Towards Data-Congruent Models of the Term Structure of Interest Rates*

Carlo A. Favero† Rubén Fernández-Fuertes‡

First Version: December, 2023
This Version: October, 2024

Abstract

Bond yields can be decomposed into two unobservable components: the expected sequence of short-term rates and term premia. The identification of these two components is crucial to understand bond pricing and the effect of monetary policy on the term structure of interest rates. This paper illustrates how M.H. Pesaran's prescription of congruency between the salient features of the data and the reduced form, explicitly derived from stochastic dynamic optimization, effectively facilitates the relevant decomposition. By examining the historical evolution of term structure models, we demonstrate that the chosen specifications have not consistently aligned with the data, presenting a missed opportunity. In fact, a data-congruent specification helps in improving forecasts of the dynamics of US short-term rates and generates stationary dynamics for the term premia.

JEL codes: E43, E52, G12.

Keywords: Data Congruency, Affine Term Structure Models, Trends and Cycles, Term Premia.

*Paper prepared for the Conference to Celebrate Professor M.H. Pesaran's Achievements, USC April 2024. We are grateful to Tobias Adrian, Michael Bauer, JS Fontaine, Francesco Giavazzi, Refet Gurkaynak, Glenn Rudebusch, Ron Smith and Guihai Zhao, the editor, Cheng Hsiao, and two anonymous referees for comments and suggestions. The paper also benefited from stimulating discussions with Daniel Gros. Financial Support by the European Union-Next Generation EU, project GRINS–Growing Resilient, INclusive and Sustainable PE00000018 (CUP B43C22000760006) is gratefully acknowledged. The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them. This paper is part of the research activities of the Baffi Centre unit on Macroeconomic Trends, Cycles, and Asset Prices.

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

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

1 Introduction

The role of theory in applied econometrics is undeniably one of the many areas where M.H. Pesaran's research has made groundbreaking contributions(Pesaran, 1990). Shortly after completing his PhD, one of us had the privilege of joining a research project led by Hashem, focused on the exploration and extraction of North Sea oil(Favero et al., 1994). After just a few days of working on the model specification, the young researcher approached Hashem, questioning the need for a non-linear specification in the oil extraction cost function, which linked costs to the residual reserves in the wells. Little did he know, he was about to receive a lesson that would stick for years to come. In fact, as soon as the question was asked, the young researcher immediately realized that it was the wrong question. Hashem, in his characteristic assertiveness, responded, "*That is a salient feature of the data. The reduced form, explicitly derived from a stochastic dynamic optimization, must also account for the physical constraints in the oil supply process*"(Pesaran and Smith, 1995). In this paper, we explore a similar line of inquiry: What salient features of the data must be considered when specifying models of the term structure of interest rates?

The dynamics of nominal government bond yields at different maturities plays a central role in shaping the response of the real economy to monetary and fiscal policy interventions. Yields can be decomposed into two unobservable components: the sequence of expected one-period rates and the term-premia (Campbell and Shiller, 1991; Duffee, 2002). The first component reflects the future expected path of monetary policy rates, while the second reflects macroeconomic fundamentals, economic policies, and the investors' attitude toward risk. Policymakers are fully aware that the market-based financing conditions that matters for the control of the business cycle and inflation depend on both components of yields (Schnabel, 2022, 2023).

Term structure models are helpful in that they allow the identification of the two components by forecasting the expected path of short-term interest rates and by then deriving forecasts for yields and term-premia at all maturities.

Spanos (1990) introduced the difference between identification (unique mapping from the reduced form to the structural form) and statistical adequacy (the validity of the reduced form as a statistical model for the data). Diebold and Rudebusch (2013), in their important book on yield curve modeling and forecasting, acknowledge the importance of the distinction made by Spanos and focus their work on the observation that most yield curve models tend to be either theoretically rigorous but empirically disappointing, or empirically successful but theoretically lacking.

We focus on the importance of a statistically adequate, or data-congruent, modeling strategy for the identification of the two unobservable components of yields.

The paper is organized as follows. Section 2 examines the US data from 1980 onward, to establish the features of a statistically adequate model for the term structure. The data show that yields are drifting, with a common drift, while both excess returns at all maturities and term spreads at all maturities are stationary. In the light of this evidence, a statistically adequate model of the term structure can be constructed by relating the drift in yields to the future expected path of monetary policy to derive stationary term-premia.

Section 3 concentrates on the early literature on testing the Expectations Theory in cointegrated VAR models (Campbell and Shiller, 1987, 1991). This literature specifies a cointegrated model for the term spread and the change in the short-term interest rates, explicitly recognizing the stationarity of the term premium. However, statistical adequacy is lost in forecasting over long-horizons, when the short-term is predicted by a deterministic trend, and this trend, given cointegration with a (-1,1) cointegrating vector between the long-term and the short-term rate, also drives the predictions for the long-term rates. This is a problem that can only be solved by the inclusion of exogenous I(1) variables determining the trend in the system (Pesaran and Smith, 1998; Pesaran et al., 2000).

Section 4 explores the statistical adequacy of dynamic factor models for the term structure with a focus on Affine Term Structure (ATS) models (Diebold et al., 2005). These models are originally designed for stationary processes in yields, as the yield dynamics is modeled as a linear function 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); vector autoregression (VAR) models are then specified for the factors and used for forecasting factors and yields. Importantly, the factor dynamics drives not only yields but also the price of risk and holding period returns. The presence of a stochastic trend in yields has several negative consequences for this approach. VAR models are inappropriate for drifting data. OLS estimates of near-unit roots are notoriously biased downward, thus overestimating the amount of mean reversion in yields, biased long-run forecasting of the dynamics of short-term rates does affect the measurement of term premia. In fact, the non-stationarity of factors often results in non-stationarity of term premia, which is not congruent with the evidence of stationarity of holding period (excess) returns and term spreads.

Section 5 describes the specification of a novel ATS model in which yields drift, 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.

Statistical adequacy of this specification follows in that it is consistent with the evidence from the data that yields are non-stationary and driven by a common trend while excess returns are stationary. The specification strategy also offers a solution for forecasting in cointegrating models, where inversion in long-term trend projections cannot be allowed for. This problem is solved by following the principle of relating macroeconomic trends to slow-moving predictable exogenous variables (Aksoy et al., 2019). Following Favero et al. (2016), Del Negro et al. (2019), and Lunsford and West (2019), short-term rates are decomposed in a trend component, related to exogenous and predictable variables such as potential output growth, the age structure of population and the inflation target of the central bank. The variables not only model successfully the observed trend in short-term rates but they are also predictable over a long-horizon, as the Bureau of Census produces forecasts for the age structure of population for long-term ahead horizon, and so does the FRED database for potential output, while the inflation target is proxied with inflation expectations at long-horizon that converge to the Central Bank's target . Given the availability of long-run forecasts, the current and future trend components of the short-term rates are constructed.

A factor model is then built by extracting factors from the cyclical components of yields, and by imposing no-arbitrage restrictions on the dynamics of holding period excess returns and term-premia.

A VAR specification is kept for the factors, but, now, it is a VAR on stationary variables. Predictions 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 therefore forward-looking) and the predictions for the cyclical components (based on the VAR for factors and, therefore, backward-looking). Bonds at any maturities are priced via pricing equations that imposes no-arbitrage restrictions. Term premia are derived as the difference between bond yields projected 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, their trend being is the average trend of short-term rates over the maturity of the bonds, while term-premia are determined by the stationary state variables.

The empirical results will show that the specification strategy statistically adequate dominates standard models in terms of forecasting performance for returns and yields at any maturity and it leads to a very different measurement of term-premia.

