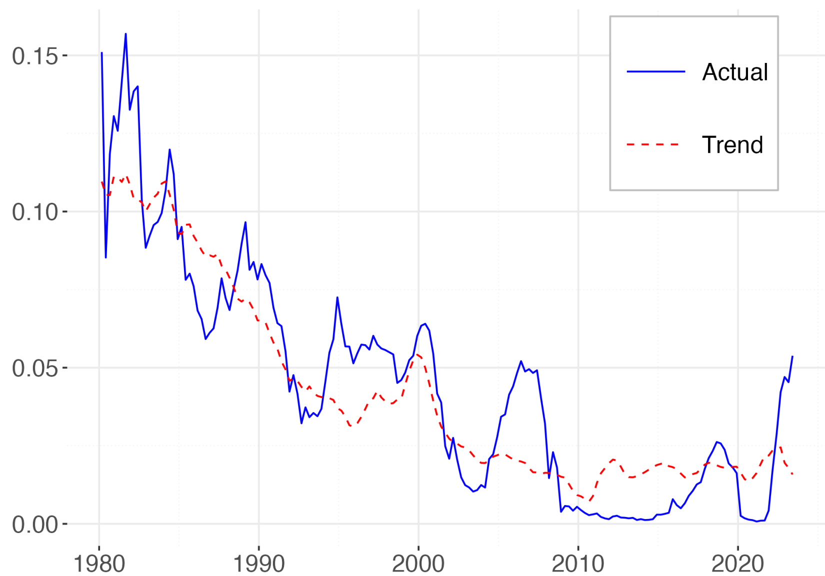
TABLE 1. Modelling the Trend in three-month yields

	Dependent variable:
	$r_t^{(1)}$
MY_t	-0.037^{***}
	(0.004)
Δy_t^{pot}	1.418^{***}
	(0.192)
π_t^{LR}	1.315^{***}
	(0.090)
Observations	174
Adjusted R^2	0.907
ADF test on residuals	-4.66^{***}
Residual Std. Error	0.017 (df = 171)
F Statistic	567.984^{***} (df = 3; 171)
Note:	*p<0.1; **p<0.05; ***p<0.01

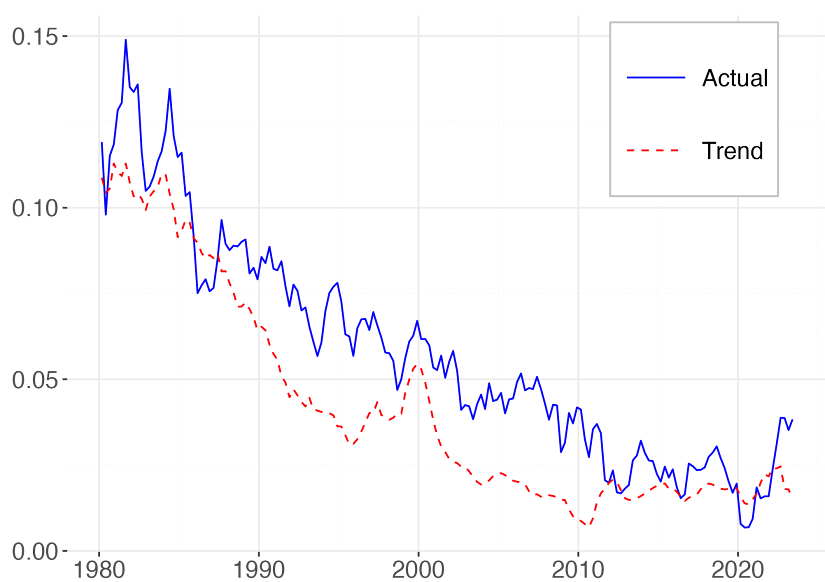
Given the trend component on one-period rates we derive the trend components for yields at any maturity as specified in Section 2. Figure 3 illustrates our results for the 3-month and the 10-year yields. Note that the cyclical components of yields contain information on the term-premia, therefore we expect them to fluctuate around a level that differs across different maturities.

FIGURE 3. Trend Components

(A) Three month yield time series against its trend.



(B) Ten year yield time series against its trend.



3.2 Consequences of TS Trend Presence in VAR factors

Table 2. Eigenvalues associated to the Principal Components' eigenvectors from the baseline model (Panel A) and our model (Panel B).

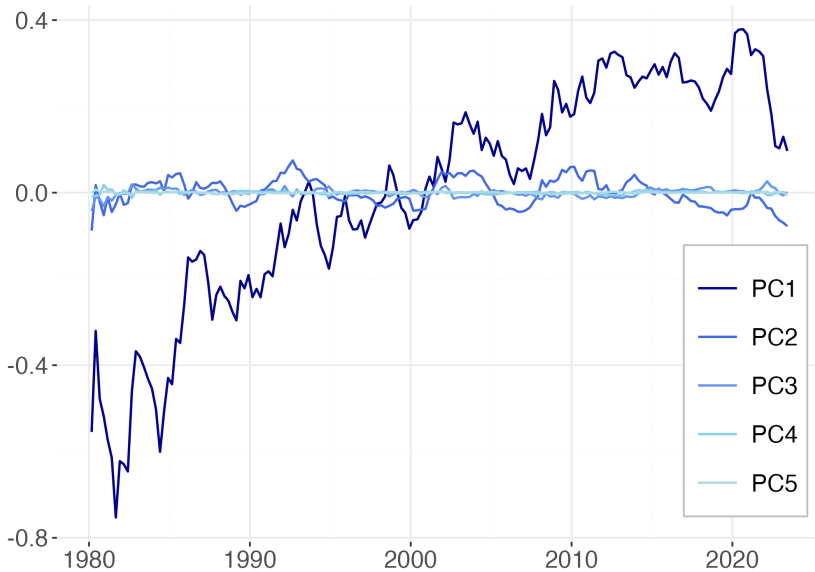
Panel A: PCs on yields (Adrian et al., 2013)				
PC1	PC2	PC3	PC4	PC5
0.9721	0.8949	0.6604	0.44	0.3306

