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

Juliusz F. Radwański

May 12, 2017

Abstract

This article proposes a convenient parametrization for the popular class of (discrete-time) essentially-affine term structure models of Duffee (2002), and Ang and Piazzesi (2003). First, I show that if bond prices are determined by N latent state variables, all their pricing information must be present in the *short end* of the term structure, i.e., one can rotate the model to an observationally equivalent form with exactly N short maturity forward rates as factors. Second, the risk-neutral transition of the rotated model is conveniently parametrized by N unrestricted real numbers (contained in a companion matrix), and only one additional parameter is needed to specify the risk neutral drift. Third, the resulting state-space representation makes it easy to estimate the model either by the Kalman filter (in one step), or treating N linear combinations of observable bond prices (or yields) as observable factors. Finally, I interpret some difficulties in fitting the essentially-affine term structure models to the data, using the standard set of Fama-Bliss discount bonds as example. The problem is the existence of (spanned) factors that significantly predict term structure movements but are virtually impossible to detect from the shortest-maturity forward rates, which is (by the results of this paper) inconsistent with no arbitrage within the discussed class of models.

Keywords: Affine Term Structure Models, Kalman Filter Estimation, Canonical Companion Form.

JEL Classification Numbers: G12, G17, E43.

*Wirtschaftswissenschaftliche Fakultt, Humboldt-Universitt zu Berlin, Dorotheenstr. 1, 10099 Berlin, Germany. Tel: +49 (0) 30 2093 5013. E-mail: juliusz.radwanski@hu-berlin.de. This paper is in preliminary stage. I would like to thank the participants of the 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 the prices of risk. 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 only the 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 counterpart, the main object of this study, was introduced by Ang and Piazzesi (2003).

There are two well-known problems associated with bringing these models to the data. One is lack of identification if the model is formulated in terms of latent factors, as explained by Dai and Singleton (2000), or Collin-Dufresne et al. (2008). This problem originates in the possibility of performing "invariant transformations" of the factors (and some parameters simultaneously) that leave all bond prices unchanged. Ruling out all such transformations may be very difficult in practice.[1] One solution to the identification problem is to use a subset of observed bond prices (or yields) as factors, along the lines of Duffie and Kan (1996), which however necessitates imposition of highly non-linear parameter restrictions (sometimes called Duffie-Kan restrictions), which essentially guarantee that the yield-based factors are consistent with the model. Under this solution, one therefore falls into the other practical difficulty of finding a convenient parametrization under which the restrictions are easy to implement, and do not result in badly-behaved likelihood functions.

In this paper I propose a relatively straightforward solution to the problems of identification and estimation. In the spirit of Duffie and Kan (1996), I use an observationally equivalent representation in terms of term structure observables. The somewhat surprising result is that under the assumption

[1] See Babbs and Nowman (1999), Collin-Dufresne et al. (2008), Aït-Sahalia and Kimmel (2010).

of no arbitrage (which is at the heart of every affine term structure model), all pricing information about the latent factors spanning the term structure must also be contained in the shortest-maturity forward rates. This result is a discrete-time counterpart of Collin-Dufresne et al. (2008) (CGJ henceforth), who show that every continuous-time model with N_f latent factors is observationally equivalent to one in which the factors are the first N_f derivatives of the term structure, evaluated at maturity zero. The discrete-time case, although conceptually similar, appears slightly more general, and differs in mathematical details (the case studied by CGJ could be obtained in the limit). From the practical point of view, it is also much easier to implement. CGJ actually attempt to measure the derivatives at maturity zero by extrapolating polynomial splines fitted to the shortest-maturity segments of empirical principal component loadings. Using discretely-spaced forward rates has the advantage that the model can be estimated by Kalman filter together with the factors, i.e., in a fully self-consistent way, without introducing measurement errors at the stage of factor measurement.[2] As explained below, it is also possible to estimate the model under the assumption that some portfolios of bonds (or yields) are priced without errors, for example the empirical PCA factor scores.[3] Overall, the canonical representation specified in terms of the short-maturity forward rates is not at all restrictive, which of course follows from the observational equivalence.

A related advantage of specifying the model in terms of the short-maturity forward rates is that the the risk-neutral dynamics can be represented by a companion matrix, and conveniently parametrized by only N_f (unrestricted) real numbers, which allows to call the proposed representation canonical. All free parameters are contained in the last row of the matrix, with all other parameters being either 0, or 1 (above the diagonal). The structure of the companion matrix reflects the fact that under no arbitrage, forward rates of longer maturities move one-to-one with risk-adjusted expectations of future forward rates. In this setup, the Duffie-Kan restrictions take the simplest possible form, and are imposed automatically. To complete the specification, one also needs to parametrize the risk-neutral drift, which turns out to be a function of factor covariance matrix, and one extra parameter.

[2]The only conceptual issue with short-maturity forward rates is how to choose the sampling frequency, but in practice this choice is dictated by the structure of the data. Discrete-time models can always be specified using the shortest maturities available in a given sample.

[3]In principle, one could assume that the short-maturity forward rates themselves are observed without error. This would produce a perfect model fit at the shortest maturities, at the cost of potentially large errors on the other side of maturity spectrum.

A related canonical representation of affine models, helpful in maximizing estimation efficiency, was proposed by Joslin et al. (2011) (from now on, JSZ), who parametrize the risk-neutral dynamics in terms of the multiset of N_f eigenvalues, and one additional parameter governing the $\mathbb{Q}$-expected long-run behavior of the short rate. The total number of free identifiable parameters under both formulations is thus the same, and both result in maximum model flexibility. However, the eigenvalue parametrization may lead to some practical problems if some eigenvalues are repeated or complex, and these cases need to be considered separately at the estimation stage.[4] In contrast, the companion parametrization is defined in terms of the *coefficients* of the characteristic polynomial of the risk-neutral transition matrix, which encode exactly the same information as its roots (the eigenvalues). Working with the coefficients appears more natural, and easier to implement, since they form an ordered set of unrestricted real numbers, making it possible for the optimization algorithm to compare all cases based on their respective likelihood values.[5]

To illustrate the convenience of the proposed normalization, I estimate a standard unrestricted three-factor specification on the monthly set of Fama-Bliss discount bonds (Fama and Bliss (1987)), both by the Kalman filter (with endogenous factors), and using empirical PCA scores as (observed) factors. In addition to the full sample spanning June 1956 – July 2015, I use a sub-sample from Cochrane and Piazzesi (2005) (CP), i.e., January 1965 – December 2003.[6] One advantage of the Fama-Bliss set is that it is constructed without prior application of smoothing methods, which makes it particularly well suited for model estimation by Kalman filter. As reported by Cochrane and Piazzesi (2005, 2009), it also appears that smoothing methods may partly remove important information about bond risk premia. On the other hand, a slight disadvantage of the Fama-Bliss set is that the shortest-maturity forward rate corresponds to one year holding period, which is quite long. Also, there are only five maturities in total, with the longest one corresponding to five years.

[4]The article of JSZ leaves an impression that the case of repeated eigenvalues may be empirically relevant. Their online supplement (Joslin et al. (2010)) shows that the arbitrage-free Nelson-Siegel specification of Christensen et al. (2011) features a repeated eigenvalue under the JSZ normalization. Since such case is borderline, one should also allow the possibility of complex eigenvalues. All these cases are consistent with no arbitrage, and it is difficult to rule them out based on economic arguments alone.

[5]Joslin et al. (2011) note the possibility of companion-form parametrization in footnote 22, attributing it to CGJ. Indeed, the transition matrix of the continuous-time model of CGJ is exactly in this form, but (as explained above) they do not make full use of this fact. The main point of JSZ is on the irrelevance of no-arbitrage restrictions for forecasting, which can be shown using any normalization, if the factors are observable – see also the discussion below.

[6]The latter choice avoids the recent zero-bound period, and well as the period prior to 1965, for which the data may be less accurate. The CP sample contains vast majority of empirically-detectable variation in bond risk premia, as shown by Radwański (2010).

The estimation results, combined with the clarity offered by the canonical companion form, shed some light on the empirical difficulties that researchers face when trying to fit no-arbitrage models to the data. As emphasized by Cochrane and Piazzesi (2005), a robust feature of the Fama-Bliss set is that the yield differential between maturities of four and five years significantly helps to forecast excess bond returns. This differential is present in the third principal component of observed data, which contributes very little to the overall variance of bond prices. At the same time it is classified as genuine by the Kalman filter, exactly due to the predictive content. On the other hand, the general results of this paper imply that any factor present in long maturities must also show up as some linear combination of the first three forward rates. The only way in which the estimation procedure is able to explain the existence of the large yield differential on the long end of maturity spectrum is to associate an extreme eigenvalue (under $\mathbb{Q}$) to a small factor that leaves the short-maturity rates virtually unchanged. The estimates of this eigenvalue are robustly below minus two in all sub-samples studied, and are significantly higher in magnitude if the model is estimated under the assumption of observable factors. In other words, if one accepts the assumption of linear factor structure in Fama-Bliss data with three factors, then the constant-volatility model of Ang and Piazzesi (2003) faces considerable difficulties in reconciling important part of predictability evidence with the assumption of no arbitrage. This also has consequences for forecasting the term structure. If the four-to-five year yield differential really predicts future bond prices (Cochrane and Piazzesi (2005)), then imposing an affine structure on the data cannot improve these forecasts. At best, the model will imply a large anomalous eigenvalue, as discussed. Attempts to estimate the model under "reasonable" eigenvalue restrictions will only make the model's forecasting performance worse.