2 The Data on the US Term Structure

At any given moment in time, bond yields vary across different maturities. This cross-sectional information is captured by the yield curve, which illustrates the relationship between bond yields and their respective maturities. However, yield curves are not static; they evolve dynamically over time. Therefore, modeling yield curves requires addressing both the cross-sectional variation in yields and their time-series evolution. Furthermore, bond yields do not fluctuate independently of each other. There exists a fundamental relationship between bond yields and one-period holding returns. Under the principle of no-arbitrage, the risk-adjusted one-period holding returns on bonds of all maturities must be equal. This implies that yields across maturities are interconnected and cannot move independently. Term structure models are used to efficiently describe the cross-section of bond yields by employing a small number of factors. These factors allow for a parsimonious representation of yield variations. By forecasting these factors using time-series models, it becomes possible to predict with parsimonious models the entire term structure of interest rates. Additionally, the no-arbitrage condition imposes further restrictions on the parameters governing both the time-series and cross-sectional dynamics of bond yields. These restrictions ensure consistency between the model's predictions and the absence of arbitrage opportunities in the market.

To illustrate the evidence from the data that it is relevant to specify statistically adequate term structure models, consider the quarterly time-series observations on the zero-coupon US Treasury yield curve estimates of the Federal Reserve Board made available by Gürkaynak et al. (2007) over the period 1980-2023.

Continuously compounded yield to maturities $r_t^{(n)}$ for bonds with residual life of n periods are related to log of bond prices, $p_t^{(n)}$, as follows

$$p_t^{(n)} = -nr_t^{(n)}.$$

The data reported in Figure 1 show that yields from the 3-month to the 15-year maturity have been trending from 1980 till the end of 2023 and they shared a common trend.

One-period holding excess returns are then defined as the difference between the returns of holding for one-quarter bonds with maturity n longer than 3-month and the yield to maturity of the 3-month bonds. Under no-arbitrage their expected value is equal to the risk-premia associated to holding for one period bonds with maturity n, $\phi_{t,t+1}^{n}$.

$$\begin{aligned}
rx_{t+1}^{(n-1)} &= p_{t+1}^{(n-1)} - p_t^{(n)} - r_t^{(1)}, \\
E_t rx_{t+1}^{(n-1)} &= \phi_{t,t+1}^{n}
\end{aligned}$$

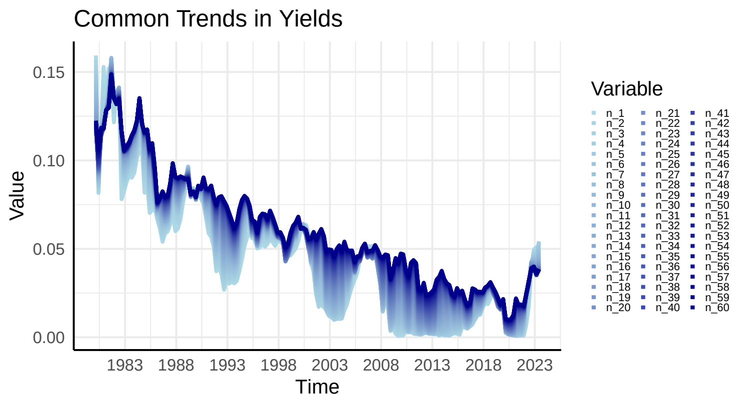


FIGURE 1. Quarterly observations on the time-series of (annualised) yields from the 3-month to the 15-year maturity. We use the same colour palette for all maturities (blue). Darkest blue indicates the highest maturity, i.e., 15 years.

The data reported in Figure 2 show that ex-post observed excess returns for bonds at all maturities are stationary.

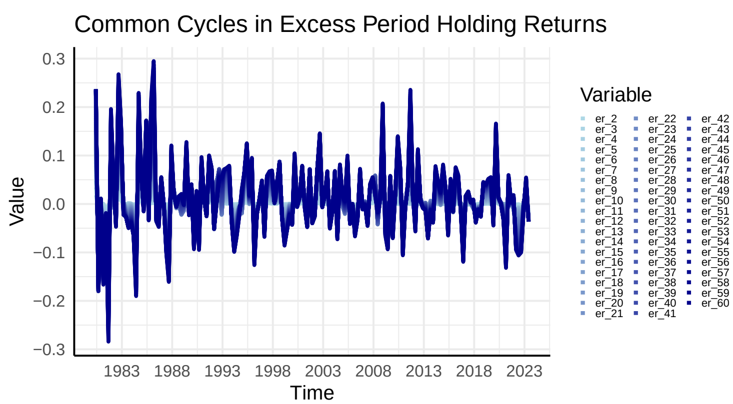


FIGURE 2. Quarterly observations on the time series of 1-quarter holding period returns for bonds at maturities between 6-months and 15 years in excess of the return on three-month Treasury Bills

By solving forward the no-arbitrage relationship, taking into account that the price of all bonds at maturity converges to one, yields at any maturities can be expressed as the sum of the average of the yields of the 1-period (3-month in our case) bonds over the life of the bonds and the term premia, i.e. the average of the risk premia over the residual life of the bonds. Term-spreads can then be expressed as a weighted sum, with declining weights, of future changes in the 3-month rates and the term premia.

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

$$r_t^{(n)} = \frac{1}{n}\sum_{i=0}^{n-1} E_t\left(r_{t+i}^{(1)} + \phi_{t+i,t+i+1}^n\right), \tag{2}$$

$$r_t^{(n)} - r_t^{(1)} = \sum_{i=1}^{n-1}\left(1 - \frac{i}{n}\right) E_t \Delta r_{t+i}^{(1)} + \frac{1}{n}\sum_{i=0}^{n-1}\phi_{t+i,t+i+1}^n. \tag{3}$$

The data reported in Figure 3 show that term-spreads are stationary. This evidence, under no-arbitrage, can only be consistent with stationarity of term premia; in fact, the weighted sum of future changes in the three month rates is stationary, as differencing removes the drift in yields.

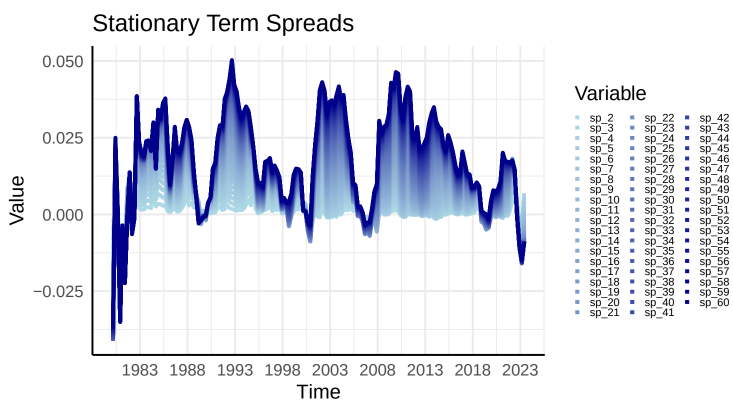


FIGURE 3. Quarterly observations on the time series of spreads of yields on bonds at maturities between 6-months and 15 years in excess of the return on three-month Treasury Bills

To sum up, the relevant features of the data are that yields are non-stationary and co-drifting while excess returns, term spreads, risk-premia and term-premia are stationary.

3 The Expectations Theory in Cointegrated VARs

Campbell and Shiller (1987) (CS) specify a term-structure model for the case of the risk-free rate $r_t^{(1)}$ and a single very long term bond, in the original paper the US Treasury 20-year, with yield $r_t^{(n)}$, that takes several features of the data into explicit account. After providing evidence for stationarity of the term spread, $S_t = r_t^{(n)} - r_t^{(1)}$ they specify a stationary VAR in the (demeaned) first difference of the short-term rate and the spread:

$$\Delta r_t^{(1)} = a(L)\Delta r_{t-1}^{(1)} + b(L)S_{t-1} + u_{1t}$$
$$S_t = c(L)\Delta r_{t-1}^{(1)} + d(L)S_{t-1} + u_{2t}$$

to show that the Expectations Theory imposes the following set of testable restrictions on on the individual coefficients of VAR:

