Estimating Discrete-Time Gaussian Term Structure Models
in Canonical Companion Form*

Juliusz F. Radwański

August 26, 2017

Abstract
This article explores a convenient parametrization for essentially-affine term structure models specified in discrete time, as in Ang and Piazzesi (2003). Under the assumption of \( N \) latent spanning state variables, all their pricing information must be contained in \( N \) short-maturity forward rates, and every class of observationally equivalent models is uniquely represented by a model with these forward rates as factors. This leads to a convenient parametrization of the risk-neutral dynamics in terms of exactly \( N \) unrestricted real numbers in the transition matrix (which takes a companion form), plus one parameter for the drift. The proposed canonical form is maximally flexible, and allows estimation by Kalman filter, or assuming that certain linear combinations of yields (e.g., PCA factors) are measured without error. The parametrization has a natural interpretation in terms of recursive hedge ratio in conditional opportunity set of excess returns implied by the model. I estimate the most popular specifications with two and three factors in the set of Fama-Bliss bonds in order to shed some light on the nature of return predictability documented by Cochrane and Piazzesi (2005). The three-factor model suggests that that important part of this predictability was associated with times of abnormal conditional Sharpe ratios. In other words, the model finds it difficult to reconcile the term structure dynamics with the assumption of no arbitrage, which indirectly supports the view that limits to arbitrage or market illiquidity played a role in shaping bond risk premia.

Keywords: Affine Term Structure Models, Canonical Companion Form.

JEL Classification Numbers: G12, G17, E43.

*Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität zu Berlin, Dorotheenstr. 1, 10099 Berlin, Germany. Tel: +49 (0) 30 2093 5013. E-mail: juliusz.radwanski@hu-berlin.de. I would like to thank the participants of AFFI 2017 (Valence), and Brown Bag seminar at HU Berlin.
1 Introduction

The class of affine term structure models has been extremely popular in many applications, such as yield forecasting, or fixed-income risk management. Although there exists no single specification that fits the data perfectly, the consensus appears to be that an empirically successful model should feature at least three factors, and a flexible specification of risk prices. Another requirement, although not directly linked to model performance, is usually analytical tractability.

One class of models that satisfies these requirements is due to Duffie and Kan (1996). Citing Duffee (2012), this class includes both homoskedastic (Gaussian), and heteroskedastic models. Dai and Singleton (2000), and Duffee (2002) combine this affine class with linear dynamics of the underlying state vector to produce the completely affine, and essentially affine classes, respectively. One of the conclusions in Duffee (2002) is that (...) Gaussian models in this class are sufficiently flexible to generate plausible forecasts of future yields. The model of Duffee (2002) is specified in continuous time. A discrete-time version is due to Ang and Piazzesi (2003).

A well-known issue associated with bringing these models to the data is lack of identification. If one starts with a set of latent factors, then it is possible to perform invariant transformations of the state vector that result in exactly the same likelihood of the observed data (i.e., alternative sets of factors and parameters are able to describe the data equally well, see also Dai and Singleton (2000)). Ruling out all such transformations in order to uniquely fix the factors may be difficult, and the estimated state variables often lack economic interpretation (Babbs and Nowman (1999), Collin-Dufresne et al. (2008), Aït-Sahalia and Kimmel (2010)).

The solution proposed in this paper builds on Duffie and Kan (1996), and complements Collin-Dufresne et al. (2008) (CGJ, henceforth). In order to guarantee identification, I use observationally equivalent representation in terms of term structure variables. The somewhat surprising result is that all pricing information in the latent factors of an $N$-factor model must also be present in the set of $N$ forward rates at the short end of the maturity spectrum (spaced according to the shortest investment horizon allowed by the model). This is a discrete-time counterpart to the result of CGJ, who show that every continuous-time model can be factorized in terms of the derivatives of the term structure (or discount function), evaluated at maturity zero.

Specifying the factors as short-maturity forward rates is not restrictive, because the question
of factor identity is distinct from that of factor measurement. The true forward rates are model-independent objects, which is precisely why the problem of identification can be solved. In fact, one is able to estimate factors and parameters jointly by Kalman filter, which uses information in all maturities available to the researcher. Alternatively, one can use any set of observed linear combinations of bond prices (or yields) that are believed to be measured without error.\(^1\)

Working with discrete-time formulation is clearly a matter of taste, but it definitely has some advantages. Every model written in continuous time needs discretization. Furthermore, discrete-time formulation in terms of short-maturity forward rates leads more naturally to a state-space representation that can be brought to the data than the formulation in terms of the derivatives of the discount function, used in CGJ.\(^2\)

A related advantage of specifying the model in terms of short-maturity forward rates as factors is that the the risk-neutral dynamics can be represented by a companion matrix, canonically parametrized by \(N\) unrestricted real numbers. All free parameters are contained in the last row, while all other entries are either 0, or 1. The structure of the companion matrix reflects the fact that under the pricing measure, forward rates of longer maturities move one-to-one with risk-adjusted expectations of future forward rates.

The total number of parameters in the \(Q\) dynamics, in addition to the parameters of the innovation covariance matrix, is \(N + 1\), since one extra real number is necessary to retain full flexibility of the risk-neutral drift (i.e., the constant vector in the VAR). This is consistent with Joslin et al. (2011) (JSZ, for short), who parametrize the risk-neutral transition in terms of its eigenvalues, and report the same number of risk-neutral parameters. However, the companion parametrization appears to be conceptually more straightforward, and even preferred under some circumstances, for example when it can be suspected that some eigenvalues are repeated, in which case the risk-neutral transition matrix may not be diagonalizable. Mathematically, under the companion parametrization, one works with the coefficients of the minimal polynomial of the transition matrix, instead of its roots. Finding the coefficients is enough to recursively solve for all bond prices of maturities encountered in practice.\(^3\)

\(^1\)The latter approach is standard in the literature, e.g., Joslin et al. (2011), Hamilton and Wu (2012). In the empirical section of the paper I adopt both methods, and show that they lead to similar conclusions.

\(^2\)In the context of CGJ, one needs to approximate the derivatives first. The authors use polynomial splines fitted to short-maturity segments of empirical PCA loadings, and compute the derivatives from the splines.

\(^3\)If some eigenvalues are repeated, the risk-neutral matrix under the JSZ parametrization is in Jordan form. This
The coefficients of the minimal polynomial can be interpreted as recursive hedge ratios, making the approach interesting from another point of view. Since the conditional investment opportunity set implied by the model is invariant to the choice of basis, one can express it in terms of realized forward premia, defined as excess returns of bonds with neighboring maturities. Then, the estimated coefficients can be used to recursively express forward premia of higher maturities in terms of linear combinations of forward premia of exactly $N$ preceding maturities. The forward premia of $N$ lowest maturities form a canonical basis for the space of excess returns, since they are directly linked to innovations in the short-maturity forward rates, i.e., they track factor innovations.

To illustrate the convenience of the proposed parametrization, I estimate unrestricted two-factor, and three-factor specifications in the monthly data set of Fama-Bliss discount bonds (Fama and Bliss (1987)), both by Kalman filter, and using empirical PCA scores as observable factors. The three-factor specification may be over-fitted in a sample with only five maturities, but the exercise is still useful to interpret some results in the literature, especially the nature of bond return predictability documented by Cochrane and Piazzesi (2005), who work with the same data set. Technically, the estimated risk-neutral dynamics of the three-factor model robustly features a very large eigenvalue, which is necessary to reconcile the observed term structure dynamics with no-arbitrage. In the data, the spread between maturities 4 and 5 years appears to move independently of the first three maturities, which is inconsistent with no-arbitrage according to the main result of the paper, and yet this spread significantly predicts excess bond returns, as documented by Cochrane and Piazzesi (2005). Intuitively, the large eigenvalue emerges as an attempt of the estimated model to link the observed variation in this spread to some linear combination of the first three maturities. This inconsistency manifests itself in very large model-implied conditional Sharpe ratios, often as high as 2-3 (maximally 6) associated with portfolio strategies that bet on term structure smoothing. One possible interpretation is that some fraction of documented return predictability may be associated with times of market stress, when limits to arbitrage allow for the existence of factors that frictionless models classify as close to arbitrage opportunities. This indirectly complements the evidence that market illiquidity may be an important driver of bond risk premia, as documented by Fontaine and Garcia (2011), or Hu et al. (2013). The large estimated

\[ \text{is the borderline case between real and complex eigenvalues. Although the probability of finding repeated eigenvalues in any data set is close to zero, the online supplement to Joslin et al. (2011) (Joslin et al. (2010)) shows that the popular arbitrage-free Nelson-Siegel specification of Christensen et al. (2011) does involve a repeated eigenvalue.} \]
Sharpe ratios are also reminiscent of the findings of Duffee et al. (2010), who shows that models of the essentially-affine class can imply extreme Sharpe ratios when the number of factors is too large, which the author interprets as sign of model over fitting. In contrast, the present paper proposes an alternative interpretation of this type of evidence.4

The rest of the paper, after the literature overview, is organized in three sections, together with concluding remarks. Section 2 presents the main results related to the existence, uniqueness, and properties of the companion parametrization. Section 3 shows two alternative ways of model estimation under the proposed parametrization, interprets the coefficients of the associated minimal polynomial, discusses the links to the continuous-time parametrization of CGJ, and explains the empirical results obtained in the real data through the lens of the main result of the paper.

1.1 Related Literature

The literature on affine term structure models usually credits the earliest developments to Vasicek (1977), and Cox et al. (1985). These models are specified in continuous time, and their main advantage is analytical tractability. Early models feature the short rate as the only factor, counter-factually predicting perfect correlation between bonds of different maturities. Multi-factor models became more popular in the 90’s. Litterman and Scheinkman (1991) show that three factors explain the cross section of yields with great precision. Duffie and Kan (1996) are credited for introducing the class of multi-factor affine models. On the other hand, the evidence of bond return predictability, especially due to Fama and Bliss (1987), and Campbell and Shiller (1991), and later Cochrane and Piazzesi (2005), spanned interest in models able to explain time variation in bond risk premia. One modeling strategy that achieves this is to incorporate stochastic volatility, as in the completely-affine class of Dai and Singleton (2000). The alternative strategy is to model factors as homoskedastic, while letting conditional Sharpe ratios move over time together with the factors, as in the essentially-affine class of Duffee (2002). In the model of Duffee (2002), the risk premia can in principle be linked to all state variables, which permits a full separation of the objective, and the risk-neutral dynamics. Most of these models are specified in continuous time. The discrete-time version of the essentially-affine model was proposed by Ang and Piazzesi (2003), building on earlier

---

4Duffee et al. (2010) uses samples with more maturities (but shorter time span), and documents extreme Sharpe ratios implied by models with more than three factors.
results of Backus and Zin (1994).

There are approaches to term structure modeling that do not require the assumption of no arbitrage. It is an empirical fact that bond prices move in just a few dimensions, and the associated factor loadings are roughly stable over time. Formulating a term structure model is then a matter of finding the factors (with the associated loadings), and postulating some factor dynamics. Standard methods of term structure fitting (McCulloch (1975), Nelson and Siegel (1987), and Svensson (1995)) completely ignore the time-series aspect, although the latter can be added (Diebold and Li (2006), Christensen et al. (2011)). Another popular approach rests on principal component analysis of yields, motivated by Litterman and Scheinkman (1991). Here, one is able to obtain a purely statistical description of the term structure, which can be complemented by assuming some (usually linear) stochastic process for factor scores.

The most important piece of motivation behind the development of the affine class of Duffie and Kan (1996) was to provide a tractable framework consistent with no arbitrage. However, flexibly specified no-arbitrage models pose practical difficulties with identification, as mentioned already in the introduction. One approach to this problem was proposed by Dai and Singleton (2000), who impose parameter normalizations intended to rule out all invariant transformations. Collin-Dufresne et al. (2008) point out some difficulties with this approach (see also references therein), and offer a solution in the spirit of Duffie and Kan (1996), by transforming the factors into a set of model-independent properties of the term structure, namely the derivatives of the discount function evaluated at maturity zero. The current paper contributes to the literature by showing that their results extend to discrete-time specifications, and that the possibility of using short-maturity forward rates as factors is a natural consequence of no arbitrage. While CGJ effectively work under the assumption of observable factors, in this paper I use the resulting canonical form to estimate model parameters and factors via maximum likelihood (Kalman filter) in a fully self-consistent way. Due to its universality, this approach can be seen as alternative to that of JSZ.