Panel B: PCs on yields' cyclical component				
PC1	PC2	PC3	PC4	PC5
0.8899	0.8899	0.7049	0.5046	0.2898

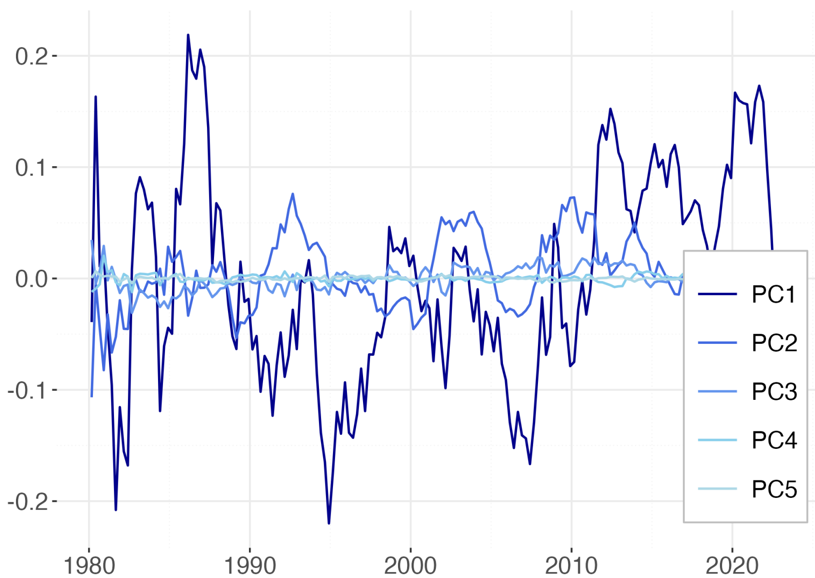
The second step in our approach is the extraction of principal components from detrended yields. We do so by following the path defined by Adrian et al. (2013), i.e,. we consider the first five principal components of a term structure of sixty cyclical components of yields with maturities from three-month (one quarter) to 15-year (sixty quarters). That means that, in align with the notation we have already defined, $N = 60$. Figure 4 illustrates the time series of these factors and compares them with those of the equivalent factors extracted from the term structure of yields wiath maturities from three-month to 15-year. The graphical evidence clearly hints at the presence of a drift in at least one of the factors estimated in the standard approach, while our proposed detrended framework seems successful in removing it.

Figure 4. Principal Components

(a) ACM



(b) FF



Indeed, if we look at the five corresponding eigenvalues, i.e., the roots of the characteristic polynomial for the two alternative VAR specification, we see that there exists a near unit-root in the VAR associated to the ACM model in which there's no correction for potential trends. As we see in Table 2, the highest eigenvalue, i.e., the one associated with the first principal component, and the one accounting by most of the variation, equals $0.9721 \approx 1$. However, this unit root is completely eliminated in our model, in which the highest eigenvalue is just 0.88. Also, notice that we do not have anymore just one principal component explaining much of the variance (which in the case of the standard model was the first one, i.e., the one associated with the unit root), but two of them, both with the same eigenvalue at least in the first four decimals[9]. Hence, it seems that correcting for the presence of trends in the term structure puts us in the good path when modelling the underlying driving factors as a vector autoregression.

3.3 Excess Returns regressions

We report in Table 3 the results of regressing excess returns on a constant, lagged pricing factors, X_t, and contemporaneous pricing factor innovations, v_{t+1}, for the standard factor specification and our factor specification in the spirit of equation (15). In particular, we consider the R^2 from the "predictive" specification in which contemporaneous pricing factor innovations are not included and the full specification and compare them with the version in which contemporaneous pricing factor innovations are included in both models, ACM and FF. It's worth highlighting that the

[9]Multiplicity higher than one may make the geometrical interpretation more challenging, since any rotation among these components would be equivalent, but this fact does not pose any threat on the final conclusions of our model.

predictive version of the model, which incorporates factors extracted from the cyclical components of yields, outperforms the standard model. Indeed R^2 is below 0.20 across all maturities and around 0.1 in almost all maturities in the standard ACM model, whilst it's higher than 0.10 in all maturities in our model. However, when we consider the full specification, the standard model achieves a nearly perfect fit, with R^2 near one in every maturity, surpassing the alternative FF model's performance..

TABLE 3. This table reports the R^2 of regresing excess returns on, either only the lagged pricing factors, X_t, or on lagged pricing factors together with contemporaneous pricing innovations, v_{t+1}.

Panel A: Standard ACM model

n	4	8	12	16	20	28	40	60
X_t	0.17	0.13	0.1	0.09	0.09	0.09	0.1	0.11
X_t and v_{t+1}	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Panel B: FF

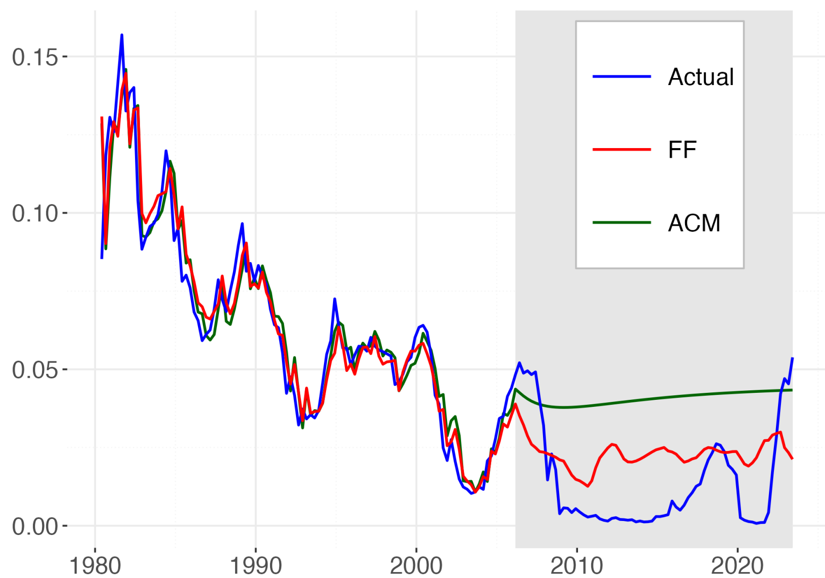
n	4	8	12	16	20	28	40	60
X_t	0.15	0.14	0.13	0.13	0.13	0.13	0.14	0.13
X_t and v_{t+1}	0.98	0.96	0.95	0.94	0.93	0.92	0.91	0.9