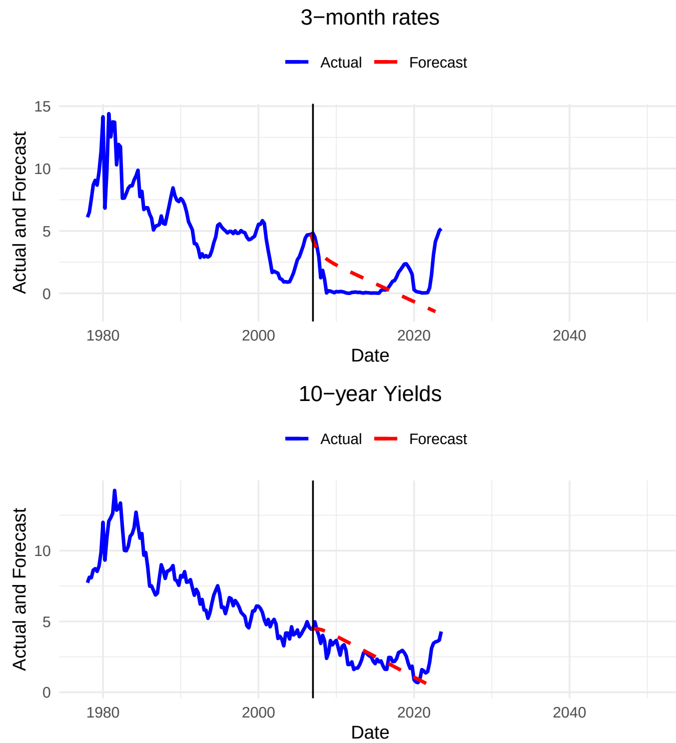
$$\{c_i = -a_i, \forall i\}, \{d_1 = -b_1 + 1/\gamma\}, \{d_i = -b_i, \forall i \neq 1\}$$

These restrictions are tested via a Wald test and rejected, thus evidence emerges for the existence of a term-premium. However, when this premium is derived as the difference between the observed term spread and that obtained by imposing the (rejected) restrictions implied by the Expectations Theory, it is shown to be small.

3.1 Are Cointegrated VAR for the Term Structure Statistically Adequate?

The CS specification is consistent with the evidence of trending yields, stationary term spreads and stationary term premia, however, it has two important limitations. First, it implies that the long-horizon short-term rate is predicted via a trend that does not allow for inversions, being determined by the estimated constant in the equation for the difference of short-term rates. Second, as a consequence of the stationarity of the spread, the long-horizon forecasts for short-term rates do "contaminate" the long-horizon forecast for the long-term rates. Figure 4 illustrates the point by reporting the out-of-sample forecasts, from 2006 onward of a cointegrated CS VAR for the the 3-month rates and the 10-year yields estimated over the sample 1980-2005. The VAR forecasts for the 3-month rates converge rather rapidly to a trend with the slope determined by the estimated constant in the first equation of the VAR. No inversion of the downward sloping trend that emerged in the latter part of the sample is possible. As the predictions for the 10-year yields are constrained by their cointegrating relationship with the 3-month rates, they also converge to the trend predicted for the 3-month rates.

Figure 4. Forecasting with a CVAR



4 Dynamic Factor Models of the Term Structure

The Cointegrated VAR approach discussed in the previous section provides a useful benchmark to illustrate the main issues in taking models to the data but it focuses on only two variables: a long-term yield and a short-term rate. Dynamic factor models of the term structure parsimoniously describe both the cross-sectional and the time-series dimensions of the entire yield curve by using a few factors as state variables. (Diebold et al., 2005).

Several approaches have been proposed to the identification of factors, that can be observed or unobserved or a mixture of the two types (see, for example, Nelson and Siegel (1987), Chen (1993), Ang and Piazzesi (2003), Kim and Wright (2005) and Adrian et al. (2015)). The common baseline for all models is the specification of a VAR for the factors (state variables),X_t, which are assumed to be stationary:

$$X_{t+1} = \mu + \Phi X_t + v_{t+1}, \tag{4}$$

$$v_{t+1}|\left\{X_s\right\}_{s=0}^{t} \sim \mathcal{N}(0,\Sigma), \tag{5}$$

This general specification encompasses a wide range of models, including those based on cross-sectional curve fitting in the popular Dynamic Nelson-Siegel approach [1]. We concentrate on Affine Term Structure Models in which the variables in X_t determine the market price of risk,λ_t in the following affine form:

$$\lambda_t = \Sigma^{-1/2}(\lambda_0 + \lambda_1 X_t), \tag{6}$$

No-arbitrage, is then assumed and there exists a pricing kernel M_t such that:

$$P_t^{(n)} = E_t\left(M_{t+1}P_{t+1}^{(n-1)}\right), \tag{7}$$

for every $n > 0$ and $t \geq 0$. Where $P_t^{(n)} = exp\left[-nR_t^{(n)}\right]$ is the price of a zero coupon bond with maturity n.

The pricing kernel is taken as exponentially affine, i.e.,

$$m_{t+1} = -R_t^1 - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\Sigma^{-1/2}v_{t+1}, \tag{8}$$

[1]see, Diebold and Li (2006), Diebold et al. (2005); no arbitrage restrictions can also been imposed in these models (Christensen et al., 2011)

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

Finally, excess period returns for bonds at all maturities$\left(rx_{t+1}^{(n-1)} = p_{t+1}^{(n-1)} - p_t^{(n)} - R_t^1\right)$ and (v_{t+1}) are jointly normally distributed with constant covariance matrix.

Under this set of assumptions, the process for excess returns is then derived as:

$$p_{t+1}^{(n-1)} - p_t^{(n)} - R_t^1 = \underbrace{\beta^{(n-1)\prime}(\lambda_0 + \lambda_1 X_t)}_{\text{Expected Return}} - \underbrace{\frac{1}{2}\left(\beta^{(n-1)\prime}\Sigma\beta^{(n-1)} + \sigma^2\right)}_{\text{Convexity Correction}} + \underbrace{\beta^{(n-1)\prime}v_{t+1}}_{\text{Return Innovation}} + e_{t+1}^{(n-1)},$$

Bond prices are derived from the process for excess returns by recursive forward substitution as,

$$p_t^n = A_n + B_n' X_t + u_t^n,$$

where, as a consequence of no-arbitrage, recursive restrictions apply to A_n and B_n. In fact, if for the one-period yield we have:

$$R_t^1 = \delta_0 + \delta_1' X_t + \epsilon_t,$$

so $A_1 = -\delta_0$ and $B_1 = -\delta_1$.

We then have:

$$\begin{aligned} A_n &= A_{n-1} + B_{n-1}'(-\lambda_0) + \frac{1}{2}(\beta^{(n-1)\prime}\Sigma\beta^{(n-1)} + \sigma^2) - \delta_0, \\ B_n' &= B_{n-1}'(\Phi - \lambda_1) - \delta_1', \\ \beta^{(n)} &= B_n', \end{aligned}$$

4.1 Are ATS Models Statistically Adequate?

The main divergence between the evidence from the data and the specification strategy of ATS model emerges as the same factors , modeled by a Vector Autoregressive Process, are the common drivers of the dynamics of both yields and excess returns, while the evidence from the data is that excess-returns are stationary while yields are (co-)drifting. Factors, extracted from drifting yields by using Principal Components, are by construction orthogonal to each other and the first ones typically capture the persistence with a near unit root. Despite this persistence, a VAR in levels is estimated for factors and the VAR-based forecasts of future one-period rates slowly converge to the mean of the sample used for estimation. As one period rates are drifting in the data, ATS models tend to generate term premia that are a-cyclical and parallel to the secular trend in yields. These features of the term-premia are not congruent with the evidence from the data. Bauer et al. (2014) observe that, as a consequence of the very high persistence in yields, term premia implied by maximum likelihood estimates of affine term structure models are misleading because of small-sample bias. They confirm that ATS models, such as that estimated by Wright (2011), tend to produce cyclical risk premium estimates, often just parallel to the secular trend in interest rates, while bias corrected term-premia show strong (counter-)cyclical behaviour. However, the bias-correction for the VAR coefficient does not address the problem that term-premia that are stationary in the data are specified as function of highly persistent state variables. This divergence becomes more important when, as in Christensen and Rudebusch (2012), the problem of non-stationarity of yields is solved by imposing a unit root in the factor capturing the level of the term structure. As a matter of fact,Christensen and Rudebusch (2012), show that this restriction produces a clear improvement in the forecasting performance of the model. However, as in the case of the CS models, first differencing of the level and projecting it with a VAR has the limitation that the level is projected as a trend with constant slope and the model cannot capture fluctuations and inversions that are observed in the data.