Finally, to offer a more concrete application of the canonical companion form, I use its computational efficiency to address a question of whether it is safe to assume observability of the factors for the purpose of term structure forecasting. The key finding of JSZ, that imposing no arbitrage does not by itself help to predict bond prices, is derived under the assumption that there exist N_f a priori known linear combinations of yields that are priced without errors, and thus reveal the factors. In practice, the combinations of yields can be obtained by principal component analysis, but other choices are possible, e.g., one can use a subset of constant-maturity yields. Intuitively, if factors are observable, then under the model assumption of VAR factor dynamics, one can do no better then running a VAR on observed portfolios of yields in order to predict their future values. Estimation of the risk-neutral parameters is then reduced to minimizing the pricing errors on combinations of yields that are not assumed to be error-free.[7] Since it can safely be assumed that factors are never actually observed without error, a natural question is how severe the induced error-in-variable bias may be. I use the benchmark set of parameter estimates obtained by the Kalman filter on the full set of Fama-Bliss forward rates (as described above) to generate 300 artificial data sets of the same size, with additional 5 years of observations used to compute the forecast errors. I then re-estimate the model twice, using the Kalman filter, and under the assumption that the empirical PCA scores can be used as observed factors. It turns out that both estimation strategies offer very similar forecasting performance, and significantly beat the benchmark random-walk forecasts.[8]

1.1 Literature Overview

This part of the Introduction contains the most important references to the literature, and a more in-depth discussion of the papers most closely related to the current research.

The literature on affine term structure models usually credits the most important early developments to Vasicek (1977), and Cox et al. (1985). These models were usually specified in continuous time, and their main advantage was analytical tractability. Early models featured the short rate as the only state variable, counterfactually predicting perfect correlation between bond prices across maturities. Multifactor models became more popular in the 90's, for example, Litterman and Scheinkman (1991) showed that three factors appear to explain the cross section of yields with great precision. Duffie and Kan (1996) are credited for introducing the class of multifactor 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 match the time variation in bond risk premia. One modeling strategy that achieves such time variation is to incorporate stochastic volatility in factor dynamics, as in the completely-affine class of Dai and Singleton (2000). The alternative strategy is to model the factors as homoskedastic, while letting the Sharpe ratios vary over time by linking them to a subset

[7] A similar estimation strategy, prescribing a VAR on observable factors, was proposed by Hamilton and Wu (2012).

[8] This is true under perfect conditions, guaranteed by the design of the experiment. Still, one method may be favored over the other in the real data, based on their relative robustness with respect to real-world failures of model assumptions.

of 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 and solved in continuous time. The discrete-time version of the essentially-affine model of Duffee (2002) was proposed by Ang and Piazzesi (2003), building on earlier results of Backus and Zin (1994).[9]

Flexibly specified models pose practical difficulties with model identification. Specification in terms of latent state variables makes it possible to transform the factors, and simultaneously change a subset of parameters in a way that leaves all pricing implications unchanged. Intuitively, the likelihood function is exactly flat in the subset of parameters subject to such changes. One approach to the identification problem, proposed by Dai and Singleton (2000), relies on imposing parameter normalizations that rule out all such invariant transformations. The paper of CGJ points to some difficulties with this approach, and argues that a superior way to insure identification is to rotate the latent factors into a set of variables that have economic meaning, and can (in principle) be measured in a model-independent way. Although CGJ argue that the level, slope and curvature at maturity zero are such model-independent variables, it is evident that all of them are functions of the observed term structure, which is exactly what a model should explain, so that a truly model-independent measurement is not guaranteed. In effect, they essentially work under the assumption of observable factors, which can be thought of as another way of securing parameter identification, dual to the approach of Dai and Singleton (2000), and already present in Duffie and Kan (1996), as CGJ acknowledge in the abstract. The main contributions of the current paper relative to CGJ are (i) in showing that their main result can be generalized to a discrete-time formulation, and used to better understand the implications of the no-arbitrage assumption in relation to the weaker assumption of linear factor structure, and (ii) in using the convenient properties of the companion-form parametrization to develop an easily implementable framework for estimating Gaussian affine models, using information in all bond prices, with the possibility of estimating economically-interpretable factors (short-maturity forward rates) simultaneously with model parameters.

Moving to model estimation, it is well known that bond prices move together in only few dimen-

[9] Further developments in the literature have mostly been about 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).

sions, and that there are some stable patterns underlying the shapes that the yield curve usually takes. Estimation of term structure models is therefore all about finding the factors, their dynamical properties, and factor loadings that together with the factors determine the yields.[10] Early methods of term structure fitting by McCulloch (1975), Nelson and Siegel (1987), and Svensson (1995) completely ignore the time-series aspect. Diebold and Li (2006) add dynamic considerations by interpreting the parameters of these fitted curves as time-varying latent factors. Another popular approach rests on principal component analysis of yields, as motivated by Litterman and Scheinkman (1991). An important feature of all these methods is that one does not need the assumption of no arbitrage, which is the main difference with respect to the class of no arbitrage models, of which the affine class is by far the most popular.

Indeed, 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. As explained above, model estimation under this assumption can pose serious challenges, even if one has solved the identification problem (either by parameter normalizations, or by expressing the factors in terms of observable variables), which is due to the presence of complicated cross-sectional restrictions on factor loadings. Part of the important contribution of JSZ is in showing that the risk-neutral dynamics of a model expressed in terms of portfolios of yields can be parametrized by a small set of parameters, without giving up model flexibility. As already mentioned in the main part of this introduction, the parametrization in terms of eigenvalues can in principle be difficult to implement due to the necessity of comparing several special cases of eigenvalue configurations. Moreover, virtually all results of JSZ are derived under the assumption that the factors can be expressed in terms of given combinations of yields. While the results of the current paper indicate that this assumption appears inconsequential in practical forecasting applications if one uses the PCA components as factors (at least in "laboratory" setup, in which the data generating process is consistent with the affine model), it clearly faces conceptual difficulties. Since the no-arbitrage assumption takes the form of Duffie-Kan restrictions affecting model-implied factor loadings, using the empirical principal-component loadings (which do not satisfy the Duffie-Kan restrictions exactly) is technically inconsistent with no arbitrage. In other words, under the assumption of

[10]The fact that the term structure observables like (log) bond prices, yields, or forward rates are all affine functions of the same state vector, and all pricing implications of an affine model can be summarized in terms of either of them, allows me to use these terms interchangeably in the contexts in which no confusion can arise.

observable PCA factors, one only achieves *relative pricing* of portfolios measured with errors in terms of the PCA scores, and the identity of the latter (together with their dynamics) is learned by means of purely statistical techniques.

Seen through the lens of this discussion, the canonical model written in terms of the shortest-maturity forward rates offers an additional advantage, namely, (iii) it can be helpful in deciding on whether the factors measured from the real data can be easily (or not) reconciled with the assumption of no arbitrage. As mentioned before, this may be problematic in the case of four-to-five year yield differential.

The most common estimation method applied to no-arbitrage models is Maximum Likelihood (ML), which motivates the usual assumption that shocks are conditionally Gaussian.[11] As pointed by Joslin et al. (2011) and Hamilton and Wu (2012), a subset of parameter estimates consistent with ML can be obtained using linear regressions if factors are observable. The article of Adrian et al. (2013) goes even further, and advocates a three step regression based procedure.[12]

2 Canonical Companion Form

The goal of this section is to formalize the most important concepts, and to derive the main results. I first explain the relation between linear factor structure and the assumption of no arbitrage, and then follow with the derivation of the companion-form parametrization.

2.1 Linear Factor Structure, and the Role of No-Arbitrage

It is assumed that the term structure is sampled at regular time intervals Δt, and that the risk-free rate corresponds to maturity Δm. By notational convenience, variables measured at two consecutive dates (maturities) are indexed by t and $t+1$ (m and $m+1$). I will also assume that $\Delta t = \Delta m$, with no loss of generality at this stage.[13] The term structure will generally be characterized in terms of log bond prices b_t^m, and continuously compounded forward rates[14]

[11] Chen and Scott (1993) is an early example.

[12] Adrian et al. (2013) also assume observable factors. They do not explicitly impose the Duffie-Kan restrictions, which is precisely the reason for the obtained simplification. However, they report that the restrictions are satisfied by the estimated parameters to a high degree of accuracy.

[13] One needs to abandon this convention at stage of practical implementation, when it is necessary to conform to the structure of the data.

[14] Continuously compounded yields and log bond prices are related to the forward rates in (1) by linear identities $b_t^m \equiv -\sum_{k=1}^m f_t^k$, $y_t^m \equiv -\frac{1}{m} b_t^m$.

$$f_t^m \equiv b_t^{m-1} - b_t^m. \tag{1}$$

The term structure at any point in time is determined by several underlying causal factors that are not directly observable. It is usually assumed that these latent factors span the term structure in a linear way, and their evolution can be modeled by a VAR process. In this paper I focus on the specifications in which the VAR is of order one (with Gaussian innovations), and all factors show up in the term structure. These assumptions can be summarized in

Assumption 1 *Linear Factor Structure with N_f Spanning Factors*

a) There exists a set of N_f latent state variables (factors) $\mathcal{X}_t$, such that the forward rates at every date t, and maturity m, are given by affine functions of the state vector[15]

$$f_t^m = f_0^m + f_1^{m\prime} \mathcal{X}_t. \tag{2}$$

b) The factors follow a first-order Gaussian vector auto-regression

$$\mathcal{X}_{t+1} = \mu_{\mathcal{X}}^{\mathbb{P}} + A_{\mathcal{X}}^{\mathbb{P}} \mathcal{X}_t + \varepsilon_{\mathcal{X},t+1}^{\mathbb{P}} \tag{3}$$

with innovation covariance matrix $\Sigma_{\mathcal{X}}$.

c) The vectors $f_1^{m\prime}$ form a matrix $F_{M \times N_f}$ that is full rank for some finite maturity M.