3.4 Fit and Forecast Performance

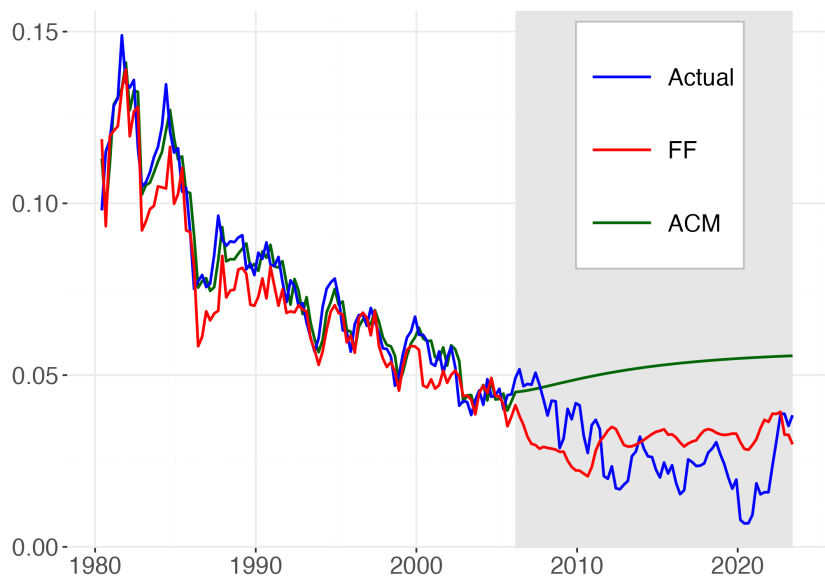
To illustrate fit and forecasting performance of the two alternative specifications, we report in Figure 5 the results of a within-sample model simulation up to 2005:Q4, where current values of the factors are used to predict yields, and of out-sample model simulation from 2006:Q1 onward, where n-step ahead forecasts of the factors (with n going from 1-quarter to 70-quarters) are used to predict yields.

Figure 5. This graph reports the fitted (1980Q1:2005Q4) and the forecasted (2006Q1:2023Q2) time series of 1Y and 10Y yields given by the standard ACM model (green) and our model (red) against the actual values (blue). The shaded area indicates our selected out-of-sample period.

(a) 3-month yield.



(b) Ten year yield.



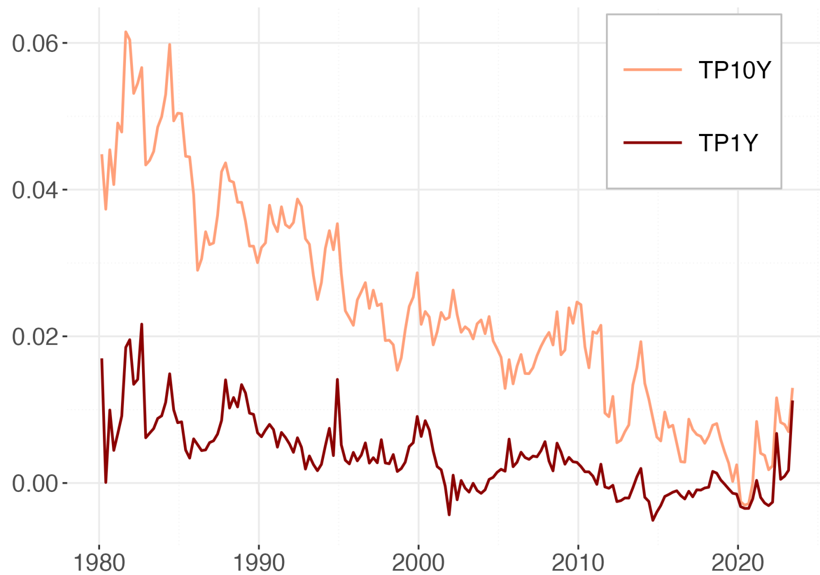
Results for 3-month and 10-year yields are reported. The within-sample performance of our models is slightly inferior to that of the standard model. The reason may be that the trend component of yields is captured with less precision, because the sample is shorter and, e.g., the demographic variable, MY_t, that aims capturing the trend has very low variability in shorter horizons. However, our detrended model clearly dominates out-of-sample, showing the capability of tracking well the long-term dynamics of yields. Indeed the standard model behaves very poorly out-of-sample, as it is evident by simply looking at the picture: the forecasted path is basically a straight line whose level is much higher than the average of the realised one. The presence of a unit root together with the very low R^2 in the regressions without innovations, v_{t+1}, compared to the near one R^2 when including them may explain these phenomena (Panel A, first row of Table 3).