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 (Del Negro et al., 2019), 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).

In particular, 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.

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. They explain this evidence by pointing out that survey forecasts of interest rates are consistent with the view that both the level and the slope of the yield curve are more persistent than under common statistical models. Zhao (2020) and Feunou and Fontaine (2023) propose structural models of trends and cycles in the term-structure capable of explaining several features of the data. However, these models rely on statistical trend-cycle decomposition that are difficult to exploit for long-term forecasting purposes.

5 A Statistically Adequate ATS model

The analysis of the data and the historical evolution of term structure models highlights several distinctive features for statistically adequate models. Co-trending yields are naturally modeled by decomposing the risk-free rate into a stochastic trend and a cyclical component. The common drift would then be captured by the stochastic trend in the short-term rates. The problem of long-horizon out-of-sample predictions of this stochastic trend could be solved by following the proposal by Aksoy et al. (2019) to introduce some slow-moving variables, exogenous and predictable, driving the trend in the short-term rates. Factors can be extracted from the stationary cyclical components of yields at all maturities obtained by the difference between actual yields and their common stochastic trend. These factors would be effectively modeled with a VAR and by applying the usual ATS assumptions excess returns would then be modeled as function of stationary factors. Finally, yields at all maturities could be derived by using no-arbitrage restrictions, and by combining stationary and cyclical risk-premia with trends consistent with the no-arbitrage restrictions.

As illustrated in detail in the Appendix, all these features can be explicitly included in a model. First, the one-period rate is decomposed in trend and cycle using the following specification:

$$
\begin{aligned}
r_t^{(1)} &= r_t^{*,(1)} + u_t^{(1)} \\
r_t^{*,(1)} &= \gamma_1 MY_t + \gamma_2 \Delta y_t^{pot} + \gamma_3 \pi_t^{LR}
\end{aligned}
$$

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. The inclusion of (log) growth rate of potential output, Δy_t^{pot}, as a variable explaining the trend come from the standard Ramsey model (Laubach and Williams, 2003).

$$
r = \frac{1}{\sigma} g + \theta. \tag{9}
$$

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 inflation 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[2] as the proxy for long-run inflation expectations.

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

This is a reasonable proxy under the assumption that the central bank is credible. The yield's cyclical part can thus be identified with the residual after regressing the short rate on those three variables, Δy_t^{pot}, MY_t, and π_t^{LR}.

Yields at all maturities are then decomposed in trends and cycles using the following strategy:

$$
\begin{aligned}
r_t^{(n)} &= r_t^{(1)} + \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) E_t \Delta r_{t+i}^{(1)} + \frac{1}{n} \sum_{i=0}^{n-1} \phi_{t+i,t+i+1}^{n} \\
r_t^{(n)} &= r_t^{*,(1)} + u_t^{(n)} \\
u_t^{(n)} &= u_t^{(1)} + \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) E_t \Delta r_{t+i}^{(1)} + \frac{1}{n} \sum_{i=0}^{n-1} \phi_{t+i,t+i+1}^{n}
\end{aligned}
$$

K factors are then be extracted by obtaining the principal components of the N cycles of the yield curve $u_t^{(j)}$, for $j = 1, \ldots, n$, which are then stacked into a $T \times N$ matrix, $\mathbf{U}$. This procedure ensures the stationarity of X_t to specify a VAR, i.e.,

$$X_{t+1} = \mu + \Phi X_t + v_{t+1} \tag{10}$$

$$v_{t+1} | \left(X_s\right)_{s=0}^{t} \sim \mathcal{N}(0, \Sigma), \tag{11}$$

where $\mu \in \mathbb{R}^K$, $\Phi \in \mathbb{R}^{K \times K}$ and $\Sigma \in \mathbb{R}^{K \times K}$.

With the usual assumptions in ATS models stacked excess returns across N maturities and T time-periods can be represented as:

$$\mathbf{rx} = \left(\lambda_0 \mathbb{1}_{T\times 1}^{\mathrm{T}} + \lambda_1 \mathbf{X}_{-}^{\mathrm{T}}\right)^{\mathrm{T}} \mathbf{B} - \frac{1}{2}\left(\mathbf{B}^{*}\mathrm{vec}\left(\Sigma\right) + \sigma^2 \mathbb{1}_{K\times 1}\right) \mathbb{1}_{T\times 1}^{\mathrm{T}} + \mathbf{V}^{\mathrm{T}}\mathbf{B} + \mathbf{E} \tag{12}$$

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. $\mathbf{B}^* = [\mathrm{vec}\,(B_1 B_1^{\mathrm{T}}) \mid \cdots \mid \mathrm{vec}\,(B_n B_n^{\mathrm{T}})]^{\mathrm{T}} \in \mathbb{R}^{K\times N^2}$,
6. $\mathbf{V} \in \mathbb{R}^{T\times K}$ and $\mathbf{E} \in \mathbb{R}^{T\times N}$.

After the estimation of all parameters of interest from the stacked representation of excess returns, bond prices at any maturity can be obtained by adding the trend and the cycle components.

The cyclical component of the one-period bond r_t^1, i.e., $u_t^{(1)} := r_t^{(1)} - r_t^{*,(1)}$, can be expressed as a linear function of the underlying factors, i.e.,

$$\begin{aligned} r_t^{(1)} &= r_t^{*,(1)} + \delta_0 + \delta_1 \cdot X_t + e_t^{(1)}, \\ p_t^{(1)} &= -r_t^{(1)}, \quad p_t^{1,*} = r_t^{*,(1)}, \end{aligned} \tag{13}$$

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.

No-arbitrage implies that bond prices at all maturities depend linearly on a trend component and on a stationary component:

$$p_t^n = p_t^{n,*} + A_n + B_n' X_t + u_t^n,$$

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 $p_t^{n,*}$.

$$\begin{aligned} A_n &= A_{n-1} + (\mu - \lambda_0)^{\mathrm{T}} B_{n-1} + \frac{1}{2}\left(B_{n-1}^{\mathrm{T}} \Sigma B_{n-1} + \sigma^2\right) - \delta_0 \\ B_n &= (\Phi - \lambda_1)^{\mathrm{T}} B_{n-1} - \delta_1 \\ p_t^{(n),*} &= p_{t+1}^{(n-1),*} - r_t^{*,(1)} \end{aligned}$$

5.1 A Statistically Adequate ATS model: The Empirical Results

Estimation and simulation[3] is performed by using the zero coupon yields provided by the FED[4] (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[5] and the measure of potential Gross Domestic Product available from the FRED database.[6] Quarterly data over the period 1980:1-2023:2 are considered. The statistical adequacy of ATS models is illustrated by comparing the results of estimation of a standard ATS model via the 3-step procedure proposed by Adrian et al. (2013), Adrian et al. (2015) (ACM) with those obtained by a 4-step procedure (FF) illustrated above and detailed in the Appendix. The three-step procedure begins after the extraction of k factors (by using principal components) from the entire term structure; in the first step a VAR for factors is estimated by OLS to obtain contemporaneous pricing factor innovations, in the second step excess returns are regressed on a constant,lagged

[3] A full replication package in `R` is available from the authors' website
[4] https://www.federalreserve.gov/econres/feds/the-us-treasury-yield-curve-1961-to-the-present.htm
[5] https://www.federalreserve.gov/econres/us-models-package.htm.
[6] https://fred.stlouisfed.org/series/GDPPOT.

pricing factors and contemporaneous pricing factor, finally, in the third-step the price of risk parameters λ_0 and λ_1 are estimated via cross-sectional regressions. Bond prices are then obtained recursively by imposing the no-arbitrage restrictions. Term-premia are derived by the difference between bond prices based on the estimated risk parameters and counterfactual bond prices obtained by setting $\lambda_0 = \lambda_1 = 0$. The four-step procedure begins with the estimation of the stochastic trend driving the short-term rates, after this trend has been identified and validated by the stationarity of the deviation of one-period rates from it, the entire term structure is detrended and k stationary factors are extracted from the cyclical components of yields. The three-step procedure is then applied on the stationary factors. Bond prices are then obtained recursively by imposing no-arbitrage restrictions both on the cyclical and on the trend components. Finally, term premia are obtained by the difference between fitted yields and counterfactual yields simulated under the restrictions $\lambda_0 = \lambda_1 = 0$.