The assumption of first-order VAR is without loss of generality, because one can always re-define the lagged state variables into a set of additional contemporaneous factors. The assumption of Gaussian shocks is commonly made in applications, essentially for analytical tractability. On the other hand, assumption c) requires that there are exactly N_f dimensions in which the term structure can move. This rules out the possibility that some factors are not present in the term structure, and only affect its dynamics through the VAR process. The current paper shares this assumption with many others, notably JSZ, and CGJ, although the main results could be extended

[15]Without loss of generality, the state variables are linearly independent, i.e., there does not exist any linear combination of the members of $\mathcal{X}_t$ that results in a constant with probability one.

to the case in which c) needs not hold.[16]

The assumption above does not rule out arbitrage by itself. It is well known that no arbitrage is equivalent to the existence of a strictly-positive pricing kernel, whose growth rate is referred to as the stochastic discount factor (SDF). The focus of this paper is on the case in which the SDF takes the essentially-affine form introduced by Duffee (2002), which achieves a great degree of flexibility in modeling the risk premia.

Assumption 2 *No Arbitrage, with Essentially-Affine SDF*

There is no arbitrage, and the stochastic discount factor takes the form

$$M_{t+1} = \exp\left\{-f_t^1 - \frac{1}{2}\Lambda\Sigma_{\mathcal{X}}\Lambda' - \Lambda\varepsilon^{\mathbb{P}}_{\mathcal{X},t+1}\right\}, \tag{4}$$

$$\Lambda_t = \Lambda_0 + \Lambda_1\mathcal{X}_t. \tag{5}$$

The consequence of the above assumption is that the factor loadings f_0^m, and f_1^m in Assumption 1 are no longer free, but instead linked across maturities. This can be summarized in the following standard result, proven in Appendix A.

Result 1 *Under Assumptions 1 and 2, the log zero-coupon bond prices are given by affine functions of the latent factors, $b_t^m = b_0^m + b_1^{m\prime}\mathcal{X}_t$, with coefficients*

$$b_1^{m\prime} = b_1^{m-1\prime}A^{\mathbb{Q}}_{\mathcal{X}} - f_1^{1\prime}, \tag{6}$$

$$b_0^m = b_0^{m-1} - f_0^1 + b_1^{m-1\prime}\mu^{\mathbb{Q}}_{\mathcal{X}} + \frac{1}{2}b_1^{m-1\prime}\Sigma_{\mathcal{X}}b_1^{m-1}, \tag{7}$$

$$\mu^{\mathbb{Q}}_{\mathcal{X}} \equiv \mu^{\mathbb{P}}_{\mathcal{X}} - \Sigma_{\mathcal{X}}\Lambda_0, \tag{8}$$

$$A^{\mathbb{Q}}_{\mathcal{X}} \equiv A^{\mathbb{P}}_{\mathcal{X}} - \Sigma_{\mathcal{X}}\Lambda_1. \tag{9}$$

[16] The possibility of hidden latent factors was pointed out by Duffee (2011b). Intuitively, such factors are consistent with no arbitrage if they predict the future risk-free rate, while at the same time determine the risk premia in a way that exactly offsets the former effect. In many implementations, macroeconomic variables that help predict the term structure are explicitly included into the state vector. See (for example) Ang and Piazzesi (2003), or Joslin et al. (2014). Another case in which b) is violated (this time by construction) is when the factors are reverse-engineered from the term structure, and modeled through VAR with lag order greater than one. This case is considered in Joslin et al. (2013).

As a consequence, the loadings of forward rates in (2) satisfy the so-called Duffie-Kan restrictions

$$f_1^{m\prime} = f_1^{1\prime} \left(A_{\mathcal{X}}^{\mathbb{Q}}\right)^{m-1}, \tag{10}$$

$$f_0^m = f_0^1 - b_1^{m-1\prime}\mu_{\mathcal{X}}^{\mathbb{Q}} - \frac{1}{2} b_1^{m-1\prime}\Sigma_{\mathcal{X}} b_1^{m-1}. \tag{11}$$

Another well-known result states that under no-arbitrage there exists a risk-neutral probability measure $\mathbb{Q}$ with the property that all traded assets are valued as if investors were risk-neutral, but formed expectations based on the probabilities under $\mathbb{Q}$. In other words, all implications of the two assumptions above can be summarized in the following model

$$\mathcal{X}_{t+1} = \mu_{\mathcal{X}}^{\mathbb{P}} + A_{\mathcal{X}}^{\mathbb{P}}\mathcal{X}_t + \varepsilon_{\mathcal{X},t+1}^{\mathbb{P}}, \tag{12}$$

$$\mathcal{X}_{t+1} = \mu_{\mathcal{X}}^{\mathbb{Q}} + A_{\mathcal{X}}^{\mathbb{Q}}\mathcal{X}_t + \varepsilon_{\mathcal{X},t+1}^{\mathbb{Q}}, \tag{13}$$

$$f_t^1 = f_0^1 + f_1^{1\prime}\mathcal{X}_t. \tag{14}$$

The system (12)-(14) can be referred to as a generic form of a no-arbitrage Gaussian dynamic term structure model (GDTSM).[17] The explicit form of the SDF (4) is no longer needed once the $\mathbb{Q}$ dynamics is defined. Appendix A shows that this model leads to exactly the same solutions for bond prices (6)-(7) and forward rates (10)-(11). In particular, the model implies a linear factor structure in bond prices and forward rates. Since the solutions depend on the exogenous loadings of the risk-free rate with respect to the factors, the latter must be specified as part of the model through equation (14). Since the no-arbitrage assumption is part of the model, it is evident that the model is equivalent to assumptions 1 and 2 taken together.

Estimating the model is equivalent to finding parameters that determine $\mu_{\mathcal{X}}^{\mathbb{P}}$, $A_{\mathcal{X}}^{\mathbb{P}}$, $\mu_{\mathcal{X}}^{\mathbb{Q}}$, $A_{\mathcal{X}}^{\mathbb{Q}}$, f_0^1, f_1^1, and $\Sigma_{\mathcal{X}}$. In order to forecast the term structure one also needs to know the state vector.

2.2 Companion-Form Parametrization

The generic model (12)-(14) with latent factors has too many parameters, not all of which are identifiable.[18] One potential solution is implicit in JSZ, who derive their main findings based on

[17] Many papers start with this specification, for example JSZ.

[18] See the Introduction, and references therein.

the fact that there exists an invariant factor transformation under which the matrix $A_{\mathcal{X}}^{\mathbb{Q}}$ takes a Jordan form, parametrized by the eigenvalues. Unfortunately, the eigenvalue parametrization has an impractical property that one needs to know the configuration of the eigenvalues (their algebraic multiplicities, and whether they are complex or not) prior to model estimation. Moreover, the factors rotated into the Jordan form do not have economic meaning, which may lead to some difficulties with their interpretation, as discussed by Collin-Dufresne et al. (2008).

This section develops the companion-form parametrization (in discrete-maturity setup) that helps to circumvent these practical disadvantages by allowing the model to be efficiently estimated together with the factors, i.e., in a way fully consistent with no arbitrage. I first show that if the term structure is sampled in discrete maturity intervals, then it is in principle possible to rotate the factors in such a way that the first N_f shortest-maturity forward rates carry the same pricing information as the latent factors themselves. I then show that the risk-neutral transition matrix under this rotation takes a simple companion form, under which the model can be estimated by an unrestricted search for N_f real parameters.[19] I complete the section by showing that the parametrization of the risk-neutral drift $\mu_{\mathcal{X}}^{\mathbb{Q}}$ requires exactly one extra parameter, and state this result in a closed-form.

Proposition 1 *Every discrete-time GDTSM with exactly N_f latent factors $\mathcal{X}_t$ is observationally equivalent to a GDTSM of the form*

$$\mathcal{Y}_{t+1} = \mu_{\mathcal{Y}}^{\mathbb{P}} + A_{\mathcal{Y}}^{\mathbb{P}} \mathcal{Y}_t + \varepsilon_{\mathcal{Y},t+1}^{\mathbb{P}}, \tag{15}$$

$$\mathcal{Y}_{t+1} = \mu_{\mathcal{Y}}^{\mathbb{Q}} + A_{\mathcal{Y}}^{\mathbb{Q}} \mathcal{Y}_t + \varepsilon_{\mathcal{Y},t+1}^{\mathbb{Q}}, \tag{16}$$

$$f_t^1 = e_1{}' \mathcal{Y}_t, \tag{17}$$

where $\mathcal{Y}_t$ is a vector of N_f short-maturity forward rates (with the risk-free rate in the first position, s.t. $e_1 = [1, 0, \ldots, 0]'$), and the covariance of the innovations $\Sigma_{\mathcal{Y}}$.