3.5 Term Premia

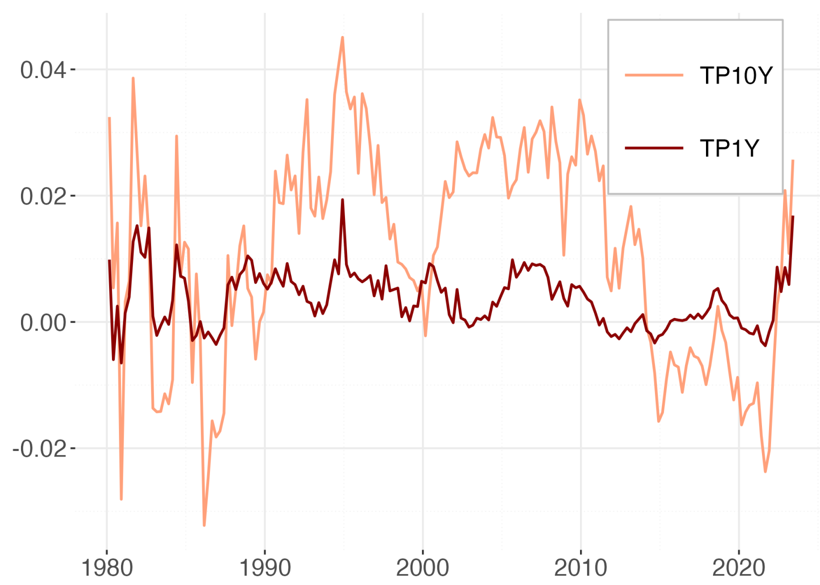
Finally, we analyse term premia in Figure 6. We consider the one-year and the 10-year horizon. The term premiums indicated by the two models appear to be quite distinct: the 10-year term premium suggested by the conventional model exhibits a noticeable trend, in contrast to the model that uses the trend-cycle decomposition of yields, where this trend is less apparent. At shorter durations, like 1 year, the term premia are relatively similar. Therefore, it is essential to analyse yields in terms of trend and cycle to accurately identify term premia over longer periods.

FIGURE 6. This figure reports the 1Y (red) and 10Y (orange) term premia in the two models.

(A) ACM.



(B) FF.



4 Conclusions

Yields to maturity are (co-)drifting and holding period excess returns are (co-)cycling. Standard Affine Term Structure model do not separate trends and cycles in the data, but use factors extracted from yields to maturity to explain holding period excess returns as well as yields to maturity. As a consequence, the empirical model has a rather disappointing performance in predicting short-term rates and generates trending risk premia. This trend is steeper at longer horizons. As risk premia are not observable, term structure models should be evaluated by their performance in predicting the future path of short-term rates. In fact, risk premia are very strongly dependent on this path. We propose a novel way to improve on the standard approach by applying the no-arbitrage restrictions to a model in which the factor structure adopted to explain holding period excess returns is extracted from de-trended yields. The trend in yields is a common trend driven by the drift in short-term rates. The drift in short-term rates in turn is not predicted by a VAR but it is related to long-term forecast for slow-moving variables such as the demographic structure of the population, potential output growth and long-term inflation forecast. A VAR structure is then adopted to model the dynamics of the stationary cyclical components. Our proposed model outperforms the standard approach in forecasting short-term rates and produces stationary risk premia, very different from those produced by the standard approach.

A Appendix

A.1 Derivations

We assume as in Adrian et al. (2013) that the systematic risk is represented by a stochastic vector, $(X_t)_{t\geq 0}$, following a stationary vector autoregression

$$X_t = \mu + \Phi X_{t-1} + v_t \tag{A.1}$$

with initial condition X_0 and whose residual terms, $(v_t)_{t\geq 0}$ follow a Gaussian distribution with variance-covariance matrix, Σ, i.e,.

$$v_t \big| \left(X_s\right)_{0\leq s\leq t} \sim \mathcal{N}\left(0, \Sigma\right). \tag{A.2}$$

Let's denote the zero coupon treasury bond price with maturity n at time t by $P_t^{(n}$. No Arbitrage Dybvig and Ross (1989) holds

$$P_t^{(n)} = \mathbb{E}_t \left[M_{t+1} P_{t+1}^{n-1} \right]. \tag{A.3}$$

Assumption 1. The pricing kernel, $m_{t+1} \coloneqq \log M_{t+1}$, is exponentially affine

$$m_{t+1} = -r_t - \frac{1}{2}||\lambda_t||^2 - \lambda_t^{\mathrm{T}} \Sigma^{-\frac{1}{2}} v_{t+1}, \tag{A.4}$$

where $r_t \coloneqq -p_t^{(1)}$ is the continuously compounded risk-free rate, and $\lambda_t \in \mathbb{R}^K$.

Assumption 2. Market prices of risk are affine

$$\lambda_t = \Sigma^{-\frac{1}{2}} \left(\lambda_0 + \lambda_1 X_t\right), \tag{A.5}$$

where $\lambda_0 \in \mathbb{R}^K$ and $\lambda_1 \in \mathbb{R}^{K\times K}$.

Assumption 3. $\left(rx_t^{(n-1)}, v_t\right)_{t\geq 0}$ are jointly normally distributed.

The excess holding return of a bond maturing in n is given by

$$rx_{t+1}^{(n-1)} := p_{t+1}^{n-1} - p_t^{(n)} - r_t \tag{A.6}$$

Now, (A.3) can be rewritten as

$$\begin{aligned}
1 &= \mathbb{E}_t\left[\exp\left\{m_{t+1} + p_{t+1}^{(n-1)} - p_t^{(1)}\right\}\right] \\
&= \mathbb{E}_t\left[\exp\left\{-r_t - \frac{1}{2}||\lambda_t||^2 - \lambda_t^{\mathrm{T}}\Sigma^{-\frac{1}{2}}v_{t+1} + rx_{t+1}^{(n)} + r_t\right\}\right] \\
&= \mathbb{E}_t\left[\exp\left\{rx_{t+1}^{(n)} - \frac{1}{2}||\lambda_t||^2 - \lambda_t^{\mathrm{T}}\Sigma^{-\frac{1}{2}}v_{t+1}\right\}\right] \\
&= \exp\left\{\mathbb{E}_t\left[\xi_{t+1}\right] + \frac{1}{2}\mathbb{V}\left[\xi_{t+1}\right]\right\},
\end{aligned} \tag{A.7}$$