Table 1 reports the results of modeling the trend in three-month rates showing that the ADF tests rejects the presence of a unit root in the residuals of the regression of three-month rates on demographics, productivity and long-run expected inflation.[7]

[7] The long-run cointegrating coefficients are extracted by a static regression. In general, as we agreed in a discussion with Ron Smith, long-run coefficients are better extracted by ARDL regressions (Smith, 2024), or by the Johansen (1995) procedure. In the case at hand, the ARDL estimates were very close both to those delivered by the static regression and to those delivered by the Johansen procedure, which rejected the null of at most zero cointegrating vectors and did not reject the null of at most one cointegrating vector. The uniqueness of the cointegrating relationship paired with the low persistence in the residual from the static regression, produces a low correlation between the long run and the short-run components in the dynamic models that justifies the result.

Table 1. Modeling 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

The estimated coefficients on the drivers of the drift on short-term rates are in line with previous studies Favero et al. (2022), Bauer and Rudebusch (2020), with a negative 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.

The stochastic trend produced by the model is plotted along with actual data in Figure 5. The model is naturally interpreted within a cointegration approach (Engle and Granger, 1987) to 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[8], implies that

[8]The term premium at time t and maturity n is given by $\frac{1}{n}\sum_{i=0}^{n-1}\phi^{n}_{t+i,t+i+1}$, which is the average of the

$u_t^{(n)}$ are stationary. Note also that, in this framework, the common stochastic trend in yields at all maturities is that 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 (9) with the appropriate scenario for the exogenous variables MY_{t+i}, Δy_{t+1}^{pot}, π_{t+i}^{LR}.

The implications of de-trending yields before the extraction of PC are highlighted by Figure 6 which compares the principal components extracted from detrended yields with those extracted from yields. The first principal component extracted from yields shows a stochastic trend which is removed from the first principal component extracted from detrended yields.

Figure 7 compares the fit and the out-of-sample forecasting performance of the two specifications for 3-month rates showing the importance of modeling the stochastic trend in one period rates. Figure 7 reports 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. The standard ATS model, estimated in three steps, performs poorly out-of-sample, except for very short horizons. In fact, the predicted path reverts to the sample mean, which is significantly higher than the observed values of the three-month rates over the forecasting horizon. Conversely, the model estimated in four steps successfully exploits cointegration and the predictability of the

expected one-period risk-premia over the residual maturity of the bond slow-moving components that drive the trend in yields, making it more effective for long-term forecasting. The out-of-sample simulation reveals that over short-forecasting horizon standard models can do better, from Figure 7 it is evident that ACM model outperforms the FF model for a forecasting horizon up to one-year ahead. To provide further evidence on this issue we have implemented an out-of sample forecasting exercise for the 1-period rate comparing the performance of the FF and the ACM models at different horizons. The results, reported in Table 2, confirms that the FF model outperforms the ACM at all forecasting horizons and the difference between the forecasting performance of the two model becomes more relevant the longer the forecasting horizon. A possible interpretation of this evidence is that the problems related to using a stationary representation for drifting data is of minor relevance for short-term forecasting. However, it is important to remark that the derivation of term-premia does always require long-run forecasts of the one period rates. Up to a (small) correction for convexity, the term premium on bonds at all maturities is derived as the difference between the observed ten-year yield and the average three-month yields over the residual life of the bonds. for example, in our data set the relevant forecasting horizon to derive the current term premium on 10-year bonds is forty steps ahead.

Model	$h = 1$	$h = 4$	$h = 8$	$h = 20$
ACM	0.083004	0.079403	0.076938	0.075291
FF	0.049388	0.046596	0.044645	0.039070
RMSFE ratio	0.59	0.58	0.58	0.52
OOS Obs.	40	40	37	25

Table 2. RMSFE for ACM and FF models at different forecast horizons when predicting the one period yield. Estimation period: 1980Q1-2012Q4.OOS Simulation 2013:1-2023:4

Finally, term premia for the 10-year maturity from the two specifications are compared in Figure 8. The known evidence of trending term premia in the three-step approach is confirmed while the four-step approach produces very different and cyclical term premia.

The a-cyclicality of term premia estimated by standard ATS models and their parallelism to the secular trend in long-term interest rates has been already noted by Bauer et al. (2014) in commenting on the estimates provided by Wright (2011). Bauer et al. (2014) attribute the acyclicality to small sample bias caused by the very high persistence in the VAR model for factors; they show that biased-adjusted estimates produce instead countercyclical term premia. In fact, adjusting for small sample-bias produces estimates that are much closer to the unit root, preventing the sequence of predicted one-period rate to converge to a biased estimate of their level. Christensen and Rudebusch (2012) build a model all (three: level, slope, curvature) unobservable latent factors but one indeed follow a mean-reverting Ornstein-Uhlenbeck process. However, they force a unit root in the first factor, letting it being a standard Brownian Motion. These generates the problem in out-of-sample forecasting that we have highlighted for the in discussing the Campbell-Shiller approach to modeling the short-term rate and the term spread. Our approach is different. Since the one-period rate's trend is captured by the long-term drivers, factors are then extracted from the deviations of yields from their drift explained by productivity demographics and long-term inflation forecasts. Our VAR for factors is much less persistent and the parameters' estimates do not require a small sample adjustment. As a result, the sequence of predicted one-period rates features much smaller forecast errors than the equivalent in standard ATS models and also our estimates of the term premia show some counter-cyclical behaviour. This evidence is in line with the empirical and theoretical research that has found support for countercyclical risk premia, including, among many others, Campbell and Shiller (1987), Cochrane and Piazzesi (2005),Campbell and Cochrane (1999), and Wachter (2006).

Figure 5. Three month yield time series against its trend.

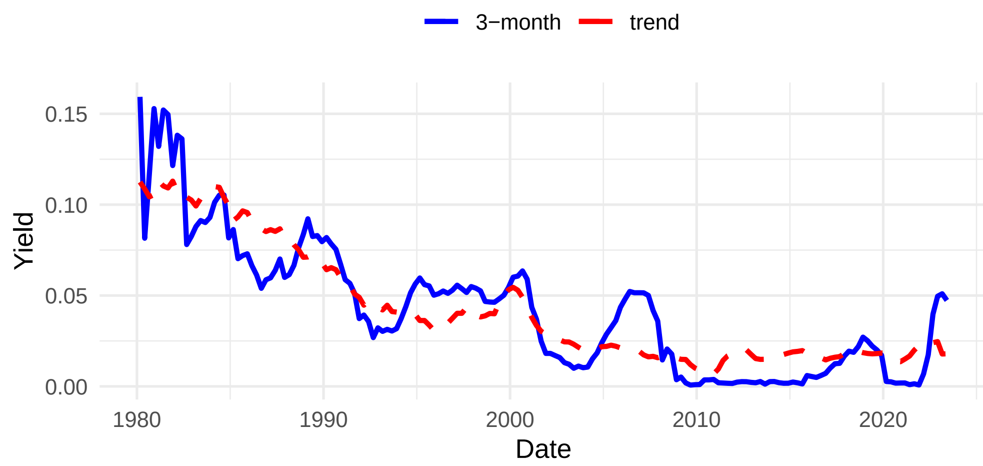
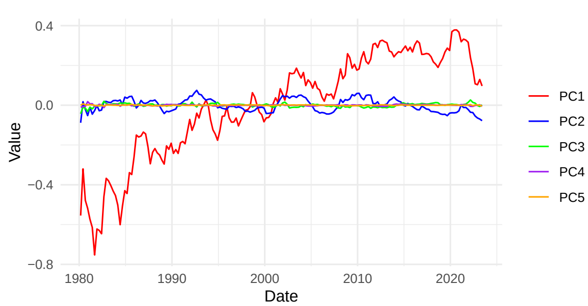


Figure 6. This graph reports the time-series of the five first principal components extracted respectively from yields, as in the ACM model, and from the cyclical components of yields, as in the FF model.