Further developments in the literature have mostly been focused on finding alternatives to Gaussian models with linear factor dynamics. Examples include Leippold and Wu (2002), Duarte (2004), Cheridito et al. (2007), Le et al. (2010), Filipović et al. (2015).
2 Discrete-Time Companion Parametrization

This section starts with an essentially affine Gaussian model in its most general form, and shows how it leads to an observationally-equivalent specification in terms of short-maturity forward rates.

2.1 General Model With Latent Factors

Let $X_t$ be a vector of latent factors, interpreted as causal determinants of the term structure. The number of factors is $N$, and it is assumed that all these dimensions are reflected in bond prices, i.e., all factors are spanned.\(^6\)

It is common to specify the general model in terms of factor dynamics under both objective and risk-neutral measures, coupled with an equation linking the short rate to the factors,

\[
\begin{align*}
X_{t+1} &= \mu_X^P + A_X^P X_t + \varepsilon_{X,t+1}^P \\
X_{t+1} &= \mu_X^Q + A_X^Q X_t + \varepsilon_{X,t+1}^Q \\
f_t^1 &= f_0^1 + f_1^1 X_t,
\end{align*}
\]

where the conditional covariance matrix of innovations $\varepsilon_{X,t+1}^P$ is $\Sigma_X$.

Standard no arbitrage arguments allow to solve for the zero-coupon (log) bond prices as $b_t^m = b_0^m + b_1^m X_t$, with coefficients

\[
\begin{align*}
b_1^m &= b_1^{m-1} A_X^Q - f_1^1, \\
b_0^m &= b_0^{m-1} - f_0^1 + b_1^{m-1} \mu_X^Q + \frac{1}{2} b_1^{m-1} \Sigma_X b_1^{m-1},
\end{align*}
\]

as shown in Appendix A, equations (31), (35). Similarly, the model-implied, one-period forward rates (continuously compounded) are $f_t^m = f_0^m + f_1^m X_t$, with

\[
\begin{align*}
f_1^m &= f_1^1 (A_X^Q)^{m-1}, \\
f_0^m &= f_0^1 - b_1^{m-1} \mu_X^Q - \frac{1}{2} b_1^{m-1} \Sigma_X b_1^{m-1},
\end{align*}
\]

\(^6\)The results of this paper can be extended to the case of unspanned factors described by Duffee (2011), in which some latent factors affect short rate expectations and bond risk premia in exactly offsetting ways. It is sometimes argued that macroeconomic variables like inflation or industrial output should be modeled as unspanned factors, see for example Joslin et al. (2014). On the other hand, Bauer and Rudebusch (2016) argue that the statistical evidence for the presence of such factors in real data is weak.
which follows Appendix A, equations (35), (40).

2.2 Short-Maturity Forward Rates as Factors

The generic model (1) has many parameters, not all of which are identifiable. Joslin et al. (2011) (JSZ) show that there exists an invariant factor transformation under which the dynamics under \( \mathbb{Q} \) are characterized by the eigenvalues of \( A_{X}^{\mathbb{Q}} \), plus one additional parameter related to the fixed component of the term structure.\(^7\) However, the eigenvalue parametrization has an unfortunate consequence that the risk-neutral transition matrix of the transformed model depends not only on the eigenvalues, but also on their configuration.\(^8\)

This section offers an alternative parametrization that appears to be conceptually more straightforward, and arguably easier to implement. I first show that all pricing information in the latent factors must be contained in the first \( N \) short-maturity forward rates, thought of as model-independent objects. The associated transition matrix is in companion form, uniquely parametrized by a vector of \( N \) real numbers. I state the main results of this subsection in a sequence of propositions.\(^9\)

**Proposition 1** Every discrete-time Gaussian term structure model (1) with \( N \) (spanned) latent factors is observationally equivalent to a model of the form

\[
Y_{t+1} = \mu_{Y} + A_{Y} Y_{t} + \varepsilon_{Y,t+1}, \\
Y_{t+1} = \mu_{\mathbb{Q}} + A_{\mathbb{Q}} Y_{t} + \varepsilon_{Y,t+1}, \\
f_{t} = e_{1}' Y_{t}, \quad e_{1} \equiv [1, 0, \ldots, 0]',
\]

in which \( Y_{t} \) is a vector of \( N \) short-maturity forward rates (including the risk-free rate).

**Proof.** By Appendix B, it is enough to find an invertible affine transformation between the two sets of factors, i.e., \( Y_{t} = \alpha + \beta X_{t} \).

Let \( \chi(t) = t^n - c_{n-1} t^{n-1} - \cdots - c_{1} t - c_{0} \) be the minimal polynomial of the risk-neutral transition matrix \( A_{X}^{\mathbb{Q}} \) of the original model, i.e., the lowest-degree monic polynomial satisfying the matrix

\(^7\)If the risk-neutral dynamics are stable, it is possible to identify this parameter to the long-run risk-neutral expectation of the short rate.

\(^8\)In practice, one usually assumes that the eigenvalues are real and distinct.

\(^9\)JSZ note the possibility of the companion parametrization in their footnote 22, p. 20. However, they choose to work with their eigenvalue parametrization, which (as they claim) is helpful to understand their main result.
equation \(\chi(A_X^Q) = 0_{N \times N}\). If the degree of \(\chi\) is \(n\), then by definition

\[(A_X^Q)^n = c_{n-1}(A_X^Q)^{n-1} + \cdots + c_1 (A_X^Q) + c_0 I. \tag{5}\]

Pre-multiplying (5) by \(f_t^{1t}\), post-multiplying by \(X_t\), and using the first equation in (3), one obtains

\[f_t^{n+1} = c_{n-1} f_t^n + \cdots + c_1 f_t^2 + c_0 f_t + \mu, \tag{6}\]

where \(\mu\) encompasses all fixed components of the model-implied forward rates needed to make the two sides equal. The implication of (6) is that the variation in \(f_t^{n+1}\) must be explained by the variation in forward rates of maturities between 1 and \(n\).

Multiplying both sides of (5) by \(A_X^Q\), repeating similar steps, and using (6) leads to the conclusion that also the variation in \(f_t^{n+2}\) must be fully explained by the forward rates of maturities between 1 and \(n\). By induction, all movements in forward rates must be spanned by the movements in the first \(n\) short maturity forward rates. \(\tag{11}\)

It follows that \(n \geq N\), or otherwise the term structure could only move in fewer dimensions than \(N\), contrary to the spanning assumption. On the other hand, the degree of the minimal polynomial is bounded by \(N\), because every square matrix satisfies its characteristic polynomial, which is of degree \(N\). \(\tag{12}\)

Stack the first \(n\) forward rates into vector \(\mathcal{Y}_t = \alpha + \beta \mathcal{X}_t\), with \(\alpha\) and \(\beta\) composed of model-implied coefficients (3). By the argument above, \(\beta\) is invertible. \(\blacksquare\)

I now turn to the parametrization of the transformed model (4). Before stating the result, define \(\Sigma_\mathcal{Y}\) as conditional covariance of shocks \(\varepsilon^\mathcal{P}_{\mathcal{Y},t+1}\). Also, let \(c = [c_0, c_1, \ldots, c_{N-1}]\)’ be a vector of \(N\) real numbers, and \(\mu\) an additional parameter. \(\tag{13}\)

**Proposition 2** The model of Proposition 1 is parametrized by \(\mu^\mathcal{P}_\mathcal{Y}, A^\mathcal{P}_\mathcal{Y}, \Sigma_\mathcal{Y}, \mu, c\). Then:

---

10. A polynomial is called monic if the coefficient at the highest-degree term is normalized to one.
11. Equation (6) looks like a non-stochastic difference equation that could be used to compute the whole term structure from the first \(n\) forward rates. This is not exactly correct, because the fixed term \(\mu\) only applies to maturity \(n + 1\) (in an \(n\)-factor model), and is generally different for other maturities.
12. The result that every matrix satisfies its characteristic polynomial is known as Cayley-Hamilton theorem. See, for example, Atiyah and Macdonald (1969).
13. The notation of \(c\) and \(\mu\) is not accidental, see equation (6) in the proof of Proposition 1. I keep the same symbols for reader’s convenience.
a) The risk-neutral drift is

\[
\mu^Q_Y = [\Delta j^1, \ldots, \Delta j^{N-1}, \Delta j^N + \mu]' \quad \text{with}
\]

\[
b^m_1 \equiv [1, \ldots, 1, 0, \ldots, 0], \quad \text{(ones in the first } m \text{ positions)},
\]

\[
\Delta j^m \equiv \frac{1}{2} (b^m_1 \Sigma_Y b^m_1 - b^{m-1}_1 \Sigma_Y b^{m-1}_1).
\] 

\(\text{(7)}\)

b) The risk-neutral transition matrix is

\[
A^Q_X = \begin{bmatrix}
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
c_0 & c_1 & c_2 & \ldots & c_{N-1}
\end{bmatrix}
\]

\(\text{(8)}\)

**Proof.** By Proposition 1, one is allowed to solve for the model-implied bond prices and forward rates starting from (4). The parametrization can then be inferred from the requirement that the model-implied forward rates of the first \(N\) maturities coincide with the new factors.

The mathematical structure of the transformed model (4) is identical to the structure of (1), so the model-implied forward rates must be affine in the new state vector, \(f^m_t = \hat{f}^0_0 + \hat{f}^m_1 Y_t\), with coefficients satisfying restrictions analogous to (3),

\[
\hat{f}^m_1 = \hat{f}^1_1 (A^Q_Y)^{m-1},
\]

\[
\hat{f}^m_0 = \frac{1}{2} \hat{b}^{m-1}_1 \hat{\mu}_Y \hat{b}^{m-1}_1 - \frac{1}{2} \hat{b}^{m-1}_1 \Sigma_Y \hat{b}^{m-1}_1,
\]

\(\text{(9)}\)

where \(\hat{b}^m_1\) are the model-implied coefficients of log zero-coupon bonds with respect to \(Y_t\).

Since the new factors are themselves forward rates, it must be that \(\hat{f}^0_0 = 0\), and \(\hat{f}^m_1 = [0, \ldots, 1, \ldots, 0]'\) (one in the \(m\)-th position), for \(m \in \{1, \ldots, N\}\). Similarly, \(\hat{b}^m_1 = -[1, \ldots, 1, \ldots, 0]'\) (ones up to the \(m\)-th position).\(^{14}\) Substituting into the first equation of (9), it is necessary that the first \(N-1\) rows of \(A^Q_Y\) must be of the form consistent with (8). On the other hand, substituting the same requirements into the second equation of (9) yields \(N\) linear restrictions on the risk-neutral

\(^{14}\text{This follows from the identity } b^m_t = -(f^t_1 + \cdots + f^m_t).\)
drift, each of the form

\[ 0 = [1, \ldots, 1, \ldots, 0] \mu_Y^Q - \frac{1}{2} [1, \ldots, 1, \ldots, 0] \Sigma_Y [1, \ldots, 1, \ldots, 0]' . \]  

Finally, we know from the proof of Proposition 1 that the \( N + 1 \)-th forward rate satisfies

\[ f_{t}^{N+1} = \mu + c_{0} f_{t}^{1} + c_{1} f_{t}^{2} + \cdots + c_{n-1} f_{t}^{N} . \]  

This completes the proof by showing that the last row of \( A_Q^Y \) must be of the form consistent with (8), and providing the \( N \)-th restriction on the risk-neutral drift, namely

\[ \mu = [1, \ldots, 1] \mu_Y^Q - \frac{1}{2} [1, \ldots, 1] \Sigma_Y [1, \ldots, 1]' . \]  

Solving the (recursive) system of equations (10), (12) results in (7). ■

Proposition 2 allows to quickly compute the number of parameters in a general (unrestricted) model. The conditional expectation of factors is determined by \((N+1)N\) numbers in \( \mu_Y^P \) and \( A_Y^P \). The factor covariance matrix is of dimension \((N+1)N/2\). The risk-neutral dynamics is additionally parametrized by \( N + 1 \) real numbers.