Proof. As shown in Appendix B, it is enough to prove the existence of an invertible affine transformation $\mathcal{Y}_t = \alpha + \beta \mathcal{X}_t$, with $\mathcal{Y}_t$ being the vector of N_f shortest-maturity forward rates.

[19] The approach of CGJ also implies a companion form of the risk-neutral transition matrix. However, their continuous-time formulation makes it difficult to notice (and appreciate) the practical advantage of this fact. Although probably less elegant, the discrete-time setup can be considered more general, as the continuous time version obtains in the special case $\Delta m \to 0$.

Let $\mu(t) = t^n - c_{n-1}t^{n-1} - \cdots - c_1 t - c_0$ be the minimal polynomial of the transition matrix under the risk-neutral measure, i.e., the lowest-degree monic polynomial satisfying $\mu(A_{\mathcal{X}}^{\mathbb{Q}}) = 0_{N_f \times N_f}$.[20] If the degree of this polynomial is n, its defining property implies

$$\left(A_{\mathcal{X}}^{\mathbb{Q}}\right)^n = c_{n-1}\left(A_{\mathcal{X}}^{\mathbb{Q}}\right)^{n-1} + \cdots + c_1\left(A_{\mathcal{X}}^{\mathbb{Q}}\right) + c_0 I. \tag{18}$$

Pre-multiplying (18) by $f_1^{1\prime}$, post-multiplying by the vector of latent factors $\mathcal{X}_t$, and using the solution for the forward rates (10), one obtains

$$f_t^{n+1} = c_{n-1}f_t^n + \cdots + c_1 f_t^2 + c_0 f_t^1 + \mu. \tag{19}$$

The constant μ encompasses all fixed terms, and is defined to make the two sides equal. The key implication of (19) is that all variation in the n-th forward rate must be fully explained by the variation in forward rates of lower maturities. Multiplying both sides of (18) by $A_{\mathcal{X}}^{\mathbb{Q}}$, repeating similar steps, and using (19) leads to the conclusion that also the variation in f_t^{n+2} must be fully explained by the same set of forward rates. By induction, all forward rates must be spanned by $f_t^1, \ldots, f_t^n$. It follows that $n \geq N_f$, or otherwise the term structure could only move in fewer dimensions than N_f, contrary to the assumption that all latent factors have effect on forward rates. On the other hand, the degree of the minimal polynomial is bounded by N_f, because every square matrix satisfies its characteristic polynomial, which always has degree equal to N_f. The only possibility is therefore that $n = N_f$ (and the minimal polynomial equals the characteristic polynomial).[21]

It follows that all variation in the latent factors must also be present in the first N_f shortest-maturity forward rates. We can stack these forward rates in vector $\mathcal{Y}_t$, and define α and β through $\mathcal{Y}_t = \alpha + \beta\mathcal{X}_t$, i.e., α and β consist of model-implied factor loadings. This transformation is affine, and it is evident by now that β must be invertible, which constructively proves the existence of the invariant factor rotation of the required form. ■

Given that the latent factors can be rotated into the vector $\mathcal{Y}_t$, one can now investigate the

[20] A polynomial is called monic if the coefficient at the highest-degree term is normalized to one.

[21] The well-known fact that every matrix satisfies its characteristic polynomial is known as Cayley-Hamilton theorem. See, for example, Atiyah and Macdonald (1969), or Birkhoff and Mac Lane (1966).

effect of this transformation on model parameters. By Proposition 1, one is allowed to start with the formulation (15)-(17) directly. In particular, the functional forms of the solutions (6)-(7) and (10)-(11) remain valid, but with parameters corresponding to the rotated model. In particular, (11) written for the first N_f forward rates ($m \in \{1, \dots, N_f\}$) takes the form

$$0 = -b_1^{m-1\prime} \mu_{\mathcal{Y}}^{\mathbb{Q}} - \frac{1}{2} b_1^{m-1\prime} \Sigma_{\mathcal{Y}} b_1^{m-1}, \tag{20}$$

which follows from the fact that these forward rates now play the role of factors, implying $f_0^m = 0$. The coefficients b_1^m take a particularly simple form under the rotated model,

$$b_1^{m\prime} = [1, \dots, 1, 0, \dots, 0], \tag{21}$$

with ones filling the first m places.[22] It follows that N_f-1 equations (20) with $m \in \{2, \dots, N_f\}$ recursively determine the first N_f-1 components of the risk-neutral drift vector $\mu_{\mathcal{Y}}^{\mathbb{Q}}$ in terms of the entries of the covariance matrix $\Sigma_{\mathcal{Y}}$ alone. Defining $j^m \equiv \frac{1}{2} b_1^{m\prime} \Sigma_{\mathcal{Y}} b_1^m$ (with $j^0 = 0$), it is straightforward to show that the m-th element of $\mu_{\mathcal{Y}}^{\mathbb{Q}}$ is just $\Delta j^m \equiv j^m - j^{m-1}$.[23]

The last term of $\mu_{\mathcal{Y}}^{\mathbb{Q}}$ can be identified by considering the constant μ in (19), which by construction equals

$$\mu = f_0^{N_f+1} - \left(c_{N_f-1} f_0^{N_f} + \dots + c_1 f_0^2 + c_0 f_0^1 \right).$$

Since all f_0 terms in the round bracket are zero, it must be that $\mu = f_0^{N_f+1}$. Now, equation (11) can be applied to the N_f+1-th forward rate,

$$\mu = -b_1^{N_f\prime} \mu_{\mathcal{Y}}^{\mathbb{Q}} - j^{N_f},$$

which together with previously-found components of $\mu_{\mathcal{Y}}^{\mathbb{Q}}$ and the form of $b_1^{N_f}$ given in (21) implies that the last element of $\mu_{\mathcal{Y}}^{\mathbb{Q}}$ is $\Delta j^m + \mu$.

The transition matrix $A_{\mathcal{X}}^{\mathbb{Q}}$ can easily be identified as the companion matrix of the minimal

[22] I slightly abuse the notation by re-using the symbols for factor loadings like f_0^m or b_1^m in the context of models with different sets of factors.

[23] For $m = 1$, equation (20) becomes an identity, containing no information about $\mu_{\mathcal{Y}}^{\mathbb{Q}}$.

polynomial $\mu(t)$ from the proof of proposition (1), i.e.,

$$A_{\mathcal{X}}^{\mathbb{Q}} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \ddots & 1 \\ c_0 & c_1 & c_2 & \dots & c_{N_f-1} \end{bmatrix} \tag{22}$$

The first N_f-1 rows reflect the fact that forward rates predict under $\mathbb{Q}$ their lower-maturity counterparts, which can be seen in condition (10). The last row contains the coefficients of the minimal polynomial $\mu(t)$, as implied by (19).

All of these findings can be summarized as a proposition:

Proposition 2 *If the generic model with latent factors $\mathcal{X}_t$ is transformed into the form of Proposition (1) (with first N_f model-implied forward rates as new factors $\mathcal{Y}_t$), then*

a) The risk-neutral drift is $\mu_{\mathcal{Y}}^{\mathbb{Q}} = [\Delta j^1, \dots, \Delta j^{N_f-1}, \Delta j^{N_f} + \mu]'$, where

$$\Delta j^m = \frac{1}{2} b_1^{m\prime} \Sigma_{\mathcal{Y}} b_1^m - \frac{1}{2} b_1^{m-1\prime} \Sigma_{\mathcal{Y}} b_1^{m-1}$$
$$b_1^m = [1, \dots, 1, 0, \dots, 0], \quad \textit{(ones in m first positions)}$$

and μ is a free parameter.

b) The risk-neutral transition matrix $A_{\mathcal{X}}^{\mathbb{Q}}$ is in companion form (22).

Proposition 2 shows that in order to parametrize the risk-neutral dynamics of every GDTSM, one needs to specify $N_f(N_f + 1)/2$ parameters of the Cholesky decomposition of the covariance matrix, plus $N_f + 1$ additional free numbers. Although it appears tempting to restrict $\mu = 0$ based on the special structure of the drift, there seems to be no economic argument supporting this.

3 Empirical Implementation

This section starts with a discussion of the data, followed by two maximum-likelihood estimation strategies for the model in the companion form. The first employs Kalman filtering in order to estimate the factors together with model parameters. The second rests on the (assumed) possibility of observing the factors with no error through empirical PCA scores.

3.1 Data

The data employed in the empirical analysis are the Fama and Bliss (1987) discount bonds from CRSP. The set consists of five full-year maturities between one and five years, sampled at monthly frequency. The maximal sample used in this paper spans June 1956 – July 2015. One advantage of this set is that no smoothing methods are applied at its construction stage, which is especially important from the point of view of the current study.[24] Important prior applications of Fama-Bliss data include Fama and Bliss (1987), and Cochrane and Piazzesi (2005, 2009), who study the failure of the Expectations Hypothesis. The latter study show that the yield-forecasting factor constructed from Fama-Bliss forward rates also captures bond return predictability in another well-known data set of Gürkaynak et al. (2007), constructed by the method of Svensson (1995).