(a) ACM



(b) FF

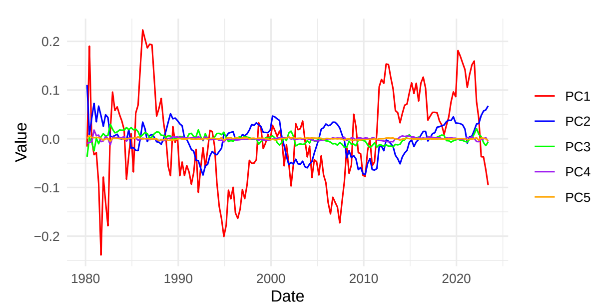


FIGURE 7. This graph reports the fitted (1980Q1:2005Q4) and the forecasted (2006Q1:2023Q2) time series of 3-month yields given by the standard ACM model (red) and FF model (green) against the actual values (blue).

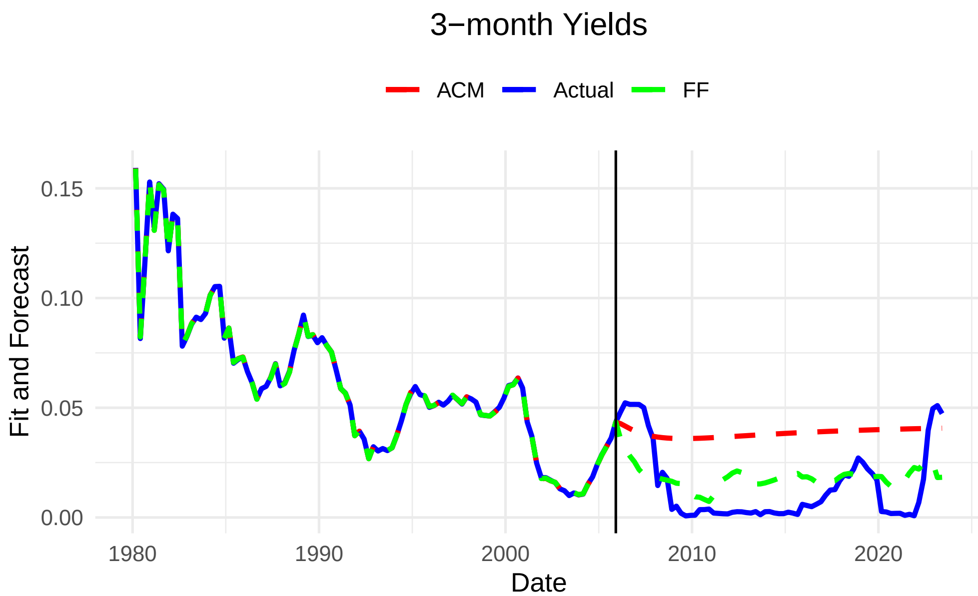
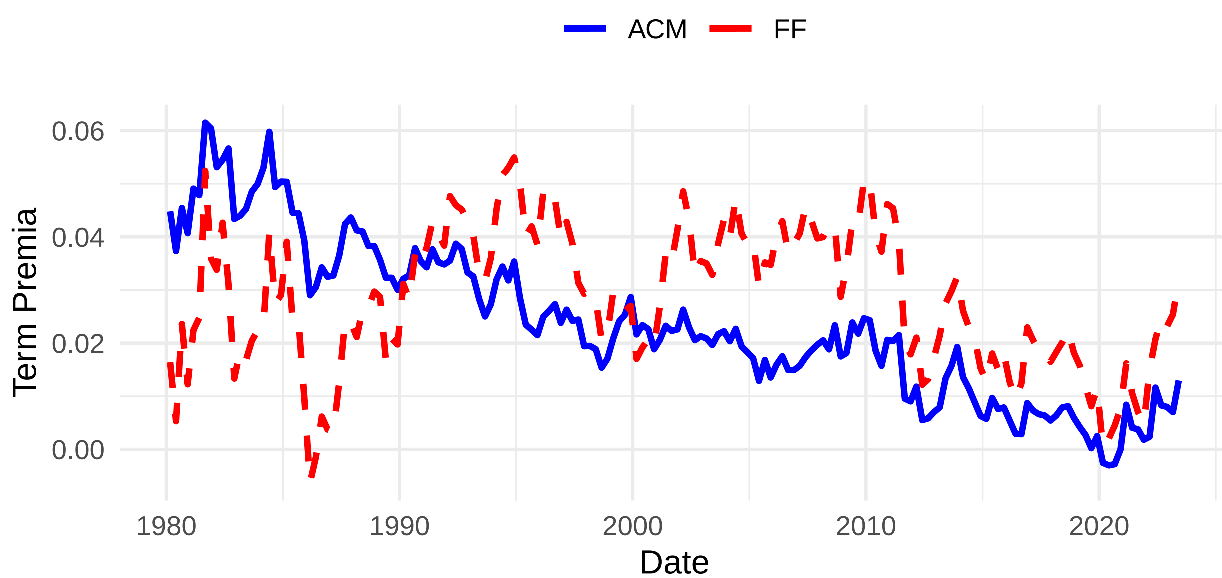


FIGURE 8. 10-year Term premia.



6 Conclusions

This paper begins by arguing that the evidence from the data implies that a statistically adequate model of the term structure should relate the drift in yields to the drifting future expected path of monetary policy and derive stationary term-premia.

We then provide an evaluation of the historical evolution of term structure models in the light of their statistical adequacy, in the sense of (Spanos, 1990). Early models, which adopted a cointegration based approach, were capable of separating trend yields from cycle from spreads. However, they suffered from a limitation in out-of-sample predictions as they did not predict the stochastic nature of the trend in yields. Affine Term Structure (ATS) models have abandoned the cointegration based approach to model the yield dynamics in terms of factors that are assumed to follow a stationary vector autoregression (VAR). However, factor extracted from yields tend to reflect their non-stationarity. As the factor dynamics not only drives yields but it also determines the price of risk and holding period returns, the presence of a stochastic trend in yields produces biased long-run forecasts of the dynamics of short-term rates and, the non-stationarity of factors often results in non-stationarity of term premia. This is not a feature congruent with the data.

The historical analysis of the discrepancies between the data and the adopted specifications naturally leads to the adoption of a statistically adequate Affine Term Structure in which i) the factor structure adopted to explain holding period excess returns is extracted from de-trended yields; ii) the drift in short-term rates is not predicted by a VAR but it is rather related to long-term forecast for slow-moving exogenous variables, such as the demographic structure of the population, potential output growth and long-term inflation forecast;

iii) excess returns are stationary as the compensation for risk depends on the cycle in yields.

We provide evidence that the statistically adequate model dominates standard models in terms of forecasting performance and produces stationary risk premia, very different from those produced by the standard approach.

We conclude that Hashem Pesaran's indication that the reduced form, explicitly derived from a stochastic dynamic optimization imposing no-arbitrage restrictions, must take also account of the physical constraints in the data, is of crucial importance also for building a valid empirical model of the term structure of interest rates.

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.

Aksoy, Yunus, Henrique S Basso, Ron P Smith, and Tobias Grasl (2019) "Demographic structure and macroeconomic trends," *American Economic Journal: Macroeconomics*, 11 (1), 193–222.

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.

Bauer, Michael D, Glenn D Rudebusch, and Jing Cynthia Wu (2014) "Term premia and inflation uncertainty: Empirical evidence from an international panel dataset: Comment," *American Economic Review*, 104 (1), 323–337.

Campbell, John Y and John H Cochrane (1999) "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of political Economy*, 107 (2), 205–251.

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

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

Chen, Ren-Raw (1993) "Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates," *Journal of Fixed Income*, 3, 14–31.