The next proposition establishes the uniqueness of canonical form (4) within each class of observationally equivalent models.

**Proposition 3** All observationally equivalent models with \( N \) factors share the same parameter values of Proposition (2).

**Proof.** By Proposition 1, any two \( N \)-factor models can be stated in terms of the first \( N \) forward rates. The observational equivalence then requires that the (conditional and unconditional) dynamics of these forward rates be identical, i.e., both models must share the same \( \mu_Y^P, A_Y^P, \) and \( \Sigma_Y \) under their respective equivalent forms (4).

The \( N + 1 \)-st (predicted) forward rates must also be identical, for every possible configuration of the factors. By (11), this is only possible if the two models agree on \( \mu \), and \( c \). ■
2.3 Difference Factors

This section proposes a further transformation that leads to factorization similar in spirit to the one adopted by Collin-Dufresne et al. (2008). The benefit relative to the already discussed formulation is mainly practical. For example, the forward rates of short maturities tend to be highly collinear, preventing precise measurement of their VAR dynamics.\(^{15}\)

Consider the following transformation of the first three (for clarity of exposition) short maturity forward rates into what will be referred to as *difference factors*,

\[
Z_t = \begin{bmatrix} g_t^1 & g_t^2 & g_t^3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} f_t^1 \\ f_t^2 \\ f_t^3 \end{bmatrix},
\]

The first element in \(Z_t\) is just the short rate. The second is the spread (difference) between the forward rate \(f_t^2\) and the short rate. The third is the difference between \(f_t^3 - f_t^2\), and the previously defined spread. In general, one can recursively define \(N\) new factors starting from \(N\)-dimensional vector of forward factors \(Y_t\) according to

\[
Z_t = (P_N)^{-1}Y_t,
\]

where \(P_N\) is a Pascal matrix of size \(N\).\(^{16}\)

Since (13) defines an invariant transformation (see Appendix B), the model (4) is observationally equivalent to a model expressed in terms of \(Z_t\),

\[
\begin{align*}
Z_{t+1} &= \mu^P_Z + A^P_Z Z_t + \varepsilon^P_{Z,t+1}, \\
Z_{t+1} &= \mu^Q_Z + A^Q_Z Z_t + \varepsilon^Q_{Z,t+1}, \\
f_t^1 &= e^1_t Z_t, \quad E(\varepsilon_Z \varepsilon_t^Z) = \Sigma_Z,
\end{align*}
\]

\(^{15}\)For example, the OLS estimate of VAR dynamics can be used as natural initial guess to initialize numerical search.

\(^{16}\)The Pascal matrix of size \(n\) is defined as

\[
P_n(i,j) = \begin{cases} \binom{i-1}{j-1}, & n \geq i \geq j \geq 1, \\
0, & \text{otherwise.}
\end{cases}
\]
with the dynamics linked to the starting model (4) by formulae (44) in Appendix B.\footnote{The model (14) is a discrete-time counterpart to the one in CGJ (the case of homoskedastic shocks). However, the difference factors introduced here are conceptually not approximations to the derivative factors of CGJ (in a discrete time model, the derivatives at maturity zero are not even defined). Instead, the results of CGJ obtain in the limit, as the length of the time intervals approaches zero, and the number of factors is held constant.}

The risk-neutral transition matrix of the transformed model has the form \( A^Q_2 = I + \hat{C} \), where \( \hat{C} \) is a companion matrix parametrized by \( \hat{c} \equiv [\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_{N-1}]' \). The parameter vector \( \hat{c} \) is linked to \( c \) of Proposition 2 in a special way. Let \( \hat{c}_{\text{ext}} \equiv [\hat{c}'; -1]' \) be an extended version of \( \hat{c} \), and similarly \( c_{\text{ext}} \equiv [c'; -1]' \). Then,

\[
\hat{c}_{\text{ext}} = (P_{N+1})'c_{\text{ext}},
\]

where \( P_{N+1} \) is a Pascal matrix of size \( N+1 \). Specifying the model in terms of the difference factors is therefore mathematically equivalent to (4), and relations (13), (15) offer a quick way of shifting between the two parametrizations.\footnote{The results of the last paragraph can be verified by applying the proposed factor transformation.}

\section{Empirical Results}

This section illustrates the usefulness of the results above in practical contexts. I first show how to estimate essentially-affine models in their canonical form, using the standard set of Fama-Bliss discount bonds as example, under alternative assumptions on the number of factors, and their observability. In addition, I provide interpretation of the results through the lens of Proposition 1, which sheds some new light on several issues in the literature.

\subsection{Data}

The standard sample of Fama and Bliss (1987) discount bonds is from CRSP.\footnote{The detailed description of the data can be found on http://www.crsp.com/products/documentation/fama-bliss-discount-bonds-%E2%80%93-monthly-only.} The set consists of five full-year maturities between one and five years, sampled at monthly frequency. The sample starts in June 1956, and ends in July 2015. One advantage of the data set is that no smoothing is applied at the stage of its construction, which is especially important from the point of view of the current study, since one application of the canonical form is to estimate factors jointly with model parameters. Another advantage is long time span, covering the period of high inflation and volatile bond risk premia of the 1980’s and 1990’s, as well as more calm periods of 1950’s and 1960’s.
Prior applications of the data set include Fama and Bliss (1987), and Cochrane and Piazzesi (2005, 2009), who study predictability of excess bond returns, and (closely related) failure of the expectations hypothesis. The latter authors show that the predictive content in Fama-Bliss data set is richer than in another well-known set constructed by Gurkaynak et al. (2007) with the use of smoothing methods of Svensson (1995).20

The disadvantage of using the FB data for the purposes of the current study is that the annual maturity intervals do not match with monthly calendar intervals. This can be circumvented in the empirical analysis by writing the models in terms of forward rates of annual (as opposed to monthly) maturities, which is valid to the extent to which there exists an invertible transformation between these rates and the true factors. While in principle it could be that the first $N$ annual rates move in fewer dimensions than $N$ (i.e., some factors do not propagate from the shortest maturities to longer rates), this appears unlikely if one is interested in models with relatively low numbers of factors. I therefore proceed with the working assumption that all estimated models can be parametrized in terms of the first $N$ annual forward rates.

3.2 Estimation by Kalman Filter

The model of Proposition 1 leads to the state-space representation

\[ Y_{t+1} = \mu_{Y} + A_{Y} Y_{t} + \varepsilon_{Y,t+1}, \]
\[ b_{t} = B_{0} + B_{1} Y_{t} + v_{t}, \]

where $b_{t}$ is a vector of noisy observations of log bond prices, $B_{0}$ and $B_{1}$ are no-arbitrage factor loadings, and $v_{t}$ is a vector of normally distributed measurement errors with covariance matrix $R = \sigma_{v}^{2} I$, uncorrelated with factor innovations.21 Appendix C explains how to derive the state-space form (16), and provides the necessary details on estimation by Kalman filter.

The set of parameters is $\Theta \equiv \{ \theta_{\mu}, \theta_{A}, \theta_{\Sigma}, \mu, c, \theta_{R} \}$, in which $\theta_{\mu}$, $\theta_{A}$, and $\theta_{\Sigma}$ parametrize the objective factor dynamics in (16), and $\mu$ and $c$ determine the $\mathbb{Q}$ dynamics according to (7), (8) of Proposition 2. $\theta_{R}$ characterizes the noise covariance matrix $R$.

---

20Cochrane and Piazzesi (2009) conjecture that term structure smoothing throws out important part of information contained in the unsmoothed yields.

21The model could be estimated under arbitrary $R$, at the cost of introducing more parameters. Author’s experience suggests that the results are reasonably insensitive to the choice of $R$. 
Following most of the literature, I assume three factors in the benchmark specification, which results in 23 parameters. I also estimate a two-factor model, for which this number drops to 13, and a four-factor specification with 36 parameters, although arguably the small number of maturities in the FB data set suggests that the latter model is almost surely over-fitted. For this reason, I only report full results for the first two cases.

The success of numerical search often depends on starting values. To obtain an initial guess of $\theta_\mu$, $\theta_A$, and $\theta_\Sigma$, I estimate a VAR on the first $N$ noisy factors. The staring vector $c$ is obtained as follows. I first perform a PCA on all forward rates, and store factor loadings associated with $N$ most important (in terms of variance) factors. Next, I exploit recursions $f_1^{m'} = f_1^{m-1} A^Q$ to estimate the columns of the (hypothetical) transition matrix $A^Q$ by running linear regressions with no constant terms. Finally, I compute the characteristic polynomial of the estimated matrix, and store its coefficients as $c$. The initial value of $\mu$ is zero, which appears a natural choice given the structure of the risk-neutral drift in Proposition 2, and the initial standard deviation of noise is set to 10 basis points.

The numerical optimization was performed using the Nelder-Mead simplex algorithm, implemented in Python/SciPy function `minimize`. The estimation time was between seven seconds (two-factor model) to about two minutes (four-factor model) per estimation loop, and the minimum number of ten repetitions (starting from previously obtained values) was required before accepting the results as final.

The estimation results for the three-factor model are presented in Table 1. The monthly $P$ dynamics of the forward factors are shown in panel A. The annualized version is in panel B, for consistency with panel C, which shows their annual dynamics under $Q$. For numerical reasons (high collinearity of the measured forward rates), I use the equivalent parametrization of section 2.3 for the $P$ dynamics of the model. The estimated forward factors $\lambda$, and their dynamics, are then obtained from the inverse of (13). The choice of the parametrization appeared to significantly improve the speed of convergence of numerical search (the maximum value of the log likelihood was obtained much faster).

**Footnotes:**

22 For numerical reasons (high collinearity of the measured forward rates), I use the equivalent parametrization of section 2.3 for the $P$ dynamics of the model. The estimated forward factors $\lambda$, and their dynamics, are then obtained from the inverse of (13). The choice of the parametrization appeared to significantly improve the speed of convergence of numerical search (the maximum value of the log likelihood was obtained much faster).

23 This method exploits the invariance of characteristic polynomial under matrix similarity, and the fact that the minimal polynomial used in the companion parametrization must equal the characteristic polynomial. In order to compute the characteristic polynomial, one can use Python/NumPy’s function `poly`, or Matlab’s function `charpoly`. Matlab’s implementation of the same algorithm is `fminsearch`. For a discussion of some of its known convergence properties, see Lagarias et al. (1998).

24 The author used 'xatol': 0.001, 'fatol': 0.01, 'maxfev': 10000 in every loop. With low numbers of factors (two, three), first convergence was obtained at once, or very quickly. The four-factor specification required 19 repetitions (27 minutes) before achieving the first (and final) convergence.

25 Choosing the annual forward rates as factors implies that the canonical companion corresponds to their annual
The last row of the companion matrix contains the coefficients $c_0, c_1, \ldots, c_{N-1}$ (in this order). From mathematical point of view, $c_0(-1)^{N-1}$ is the determinant of $A_Q$, and $c_{N-1}$ is its trace, although this is of limited importance here. The numbers also have an economic interpretation in terms of a replicating portfolio of higher-maturity excess bond returns (or forward premia), on which I will comment below. The three estimates in Table 1 are considerably away from their starting values computed from empirical PCA loadings (3.11, −9.33, 7.58, respectively).

The pricing errors in panel D are reasonably small, ranging between 5.8-15.8 basis points, depending on maturity, and computation method (root-mean-square vs. mean-absolute). This panel also shows the estimated noise standard deviation of 17.7 b.p. These numbers are comparable to other studies fitting affine models to Fama-Bliss data, e.g., Cochrane and Piazzesi (2009).

The eigenvalues implied by the estimated parameters are in panel E, in decreasing order (with the exception of the last row). The $P$ dynamics appear to be characterized by a stable VAR process, with all eigenvalues positive, real, and distinct. Somewhat strikingly, the $Q$ dynamics features a large (in absolute terms) eigenvalue of 2.48, implying an explosive behavior of some linear combination of forward rates under the pricing measure. This finding is robust across models (and methods of their estimation) whenever the number of factors is above two, and will be given special attention further in the paper.

The model-implied forward factors (for the three-factor model) are in Figure 1, together with their noisy data counterparts, and the corresponding Kalman-filter errors. Figure 2 presents the same information in terms of the $Z_t$ factors of section 2.3. It is evident that the difference factors of higher order require relatively more smoothing in order to be consistent with the three-factor model with no arbitrage.