where $\xi_{t+1} := rx_{t+1}^{(n)} - \frac{1}{2}||\lambda_t||^2 - \lambda_t^{\mathrm{T}}\Sigma^{-\frac{1}{2}}v_{t+1}$, and

$$\mathbb{E}_t\left[rx_{t+1}^{(n-1)}\right] = \mathbb{E}_t\left[rx_{t+1}^{(n-1)}\right] - \frac{1}{2}||\lambda_t||^2 \tag{A.8}$$

$$\begin{aligned}
\mathbb{V}_t\left[rx_{t+1}^{(n-1)}\right] &= \mathbb{V}_t\left[rx_{t+1}^{(n-1)} - \lambda_t^{\mathrm{T}}\Sigma^{-\frac{1}{2}}v_{t+1}\right] \\
&= \mathbb{V}_t\left[rx_{t+1}^{(n-1)}\right] + \mathbb{V}_t\left[\lambda_t^{\mathrm{T}}\Sigma^{-\frac{1}{2}}v_{t+1}\right] - 2\mathrm{cov}\left(rx_{t+1}^{(n-1)}, \lambda_t^{\mathrm{T}}\Sigma^{-\frac{1}{2}}v_{t+1}\right) \\
&= \mathbb{V}_t\left[rx_{t+1}^{(n-1)}\right] + \lambda_t^{\mathrm{T}}\Sigma^{-\frac{1}{2}}\mathbb{V}_t\left[v_{t+1}\right]\Sigma^{-\frac{1}{2}}\lambda_t - 2\lambda_t^{\mathrm{T}}\Sigma^{-\frac{1}{2}}\mathrm{cov}_t\left(rx_{t+1}^{(n-1)}, v_{t+1}\right) \\
&= \mathbb{V}_t\left[rx_{t+1}^{(n-1)}\right] + ||\lambda_t||^2 - 2\lambda_t^{\mathrm{T}}\beta_t^{(n-1)}.
\end{aligned} \tag{A.9}$$

where

$$\beta_t^{(n-1)} := \Sigma^{-1}\mathrm{cov}_t\left(rx_{t+1}^{(n-1)}, v_{t+1}\right) \in \mathbb{R}^K. \tag{A.10}$$

Therefore, (A.3) is equivalent to

$$0 = \mathbb{E}_t \left[rx_{t+1}^{(n-1} \right] + \frac{1}{2}\mathbb{V}_t \left[rx_{t+1}^{(n)} \right] - \lambda_t^{\mathrm{T}} \beta_t^{(n-1)}, \tag{A.11}$$

which gives us a nice expression for the expected returns:

$$E_t \left[rx_{t+1}^{(n-1)} \right] = \lambda_t^{\mathrm{T}} \beta_t^{(n-1)} - \frac{1}{2}\mathbb{V}_t \left[rx_{t+1}^{(n)} \right]. \tag{A.12}$$

Assumption 5. $\beta_t^{(n)} = \beta^{(n)}$ for every $t \geq 0$.

We can decompose the unexpected excess return, $rx_{t+1}^{(n-1)} - \mathbb{E}_t \left[rx_{t+1}^{(n-1)} \right]$ into a component that is correlated with v_{t+1} and another component which is conditionally orthogonal, $\varepsilon_{t+1}^{(n-1)}$ (return pricing error):

$$rx_{t+1}^{(n-1)} - \mathbb{E}_t \left[rx_{t+1}^{(n-1)} \right] = v_{t+1}^{\mathrm{T}} \gamma^{(n-1)} + \varepsilon_{t+1}^{(n-1)}. \tag{A.13}$$

Notice that

$$\beta_t^{(n-1)} = \Sigma^{-1} \left(\mathbb{E} \left[rx_{t+1}^{(n-1)} v_{t+1} \right] - \mathbb{E} \left[rx_{t+1}^{(n-1)} \right] \mathbb{E}_t \left[v_{t+1} \right] \right) = \Sigma^{-1} \mathbb{E} \left[rx_{t+1}^{(n-1)} v_{t+1} \right]$$

and

$$\gamma^{(n-1)} = \left(\mathbb{E} \left[v_{t+1}^{\mathrm{T}} v_{t+1} \right] \right)^{-1} \mathbb{E} \left[v_{t+1} rx_{t+1}^{(n-1)} \right] = \Sigma^{-1} \mathbb{E} \left[rx_{t+1}^{(n-1)} v_{t+1} \right],$$

because $\mathbb{E}\left[v_{t+1}^{\mathrm{T}} v_{t+1}\right] = \Sigma$. So actually $\gamma^{(n)} = \beta^{(n)}$ for every $n \geq 0$. Hence,

$$
\begin{aligned}
\mathbb{V}\left[rx_{t+1}^{(n-1)}\right] &= \mathbb{E}_t\left[\left(rx_{t+1}^{(n-1)} - \mathbb{E}_t\left[rx_{t+1}^{(n-1)}\right]\right)^2\right] \\
&= \mathbb{E}_t\left[\left(v_{t+1}^{\mathrm{T}}\beta^{(n-1)} + \varepsilon_{t+1}^{n-1}\right)^2\right] \\
&= \mathbb{E}_t\left[\left(v_{t+1}^{\mathrm{T}}\beta^{(n-1)}\right)^2 + 2v_{t+1}^{\mathrm{T}}\beta^{(n-1)}\varepsilon_{t+1}^{(n-1)} + \left(\varepsilon_{t+1}^{(n-1)}\right)^2\right] \\
&= \left(\beta^{(n-1)}\right)^{\mathrm{T}} \mathbb{E}_t\left[v_{t+1}v_{t+1}^{\mathrm{T}}\right]\beta^{(n-1)} + \sigma^2 \\
&= \left(\beta^{(n-1)}\right)^{\mathrm{T}} \Sigma\beta^{(n-1)} + \sigma^2.
\end{aligned}
$$