Christensen, Jens HE, Francis X Diebold, and Glenn D Rudebusch (2011) "The affine arbitrage-free class of Nelson–Siegel term structure models," *Journal of Econometrics*, 164 (1), 4–20.

Christensen, Jens HE and Glenn D Rudebusch (2012) "The response of interest rates to US and UK quantitative easing," *The Economic Journal*, 122 (564), F385–F414.

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

Cochrane, John H and Monika Piazzesi (2005) "Bond risk premia," *American economic review*, 95 (1), 138–160.

Del Negro, Marco, Domenico Giannone, Marc P Giannoni, and Andrea Tambalotti (2019) "Global trends in interest rates," *Journal of International Economics*, 118, 248–262.

Diebold, Francis X and Canlin Li (2006) "Forecasting the term structure of government bond yields," *Journal of econometrics*, 130 (2), 337–364.

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

Diebold, Francis X and Glenn D Rudebusch (2013) *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*: Princeton University Press.

Duffee, Gregory R (2002) "Term premia and interest rate forecasts in affine models," *The Journal of Finance*, 57 (1), 405–443.

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

Favero, Carlo A, M Hashem Pesaran, and Sunil Sharma (1994) "A duration model of irreversible oil investment: Theory and empirical evidence," *Journal of Applied Econometrics*, 9 (S1), S95–S112.

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.

Feunou, Bruno and Jean-Sébastien Fontaine (2023) "Secular economic changes and bond yields," *The Review of Economics and Statistics*, 105 (2), 408–424.

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.

Johansen, Soren (1995) "Likelihood-Based Inference in Cointegrated Vector Autoregressive Models Oxford University Press," *New York*.

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).

Nelson, Charles R and Andrew F Siegel (1987) "Parsimonious modeling of yield curves," *Journal of business*, 473–489.

Pesaran, M Hashem (1990) "An econometric analysis of exploration and extraction of oil in the UK continental shelf," *The Economic Journal*, 100 (401), 367–390.

Pesaran, M Hashem, Yongcheol Shin, and Richard J Smith (2000) "Structural analysis of vector error correction models with exogenous I (1) variables," *Journal of econometrics*, 97 (2), 293–343.

Pesaran, M Hashem and Ron Smith (1995) "The role of theory in econometrics," *Journal of econometrics*, 67 (1), 61–79.

Pesaran, M Hashem and Ron P Smith (1998) "Structural analysis of cointegrating VARs," *Journal of economic surveys*, 12 (5), 471–505.

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.

——— (2023) "The Last Mile. Keynote speech at the annual Homer Jones Memorial Lecture," https://www.ecb.europa.eu/press/key/date/2023/html/ecb.sp231102_1~4bb07ebef7.en.html, Accessed on 2023/11.

Smith, Ron (2024) "Issues in the Use of ARDL Models," working paper, Birkbeck College.

Spanos, Aris (1990) "The simultaneous-equations model revisited: Statistical adequacy and identification," *Journal of Econometrics*, 44 (1-2), 87–105.

Wachter, Jessica A (2006) "A consumption-based model of the term structure of interest rates," *Journal of Financial economics*, 79 (2), 365–399.

Wright, Jonathan H (2011) "Term premia and inflation uncertainty: Empirical evidence from an international panel dataset," *American Economic Review*, 101 (4), 1514–1534.

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

A Online Appendix

A.1 Derivations

As Adrian et al. (2013), we assume that the systematic risk is represented by a stochastic vector, $(X_t)_{t\geq 0}$, that follows 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,.

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

Let's denote the zero coupon treasury bond price with maturity n at time t by $P_t^{(n)}$. We take the following assumptions:

Assumption 1. No-arbitrage condition holds (Dybvig and Ross, 1989), i.e.,

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

Assumption 2. The pricing kernel, $m_{t+1} \coloneqq \log M_{t+1}$, is exponentially affine

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

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

Assumption 3. Market prices of risk are affine

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

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

Assumption 4. $\left(xr_t^{(n-1)}, v_t\right)_{t\geq 0}$ are jointly normally distributed for $n \geq 2$.

Thanks to all these assumptions, we can continue our modeling by recalling the definition of the excess holding return of a bond maturing in n periods, i.e.,

$$xr_{t+1}^{(n-1)} \coloneqq p_{t+1}^{(n-1)} - p_t^{(n)} - r_t^{(1)}, \tag{A.6}$$

where $n-1$ indicates the $n-1$ periods remaining since time $t+1$ with respect to which the return is computed. Now, (A.3) can be rewritten as

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

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

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

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

where

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

Therefore, no-arbitrage condition (A.3) is equivalent to

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

which gives us the following expression for the expected returns:

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

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

If we were to decompose the unexpected excess return, $xr_{t+1}^{(n-1)} - \mathbb{E}_t\left[xr_{t+1}^{(n-1)}\right]$ 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), we could simply write the following OLS-wise form

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

and try to figure out who the $\gamma^{(n-1)}$ is. To do so, notice that

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

and

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

because $\mathbb{E}\left[v_{t+1}^{\mathrm{T}}v_{t+1}\right] = \Sigma$. Therefore, $\gamma^{(n)} = \beta^{(n)}$ for every $n \geq 0$. With this identity in our hands,

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

Finally,

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

A.2 Estimation

We can then rewrite (A.14) as

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

and therefore have a vectorial form:

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

where

1. $\mathbf{xr} \in \mathbb{R}^{T\times N}$.
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. $\mathbf{B}^* = [\mathrm{vec}\left(B_1 B_1^{\mathrm{T}}\right) \mid\cdots\mid \mathrm{vec}\left(B_n B_n^{\mathrm{T}}\right)]^{\mathrm{T}} \in \mathbb{R}^{K^2\times N}$,
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 four stepsby extending Adrian et al. (2013) procedure:

1. Construct the pricing factors $(X_t)_{t=1}^T$. First, model 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.

Second, derive the cyclical components of yields at any maturity by considering the difference between yields and the trend in the three-month rate. Third consider as price factors the first k principal components of de-trended yields.

2. Model the pricing factors, $(X_t)_{t=1}^{T}$ via a VAR 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 variance-covariance matrix is thus

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

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

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

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 $\mathbf{xr} = \mathbf{Z}d + \mathbf{E}$ and therefore

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

Then, collect the residuals from this regression into the matrix

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

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 $\hat{b}$.

4. 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 for the Term Structure

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

$$xr_{t+1}^{(n-1)} = (\lambda_0 + \lambda_1 X_t)^{\mathrm{T}} \beta^{(n-1)} - \frac{1}{2}\left(\left(\beta^{(n-1)}\right)^{\mathrm{T}} \Sigma \beta^{(n-1)} + \sigma^2\right) + v_{t+1}^{\mathrm{T}} \beta^{(n-1)} + \varepsilon_{t+1}^{(n-1)}. \tag{A.28}$$

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.,

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

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

$$p_t^{(1)} = -r_t^{(1)} = -r_t^{*,(1)} - \delta_0 - X_t^{\mathrm{T}} \delta_1, \tag{A.30}$$

motivating that $A_1 = -\delta_0$, $B_1 = -\delta_1$, and $p_t^{1,*} = -r_t^{*,(1)}$. For any $n > 1$,

$$\begin{aligned}
xr_{t+1}^{(n-1)} &= p_{t+1}^{*,(n-1)} + A_{n-1} + X_{t+1}^{\mathrm{T}} B_{n-1} + e_{t+1}^{(n-1)} \\
&- p_t^{*,(n)} - A_n - X_t^{\mathrm{T}} B_n - e_t^{(n)} \\
&+ p_t^{*,(1)} + A_1 + X_t^{\mathrm{T}} B_1 + e_t^{(1)} \\
&= p_{t+1}^{*,(n-1)} + A_{n-1} + (\mu + \Phi X_t + v_{t+1})^{\mathrm{T}} B_{n-1} + e_{t+1}^{(n-1)} \\
&- p_t^{*,(n)} - A_n - X_t^{\mathrm{T}} B_n - e_t^{(n)} \\
&+ p_t^{*,(1)} + A_1 + X_t^{\mathrm{T}} B_1 + e_t^{(1)} \\
&= xr_{t+1}^{*,(n-1)} + (A_{n-1} - A_n + A_1 + \mu^{\mathrm{T}} B_{n-1}) \\
&+ X_t^{\mathrm{T}} \left(\Phi^{\mathrm{T}} B_{n-1} - B_n + B_1\right) + \left(e_{t+1}^{n-1} - e_t^{(n)} + e_t^{(1)}\right) + v_{t+1}^{\mathrm{T}} B_{n-1}.
\end{aligned} \tag{A.31}$$