The two-factor model is presented in Table 2 (with the same structure as Table 1). It appears that the $P$ dynamics are somewhat better behaved, which is likely the consequence of reduced amount of collinearity. The coefficients in the companion matrix are also smaller in absolute values, with its determinant being 0.64 instead of 1.69. Consistently, all eigenvalues are below one. All this comes at a cost of higher pricing errors (although still uniformly below 20 b.p.), and larger estimated noise standard deviation of 23.5 b.p.
3.3 Estimation Under Observable Factors

Much of the literature on estimating term structure models assumes that factors are observable, i.e., that there exists at least $N$ linear combinations of term structure observables (yields, bond prices, or forward rates) that can be inverted for the pricing factors.\(^{28}\) Intuitively, if factors are observable, no filtering is needed to extract information about the precise position of the state vector, which significantly reduces estimation time. This section presents estimation results of unconstrained models under this assumption.

Suppose a known matrix $W$ produces a vector of bond portfolios $P_t$ priced without error. In applications, $W$ usually coincides with the first $N$ PCA loadings, and this assumption is adopted here. The assumption of no errors on $P_t$ is equivalent to the statement that there exists an invariant transformation allowing to restate the model in terms of $P_t$ in place of the latent factors. In the context of this paper, the pricing information in $P_t$ is supposed to be the same as in the forward factors $X_t$, without the need of filtering.

Based on the results of Joslin et al. (2011), and Hamilton and Wu (2012), if factors are observable (and the model is otherwise unconstrained), the ML estimate of the $P$ dynamics coincides with the OLS estimate, which allows to exclude $N(N + 1)$ from numerical search.\(^{29}\) Essentially, the assumption pins down the factors, together with their conditional expectations, and the remaining task is to find factor loadings consistent with no arbitrage.

The canonical parametrization of Proposition (2) is particularly convenient in the context of observable factors, since the number of parameters subject to numerical search is reduced to $N + 1$ in the risk-neutral dynamics, $N(N - 1)/2$ in the innovation covariance matrix, plus any parameters describing the measurement errors. These numbers are 7, 11, and 16 for the models with two, three, and four factors (assuming a single noise parameter).

Estimation under the assumption of observable factors is explained in detail in Appendix D.\(^{30}\) The initial parameters $\mu$, $c$ of the $Q$ dynamics are computed in the same way as in the previous section, and the starting value for the innovation covariance matrix (its Cholesky decomposition)

\(^{28}\)Notable examples are Joslin et al. (2011), and Hamilton and Wu (2012).
\(^{29}\)The OLS residuals can also be used to construct a good initial guess of the covariance of shocks.
\(^{30}\)This appendix can also be regarded as an alternative explanation (informal proof) of the main result of Joslin et al. (2011), namely that imposing no arbitrage is not helpful in factor forecasting if the model is unconstrained, and the factors are observable.
is obtained from VAR residuals of PCA factor scores.\textsuperscript{31}

The estimation time is considerably faster than with Kalman filter. A single estimation loop takes about 3, 12, or 33 seconds, depending on the number of factors.\textsuperscript{32}

The results for the 3-factor model are displayed in Table 3. The monthly and annualized $\mathbb{P}$ dynamics of the implied forward factors appear rather peculiar, which is the direct consequence of the assumption of observable factors, i.e., that there exists an exact mapping between the empirical PCA scores and the three short-maturity forward rates. In fact, the determinant of this mapping implied by the estimated parameters is $-0.0019$, which suggests that at least one linear combination of the observable factors is difficult to detect from the short-maturity forward rates. However, in spite of seemingly unreasonable numbers, the eigenvalues of the $\mathbb{P}$ dynamics (panel E) appear quite sensible. The only obvious difference from Table 1 is that the smallest one implies an almost full decay of some linear combination of the forward factors after one year.

The estimated $\mathbb{Q}$ dynamics in panel C of Table 3, characterized by the $c$ vector contained in the bottom row of the companion matrix, are not fact from the initial guess described in the previous section. The small difference is the consequence of the fact that the ML estimate depends on other model parameters, which is not true of the initial guess, obtained from empirical PCA loadings alone. Interestingly, one of the eigenvalues under $\mathbb{Q}$ takes a very high value of 7.16, thus being much higher than the already anomalous value obtained by the Kalman filter.

The pricing errors are comparable in magnitude to the ones reported in Table 1, and range between 0.9 and 16.0 basis points. The model appears to extremely well explain the long end of the term structure, which (by comparison with the Kalman filter results) must be solely due to the assumption of observable factors.

The results in Table 3 for the two-factor model are very similar to the ones obtained by the Kalman filter (Table 2). In particular, the dynamics under both objective and pricing measures appear well-behaved, and the eigenvalues are all below one.\textsuperscript{33}

\textsuperscript{31}I parametrize the likelihood in terms of this covariance matrix, instead of the covariance of innovations to $\mathbb{Y}_t$, for numerical convenience. Given a set of parameters subject to the numerical search, there is an endogenous mapping between the two sets of factors, which allows to compute one covariance matrix from the other.

\textsuperscript{32}The convergence options are the same as in footnote 26. Estimation results are accepted as final after no less than five repetitions of the optimization routine, each starting from previously obtained values. Subsequent repetitions are typically much faster then the initial ones, and the total estimation time, depending on the number of factors, is about 5, 20, and 85 seconds.

\textsuperscript{33}The determinant of the mapping between the two PCA scores and the model-implied forward factors (not reported in the table) was $-0.06$, about 30 times greater than with three factors.
Figures 3 and 4 present the canonical factors estimated under the assumption that they can be obtained by inverting the PCA scores, and compare them to their Kalman-filter counterparts. The series are seemingly very close to each other, with differences not greater than the magnitudes of measurement errors reported above. It is virtually impossible to compare the estimation methods based on these figures alone, and they are presented mainly for completeness.

3.4 Portfolio Interpretation of the Minimal Polynomial

The minimal polynomial associated with the companion matrix $A^Q_Y$ can be (approximately) interpreted as recursive replicating portfolio of long-maturity forward premia, defined below, in terms of factor-mimicking excess returns.

One can define the realized forward premium of maturity $m$ ($m \geq 2$) as

$$f_{t+1}^m = f_t^m - f_{t+1}^{m-1},$$

i.e., the surprise with respect to the (unbiased) expectations hypothesis associated with maturity $m$, which is also the log excess return between bonds of two neighboring maturities,

$$f_{t+1}^m = r_{t+1}^m - r_{t+1}^{m-1}.$$

The model implies a factor structure in bond returns, i.e., every realized log return of $m$-maturity bond is of the form

$$r_t^m = f_t^1 - \frac{1}{2} b_1^{m-1}' \Sigma_Y b_1^{m-1} + b_1^{m-1}' (\Sigma_Y \Lambda_t + \varepsilon_Y^{p,Y,t+1}),$$

(17)

where $b_1^{m-1}$ is the vector of no-arbitrage coefficients with respect to the forward factors $Y_t$, which are also the loadings with respect to the factor innovations, and $\Lambda_t$ is the price of risk associated with the explicit stochastic discount factor (see Appendix A).

The realized forward premia can be computed by differencing (17), which results in

$$f_{t+1}^m = \frac{1}{2} (b_1^{m-1} - b_1^{m-2}') \Sigma_Y (b_1^{m-1} - b_1^{m-2})' (\Sigma_Y \Lambda_t + \varepsilon_Y^{p,Y,t+1}).$$

(18)
Under the companion parametrization, the differenced bond coefficients \( (b_{1}^{m-1} - b_{1}^{m-2})' \) are exactly the same as the rows of

\[
F \equiv \left[(1, 0, \ldots, 0)(A_{Y}^{Q})^{m-1}\right]_{M \times N}
\]

for some maximal maturity of interest \( M \). Since \( A_{Y}^{Q} \) is a companion matrix, the first \( N \) rows of \( F \) form an identity block \( I_{N \times N} \), and every subsequent row is a linear combination of \( N \) preceding rows, with coefficients given by the minimal polynomial, contained in \( A_{Y}^{Q} \) as its last row. Combining the forward premia (18) into a vector (of length \( M - 1 \)),

\[
f_{p r_{t+1}} = -\tilde{\mu}^{Q} - F(\Sigma_{t}^{Y} \Lambda_{t} + \varepsilon_{t}^{P_{Y}, t+1}),
\]

where \( \tilde{\mu}^{Q} \) is a vector of Jensen terms, one can see that only the premia of the first \( N \) maturities can move independently (and they exactly mimic factor innovations), while all subsequent premia are recursively spanned by the preceding ones. Alternatively, all forward premia can be replicated by the \( N \) factor-tracking excess returns, using the numbers in the appropriate rows of \( F \).\(^{34}\)

This interpretation of the minimal polynomial is of course only approximate if one would like to compute excess returns from simple bond returns, as would be done in practice.

Going back to the empirical results, Table 1 implies the following relationship (in a three-factor model, estimated by Kalman filter) linking the forward premium of maturity 5 years with forward premia of lower maturities

\[
f_{p r_{t+1}}^{5} = 1.69f_{p r_{t+1}}^{2} - 4.79f_{p r_{t+1}}^{3} + 4.14f_{p r_{t+1}}^{4}.
\]

The coefficients are large in absolute values, and partly offsetting.\(^{35}\) Evidently, the estimated model implies a rather peculiar replicating portfolio of the excess return \( f_{p r_{t+1}}^{5} = r_{t+1}^{5} - r_{t+1}^{4} \) in terms of factor-tracking excess returns of lower maturities. The two-factor model (Table 2) appears to perform much better in this respect, i.e., from the point of view of the plausibility of the estimated

\(^{34}\)The notation for the Jensen terms vector reflects the fact that the first \( N \) components of \( \tilde{\mu}^{Q} \) almost coincide with the risk-neutral drift of the factors in (7). The difference is only the term \( \mu \), which needs to be added to the last component of the risk-neutral drift to preserve model flexibility, but is not part of \( \tilde{\mu}^{Q} \). Also, the risk-neutral drift is only defined for the factors, while \( \tilde{\mu}^{Q} \) can be computed for any number of maturities.

\(^{35}\)These are not portfolio weights, so they do not add to one. The sum is 1.04, reflecting non-stationary factor dynamics under \( Q \).
replicating coefficients.

3.5 Anomalous Eigenvalues Imply Extreme Sharpe Ratios

The anomalous eigenvalues associated with $Q$ dynamics of the estimated three-factor models (using either estimation method) appear puzzling, but it is possible to justify their presence using the insights of Proposition 1. According to this proposition, all pricing information in latent factors must be present in short-maturity forward rates. By contraposition, there must exist arbitrage opportunities whenever there are factors that only affect long-term rates. If one imposes a model structure with no-arbitrage, the latter is ruled out by assumption, but even in this case there remains a possibility of observing very large Sharpe ratios which the model interprets as high prices of risk.\footnote{The presence of arbitrage is equivalent to infinite maximum Sharpe ratio. Large but finite Sharpe ratios are often referred to as arbitrage opportunities by practitioners.}

In this section I show that the anomalous eigenvalues imply existence of bond portfolios that offer Sharpe ratios that appear too large from the point of view of real-world experience. This finding is consistent with Duffee et al. (2010), who reports extreme Sharpe ratios estimated with four-, and five-factor specifications (estimated on shorter data sets, but with more maturities).

Looking at Figure 5, it appears that one of the model-implied PCA factors (curvature) mainly affects the spread between maturities of four and five years, thus being potentially hard to reconcile with no-arbitrage. Interestingly, the top left panel shows that the shape of this factor is very similar whether estimated by Kalman filter (without factor pre-measurement), or not. The factor appears genuine, which resounds the findings of Cochrane and Piazzesi (2005), who show that the 'curvature' PCA factor plays an important role in forecasting excess bond returns, contributing significantly to their return-predicting linear combination of forward rates.

In order to uncover the link between the estimated eigenvalues and maximal Sharpe ratios, it is useful to first identify the dimensions of factor shocks that cause term structure movements exactly corresponding to these eigenvalues. This can be done by applying the following factor transformation, taking the advantage of the fact that the estimated eigenvalues are distinct.