Arguably, the long time span may result in problems if there are important structural breaks in the data-generating process. For this reason I also use a subset of the data corresponding to the study in Cochrane and Piazzesi (2005), i.e., spanning January 1965 – December 2003. It is known that the time variation in empirical bond risk premia was particularly pronounced in that sample. Moreover, this choice avoids the most recent period (starting from 2009), in which the short-maturity interest rates have remained very close to the zero bound.[25] I refer to this subset simply as the CP sample.

3.2 Estimation by Kalman Filter

The model in companion form leads to the state-space representation

$$\mathcal{Y}_{t+1} = \mu^{\mathbb{P}}_{\mathcal{Y}} + A^{\mathbb{P}}_{\mathcal{Y}}\mathcal{Y}_t + \varepsilon^{\mathbb{P}}_{\mathcal{Y},t+1}, \tag{23}$$

$$b^o_t = B_0 + B_1\mathcal{Y}_t + v_t, \tag{24}$$

[24] The detailed description of the data construction is provided on `http://www.crsp.com/products/documentation/fama-bliss-discount-bonds-%E2%80%93-monthly-only`.

[25] I do not explicitly model the zero bound, although the results could likely be extended to that case. A list of representative papers that explicitly take the zero bound into account can be found in the Introduction.

in which b_t^o is a vector of noisy observations of bond prices at time t, and B_0, B_1 are model-implied coefficients satisfying no arbitrage. I assume that the covariance matrix R of the measurement errors v_t is diagonal, with equal variances σ_v^2 of individual terms. The state-space representation could equivalently be defined in terms of other term-structure observables (forward rates, yields), but this would require a modification of the measurement error covariance matrix in order to reflect the particular transformation of the data.[26]

The parameters are estimated by ML, using Kalman filter to compute the sequence of observable innovations for every set of parameter values, and choosing the set of parameters that maximizes the log likelihood function. Appendix C presents the standard recursive steps needed to compute the likelihood. I use the stationary version of the filter, which is consistent with the constant-volatility structure of the model.

The set of parameters is $\Theta = \{\theta_\mu, \theta_A, \theta_\Sigma, \mu, c, \theta_R\}$, in which θ_μ and θ_A are vectors of parameters of the $\mathbb{P}$ dynamics, corresponding to $\mu_{\mathcal{Y}}^{\mathbb{P}}$ and $A_{\mathcal{Y}}^{\mathbb{P}}$, respectively. θ_Σ consists of parameters governing the innovation covariance matrix $\Sigma_{\mathcal{Y}}$ (obtained by the Cholesky decomposition), μ and c are parameters of the $\mathbb{Q}$ dynamics under the companion parametrization (see Proposition 2), and θ_R is a vector that determines the noise covariance matrix (in the current context, just one number σ_v^2).

I assume three factors, which results in 23 parameters (under the assumption that only one parameter determines R).[27] It is well known that success of numerical search algorithms often depends on the right choice of the starting parameter values. To obtain initial estimates of θ_μ, θ_A, and θ_Σ, I use an unconstrained VAR on the shortest-maturity forward rates, which are the noisy versions of the true factors under the companion parametrization. The staring values for parameters in the c vector (coefficients of the minimal polynomial of $A_{\mathcal{X}}^{\mathbb{Q}}$) are obtained as follows. I first perform the principal-component analysis of the unconditional covariance matrix of all observed forward rates, store the factor loadings $f_1^{m\prime}$, and exploit the Duffie-Kan restrictions (10) in the form $f_1^{m\prime} = f_1^{m-1\prime} A^{\mathbb{Q}}$ to estimate the columns of the transition matrix $A^{\mathbb{Q}}$ by linear regressions (with no constant terms). The initial value of c is then obtained by computing the characteristic polynomial

[26] The elements of v_t, to which I refer to as measurement errors, in fact summarize all deviations of measured bond prices from their model-implied values, occurring for any reason, e.g., micro-structure effects. It is unclear what data transformation should be considered canonical in the sense of being consistent with a diagonal R matrix. I assume that this is true for bond prices, because they reflect most directly what should matter to investors. The model can be estimated under arbitrary R matrix, at the cost of introducing new unknown parameters. The experience of the author suggests that estimation results are quite insensitive to the particular choice of R.

[27] With two (four) factors, this number would become 13 (35).

of $A^{\mathbb{Q}}$.[28] The initial value of μ is set to zero, which appears to be a natural choice given the structure of the risk-neutral drift in Proposition 2. Finally, the variance of the measurement error is set such that the implied standard deviation of each component of v_t is 10 basis points.

The numerical optimization is performed using the Nelder-Mead algorithm, implemented in Matlab function `fminsearch`. In the case of the 3-factor model estimated on the full sample, the procedure converges in about 2-3 minutes, depending on computer speed. [29]

The results for the benchmark three-factor model estimated using all data are presented in Table 1. Panel A. displays the drift, transition matrix, and the covariance matrix of innovations for the estimated monthly $\mathbb{P}$ dynamics. Panel B. contains the same information in annualized form, which makes it more comparable with the annual $\mathbb{Q}$ dynamics, presented in Panel C. Of special importance is the companion matrix $A_{\mathcal{Y}}^{\mathbb{Q}}$, with all parameters contained in the last row. These parameters are the coefficients of its minimal (and at the same time characteristic) polynomial, c_0, c_1, and c_2. The estimated values are not easy to interpret directly. Mathematically, c_0 is the determinant of $A_{\mathcal{Y}}^{\mathbb{Q}}$, and c_2 is its trace. All three numbers are considerably away from the starting values computed from empirical PCA loadings, which took more extreme values of -4.3, 10.4, and -5.3, respectively (not reported in the table).

Panel D. of Table 1 shows the magnitudes of the pricing errors between the observed, and the model-implied bond prices (in logs). The errors are reasonably small, ranging between 6.8-19.4 basis points, depending on maturity and the method used to compute them (root-mean-square vs. mean-absolute-value). This panel also shows the estimated noise standard deviation of 19.3 b.p. These pricing errors are comparable to other studies fitting affine models to Fama-Bliss data, e.g., Cochrane and Piazzesi (2009).

The eigenvalues implied by these estimates are shown in panel E., in decreasing order. Not surprisingly, all of them are positive for the $\mathbb{P}$ dynamics, and below one. The annualized values are the 12-th powers of the respective monthly numbers. The two positive eigenvalues of the $\mathbb{Q}$ dynamics

[28] One can use the Matlab function `charpoly`. The characteristic polynomial is invariant to changes of basis, and therefore independent of invariant factor transformations. If the PCA factors were the true factors, and if the model exactly satisfied the no-arbitrage restrictions, the regressions used to obtain $A^{\mathbb{Q}}$ would fit perfectly, and the empirical characteristic polynomial would uncover the true parameters.

[29] I set the maximum numbers of iterations, and function evaluations to 10000, and the tolerance level to 1e-3. If the procedure does not converge within these limits, the resulting parameters are used as starting values for the next iteration, until convergence. The numerical optimizer of Matlab is quite popular among other authors. For a technical discussion of the convergence properties of the Nelder-Mead algorithm, see Lagarias et al. (1998).

are slightly higher than their $\mathbb{P}$ counterparts, suggesting positive risk premia for the innovations to the associated dimensions of factor movements. On the other hand, one of the eigenvalues under $\mathbb{Q}$ is negative (and large in absolute value), which is somewhat counter-intuitive, although by itself not inconsistent with no arbitrage. As will be discussed later, the negative eigenvalue appears to be a robust property of the data, and simply indicates that the model faces difficulties in fitting the term structure with a more "reasonable" number.

The dashed lines in every sub-figure of Figure 2 show the canonical factors, i.e., the three shortest-maturity forward rates estimated by the Kalman filter in the full set of Fama-Bliss data (the lines are shifted by 50 b.p. relative to observed forward rates in order to improve readability). Evidently, the estimated factors closely track their noisy counterparts, confirming the reliability of the filtering procedure.

Table 2 shows the results obtained by the Kalman filter on the CP sample. It appears that the parameters of the $\mathbb{P}$ dynamics, both in monthly, and in annualized terms, are quite similar to their full-sample estimates. Also, the $\mathbb{Q}$ dynamics does not differ by much, which is also true for the pricing errors (which are marginally smaller). The large negative eigenvalue of $A_{\mathcal{Y}}^{\mathbb{Q}}$ is present also in the CP sub-sample.

3.3 Estimation Under Observable Factors

The literature on term structure modeling appears to have reached a consensus that the factors can be treated as observable for the purpose of model estimation and forecasting. For example, Joslin et al. (2011), and Hamilton and Wu (2012) mostly work under the assumption that there are exactly N_f known linear combinations of observed yields that the model is able to price without error. In other words, measuring some well-chosen linear combinations of term structure observables uncovers the factors. Intuitively, if factors are observable, no filtering is needed to extract information about the precise position of the state vector, which significantly reduces computational complexity of the problem.

Suppose that such combinations indeed exist, and that one knows the matrix W that produces a vector of bond (or yield) portfolios $\mathcal{P}_t$ priced without error. In applications, W usually corresponds to the (orthogonal) principal-component loadings, or to some well-chosen constant-maturity yields. The assumption of no errors on $\mathcal{P}_t$ is equivalent to a statement that there exists an invariant transformation allowing to re-state the model in terms of $\mathcal{P}_t$ instead of the original factors.