Hence, the following must hold

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

i.e.,

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

and therefore

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

The last equation simplifies everything even more:

$$A_n = A_{n-1} + (\mu - \lambda_0)^{\mathrm{T}} B_{n-1} + \frac{1}{2}\left(B_{n-1}^{\mathrm{T}} \Sigma B_{n-1} + \sigma^2\right) - \delta_0 \tag{A.32}$$

$$B_n = (\Phi - \lambda_1)^{\mathrm{T}} B_{n-1} - \delta_1 \tag{A.33}$$

$$p_t^{(n),*} = p_{t+1}^{(n-1),*} - r_t^{*,(1)} \tag{A.34}$$

Equation (A.34) for the price stochastic trend implies that

$$r_t^{*,(n)} = \frac{1}{n} \sum_{i=0}^{n-1} r_{t+i}^{*,(1)}. \tag{A.35}$$

On the other hand, these equations are fully deterministic, meaning that one can iterate all the equations back to get expressions that depend only on the initial values, A_1 and B_1. First,

$$\begin{aligned}
B_n &= (\Phi - \lambda_1)^{\mathrm{T}} \left((\Phi - \lambda_1)^{\mathrm{T}} B_{n-2} - \delta_1\right) - \delta_1 \\
&= \cdots \\
&= \left[(\Phi - \lambda_1)^{\mathrm{T}}\right]^{n-1} B_1 - \sum_{j=1}^{n-2} \left[(\Phi - \lambda_1)^{\mathrm{T}}\right]^j \delta_1. \qquad \text{(A.36)} \\
&= -\sum_{j=1}^{n-1} \left[(\Phi - \lambda_1)^{\mathrm{T}}\right]^j \delta_1
\end{aligned}$$

Second,

$$\begin{aligned}
A_n &= A_{n-2} + (\mu - \lambda_0)^{\mathrm{T}} (B_{n-1} + B_{n-2}) + \frac{1}{2}\left(B_{n-1}^{\mathrm{T}} \Sigma B_{n-1} + B_{n-2}^{\mathrm{T}} \Sigma B_{n-2}\right) + 2\left(\frac{1}{2}\sigma^2 - \delta_0\right) \\
&= A_{n-2} + (\mu - \lambda_0)^{\mathrm{T}} (B_{n-1} + B_{n-2}) \\
&\quad + \frac{1}{2}\left([B_{n-1} + B_{n-2}]^{\mathrm{T}} \Sigma [B_{n-1} + B_{n-2}]\right) + 2\left(\frac{1}{2}\sigma^2 - \delta_0\right) \\
&= A_1 + (\Phi - \lambda_1)^{\mathrm{T}} \sum_{j=1}^{n-1} B_{n-j} + \frac{1}{2}\left(\sum_{j=1}^{n-1} B_{n-j}\right)^{\mathrm{T}} \Sigma \left(\sum_{j=1}^{n-1} B_{n-j}\right) + (n-1)\left(\frac{1}{2}\sigma^2 - \delta_0\right)
\end{aligned}$$

It's not difficult to see that

$$\sum_{j=1}^{n-1} B_{n-j} = \sum_{j=1}^{n-1}\sum_{k=1}^{n-j} \left[(\Phi - \lambda_1)^{\mathrm{T}}\right]^j \delta_1 = \sum_{j=1}^{n-1}(n-j) \left[(\Phi - \lambda_1)^{\mathrm{T}}\right]^j \delta_1. \tag{A.37}$$

That allows us to write

$$\begin{aligned} A_n &= (\Phi - \lambda_1)^{\mathrm{T}} \sum_{j=1}^{n-1}(n-j) \left[(\Phi - \lambda_1)^{\mathrm{T}}\right]^j \\ &+ \frac{1}{2}\left(\sum_{j=1}^{n-1}(n-j)(\Phi - \lambda_1)^j\right) \Sigma \left(\sum_{j=1}^{n-1}(n-j) \left[(\Phi - \lambda_1)^{\mathrm{T}}\right]^j\right) \\ &+ n\left(\frac{1}{2}\sigma^2 - \delta_0\right). \end{aligned} \tag{A.38}$$

A.4 Recursion for Term Premia

Remember that

$$TP_t^{(n)} = u_t^{(n)} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_t\left[u_{t+i}^{(1)}\right], \tag{A.39}$$

where $u_t^{(n)} = r_t^{(n)} - r_t^{*,(n)}$. The affine model implies that

$$u_t^{(n)} = -n\left(A_n + X_t^{\mathrm{T}} B_n + e_t^{(n)}\right). \tag{A.40}$$

In particular, for $n = 1$,

$$u_t^{(1)} = -A_1 - X_t^{\mathrm{T}} B_1 - e_t^{(1)}. \tag{A.41}$$

Hence,

$$\mathbb{E}_t\left[u_{t+i}^{(1)}\right] = -A_1 - \mathbb{E}_t\left[X_{t+i}^{\mathrm{T}}\right] B_1. \tag{A.42}$$

Now, since $X_{t+i} = \mu + \Phi X_{t+i-1} + v_{t+i}$, then, we can iterate backwards to get

$$\begin{aligned} X_{t+i} &= \mu + \Phi X_{t+i-1} + v_{t+i} \\ &= \mu + \Phi\left(\mu + \Phi X_{t+i-2} + v_{t+i-1}\right) + v_{t+i} \\ &= (1+\Phi)\mu + \Phi^2 X_{t+i-2} + \Phi v_{t+i-1} + v_{t+i} \\ &= \cdots \\ &= \left(\sum_{j=0}^{i-1} \Phi^j\right)\mu + \Phi^i X_t + \sum_{j=0}^{i-1} \Phi^j v_{t+i-j}. \end{aligned} \tag{A.43}$$

Since $\mathbb{E}_t\left[v_s\right] = 0$ for every $s > t$, then

$$\mathbb{E}_t\left[X_{t+i}\right] = \widetilde{\Phi}_i \mu + \Phi^i X_t, \tag{A.44}$$

where

$$\widetilde{\Phi}_i = \left(\sum_{j=0}^{i-1} \Phi^j \right). \tag{A.45}$$

Hence,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_t \left[u_t^{(1)} \right] &= -A_1 - \frac{1}{n} \sum_{i=1}^{n} \left(\widetilde{\Phi}_i \mu + \Phi^i X_t \right)^{\mathrm{T}} B_1 \\
&= -A_1 - \frac{1}{n} B_1^{\mathrm{T}} \left(\sum_{i=1}^{n} \widetilde{\Phi}_i \right) \mu - \frac{1}{n} B_1^{\mathrm{T}} \left(\sum_{i=1}^{n} \Phi^i \right) X_t \\
&= -A_1 - \frac{1}{n} B_1^{\mathrm{T}} \left(\sum_{i=1}^{n} \widetilde{\Phi}_i \right) \mu - \frac{1}{n} B_1^{\mathrm{T}} \widetilde{\Phi}_n X_t \\
&= \Xi_n + \Psi_n X_t
\end{aligned} \tag{A.46}$$

where

$$\Xi_n = -\frac{1}{n} A_1 - \frac{1}{n} B_1^{\mathrm{T}} \left(\sum_{i=1}^{n} \widetilde{\Phi}_i \right) \mu \tag{A.47}$$

$$\Psi_n = -\frac{1}{n} B_1^{\mathrm{T}} \widetilde{\Phi}_i \tag{A.48}$$

Hence,

$$TP_t^{(n)} = u_t^{(n)} + \Xi_n + \Psi_n X_t \tag{A.49}$$