The model under the companion parametrization implies that the forward rates are

$$f_t = f_0 + FY_t,$$

(22)
with $F$ of the same form as in (19). The companion matrix $A^Q$ can be diagonalized into $A^Q = VLV^{-1}$, where $V$ is a Vandermonde matrix

$$V \equiv [l_1^{m-1}, l_2^{m-1}, \ldots, l_N^{m-1}]_{m \in 1, \ldots, N}$$

composed of row vectors of the eigenvalues, raised element-wise to consecutive integer powers, up to $N - 1$. Defining $V_t = V^{-1}Y_t$, the forward rates (22) are equivalently represented as

$$f_t = f_0 + \tilde{V}V_t,$$

with $\tilde{V}$ being an extended Vandermonde matrix of size $M \times N$, containing all powers of the eigenvalues up to $M - 1$. Similarly, the model-implied bond prices with respect to the transformed state vector $V_t$ are

$$b_t = b_0 + \tilde{B}V_t,$$

where $\tilde{B}$ is the negative of the cumulative sum (over the rows) of $\tilde{V}$.

By construction, representation (24) captures the effects of term structure movements along factor dimensions encoded in $V_t$, which are directly related to the distinct eigenvalues. Figure 6 graphs the columns of $\tilde{B}$ in the left panel, where the extreme effect of the anomalous eigenvalue is evident. The right panel scales the factor loadings in a way that they can be interpreted as reactions to conditional (annualized) standard-deviation shocks to components of $V_t$, holding the other components fixed. Here, one can even more clearly see that the factor dimension corresponding to the anomalous eigenvalue mainly affects bonds of maturities 4 and 5 years, and is responsible for only about 0.1-percentage point price change in the latter, leaving bonds of the first three maturities affected by no more than typical magnitudes of measurement errors.\footnote{The right panel cannot be interpreted in terms of impulse responses to structural shocks, though.}

With the help of the factor transformation described above, one can also investigate the extent to which the existence of the anomalous eigenvalue (and its associated factor dimension) affects the maximum Sharpe ratio. Under the factor model, the first $N$ rows of (20) completely summarize the conditional investment opportunity set. In order to express this set in terms of $V_t$ innovations,
\[ \tilde{F} = FV, \Sigma_Y = V^{-1} \Sigma Y V^{-1}', \tilde{\Lambda}_t = V' \Lambda_t, \tilde{\varepsilon}_{V,t+1}^P = V^{-1} \varepsilon_{V,Y,t+1}^P, \] and write (20) as

\[ \text{fpr}_{t+1} = -\tilde{\mu}^Q - \tilde{F}(\Sigma_Y \tilde{\Lambda}_t + \varepsilon_{V,Y,t+1}^P), \tag{25} \]

where again it is enough to focus on the first \( N \) rows. The components of \( \varepsilon_{V,Y,t+1}^P \) can be ordered by the eigenvalues, and I adapt an ordering from the largest to the smallest, with the exception of the anomalous eigenvalue, whose component is placed in the last position.

Repeatedly removing dimensions of excess returns from the conditional opportunity set must necessarily decrease the maximum attainable Sharpe ratio. In particular, one can orthogonalize the components of \( \varepsilon_{V,Y,t+1}^P \) by applying a Cholesky decomposition of the covariance matrix \( \Sigma_Y \). If \( C \) is such that \( \Sigma_Y = CC' \), then the square of the maximum Sharpe ratio for the unrestricted opportunity set is

\[ \text{SR}_{\text{max}}^2 = \lambda_t' \Sigma_Y \Lambda_t = (C' \lambda_t)'(C' \lambda_t). \tag{26} \]

The vector \( C' \lambda_t \) has the interpretation of conditional prices of risk associated with the orthogonalized shocks to \( V_t \), and the unrestricted maximum (squared) Sharpe ratio is simply its Euclidean norm. Setting the last \( k \) components of \( C' \lambda_t \) to zero results in a lower maximum Sharpe ratio that corresponds to the restricted opportunity set.

Figure 7 plots the sequentially restricted, as just described, maximum conditional Sharpe ratios for one-year investments, \( \sqrt{\text{SR}_{\text{max}}^2} \), implied by the estimated three- and two-factor models estimated by the Kalman filter. The top panel confirms that the three-factor model, although free of arbitrage in the strict sense, admits extreme Sharpe ratios of magnitudes about 6 in the years around 1980. Although this period can be considered anomalous due to the monetary policy experiment that largely destabilized nominal interest rates, this by itself should not justify the existence of close-to-arbitrage investment opportunities. Moreover, the values of maximum Sharpe ratio of about 2-3 are not uncommon throughout the whole sample, and such numbers are well beyond what has been considered reasonable in the literature.\(^{38}\) The figure also makes it evident that it is exactly the existence of the factor associated with the anomalous eigenvalue that expands the conditional opportunity set in such extreme way. The two remaining dimensions of excess returns only allow

\(^{38}\)Acceptable bounds for the maximum Sharpe ratio have risen over the years. Duffee et al. (2010) cites values from 0.25 used by Ross (1976), to 1.0 in Cochrane and Saa-Requejo (2000).
Sharpe ratios much closer to the acceptable standards, although also still above 1 around 1980’s.\textsuperscript{39}

3.6 Comments on Term Structure Forecasting

According to Joslin et al. (2011), the assumption of no arbitrage does not by itself improve factor forecasts obtained from estimated essentially-affine models, if one does not restrict the prices of risk, and under the assumptions that there exist linear combinations of yields that can be used to measure the factors exactly.\textsuperscript{40}

A natural question is whether it is worth to make the simplifying assumption of observable factors, which is guaranteed to fail in any given data set. On the other hand, using methods such as principal component analysis can be quite effective in measuring the factors prior to model estimation, if the assumed linear factor structure is also present in the real data.

To address this issue, I conducted a Monte-Carlo analysis of the forecasting performance of the two estimation strategies of sections 3.2, and 3.3. Based on simulations of 500 artificial samples of size exactly the same as the Fama-Bliss set, and using the three-factor model of Table 1 as the data-generating process, I found essentially no difference between the two methods in terms of out-of-sample forecasting, although the results may not be very general, given that all model assumptions were satisfied in the artificial samples by construction.\textsuperscript{41} In real data, one or more of the assumptions may fail (e.g., time-varying the number of factors, non-Gaussian shocks), favoring one method over another. This issue is, however, outside of the scope of the current paper.

\textsuperscript{39}The Sharpe ratios computed in this paper all correspond to log returns. Duffee et al. (2010) shows that the corresponding model-implied Sharpe ratios for simple returns (i.e., ones that are of more direct relevance to the investors) are always higher, and usually by much. On the other hand, the models considered in this paper are homoskedastic, so they assign the whole variation in maximum Sharpe ratios to time varying in expected excess returns.

\textsuperscript{40}Intuitively, essentially-affine models specify linear (VAR) dynamics in the factors, so that equation-by-equation ordinary least squares consistently recovers the parameters, provided that it is not possible to use any additional statistical information. If the prices of risk are fully flexible, than the $Q$ dynamics only informs about the prices of risk given the estimated $P$ dynamics, but does not contain any additional information about parameters of the latter. This is true in the Gaussian setup. Mathematically, the likelihood of observing a given sample of factors is flat in the parameters that determine the $Q$ dynamics, once it has been maximized with respect to parameters of $P$ that determine conditional factor forecasts. Appendix D to this paper can be seen as an alternative (w.r.t. Joslin et al. (2011)) explanation of this fact.

\textsuperscript{41}The results are available upon request, and in the previous draft of the paper.
3.7 Discussion of the Empirical Results

The theoretical results of section 2, combined with the empirical evidence of this section, together shed some light on the nature of return predictability in historical US bond data, at least when interpreted through the lens of the estimated model. More concretely, the evidence suggests that high bond risk premia are very often associated with conditional investment opportunities very close to arbitrage, which is broadly consistent with the literature on limits to arbitrage, starting with Shleifer and Vishny (1997). This literature has so far documented numerous cases in which no-arbitrage relations can seemingly fail in the context of fixed-income markets (e.g., Duffie (1996), Fleckenstein et al. (2014)), and it is by now well understood that full market efficiency is not always guaranteed. Moreover, there is evidence that funding liquidity of financial intermediaries is an important driver of asset prices, and bond risk premia in particular (Fontaine and Garcia (2011), Adrian et al. (2014), He et al. (2016)). The evidence presented here is conceptually consistent with Hu et al. (2013), who interpret term structure noise (with respect to smoothed yield curves) in terms of market illiquidity. However, the current paper suggests that the clear distinction between factors and noise may sometimes be difficult.

4 Concluding Remarks

In this paper, I develop a canonical parametrization for the very popular class of Gaussian dynamic term structure models with essentially-affine prices of risk, specified in discrete time, as in Ang and Piazzesi (2003). The possibility of the proposed parametrization rests upon the result that if there is no arbitrage, and if the term structure is driven by \( N \) spanning factors, then all information about these factors must be contained in the short end of the maturity spectrum, i.e., in the first \( N \) forward rates spaced at the same maturity intervals as the minimal time horizon of the investors.

This result complements Collin-Dufresne et al. (2008), who work with continuous-time models (under the assumption of a continuum of observed maturities), and show that the factor dynamics can be mapped into the dynamics of the first \( N \) derivatives of the term structure, evaluated at maturity zero.

This result leads to a natural parametrization of the \( Q \) dynamics in terms of the minimal polynomial of the transition matrix. After transforming the state vector into the set of \( N \) short-maturity...
forward rates, the latter is in companion form, with only the last row subject to estimation. Moreover, the last row contains exactly the coefficients of the minimal polynomial, which form a vector of unrestricted real numbers, and the estimation (of an unrestricted model) can be implemented without any knowledge of the eigenvalues. The full characterization of the risk-neutral dynamics is completed by specifying one more parameter for the risk-neutral drift, in addition to the parameters that define the innovation covariance matrix.

The companion parametrization is akin to the eigenvalue parametrization employed by Joslin et al. (2011). Both minimize the number of parameters that can be identified from the data, while preserving full model flexibility. The companion parametrization, however, appears conceptually more straightforward, and can be easier to implement, especially in situations where one suspects that some eigenvalues (roots of the minimal polynomial) may be repeated or complex, which makes numerical search over them harder to implement.

Due to relatively low number of unknown parameters, the model can be estimated via Kalman filter in time-efficient manner. In this case, the factors and factor loadings are fully consistent with internal no-arbitrage restrictions. The fact that only short-maturity forward rates are used as factors does not matter, since all observed maturities are treated symmetrically in the observation equation (i.e., factor identity is a distinct issue from factor measurement).

Alternatively, one can assume that factors are observable from some pre-measured linear combinations of yields, like the principal components of the term structure. This assumption is much more common in the literature, and I confirm in the empirical section that it leads to a great increase in the estimation speed.

Both estimation methods are tested on the sample of Fama-Bliss bonds, using specifications with two and three factors, and fully flexible prices of risk. I find that the assumption of no arbitrage is quite hard to reconcile with the data under the three-factor model. Technically, one of the factors picks up movements in long-maturities that are hardly present in the short end of the term structure, and thus provides investment opportunities with extreme Sharpe ratios (which is also manifested by a large eigenvalue in the estimated pricing measure). This appears to bridge the gap between the empirical findings of Duffee et al. (2010), who also documents extreme Sharpe ratios, although in models with more than three factors, and Cochrane and Piazzesi (2005), who document that a factor that loads heavily on the spread between yields of maturities 4 and 5 years.
significantly helps predict excess bond returns in the Fama-Bliss sample.

There are several extension possibilities. For example, one could introduce unspanned factors linked to macroeconomic variables in the empirical analysis, as in Joslin et al. (2014). Also, by focusing on unrestricted models, I effectively assumed that the dimensionality of risk premia is equal to the number of factors. For practical reasons (e.g., related to the efficiency of forecasting) one might want to restrict this dimensionality, which can easily be done by an appropriate parametrization of the $P$ dynamics. Another set of extensions is more directly linked to the proposed canonical form, under which it is straightforward to place, and test linear restrictions on the coefficients of the minimal (and characteristic) polynomial of the $Q$ dynamics. For example, a hypothesis of exactly unit root under the pricing measure can be formulated as a requirement that these coefficients sum to one. More generally, the companion parametrization might be a natural starting point whenever one is interested in testing hypotheses that can be formulated as linear restrictions on the estimated polynomial coefficients.
References


Table 1: Estimation results of a three-factor model by ML, using the Kalman filter. The sample is from June 1952 to July 2015.