What we get is

$$
\begin{aligned}
rx_{t+1}^{(n-1)} = (\lambda_0 + \lambda_1 X_t)^{\mathrm{T}}\beta^{(n-1)} - \frac{1}{2}\left(\left(\beta^{(n-1)}\right)^{\mathrm{T}}\Sigma\beta^{(n-1)} + \sigma^2\right) \\
+ v_{t+1}^{\mathrm{T}}\beta^{(n-1)} + \varepsilon_{t+1}^{(n-1)}.
\end{aligned}
\tag{A.14}
$$

A.2 Estimation

We can then rewrite (A.14) as

$$
rx_{t+1}^{(n-1)} = (\lambda_0 + \lambda_1 X_t)B_{n-1} - \frac{1}{2}\left(B_{n-1}^{\mathrm{T}}\Sigma B_{n-1} + \sigma^2\right) + v_{t+1}^{\mathrm{T}}B_n + e_{t+1}^{(n-1)} \tag{A.15}
$$

and therefore have a vectorial form:

$$
\mathbf{rx} = \left(\lambda_0 \mathbb{1}_{T\times 1}^{\mathrm{T}} + \lambda_1 \mathbf{X}_{-}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{B} - \frac{1}{2}\left(\mathbf{B}^{*}\mathrm{vec}\left(\Sigma\right) + \sigma^2\mathbb{1}_{K\times 1}\right)\mathbb{1}_T^{\mathrm{T}} + \mathbf{V}^{\mathrm{T}}\mathbf{B} + \mathbf{E} \tag{A.16}
$$

where

1. $\mathbf{rx} \in \mathbb{R}^{T\times N}$.
2. $\lambda_0 \in \mathbb{R}^K$, $\lambda_1 \in \mathbb{R}^{K\times K}$,

3. $\mathbf{X}_{-} = [X_1 \mid X_2 \mid \cdots \mid X_{T-1}]^{\mathrm{T}} \in \mathbb{R}^{T \times K}$,
4. $\mathbf{B} \in \mathbb{R}^{K \times N}$,
5. $\mathbf{B}^{*} = [\mathrm{vec}\,(B_1 B_1^{\mathrm{T}}) \mid \cdots \mid \mathrm{vec}\,(B_n B_n^{\mathrm{T}})]^{\mathrm{T}} \in \mathbb{R}^{K \times N^2}$,
6. $\mathbf{V} \in \mathbb{R}^{T \times K}$ and $\mathbf{E} \in \mathbb{R}^{T \times N}$.

So we take (A.16) as our reference point in the estimation process that we do in three steps following Adrian et al. (2013) procedure:

1. Construct the pricing factors, $(X_t)_{t=1}^{T}$ and estimate the VAR coefficients $\mu \in \mathbb{R}^K$ and $\Phi \in \mathbb{R}^K$ in (A.1) using OLS. Then take $(\hat{v}_t)_{t=1}^{T}$ from $\hat{v}_t := X_t - \hat{X}_t \in \mathbb{R}^K$, where $\hat{X}_t = \mu + \Phi X_{t-1}$ for every $t = 1, \ldots, T$. Stack the time series $(v_t)_{t=1}^{T}$ into the matrix $\hat{\mathbf{V}} \in \mathbb{R}^{T \times K}$. The vatiance-covariance matrix is thus

$$\hat{\Sigma} = \frac{\hat{\mathbf{V}}^{\mathrm{T}} \hat{\mathbf{V}}}{T} \tag{A.17}$$

2. Perform the regression according to (A.16), i.e.,

$$\mathbf{rx} = a \mathbb{1}_{T \times K} \mathbb{1}_{K \times N} + \hat{\mathbf{V}} b + \mathbf{X}_{-} c + \mathbf{E} \tag{A.18}$$

where $a \in \mathbb{R}$, $b, c \in \mathbb{R}^{K \times N}$. Collect everything into single matrices

$$\mathbf{Z} = \left[\mathbb{1}_{T \times 1} \mid \hat{\mathbf{V}} \mid \mathbf{X}_{-} \right] \in \mathbb{R}^{T \times (2K+1)} \tag{A.19}$$

$$d = [a \mathbb{1}_{K \times 1} \mid b \mid c]^{\mathrm{T}} \in \mathbb{R}^{(2K+1) \times N} \tag{A.20}$$

so we can write $\mathbf{rx} = \mathbf{Z} d + \mathbf{E}$ and therefore

$$\hat{d} = \left(\mathbf{Z}^{\mathrm{T}} \mathbf{Z} \right)^{-1} \mathbf{Z}^{\mathrm{T}} \mathbf{rx}. \tag{A.21}$$

Then, collect the residuals from this regression into the matrix

$$\hat{\mathbf{E}} = \mathbf{rx} - \mathbf{Z}\hat{d} \in \mathbb{R}^{T\times N}. \tag{A.22}$$

and estimate

$$\hat{\sigma}^2 = \frac{\mathrm{tr}\left(\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}}\right)}{NT}. \tag{A.23}$$

Finally, we construct $\hat{\mathbf{B}}^*$ from $\hat{b}$.