In particular, the baseline model can be stated in terms of the first N_f shortest-maturity forward rates, i.e., using the companion-form parametrization, and rotated into a model with observable bond portfolios as factors. For concreteness, assume that the W matrix consists of PCA loadings, in which case $\mathcal{P}_t$ have the interpretation of empirical PCA factor scores. Based on the results of Joslin et al. (2011), and Hamilton and Wu (2012), the ML parameter estimates governing the conditional expectation of the factors $\mathcal{P}_t$ under the $\mathbb{P}$ measure can be recovered by OLS, running an unrestricted VAR on these factors, without resorting to numerical search.[30] As a result, the assumption pins down both the factors, and their $\mathbb{P}$ dynamics before other parameters are found. The estimation of the model is then reduced to finding the innovation covariance matrix, parameters controlling the conditional factor expectations under $\mathbb{Q}$, and the covariance matrix of the measurement errors. Intuitively, this is performed by finding parameters that produce model-implied bond prices (or yields) that are assumed to be measured *with* error, in terms of the observed factors.

The companion-form parametrization is particularly convenient also in the context of observable factors, and Appendix D formally derives the likelihood function.[31] The initial parameters for the $\mathbb{Q}$ dynamics are found in exactly the same way as for the Kalman filter case. The starting value of the innovation covariance matrix can be computed from the VAR residuals of measured PCA scores. In fact, it is possible (and also numerically efficient) to parametrize the likelihood directly in terms of the dynamics (and covariance) of the PCA factors, which is the way in which I proceed.

The results are displayed in Table 3 (full sample), and 4 (CP sample). Panels A. of these tables report the estimated monthly $\mathbb{P}$ dynamics of empirical PCA scores, to which I will refer by their traditional names, i.e., level, slope, and curvature. Panels B. present the annualized dynamics. An interesting property of the transition matrices is that they are approximately upper-triangular, implying that the slope and the curvature predict the level, but not the other way. Another consequence of this structure is that the diagonal elements essentially correspond to the eigenvalues, which is confirmed in panels E. The estimated $\mathbb{Q}$ dynamics are qualitatively similar to the Kalman filter cases, although the numbers look more extreme. Also the negative eigenvalues of

[30] OLS residuals can also be used to construct a very good set of starting values for the parameters that define factor innovation covariance matrix.

[31] The steps in Appendix D can also be regarded as an alternative explanation of the main result in JSZ, namely that under observable factors estimating the risk-neutral parameters by itself does not improve the ability of the model to predict the term structure factors.

the risk-neutral transition matrices are much greater in absolute values, reaching -13.3, and -5.6, depending on the sample. The estimated covariance matrices of forward-rate factor innovations ($\Sigma_{\mathcal{Y}}$) imply larger conditional volatilities than under the Kalman filter, which is consistent with the interpretation that 100% of factor innovations must be considered genuine, and not due to measurement errors.

Interestingly, the pricing errors reported in panels D. of both tables are slightly lower than their Kalman filter counterparts. Especially the 5-year bond appears to be explained with virtually no error. As will be explained below, this seemingly perfect precision is a consequence of the special role that curvature plays under the assumption of observable factors, and is related to the presence of an extreme eigenvalue in the spectrum of the estimated $\mathbb{Q}$ dynamics.

Finally, Figure 2 presents the time-series plots of the model-implied forward rates of maturities one, two, and three years (dotted lines), inverted from the PCA factors using a model-implied mapping. Also in this case the factors are almost perfectly aligned with the observed forward rates.

3.4 Interpreting the Anomalous Eigenvalue

A large negative eigenvalue associated with the estimated $\mathbb{Q}$ dynamics appears to be a persistent feature of the data, independently of whether the model is estimated by the Kalman filter, or under the assumption of observable factors. The top of Figure 3 displays the loadings of all forward rates with respect to the three PCA factors extracted from the data, and the two other rows of that figure present analogous loadings implied by the estimated models, under both sets of assumptions discussed. The dotted lines correspond to the loadings on the smallest factor, responsible for 0.32%-0.52% of the total variance in bond prices.

Interestingly, the dotted lines in all pictures feature a large jump between maturities of four and five years, which is not associated by any clear pattern in the short-maturity loadings relative to the other factors. In other words, there is a factor that only affects long-maturity forward rates, while being barely distinguishable from the other factors if one looks at the shortest maturities in isolation. This is especially true in the graphs at the bottom of Figure 3, which correspond to the estimation method that is instructed (by its design) to treat the empirical PCA scores as given, and only adjust model parameters (the $\mathbb{Q}$ dynamics) in order to produce factor loadings consistent with the Duffie-Kan restrictions. The role of the anomalous factor is then only to produce the spike on the last bond price relative to the dashed line (the slope factor). This can only be achieved by prescribing a very large eigenvalue, making use of the fact that model-implied factor loadings are power functions of maturities.

On the other hand, by the theoretical results of this paper, such factor behavior is difficult to reconcile with no arbitrage, because all information about the state variables should be detectable from short-maturity forward rates. Why is the anomalous factor not classified as noise, at least by the Kalman filter? The answer is that it contains important information about the future behavior of the yield curve. Indeed, the return-forecasting factor of Cochrane and Piazzesi (2005) loads heavily on the four-to-five yield differential, as emphasized by these authors.[32] The anomalous factor must be therefore accepted as a robust feature of the data, and using Kalman filter at the estimation stage cannot help significantly.[33]

These findings may shed some light on the question of whether assuming no arbitrage can help forecast the term structure. One can (correctly) argue that assuming no arbitrage should in principle improve forecasts, because it summarizes very strong incentives of market participants. At the same time, several papers, notably Joslin et al. (2011), and Duffee (2011a) dispute the usefulness of no-arbitrage assumptions in real-world forecasting applications. Based on the results of this paper, these arguments can even be extended to saying that imposing no arbitrage (or more precisely, restricting "unreasonable" parameter values at the estimation stage) may even lead to *worse* forecasts. Clearly, in the current context this implication probably mostly follows from model inflexibility, and should not be taken as strong evidence of arbitrage in historical data.[34]

4 Simulated Out-of-Sample Performance

It can be argued that the assumption of observable factors may lead to a bias in OLS factor dynamics due to errors in variables. On the other hand, if the W matrix (that determines observable portfolios of yields) is constructed from the PCA decomposition of the unconditional covariance matrix, and

[32] See Cochrane and Piazzesi (2005), p. 146, and Fig. 2 on p. 141.

[33] Kalman filter collects predictive information also from the smaller empirical factors in observed bond prices that are classified as pure noise by the PCA. This explains why the dotted lines in the middle row of Figure 3 are less smooth. Intuitively, the lower magnitudes of the eigenvalues are possible because the factor loadings feature more visible oscillations, which makes the factor somewhat better visible in the shortest maturities.

[34] It is possible that the anomalous results would largely go away if one had access to a richer data set with more maturities. This possibility, if true, would be a warning against model over-fitting by specifying a three-factor model in a data set with only 5 maturities. To my defense, such practice is standard in the literature.

if bond prices really follow a linear factor structure, then one may hope to obtain relatively good factor estimates in this way.

So are there any practical advantages of estimating the model together with the factors (and their corresponding loadings), in a way fully consistent the model, and with no arbitrage at the same time? This section addresses this issue by performing Monte-Carlo simulations, which is feasible thanks to the computational efficiency of the estimation methods under the companion parametrization.

4.1 Artificial Data

Seemingly, the most natural way to assess the out-of-sample performance of two estimation methods would be to perform a rolling-window estimation in the real data. The results of such exercise would, however, be hard to interpret. For example, the assumption of parameter stability is likely to be violated due to structural breaks. Also, the exact nature of the measurement error, and the precise number of factors (possibly time-varying) are unknown. These are only a few examples of violations of model assumptions which may favor one method over the other, depending on which of them happens to be more robust under given circumstances.

In order to assess the severity of the bias, I therefore simulate 300 panels of bond prices, using the baseline parameter values taken from Table 1. Every artificial data set is of the same size as the full Fama-Bliss sample, plus extra 60 months used to compute out-of-sample forecast errors. To every sample I add i.i.d. noise generated under the assumption of diagonal covariance matrix, with variances equal across maturities.

In every simulated sample the model is estimated twice, without and with the assumption of observable factors, as explained in the previous section.

In the first case, the estimation procedure relies on Kalman filtering, with first estimating a VAR on noisy short-maturity forward rates to obtain the starting values of the $\mathbb{P}$ dynamics. In the second case, I first obtain a PCA decomposition of the unconditional covariance of (noisy) bond prices, and use the associated factor loadings to construct bond portfolios supposedly measured without error. I then compute their VAR dynamics by OLS, and store the innovation covariance matrix in order to use it as a starting value for the ML estimate. The starting parameter values of the $\mathbb{Q}$ dynamics are computed in the same way under both methods, which also share the same initial value of measurement error variance.[35] The forecast errors are defined as differences between the realized (noise-free) log bond prices, and their model-based forecasts.

4.2 Out-of-Sample Results

Table 6 presents the root mean square errors (RMSE) over all simulations, for all bond maturities between one and five years, and selected horizons between one month and five years. Panels A., B., and C. correspond to the Kalman filter, ML with observable factors, and random walk, respectively. Both estimation methods produce forecasts that are superior to the random walk, especially at short horizons (1-3 months), and for short-maturity bonds (1-2 years), although the gain disappears in other cases. Table 6 presents comparative benefits of using pairs of forecasting methods relative to each other. The most important information from the point of view of this study is contained in panel C. It can be seen that both methods under investigation are able to produce almost equally precise forecasts. There seems to be a marginal benefit in using Kalman filter for short-term forecasts of the one-year bond, but in other cases the benefit disappears.