<table>
<thead>
<tr>
<th></th>
<th>$\mu_Y$</th>
<th>$A_Y$</th>
<th>$\Sigma_Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Monthly Dynamics ($P$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.00055$</td>
<td>0.9478</td>
<td>0.1037</td>
<td>0.1037 -0.0624 1e-4</td>
</tr>
<tr>
<td>$-0.00023$</td>
<td>0.2764</td>
<td>0.3547</td>
<td>0.3715</td>
</tr>
<tr>
<td>$-0.00046$</td>
<td>0.3935</td>
<td>-0.8945</td>
<td>1.5085</td>
</tr>
<tr>
<td><strong>B. Annualized Dynamics ($P$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.00544$</td>
<td>0.6668</td>
<td>0.5425</td>
<td>-0.3222</td>
</tr>
<tr>
<td>$0.00257$</td>
<td>1.3550</td>
<td>-2.0277</td>
<td>1.6265</td>
</tr>
<tr>
<td>$0.00232$</td>
<td>1.8816</td>
<td>-3.9419</td>
<td>3.0362</td>
</tr>
<tr>
<td><strong>C. Annual Dynamics ($Q$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.00010$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$0.00038$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$-0.00206$</td>
<td>1.6946</td>
<td>-4.7895</td>
<td>4.1373</td>
</tr>
<tr>
<td><strong>D. Pricing Errors, and Noise St. Dev.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (b.p.)</td>
<td>12.54</td>
<td>11.66</td>
<td>15.67 15.83</td>
</tr>
<tr>
<td>MAPE (b.p.)</td>
<td>9.65</td>
<td>8.80</td>
<td>10.69 10.34</td>
</tr>
<tr>
<td><strong>E. Eigenvalues</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly $P$ Dynamics</td>
<td>$0.9910$</td>
<td>$0.9604$</td>
<td>0.8597</td>
</tr>
<tr>
<td>Annualized $P$ Dynamics</td>
<td>$0.8968$</td>
<td>$0.6155$</td>
<td>0.1630</td>
</tr>
<tr>
<td>Annual $Q$ Dynamics</td>
<td>$0.8644$</td>
<td>$0.7894$</td>
<td>2.4835</td>
</tr>
</tbody>
</table>
Table 2: Estimation results of a two-factor model by ML, using the Kalman filter. The sample is from June 1952 to July 2015.

<table>
<thead>
<tr>
<th></th>
<th>A. Monthly Dynamics (P)</th>
<th>B. Annualized Dynamics (P)</th>
<th>C. Annual Dynamics (Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_Y^p$</td>
<td>$A_Y^p$</td>
<td>$\Sigma_Y$</td>
<td>$\mu_Y^q$</td>
</tr>
<tr>
<td>0.00031</td>
<td>0.9849</td>
<td>0.0074</td>
<td>0.00011</td>
</tr>
<tr>
<td>0.00044</td>
<td>0.0219</td>
<td>0.9714</td>
<td>0.00011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>D. Pricing Errors, and Noise St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (b.p.)</td>
<td>18.40 14.62 18.82 24.67 18.50</td>
</tr>
<tr>
<td>MAPE (b.p.)</td>
<td>14.06 10.87 13.45 17.00 13.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>E. Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly P</td>
<td>Monthly P Dynamics 0.9926 0.9638</td>
</tr>
<tr>
<td>Annualized P</td>
<td>Annualized P Dynamics 0.9145 0.6421</td>
</tr>
<tr>
<td>Annual Q</td>
<td>Annual Q Dynamics 0.9077 0.7068</td>
</tr>
</tbody>
</table>

Noise St.Dev. (b.p.) 23.5
Table 3: Estimation results of a three-factor model by ML, under observable PCA factors. The sample is from June 1952 to July 2015.

<table>
<thead>
<tr>
<th></th>
<th>A. Monthly Dynamics (P)</th>
<th>B. Annualized Dynamics (P)</th>
<th>C. Annual Dynamics (Q)</th>
<th>D. Pricing Errors, and Noise Std</th>
<th>E. Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{Y}$</td>
<td>$A_{Y}$</td>
<td>$\Sigma_Y$</td>
<td>$\mu_{Y}$</td>
<td>RMSE (b.p.)</td>
<td>Monthly P Dynamics</td>
</tr>
<tr>
<td>0.00685</td>
<td>-0.9253</td>
<td>4.7425</td>
<td>-2.9117</td>
<td>12.76</td>
<td>0.9922</td>
</tr>
<tr>
<td>-0.00470</td>
<td>1.5318</td>
<td>-2.7808</td>
<td>2.3101</td>
<td>10.27</td>
<td>0.9459</td>
</tr>
<tr>
<td>-0.01118</td>
<td>3.4758</td>
<td>-8.5672</td>
<td>6.2400</td>
<td>16.03</td>
<td>0.5958</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1e-4 $\times$</td>
<td></td>
<td>10.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1990</td>
<td>0.1480</td>
<td>0.1100</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1480</td>
<td>0.1350</td>
<td>0.1350</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1100</td>
<td>0.1350</td>
<td>0.1450</td>
<td></td>
</tr>
<tr>
<td>$\mu_{Y}$</td>
<td>$A_{Y}$</td>
<td>$\Sigma_Y$</td>
<td>$\mu_{Y}$</td>
<td>MAPE (b.p.)</td>
<td>Annualized P Dynamics</td>
</tr>
<tr>
<td>0.01440</td>
<td>-2.3877</td>
<td>7.9491</td>
<td>-4.7960</td>
<td>9.95</td>
<td>0.9100</td>
</tr>
<tr>
<td>-0.00091</td>
<td>2.0665</td>
<td>-3.8922</td>
<td>2.8225</td>
<td>7.67</td>
<td>0.5133</td>
</tr>
<tr>
<td>-0.00861</td>
<td>4.9091</td>
<td>-11.4962</td>
<td>7.7052</td>
<td>11.04</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1e-3 $\times$</td>
<td></td>
<td>7.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1851</td>
<td>0.1567</td>
<td>0.1441</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1567</td>
<td>0.1518</td>
<td>0.1441</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1335</td>
<td>0.1441</td>
<td>0.1465</td>
<td></td>
</tr>
<tr>
<td>$\mu_{Q}$</td>
<td>$A_{Q}$</td>
<td>$\Sigma_Y$</td>
<td>$\mu_{Q}$</td>
<td>Noise St.Dev. (b.p.)</td>
<td>Annual Q Dynamics</td>
</tr>
<tr>
<td>0.00009</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>17.8</td>
<td>0.8868</td>
</tr>
<tr>
<td>0.00023</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.7360</td>
</tr>
<tr>
<td>-0.01375</td>
<td>4.6702</td>
<td>-12.2643</td>
<td>8.7781</td>
<td></td>
<td>7.1553</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1e-3 $\times$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1851</td>
<td>0.1567</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1567</td>
<td>0.1518</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1335</td>
<td>0.1441</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RMSE (b.p.) | 12.76 | 10.27 | 16.03 | 10.29 | 1.23 | Noise St.Dev. (b.p.) | 17.8
MAPE (b.p.) | 9.95 | 7.67 | 11.04 | 7.31 | 0.86 |
Table 4: Estimation results of a two-factor model by ML, under observable PCA factors. The sample is from June 1952 to July 2015.

<table>
<thead>
<tr>
<th></th>
<th>A. Monthly Dynamics ($\mathbb{P}$)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_Y^p$</td>
<td>$A_Y^p$</td>
<td>$\Sigma_Y$</td>
</tr>
<tr>
<td>$0.00028$</td>
<td>0.9709</td>
<td>0.0209</td>
<td>1e-4 × 0.1990</td>
</tr>
<tr>
<td>$0.00051$</td>
<td>0.0261</td>
<td>0.9664</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B. Annualized Dynamics ($\mathbb{P}$)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_Y^p$</td>
<td>$A_Y^p$</td>
<td>$\Sigma_Y$</td>
</tr>
<tr>
<td>$0.00351$</td>
<td>0.7284</td>
<td>0.1787</td>
<td>1e-3 × 0.1849</td>
</tr>
<tr>
<td>$0.00557$</td>
<td>0.2226</td>
<td>0.6897</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>C. Annual Dynamics ($\mathbb{Q}$)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_Y^Q$</td>
<td>$A_Y^Q$</td>
<td>$\Sigma_Y$</td>
</tr>
<tr>
<td>$0.00009$</td>
<td>0</td>
<td>1</td>
<td>1e-3 × 0.1849</td>
</tr>
<tr>
<td>$0.00194$</td>
<td>-0.6454</td>
<td>1.6183</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>D. Pricing Errors, and Noise St. Dev.</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (b.p.)</td>
<td>17.90</td>
<td>13.28</td>
<td>17.88</td>
</tr>
<tr>
<td>MAPE (b.p.)</td>
<td>13.62</td>
<td>9.44</td>
<td>12.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>E. Eigenvalues</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly $\mathbb{P}$ Dynamics</td>
<td>0.9921</td>
<td>0.9452</td>
<td></td>
</tr>
<tr>
<td>Annualized $\mathbb{P}$ Dynamics</td>
<td>0.9094</td>
<td>0.5086</td>
<td></td>
</tr>
<tr>
<td>Annual $\mathbb{Q}$ Dynamics</td>
<td>0.9058</td>
<td>0.7125</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Fama-Bliss forward rates of maturities 1-3 years (first panel), and corresponding forward factors $\gamma_i$ in a three-factor version of canonical model (4), estimated with Kalman filter (second panel). The three bottom panels present differences between the respective series (empirical measurement errors). The sample is June 1956 to July 2015.
Figure 2: The difference factors $Z_t$, connected to forward factors $Y_t$ by transformation (13), and conceptually related to derivative factors of Collin-Dufresne et al. (2008). The two lines in each panel correspond to the model-implied factors (estimated by the Kalman filter), and their empirical counterparts. The sample is June 1956 to July 2015.
Figure 3: The top panel shows the forward factors $Y_t$ estimated in a three-factor model under the assumption of observable factors (PCA scores). The second panel plots the same factors, estimated by the Kalman filter (as in Figure 1). The three bottom panels present differences between the respective series, reflecting the amount of disagreement between the two estimation methods. The sample is June 1956 to July 2015.
Figure 4: The difference factors $Z_t$, connected to forward factors $Y_t$ by transformation (13), and conceptually related to derivative factors of Collin-Dufresne et al. (2008). The solid lines are estimated under the assumption of observable factors (PCA scores). The dashed lines correspond to the Kalman-filter estimates, the same as in Figure 2. The sample is June 1956 to July 2015.
Figure 5: Principal component loadings in the cross section of log bond prices (five annual maturities) implied by a three-factor model estimated by Kalman filter (left), and under observable factors (right). The thicker lines correspond to model-implied log price loadings, and the thinner (dotted) lines to the data (raw PCA, not necessarily free of arbitrage). The top panels report the orthonormal loadings, and the bottom panels show the same loadings scaled to reflect the effect of unconditional one-standard-deviation changes in the factors. The bottom panels have units approximately reflecting percentage changes in bond prices.
Figure 6: Loadings of bond prices (of five annual maturities) with respect to eigenvalue-tracking factors $\mathcal{V}_t$, defined in section 3.5, in the three-factor model estimated by Kalman filter on the full sample of Fama-Bliss bonds (June 1956 – July 2015). The left panel plots the columns of $\tilde{B}$ in representation $b_t = b_0 + \tilde{B}_t \mathcal{V}_t$. The right panel scales the same loadings such that the units correspond to (approximate) percentage-point price changes in reaction to annual factor shocks in $\mathcal{V}_t$ of one-standard-deviation magnitudes. The legend identifies the loadings with the corresponding eigenvalues.
Figure 7: Cumulative maximum Sharpe ratios (for log excess returns) implied by three-factor (top), and two-factor (bottom) models, estimated by Kalman filter on the full sample of Fama-Bliss bonds. The lines correspond to Sharpe ratios attainable in unrestricted, and restricted investment opportunity sets, by repeatedly removing dimensions of excess returns corresponding to distinct eigenvalues under $\mathbb{Q}$. See section 3.5 for details.
Appendix A Solving Affine Term Structure Models

This appendix contains some basic results on discrete-time affine term structure models with constant volatilities and flexible market prices of risk, as in Ang and Piazzesi (2003). The material is standard, and can be found in many other papers. The starting point here is the formulation with explicit stochastic discount factor, which naturally leads to an alternative formulation in terms of the risk-neutral dynamics, used as the starting point in the main body of the paper.