3. Estimate the price of risk parameters, λ_0 and λ_1 via cross-sectional regression. Recall from (A.16) that

$$a = \left(\lambda_0 \mathbb{1}_{T\times 1}^{\mathrm{T}}\right)^{\mathrm{T}} \mathbf{B} - \frac{1}{2}\left(\mathbf{B}^* \mathrm{vec}\left(\Sigma\right) + \sigma^2 \mathbb{1}_{K\times 1}\right) \mathbb{1}_T^{\mathrm{T}} \tag{A.24}$$

$$c = \lambda_1^{\mathrm{T}} \mathbf{B} \tag{A.25}$$

If we transpose them, we can estimate λ_0 and λ_1 via OLS, i.e.,

$$\hat{\lambda}_0 = \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^{\mathrm{T}}\right)^{-1} \hat{\mathbf{B}} \left[\hat{a}^{\mathrm{T}} + \frac{1}{2}\mathbb{1}_{T\times 1}\left(\mathbf{B}^* \mathrm{vec}\left(\Sigma\right) + \sigma^2 \mathbb{1}_{N\times 1}\right)^{\mathrm{T}}\right] \tag{A.26}$$

$$\hat{\lambda}_1 = \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^{\mathrm{T}}\right)^{-1} \hat{\mathbf{B}}\hat{c}^{\mathrm{T}} \tag{A.27}$$

A.3 Recursion

Consider the generating process for log excess returns in our model:

$$rx_{t+1}^{(n-1)} = (\lambda_0 + \lambda_1 X_t)^{\mathrm{T}} \beta^{(n-1)} - \frac{1}{2}\left(\left(\beta^{(n-1)}\right)^{\mathrm{T}} \Sigma \beta^{(n-1)} + \sigma^2\right) + v_{t+1}^{\mathrm{T}} \beta^{(n-1)} + \varepsilon_{t+1}^{(n-1)}.$$

(A.28)

We need now to find two sequences of coefficients, $(A_n)_{n=1}^N$ and $(B_n)_{n=1}^N$, that allow us to express bond prices as exponentially affine in the vector of state variables, X_t, plus a trend term, $p_t^{*,(n)}$, i.e.,

$$p_t^{(n)} = p_t^{(n),*} + A_n + X_t^{\mathrm{T}} B_n + u_t^{(n)}, \tag{A.29}$$

where $p_t^{(n)} \coloneqq \log P_t^{(n)}$. Notice that

$$p_t^{(1)} = -r_t = -r_t^* - X_t^{\mathrm{T}} e_1, \tag{A.30}$$

motivating that $A_1 = 0$, $B_1 = -e_1$, and $p_t^{1,*} = -r_t^*$. For any $n > 1$,

$$\begin{aligned}
rx_{t+1}^{(n-1)} &= p_{t+1}^{(n-1),*} + A_{n-1} + X_{t+1}^{\mathrm{T}} B_{n-1} + u_{t+1}^{(n-1)} \\
&- p_t^{(n),*} - A_n - X_t^{\mathrm{T}} B_n - u_t^{(n)} \\
&+ p_t^{(1),*} + A_1 + X_t^{\mathrm{T}} B_1 + u_t^{(1)} \\
&= p_{t+1}^{(n-1),*} + A_{n-1} + (\mu + \Phi X_t + v_{t+1})^{\mathrm{T}} B_{n-1} + u_{t+1}^{(n-1)} \\
&- p_t^{(n),*} - A_n - X_t^{\mathrm{T}} B_n - u_t^{(n)} \\
&+ p_t^{(1),*} + A_1 + X_t^{\mathrm{T}} B_1 + u_t^{(1)} \\
&= rx_{t+1}^{(n-1),*} + (A_{n-1} - A_n + A_1 + \mu^{\mathrm{T}} B_{n-1}) \\
&+ X_t^{\mathrm{T}} \left(\Phi B_{n-1} - B_n + B_1\right) + \left(u_{t+1}^{n-1} - u_t^{(n)} + u_t^{(1)}\right) + v_{t+1}^{\mathrm{T}} B_{n-1}
\end{aligned}$$

Hence, the following must hold

$$
\begin{aligned}
& rx_{t+1}^{(n-1),*} + \left(A_{n-1} - A_n + A_1 + \mu^{\mathrm{T}} B_{n-1}\right) \\
& + X_t^{\mathrm{T}} \left(\Phi B_{n-1} - B_n + B_1\right) + \left(u_{t+1}^{n-1} - u_t^{(n)} + u_t^{(1)}\right) \\
& = (\lambda_0 + \lambda_1 X_t^{\mathrm{T}} \beta^{(n-1)} - \frac{1}{2}\left(\left(\beta^{(n-1)}\right)^{\mathrm{T}} \Sigma \beta^{(n-1)} + \sigma^2\right) + v_{t+1}\beta^{(n-1)} + \varepsilon_{t+1}^{(n-1)}
\end{aligned}
$$

i.e.,

$$
\begin{aligned}
A_{n-1} - A_n + A_1 + \mu^{\mathrm{T}} B_{n-1} &= \lambda_0^{\mathrm{T}} \beta^{(n-1)} - \frac{1}{2}\left(\left(\beta^{(n-1)}\right)^{\mathrm{T}} \Sigma \beta^{(n-1)} + \sigma^2\right) \\
\Phi^{\mathrm{T}} B_{n-1} - B_n + B_1 &= \lambda_1^{\mathrm{T}} \beta^{(n-1)} \\
u_{t+1}^{n-1} - u_t^{(n)} + u_t^{(1)} + v_{t+1}^{\mathrm{T}} B_{n-1} &= \varepsilon_{t+1}^{(n-1)} \\
rx_{t+1}^{(n-1),*} &= 0 \\
v_{t+1}^{\mathrm{T}} \beta^{(n-1)} &= v_{t+1}^{\mathrm{T}} B_{n-1}
\end{aligned}
$$

and therefore

$$
\begin{aligned}
A_n &= A_{n-1} + \mu^{\mathrm{T}} B_{n-1} - \lambda_0^{\mathrm{T}} \beta^{(n-1)} + \frac{1}{2}\left(\left(\beta^{(n-1)}\right)^{\mathrm{T}} \Sigma \beta^{(n-1)} + \sigma^2\right) \\
B_n &= \Phi^{\mathrm{T}} B_{n-1} + B_1 - \lambda_1^{\mathrm{T}} \beta^{(n-1)} \\
p_t^{(n),*} &= p_{t+1}^{(n-1),*} - r_t^* \\
\beta^{(n)} &= B_n
\end{aligned}
$$

The last equation simplifies everything even more:

$$A_n = A_{n-1} + (\mu - \lambda_0)^{\mathrm{T}} B_{n-1} + \frac{1}{2}\left(B_{n-1}^{\mathrm{T}} \Sigma B_{n-1} + \sigma^2\right) \tag{A.31}$$

$$B_n = (\Phi - \lambda_1)^{\mathrm{T}} B_{n-1} - e_1 \tag{A.32}$$

$$p_t^{(n),*} = p_{t+1}^{(n-1),*} - r_t^* \tag{A.33}$$

References

Adrian, Tobias, Richard K Crump, and Emanuel Moench (2013) "Pricing the term structure with linear regressions," *Journal of Financial Economics*, 110 (1), 110–138.