Table 7 and Table 8 are constructed in an identical way, but present the mean absolute forecast errors (MAE), and confirm all conclusions drawn above. Summing up, this exercise has not provided a conclusive argument for the uniform acceptance of one estimation method over the other.

5 Concluding Remarks

In this paper, I build on the insights of Joslin et al. (2011) (JSZ), and Collin-Dufresne et al. (2008) (CGJ) to develop a canonical parametrization (and estimation strategy based on it) for the very popular class of Gaussian Dynamic Term Structure Models (GDTSM) with essentially-affine prices of risk of Duffee (2002), and Ang and Piazzesi (2003), specified in discrete time.

The convenience of the proposed parametrization rests upon the proven result that if there is no arbitrage, and if the term structure is driven by N_f spanning factors, then all the information about the factors must be contained in the shortest end of the maturity spectrum. In the limiting case, studied by CGJ, the factor dynamics can be mapped into the dynamics of the first N_f derivatives of the term structure evaluated at maturity zero. In the discrete-time case, all information about

[35] I use the same value as in the case of real-data estimation, i.e., 1e-6.

the factors must be contained in the first N_f shortest-maturity forward rates.

This leads to a natural parametrization of the $\mathbb{Q}$ dynamics in terms of the coefficients of the characteristic polynomial of the transition matrix. Actually, if the model is stated in terms of the short-maturity forward rates as factors, then its $\mathbb{Q}$ transition dynamics assumes the form of a companion matrix (associated with the characteristic polynomial), in which only the last row is unrestricted, and all other entries fixed to 0 or 1 by the Duffie-Kan restrictions. The characterization of the risk-neutral dynamics can be completed by specifying one extra parameter governing the risk-neutral drift of the factors (in addition to the factor covariance matrix).

The companion-form parametrization is akin to the eigenvalue parametrization employed by JSZ. Both of them minimize the number of parameters that can be identified from the data, while at the same time preserve full model flexibility. The companion-form parametrization is arguably easier to implement, because one does not need any prior knowledge of how the characteristic polynomial factorizes. If the eigenvalues are repeated (or complex) the mathematical form of the $\mathbb{Q}$ transition matrix changes, and these cases have to be considered separately at the estimation stage. It is also unclear whether one can achieve smooth transition between these cases during a numerical evaluation of the likelihood function. [36]

Due to the relatively low number of unknown parameters, the model can be efficiently estimated by the Kalman filter, in which case the factors and factor loadings are fully consistent with the no-arbitrage restrictions imposed by the model on the data. While this approach appears to be computationally efficient (in the set of Fama-Bliss bond prices), it is possible to increase the estimation speed even more by assuming that factors are observable. For example, one can use the principal components of bond prices (or yields) in an attempt to uncover the factors, thus avoiding Kalman filtering. This method can be applied irrespective of the chosen canonical form, since one can always back out the fundamental factors from the model-implied mapping between the PCA scores and the underlying factors, for example the short-maturity forward rates.[37]

Consistent with the main point of JSZ, under the assumption of observable factors it should

[36]Apart from this property of estimation convenience, the companion-form parametrization is more useful if one wishes to impose extra conditions on the characteristic polynomial. For example, a condition that the determinant of $A^{\mathbb{Q}}$ is zero can be easily imposed on the model in companion form, since the determinant is simply one of the parameters. On the other hand, any constraints on the eigenvalues would be easier to impose under the JSZ normalization. I do not consider any such additional constrains in the current paper.

[37]Although the factors under the canonical companion form are the shortest-maturity forward rates, one *needs not* assume that they are measured without error.

be impossible to improve factor predictions over those obtained by an unrestricted VAR, at least if the model is true in the given data set.[38] Since it is only slightly more time-intensive to apply the Kalman filter, a natural question is whether it is worth to make the simplifying assumption of observable factors, which is guaranteed not to hold exactly in any given data. To address this question, I perform a Monte-Carlo analysis of the forecasting performance of the two just-described estimation strategies. To put both methods on equal grounds, I simulate many data sets using benchmark parameter values estimated in the real data, and add noise to the generated bond prices. The results indicate that both Kalman filtering, and estimation under observable factors offer virtually identical out-of-sample precision.

However, this conclusion may change in real-world applications, because some model assumptions may fail. For example, the canonical companion form provides the lens through which one can see problems with fitting essentially-affine models to the well-known set of Fama-Bliss discount bonds, in which small factors that predict bond returns seem to only be present in long-maturity forward rates, without significantly (if at all) affecting the short end. Based on the results of the current paper, this can only be reconciled with essentially-affine model structure if one is ready to accept extreme eigenvalue estimates associated with such factors. The well-known bond return forecasting factor of Cochrane and Piazzesi (2005) relies significantly on such small empirical factors. This does not necessarily imply that a large part of the predictability evidence was due to arbitrage possibilities.[39] Rather, it signals that the current workhorse no-arbitrage framework may need further development.

[38] Appendix D of this paper can be seen as an alternative proof of this fact.

[39] However, the recent financial crisis has proven that even well-established no-arbitrage relations may fail under conditions of market stress.

Appendix A Solving Generic GDTSM

This appendix summarizes the standard steps in solving a Gaussian term structure model starting from the form with explicit stochastic discount factor (SDF), and then shows that the model can be equivalently stated in terms of the risk-neutral dynamics of the factors.

Explicit SDF

Assume that the dynamics of the latent factors $\mathcal{X}_t$, and the risk-free rate are

$$\mathcal{X}_{t+1} = \mu_{\mathcal{X}}^{\mathbb{P}} + A_{\mathcal{X}}^{\mathbb{P}} \mathcal{X}_t + \varepsilon_{\mathcal{X},t+1}^{\mathbb{P}}, \tag{25}$$

$$f_t^1 = f_0^1 + f_1^{1\prime} \mathcal{X}_t, \tag{26}$$

and the SDF is of the form

$$logM_{t+1} \equiv m_{t+1} = -f_t^1 - \frac{1}{2}\Lambda \Sigma_{\mathcal{X}} \Lambda' - \Lambda \varepsilon_{\mathcal{X},t+1}^{\mathbb{P}}, \tag{27}$$

$$\Lambda_t = \Lambda_0 + \Lambda_1 \mathcal{X}_t. \tag{28}$$

Solving the model is equivalent to finding closed-form solutions for zero-coupon bond prices for every maturity m. Normalizing the face value of every bond to one, it is conjectured (and later verified) that the log zero-coupon bond prices are affine in the latent state vector,

$$b_t^m = b_0^m + b_1^{m\prime} \mathcal{X}_t. \tag{29}$$

The one-period return accruing to the holder of a bond is

$$\begin{aligned} r_{t+1}^m &= b_{t+1}^{m-1} - b_t^m \\ &= b_0^{m-1} - b_0^m + b_1^{m-1\prime} \mu_{\mathcal{X}}^{\mathbb{P}} + \left(b_1^{m-1\prime} A_{\mathcal{X}}^{\mathbb{P}} - b_1^{m\prime} \right) \mathcal{X}_t + b_1^{m-1\prime} \varepsilon_{\mathcal{X},t+1}^{\mathbb{P}}, \end{aligned} \tag{30}$$

which follows from (29), together with the factor dynamics (25).

The asset-pricing condition for log-normally distributed returns is[40]

$$E_t^{\mathbb{P}}\left(m_{t+1}+r_{t+1}^m\right)+\frac{1}{2}Var_t\left(m_{t+1}+r_{t+1}^m\right)=0. \tag{31}$$

Substituting 27 and 30 into (31), and using the condition that the pricing equation must hold for every possible value of the state vector, one obtains a system of recursive conditions

$$\begin{aligned} b_1^{m\prime} &= b_1^{m-1\prime}\left(A_{\mathcal{X}}^{\mathbb{P}}-\Sigma_{\mathcal{X}}\Lambda_1\right)-f_1^{1\prime}, \\ b_0^m &= b_0^{m-1}-f_0^1+b_1^{m-1\prime}\left(\mu_{\mathcal{X}}^{\mathbb{P}}-\Sigma_{\mathcal{X}}\Lambda_0\right)+\frac{1}{2}b_1^{m-1\prime}\Sigma_{\mathcal{X}}b_1^{m-1}. \end{aligned}$$

In order to guarantee that the shortest-maturity bond price corresponds to the assumed form of the risk-free rate f_t^1, one needs to define $b_0^0=0$ and $b_1^{0\prime}=0'$ as part of the solution, which at the same time allows to solve the recursions for every maturity m.