Explicit Stochastic Discount Factor

The model consists of two ingredients, the dynamics of state variables, and a specification of the stochastic discount factor (SDF). The factors $X_t$ are assumed to follow a vector auto-regressive (VAR) process

$$X_{t+1} = \mu^X + A^X_t X_t + \varepsilon^X_{t+1},$$

with normally distributed innovations characterized by conditional covariance matrix $\Sigma^X$.

The SDF is specified in log form as

$$\log M_{t+1} \equiv m_{t+1} = -f^1_t - \frac{1}{2} \Lambda^t \Sigma^X \Lambda^t - \Lambda^t \varepsilon^X_{t+1},$$

$$f^1_t = f^1_0 + f^1_1' X_t,$$

$$\Lambda^t = \Lambda^0 + \Lambda^1 X_t,$$

where $f^1_t$ is continuously compounded risk-free rate between $t$ and $t+1$ (the short rate), and $\Lambda_t$ is the vector of market prices of risk.

This combination of assumptions leads to tractable solutions for zero-coupon bond prices and forward rates, which turn out to be affine functions of the state vector. One conjectures that the log price of a unit cash flow to be paid at maturity $m$ is of the form

$$b^m_t = b^m_0 + b^m_1' X_t,$$

and therefore the log return on holding this bond for one period is

$$r^m_{t+1} = b^m_{t+1} - b^m_t = b^m_0 + b^m_1' \mu^X + (b^m_{t+1} - b^m_0) X_t + b^m_1' \varepsilon^X_{t+1},$$

which follows directly from (31), combined with factor dynamics (27).

In order to verify the conjecture (31), one applies the standard no arbitrage condition $E^p_t (e^{m_{t+1} + r^m_{t+1}}) = 1$, which must hold for all $m$. Since both $m_{t+1}$ and $r^m_{t+1}$ are normal random variables, this is equivalent to

$$E^p_t (m_{t+1} + r^m_{t+1}) + \frac{1}{2} Var_t (m_{t+1} + r^m_{t+1}) = 0.$$ (33)

Substituting (28) and (32) into (33), and using the condition that the pricing equation must hold for every possible value of the state vector at date $t$, one obtains a system of recursive conditions

$$b^m_1 = b^m_0 - f^1_0 + b^m_1' \mu^X + \frac{1}{2} b^m_1' \Sigma^X b^m_1,$$

$$b^m_0 = b^m_0 - f^1_0 + b^m_1' \mu^X + \frac{1}{2} b^m_1' \Sigma^X b^m_1.$$ (34)

Defining $b^m_0 = 0$ and $b^m_1' = 0$ as part of the solution (which is consistent with the natural requirement
that unit cash flows at maturity zero are valued at par) allows to initialize these recursions, and find bond pricing formulae for every maturity \( m \).

Continuously compounded, one-period forward rates are related to log bond prices by identities \( f_t^m \equiv b_t^{m-1} - b_t^m \), and therefore must be of the form

\[
f_t^m = f_0^m + f_1^m \mathcal{X}_t,
\]

with loadings

\[
\begin{align*}
  f_1^m' &= f_1^1(A_X^p - \Sigma_X \Lambda_1)^{m-1}, \\
  f_0^m &= f_0^1 - b_1^{m-1}(\mu_X^p - \Sigma_X \Lambda_0) - \frac{1}{2} b_1^{m-1} \Sigma_X b_1^{m-1}.
\end{align*}
\]

### Risk-Neutral Dynamics

For every random variable \( z_{t+1} \) (measurable with respect to time \( t+1 \) information set) one can define conditional risk-neutral expected value as \( \mathbb{E}_t^Q(z_{t+1}) \equiv \mathbb{E}_t^P(z_{t+1}) + \text{Cov}_t(m_{t+1}, z_{t+1}) \), where \( m_{t+1} \) is the log SDF.\(^ {42} \) Intuitively, the objective expectation of \( z_{t+1} \) is corrected for risk, as measured by the covariance with the SDF. In particular, the risk-neutral expectation of factor innovation \( \varepsilon_{X,t+1}^p \) is just \(-\Sigma_X \Lambda_t\), and therefore the risk-neutral expectation of \( \varepsilon_{X,t+1}^p \equiv \varepsilon_{X,t+1}^p + \Sigma_X \Lambda_t \) is zero.

Adding and subtracting the term \( \Sigma_X \Lambda_t = \Sigma_X (\Lambda_0 + \Lambda_1 \mathcal{X}_t) \) to the right-hand side of factor dynamics (27), defining \( \mu_X^p = \mu_X^p - \Sigma_X \Lambda_0 \), \( A_X^p = A_X^p - \Sigma_X \Lambda_1 \), and using the definition of \( \varepsilon_{X,t+1}^p \) above, one can write

\[
\mathcal{X}_{t+1} = \mu_X^p + A_X^p \mathcal{X}_t + \varepsilon_{X,t+1}^p,
\]

which is referred to as the risk-neutral factor dynamics.

Evidently, specifying the objective dynamics (27) together with its risk-neutral counterpart (37) is sufficient to fully characterize the risk adjustments to all factors due to their covariances with the SDF, i.e., the market prices of risk \( \Lambda_t = \Lambda_0 + \Lambda_1 \mathcal{X}_t \). In order to complete the model, one therefore only needs to add the short rate equation (29). Summing up, the term structure model can be re-stated as

\[
\begin{align*}
  \mathcal{X}_{t+1} &= \mu_X^p + A_X^p \mathcal{X}_t + \varepsilon_{X,t+1}^p \\
  \mathcal{X}_{t+1} &= \mu_X^Q + A_X^Q \mathcal{X}_t + \varepsilon_{X,t+1}^Q \\
  f_t^1 &= f_0^1 + f_1^1 \mathcal{X}_t.
\end{align*}
\]

The definitions of \( \mu_X^Q \), and \( A_X^Q \) also allow to re-write the bond pricing solutions (34),

\[
\begin{align*}
  b_1^{m'} &= b_1^{m-1'} A_X^Q - f_1^1', \\
  b_0^m &= b_0^m - f_0^1 + b_1^{m-1'} \mu_X^Q + \frac{1}{2} b_1^{m-1'} \Sigma_X b_1^{m-1},
\end{align*}
\]

and the forward coefficients in equations (36) as

\[
\begin{align*}
  f_1^m' &= f_1^1(A_X^Q)^{m-1}, \\
  f_0^m &= f_0^1 - b_1^{m-1'} \mu_X^Q - \frac{1}{2} b_1^{m-1'} \Sigma_X b_1^{m-1}.
\end{align*}
\]

\(^ {42} \) This discussion is kept at an informal level. One needs to restrict the space of random variables such that both expected values exist.
It is evident that these solutions only depend on parameters specified in system (38), so that the latter indeed captures all pricing implications of the model.\footnote{The asset-pricing condition (33) can also be re-stated in a risk-neutral form. Using the fact that \(E^\mathbb{P}_t (m_{t+1}) + \frac{1}{2} \text{Var}_t (m_{t+1}) = -f^1_t\) and \(E^\mathbb{Q}_t (r^m_{t+1}) = E^\mathbb{Q}_t (r^m_{0,t+1}) + \text{Cov}_t (m_{t+1}, r^m_{0,t+1}),\) (33) takes a notationally simpler form \(E^\mathbb{Q}_t (r^m_{t+1}) - f^1_t + \frac{1}{2} \text{Var}_t (r^m_{0,t+1}) = 0.\) In words, under the risk-neutral expectation every bond must offer return equal to the risk-free rate, up to a convexity adjustment.}

## Appendix B  
**Invariant Transformations**

This appendix shows that two \(N\)-factor models are observationally equivalent (i.e., share exactly the same predictions for the term-structure) whenever there exists an invertible affine transformation between their respective state vectors.\footnote{The content of this appendix is a slightly more detailed version of Appendix B. in Joslin et al. (2011).} Let \(\mathcal{X}_t\) and \(\mathcal{Y}_t\) be linked by an invertible relation

\[
\mathcal{Y}_t = \alpha + \beta \mathcal{X}_t. \tag{41}
\]

Start with a general model stated in terms of \(\mathcal{X}_t\),

\[
\begin{align*}
\mathcal{X}_{t+1} &= \mu^\mathbb{P}_\mathcal{X} + A^\mathbb{P}_\mathcal{X} \mathcal{X}_t + \epsilon^\mathbb{P}_{\mathcal{X},t+1}, \\
\mathcal{X}_{t+1} &= \mu^\mathbb{Q}_\mathcal{X} + A^\mathbb{Q}_\mathcal{X} \mathcal{X}_t + \epsilon^\mathbb{Q}_{\mathcal{X},t+1}, \\
f^1_t &= f^1_0 + f^1_t \mathcal{X}_t, \quad E(\epsilon_{\mathcal{X}} \epsilon'_{\mathcal{X}}) = \Sigma_{\mathcal{X}}. 
\end{align*} \tag{42}
\]

By (41), \(\mathcal{X}_t = \beta^{-1}(\mathcal{Y}_t - \alpha)\). Substituting this into the original model one obtains a representation in terms of the new factors,

\[
\begin{align*}
\mathcal{Y}_{t+1} &= \mu^\mathbb{P}_\mathcal{Y} + A^\mathbb{P}_\mathcal{Y} \mathcal{Y}_t + \epsilon^\mathbb{P}_{\mathcal{Y},t+1}, \\
\mathcal{Y}_{t+1} &= \mu^\mathbb{Q}_\mathcal{Y} + A^\mathbb{Q}_\mathcal{Y} \mathcal{Y}_t + \epsilon^\mathbb{Q}_{\mathcal{Y},t+1}, \\
f^1_t &= f^1_0 + f^1_t \mathcal{Y}_t, \quad E(\epsilon_{\mathcal{Y}} \epsilon'_{\mathcal{Y}}) = \Sigma_{\mathcal{Y}}.
\end{align*} \tag{43}
\]

The relations between the parameters are

\[
\begin{align*}
A^i_{\mathcal{Y}} &= \beta A^i_{\mathcal{X}} \beta^{-1}, \quad \text{for } i \in \{\mathbb{P}, \mathbb{Q}\} \\
\mu^i_{\mathcal{Y}} &= \beta \mu^i_{\mathcal{X}} + (I - A^i_{\mathcal{Y}}) \alpha, \quad \text{for } i \in \{\mathbb{P}, \mathbb{Q}\} \\
f^1_0 &= f^1_0 \beta^{-1}, \\
\hat{f}^1_0 &= f^1_0 - \hat{f}^1_1 \alpha, \\
\Sigma_{\mathcal{Y}} &= \beta \Sigma_{\mathcal{X}} \beta'. \tag{44}
\end{align*}
\]

Bond prices (of all maturities) are affine in the original factors, \(b^m_t = b^m_0 + b^m_1 \mathcal{X}_t\), and therefore also in the transformed ones, \(\hat{b}^m_t = \hat{b}^m_0 + \hat{b}^m_1 \mathcal{Y}_t\), with

\[
\begin{align*}
\hat{b}^m_1 &= b^m_1 \beta^{-1}, \\
\hat{b}^m_0 &= b^m_0 - \hat{b}^m_1 \alpha. \tag{45}
\end{align*}
\]

By the results of Appendix A, the original coefficients satisfy (39). Using this fact, together with the links between parameters (44) and relations (45), it is straightforward to verify that the transformed...
coefficients satisfy analogous conditions with respect to the transformed model (43),

\[
\begin{align*}
\hat{b}_1^{m'} &= \hat{b}_1^{m-1} A_\text{Y}^Q - f_1^t, \\
\hat{b}_0^{m} &= \hat{b}_0^{m-1} - f_0^t + \hat{b}_1^{m-1} \mu_\text{Y} + \frac{1}{2} \hat{b}_1^{m-1} \Sigma_\text{Y} \hat{b}_1^{m-1},
\end{align*}
\] (46)

which shows that the latter prices the term structure by no arbitrage, and the transformation is indeed invariant in the sense that it leaves all bond prices unchanged.