——— (2015) "Regression-based estimation of dynamic asset pricing models," *Journal of Financial Economics*, 118 (2), 211–244.

Ang, Andrew and Monika Piazzesi (2003) "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary economics*, 50 (4), 745–787.

Bakshi, Gurdip S and Zhiwu Chen (1994) "Baby boom, population aging, and capital markets," *Journal of business*, 165–202.

Bauer, Michael D. and Glenn D. Rudebusch (2020) "Interest Rates under Falling Stars," *American Economic Review*, 110 (5), 1316–54.

Campbell, John Y and Robert J Shiller (1987) "Cointegration and tests of present value models," *Journal of political economy*, 95 (5), 1062–1088.

Cieslak, Anna and Pavol Povala (2015) "Expected returns in Treasury bonds," *The Review of Financial Studies*, 28 (10), 2859–2901.

Diebold, Francis X, Monika Piazzesi, and Glenn D Rudebusch (2005) "Modeling bond yields in finance and macroeconomics," *American Economic Review*, 95 (2), 415–420.

Dybvig, Philip H and Stephen A Ross (1989) "Arbitrage," in *Finance*, 57–71: Springer.

Engle, Robert F and Clive WJ Granger (1987) "Co-integration and error correction: Representation, estimation, and testing," *Econometrica*, 55 (2), 251–276.

Fama, Eugene F (2006) "The behavior of interest rates," *The Review of Financial Studies*, 19 (2), 359–379.

Favero, Carlo A and Ruben Fernandez-Fuertes (2023) "Monetary Policy in the COVID Era and Beyond: The Fed vs the ECB," *Available at SSRN 4557795*.

Favero, Carlo A, Arie E Gozluklu, and Haoxi Yang (2016) "Demographics and the behavior of interest rates," *IMF Economic Review*, 64 (4), 732–776.

Favero, Carlo A, Alessandro Melone, and Andrea Tamoni (2022) "Monetary policy and bond prices with drifting equilibrium rates," *Journal of Financial and Quantitative Analysis*, 1–26.

Golinski, Adam and Paolo Zaffaroni (2016) "Long memory affine term structure models," *Journal of Econometrics*, 191 (1), 33–56, https://EconPapers.repec.org/RePEc:eee:econom:v:191:y:2016:i:1:p:33-56.

Gürkaynak, Refet S, Brian Sack, and Eric Swanson (2005) "The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models," *American economic review*, 95 (1), 425–436.

Gürkaynak, Refet S, Brian Sack, and Jonathan H Wright (2007) "The US Treasury yield curve: 1961 to the present," *Journal of monetary Economics*, 54 (8), 2291–2304, https://www.federalreserve.gov/econres/feds/the-us-treasury-yield-curve-1961-to-the-present.htm, Accessed on 2023/10.

Jardet, Caroline, Alain Monfort, and Fulvio Pegoraro (2013) "No-arbitrage Near-Cointegrated VAR(p) term structure models, term premia and GDP growth," *Journal of Banking & Finance*, 37 (2), 389–402, 10.1016/j.jbankfin.2012.0.

Jordà, Òscar and Alan M Taylor (2019) "Riders on the Storm,"Technical report, National Bureau of Economic Research.

Jørgensen, Kasper (2018) "How Learning from Macroeconomic Experiences Shapes the Yield Curve," Finance and Economics Discussion Series 2018, Board of Governors of the Federal Reserve System, https://ideas.repec.org/p/fip/fedgfe/2015-77.html.

Kim, Don H and Jonathan H Wright (2005) "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates."

Kozicki, Sharon and P. A. Tinsley (2001) "Shifting endpoints in the term structure of interest rates," *Journal of Monetary Economics*, 47 (3), 613–652.

Laubach, Thomas and John C Williams (2003) "Measuring the natural rate of interest," *Review of Economics and Statistics*, 85 (4), 1063–1070.

Lunsford, Kurt G and Kenneth D West (2019) "Some evidence on secular drivers of US safe real rates," *American Economic Journal: Macroeconomics*, 11 (4), 113–39.

Mian, Atif R, Ludwig Straub, and Amir Sufi (2021) "What Explains the Decline in r*? Rising Income Inequality Versus Demographic Shifts," *Becker Friedman Institute for Economics Working Paper* (2021-104).

Negro, Marco Del, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti (2017) "Safety, Liquidity, and the Natural Rate of Interest," *Brookings Papers on Economic Activity*, 48 (1), 235–316.

Piazzesi, Monika, J.Salomao, and Martin Schneider (2015) "Trend and cycle in bond premia," https://web.stanford.edu/~piazzesi/trendcycle.pdf, Accessed on 2023/10.

Schnabel, Isabel (2022) "United in diversity-Challenges for monetary policy in a currency union," https://www.ecb.europa.eu/press/key/date/2022/html/ecb.sp220614~67eda62c44.en.html, Accessed on 2023/08.

Zhao, Guihai (2020) "Learning, equilibrium trend, cycle, and spread in bond yields,"Technical report, Bank of Canada Staff Working Paper.