### Appendix C  Details on Estimation by Kalman Filter

Propositions 1 and 2 show that every no-arbitrage GDTSM can be transformed into one in which the factors are the shortest-maturity forward rates \( Y_t \), and the transition matrix under \( Q \) is in companion form (8). The state-space representation of the model is

\[
\begin{align*}
Y_{t+1} &= \mu_\text{Y} + A_\text{Y} Y_t + \varepsilon_{Y,t+1}^P, \\
b_t^y &= B_0 + B_1 Y_t + v_t,
\end{align*}
\] (47)

where \( b_t^y \) is a vector of noisy observations of the term structure at time \( t \), \( B_0, B_1 \) are model-implied coefficients (satisfying the no arbitrage restrictions), and \( v_t \) is an i.i.d. vector of measurement errors with covariance matrix \( R \), independent of factor innovations \( \varepsilon_{Y,t+1}^P \). The covariance matrix of \( \varepsilon_{Y,t+1}^P \) is \( \Sigma_\text{Y} \).

The Kalman filter iteratively estimates the forward rates in \( Y_t \) based on the history of observed bond prices. Define \( \hat{Y}_t \equiv E[Y_t|b_t^y, \ldots, b_{t-1}^y] \) as the (prior) estimate of the state vector, and \( \hat{\Sigma}_t \equiv E[(Y_t - \hat{Y}_t)(Y_t - \hat{Y}_t)'] \) as the matrix measuring the uncertainty in the estimate. At every point in time the filter computes the innovation \( a_t \equiv b_t^y - B_0 - B_1 \hat{Y}_t \), and uses it to update the current state, and to form next period’s prior \( \hat{Y}_{t+1} \), according to the transition equation (47). The filter equations also include the updating rule for transforming innovations into updates of the state, and the description of the time evolution of the uncertainty matrix \( \hat{\Sigma}_t \), as summarized below.\(^{45}\)

Suppose one knows the initial (multivariate Gaussian) distribution of the state vector, \( Y_1 \sim N(\hat{Y}_1, \hat{\Sigma}_1) \), and observes a sample of bond prices \( b_1^y, \ldots, b_t^y \). The filter equations are:

\[
\begin{align*}
a_t &= b_t^y - B_0 - B_1 \hat{Y}_t, \\
\hat{K}_t &= A_\text{Y}^T \hat{\Sigma}_t B_1 + R)^{-1}, \\
\hat{Y}_{t+1} &= \mu_\text{Y} + A_\text{Y} \hat{Y}_t, \\
\hat{\Sigma}_{t+1} &= \Sigma_\text{Y} + \hat{K}_t R \hat{K}_t' \Sigma_t (A_\text{Y}^T - \hat{K}_t B_1)' + (A_\text{Y}^T - \hat{K}_t B_1) \hat{\Sigma}_t (A_\text{Y}^T - \hat{K}_t B_1)' \hat{\Sigma}_t (A_\text{Y}^T - \hat{K}_t B_1). (52)
\end{align*}
\]

In the current application, the filter can be initiated with the first \( N \) shortest-maturity forward rates observed at the very beginning of the sample. If one uses the stationary version of the filter by first iterating equations (50) and (52) until convergence (for a given set of parameters), the matrices \( \hat{K} \), and \( \hat{\Sigma} \) become constant (equal their steady-state values), and filtering is reduced to computing the innovations (49), and predicting the state (51).\(^{46}\)

The likelihood of observing the data for a fixed set of model parameters is the same as the

\(^{45}\) A very good presentation of the Kalman filter is contained in Ljungqvist and Sargent (2012), p. 56.

\(^{46}\) Anderson et al. (1996) summarize the conditions under which the steady-state filter can be applied.
likelihood of observing a sequence of innovations $a_t$. For a Gaussian model with $N$ factors, the
log-likelihood function implied by the stationary filter is

$$L = -rac{TN}{2} \log 2\pi - rac{T}{2} \log |\Omega| - \frac{1}{2} \sum_{t=1}^{T} a_t \Omega^{-1} a_t',$$

(53)

where $\Omega$ is the covariance matrix of innovations, $\Omega = R + B_1\hat{\Sigma}B_1'$ (as is evident from (49)).

The model is estimated by finding a set of parameters that produce the innovations that are
the most likely, given the assumed model structure, and the observed data.

**Appendix D  Details on Estimation Under Observable Factors**

This appendix discusses GDTSM estimation under the assumption that there exist exactly $N$
known linear combination of term-structure observables (yields, forward rates, or discount bonds
prices) explained by the model perfectly. Keeping the other assumptions and notation consistent
with Appendix C, and working with log bond prices, the state-space representation under the
companion-form parametrization defined in Propositions 1 and 2 is

$$Y_{t+1} = \mu P Y_t + A P Y_t + \varepsilon_{t+1},$$

(54)

$$b_t^\phi = B_0 + B_1 Y_t + v_t.$$  (55)

The extra assumption of observable factors (referred to as ”Case P” by Joslin et al. (2011)) can be
summarized as follows.

**Assumption 1** For a $N$-factor model, there exist exactly $N$ a priori known linear combinations
of observed bond prices that uncover the factors, i.e., there exists a matrix $W$ such that

$$Wv_t = 0,$$

(56)

Without loss of generality, it can be assumed that the rows of $W$ have unit norm, and that they
are linearly independent. \footnote{One can always re-scale the bond prices such that the normalization holds. Linear independence of the rows is a necessary consequence of the assumption.} One can define the bond portfolios given by $W$ as $P_t \equiv Wb_t^\phi$. In applications, $W$ is usually chosen as the PCA loading matrix, corresponding to the scores of maximum unconditionail variance, and the portfolios $P_t$ correspond to the PCA factor scores. \footnote{This choice is particularly convenient for the reason that the rows of $W$ are orthonormal, and the resulting factors are unconditionally uncorrelated in every given sample. Other choices of $W$ are possible. For example, one could use bond prices of several selected maturities as observable factors.}

By Assumption 1, the observation equation (55) can be used to measure the factors in every
given sample,

$$P_t = Wb_t^\phi = WB_0 + (WB_1)Y_t.$$  (57)

The above equation effectively defines an invariant transformation that can be used to express
the model in terms of the bond portfolios as factors. On the other hand, the other bond portfolios
(priced with error) are still informative about model parameters. To find all observable implications
of the model under Assumption 1, define $W^\perp$ as a matrix consisting of orthonormal rows spanning
the null space of $W$. This matrix is also known a priori, and can be used to derive the state-space
representation of the rotated model,
\[ P_{t+1} = \mu_P^P + A_P^P P_t + \varepsilon_{P,t+1}^P, \quad (58) \]
\[ b_t^i = H_0 + H_1 P_t + v_t^i, \quad (59) \]
where the parameters are related to those of the original model through
\[ \mu_P^P = W B_1 \mu_Y^P + [I - (W B_1) A_Y^P (W B_1)^{-1}], \quad (60) \]
\[ A_P^P = (W B_1) A_Y^P (W B_1)^{-1}, \quad (61) \]
\[ H_0 = W^\perp B_0 - W^\perp B_1 (W B_1)^{-1} W B_0, \quad (62) \]
\[ H_1 = W^\perp B_1 (W B_1)^{-1}. \quad (63) \]

The covariance matrices of the random terms in (58) and (59) are, respectively
\[ \Sigma_P \equiv (W B_1) \Sigma_Y (W B_1)', \quad (64) \]
\[ R^\perp \equiv W^\perp R (W^\perp)', \quad (65) \]

Since the factors in the new transition equation (58) are observable, one does not need to use the Kalman filter. Moreover, by the standard result of Zellner (1962) (used in the same context by Joslin et al. (2011)), the maximum-likelihood estimates of coefficients \( \mu_P^P \), and \( A_P^P \) coincide with their OLS estimates, and do not depend on the covariance matrix \( \Sigma_P \), which allows to exclude many model parameters from the numerical search. For example, in the case of a 3-factor model, the dimensionality of the parameter space drops by 12, which greatly reduces the estimation time.

Recall that under the companion-form parametrization, one is interested in finding \( \Theta = \{ \theta_\mu, \theta_A, \theta_\Sigma, \mu, c, \theta_R \} \), where \( \theta_\mu \) and \( \theta_A \) determine the conditional expectation in (54), \( \theta_\Sigma \) is the vectorized triangular matrix in the Cholesky decomposition of \( \Sigma_Y \), \( \mu \) and \( c \) parametrize the conditional Q dynamics of factors \( Y_t \), and \( \theta_R \) describes the noise covariance matrix. The rest of this Appendix describes the construction of the log-likelihood function, and shows that it does not depend on \( \theta_\mu \), and \( \theta_A \).

Given the equivalence of the state-space representations (54)-(55), and (58)-(59) under Assumption 1, the probability of observing a given sample of bond prices \( \{b_t^i\}_{t \in (1, \ldots, T)} \), conditional on the parameters in \( \Theta \), can be factored into
\[ \text{prob}(\{b_t^i\} | \Theta) = \text{prob}(\{P_t\} | \Theta) \times \text{prob}(\{b_t^i\} | \{P_t\}; \Theta). \quad (66) \]

In light of (58), the first term on the right is the likelihood of observing the sequence of (T-1) VAR innovations to the observable bond portfolios. The log of this term, suppressing the \( 2\pi \) part, is
\[ \mathcal{L}_1 \equiv -\frac{(T-1)}{2} \log |\Sigma_P| - \frac{1}{2} \sum_{t=1}^{T-1} (P_{t+1} - \mu_P^P - A_P^P P_t) \Sigma_P^{-1} (P_{t+1} - \mu_P^P - A_P^P P_t)', \quad (67) \]
where \( \Sigma_P \) is given in (64). As noted above, \( \mu_P^P \) and \( A_P^P \) can be estimated by OLS, and treated as fixed in every given sample. Effectively, this part of the log likelihood only depends on parameters that determine \( \Sigma_P \), i.e., \( \Theta_1 \equiv \{ \theta_\Sigma, c \} \), as is evident from (64).

The other term in the factorization (66) is the probability of observing the sequence of T realizations of \( v_t^i \), for given \( P_t \), and model parameters that determine \( H_0, H_1, \) and \( R^\perp \), as indicated
by (59). The log of this probability is (again ignoring the constant part)

\[ L_2 \equiv -\frac{T}{2} \log |R^\perp| - \frac{1}{2} \sum_{t=1}^{T} (b_t^\perp - H_0 - H_1 P_t)(R^\perp)^{-1} (b_t^\perp - H_0 - H_1 P_t)'. \]  \hspace{1cm} (68)

Equations (62), (63), and (65) indicate that in order to compute this part of the likelihood, one only needs to know the model-implied bond pricing matrices \( B_0 \) and \( B_1 \). Under the companion-form parametrization, the latter only depends on \( c \). To find \( B_0 \), one first needs to find the risk-neutral drift \( \mu^Q \), which according to Proposition 2 depends on parameters \( \theta_\Sigma, \) and \( \mu \). Given the risk-neutral drift, one is able to complete the solution of model-implied bond prices by finding \( B_0 \). Overall, the second part of the likelihood is parametrized by \( \Theta_2 \equiv \{ \theta_\Sigma, c, \mu, \theta_R \} \).

The total log likelihood is the sum of (67) and (68), and the corresponding parameter set is \( \Theta_1 \cup \Theta_2 \), which does not depend on \( \theta_\mu \) and \( \theta_A \).

\textsuperscript{49}Consistent with Assumption 1, the matrix \( W \) of bond portfolios measured without error is known. The matrix \( W^\perp \) can be defined given \( W \), and also treated as known (for example, using Matlab function \texttt{null}, or Python function \texttt{nullspace}, contained in a module available at http://scipy-cookbook.readthedocs.io/items/RankNullspace.html).