If one defines $\mu_{\mathcal{X}}^{\mathbb{Q}}\equiv\mu_{\mathcal{X}}^{\mathbb{P}}-\Sigma_{\mathcal{X}}\Lambda_0$ and $A_{\mathcal{X}}^{\mathbb{Q}}\equiv A_{\mathcal{X}}^{\mathbb{P}}-\Sigma_{\mathcal{X}}\Lambda_1$, these conditions take the form

$$b_1^{m\prime}=b_1^{m-1\prime}A_{\mathcal{X}}^{\mathbb{Q}}-f_1^{1\prime}, \tag{32}$$

$$b_0^m=b_0^{m-1}-f_0^1+b_1^{m-1\prime}\mu_{\mathcal{X}}^{\mathbb{Q}}+\frac{1}{2}b_1^{m-1\prime}\Sigma_{\mathcal{X}}b_1^{m-1}. \tag{33}$$

It follows that the forward rates are of the form $f_t^m=f_0^m+f_1^{m\prime}\mathcal{X}_t$, with loadings

$$f_1^{m\prime}=f_1^{1\prime}\left(A_{\mathcal{X}}^{\mathbb{Q}}\right)^{m-1}, \tag{34}$$

$$f_0^m=f_0^1-b_1^{m-1\prime}\mu_{\mathcal{X}}^{\mathbb{Q}}-\frac{1}{2}b_1^{m-1\prime}\Sigma_{\mathcal{X}}b_1^{m-1}. \tag{35}$$

Risk-Neutral Factor Dynamics

No-arbitrage GDTSMs are sometimes stated in terms of the risk-neutral dynamics of the factors. Define the $\mathbb{Q}$ expectation $E_t^{\mathbb{Q}}(z_{t+1})$ of a generic random variable z_{t+1} as $E_t^{\mathbb{Q}}(z_{t+1})\equiv E_t^{\mathbb{P}}(z_{t+1})+Cov_t(m_{t+1},z_{t+1})$, where m_{t+1} is the log SDF.[41]

The VAR factor dynamics (25), after subtracting and adding the term $\Sigma_{\mathcal{X}}\Lambda_t$ defined in (28),

[40] This is a special case of the condition that the price P_t of a payoff P_{t+1} must satisfy $P_t=E_t^{\mathbb{P}}(M_{t+1}P_{t+1})$ if the SDF is M_{t+1}.

[41] Intuitively, the objective expected value under $\mathbb{P}$ is being adjusted due to risk, as measured by the covariance with the SDF.

becomes

$$\mathcal{X}_{t+1} = \left(\mu_{\mathcal{X}}^{\mathbb{Q}} - \Sigma_{\mathcal{X}}\Lambda_0\right) + \left(A_{\mathcal{X}}^{\mathbb{Q}} - \Sigma_{\mathcal{X}}\Lambda_1\right)\mathcal{X}_t + \varepsilon_{\mathcal{X},t+1}^{\mathbb{P}} + \Sigma_{\mathcal{X}}\left(\Lambda_0 + \Lambda_1\mathcal{X}_t\right).$$

Using the definitions of $\mu_{\mathcal{X}}^{\mathbb{Q}}$ and $A_{\mathcal{X}}^{\mathbb{Q}}$ as in (32)-(33), and defining a new shock $\varepsilon_{\mathcal{X},t+1}^{\mathbb{Q}} \equiv \varepsilon_{\mathcal{X},t+1}^{\mathbb{P}} + \Sigma_{\mathcal{X}}\Lambda_t$, the previous equation can be written more compactly as

$$\mathcal{X}_{t+1} = \mu_{\mathcal{X}}^{\mathbb{Q}} + A_{\mathcal{X}}^{\mathbb{Q}}\mathcal{X}_t + \varepsilon_{\mathcal{X},t+1}^{\mathbb{Q}}, \tag{36}$$

where the shock is Gaussian, with zero expectation under $\mathbb{Q}$. By the Girsanov's theorem, the covariance of $\varepsilon_{\mathcal{X},t+1}^{\mathbb{Q}}$ is $\Sigma_{\mathcal{X}}$, i.e., the risk adjustment only affects the expected values.

The asset-pricing condition (31) can similarly be re-written in terms of the risk-neutral expected value. Using $E_t^{\mathbb{P}}\left(m_{t+1}\right) + \frac{1}{2}Var_t\left(m_{t+1}\right) = -f_t^1$ (the one-period risk-free bond satisfies the pricing condition), and $E_t^{\mathbb{Q}}\left(r_{t+1}^m\right) = E_t^{\mathbb{P}}\left(r_{t+1}^m\right) + Cov_t\left(m_{t+1}, r_{t+1}^m\right)$, (31) takes a simpler form

$$E_t^{\mathbb{Q}}\left(r_{t+1}^m\right) - f_t^1 + \frac{1}{2}Var_t\left(r_{t+1}^m\right) = 0. \tag{37}$$

This condition is nothing else then the requirement that under the risk-neutral measure, the expected return on every bond must equal the risk-free rate, up to the convexity adjustment due to the log-Normal formulation.

The generic no-arbitrage GDTSM can now be stated as a collection of three equations

$$\mathcal{X}_{t+1} = \mu_{\mathcal{X}}^{\mathbb{P}} + A_{\mathcal{X}}^{\mathbb{P}}\mathcal{X}_t + \varepsilon_{\mathcal{X},t+1}^{\mathbb{P}} \tag{38}$$

$$\mathcal{X}_{t+1} = \mu_{\mathcal{X}}^{\mathbb{Q}} + A_{\mathcal{X}}^{\mathbb{Q}}\mathcal{X}_t + \varepsilon_{\mathcal{X},t+1}^{\mathbb{Q}} \tag{39}$$

$$f_t^1 = f_0^1 + f_1^{1\prime}\mathcal{X}_t. \tag{40}$$

Since the $\mathbb{Q}$ dynamics and the risk-neutral pricing condition (37) simply follow from re-defining the expected value, the bond prices and forward rates implied by the model must be exactly of the form (32)-(33), and (34)-(35). The latter claim can also be confirmed directly, by substituting the conjectured form of the bond pricing formula into (37), and using (39) to compute the expectations.

Appendix B Invariant Transformations

This appendix explains the idea of invariant transformations, and shows that two models are observationally equivalent (i.e., share the same predictions for the term-structure observables) whenever there exists an invariant transformation between their respective sets of factors.

Start with a model in the generic form

$$
\begin{aligned}
\mathcal{X}_{t+1} &= \mu^{\mathbb{P}}_{\mathcal{X}} + A^{\mathbb{P}}_{\mathcal{X}} \mathcal{X}_t + \varepsilon^{\mathbb{P}}_{\mathcal{X},t+1}, \\
\mathcal{X}_{t+1} &= \mu^{\mathbb{Q}}_{\mathcal{X}} + A^{\mathbb{Q}}_{\mathcal{X}} \mathcal{X}_t + \varepsilon^{\mathbb{Q}}_{\mathcal{X},t+1}, \\
f^1_t &= f^1_{0,\mathcal{X}} + f^1_{1,\mathcal{X}} \mathcal{X}_t,
\end{aligned}
$$

with shock covariance matrix $\Sigma_{\mathcal{X}}$.

Assume that there exists a vector of factors $\mathcal{Y}_t$, related to $\mathcal{X}_t$ by an invertible affine transformation (called invariant transformation) $\mathcal{Y}_t = \alpha + \beta \mathcal{X}_t$. It follows that $\mathcal{X}_t = \beta^{-1}(\mathcal{Y}_t - \alpha)$. Substituting this into the original model, we obtain a representation in terms of the new factors,

$$
\begin{aligned}
\mathcal{Y}_{t+1} &= \mu^{\mathbb{P}}_{\mathcal{Y}} + A^{\mathbb{P}}_{\mathcal{Y}} \mathcal{Y}_t + \varepsilon^{\mathbb{P}}_{\mathcal{Y},t+1}, \\
\mathcal{Y}_{t+1} &= \mu^{\mathbb{Q}}_{\mathcal{Y}} + A^{\mathbb{Q}}_{\mathcal{Y}} \mathcal{Y}_t + \varepsilon^{\mathbb{Q}}_{\mathcal{Y},t+1}, \\
f^1_t &= f^1_{0,\mathcal{Y}} + f^1_{1,\mathcal{Y}} \mathcal{Y}_t.
\end{aligned}
$$

The relationships between the new and old model parameters are

$$
\begin{aligned}
A^{\mathbb{X}}_{\mathcal{Y}} &= \beta A^{\mathbb{X}}_{\mathcal{X}} \beta^{-1}, \quad \text{for } \mathbb{X} \in \{\mathbb{P}, \mathbb{Q}\} \\
\mu^{\mathbb{X}}_{\mathcal{Y}} &= \beta \mu^{\mathbb{X}}_{\mathcal{X}} + (I - A^{\mathbb{X}}_{\mathcal{Y}})\alpha, \quad \text{for } \mathbb{X} \in \{\mathbb{P}, \mathbb{Q}\} \\
f^1_{1,\mathcal{Y}} &= f^1_{1,\mathcal{X}} \beta^{-1}, \\
f^1_{0,\mathcal{Y}} &= f^1_{0,\mathcal{X}} - f^1_{1,\mathcal{Y}} \alpha, \\
\Sigma_{\mathcal{Y}} &= \beta \Sigma_{\mathcal{X}} \beta'.
\end{aligned}
$$

Similarly, if the bond prices (or other term structure observables) are affine in $\mathcal{X}$ with $b_t =$

$f^1_{0,\mathcal{X}} + f^1_{1,\mathcal{X}}\mathcal{X}_t$, then they are also affine in the transformed state vector $\mathcal{Y}$, with coefficients

$$f^1_{1,\mathcal{Y}} = f^1_{1,\mathcal{X}}\beta^{-1},$$
$$f^1_{0,\mathcal{Y}} = f^1_{0,\mathcal{X}} - f^1_{1,\mathcal{Y}}\alpha.$$

By construction, these formulas for the bond prices are consistent with no-arbitrage in both models. Both formulations are therefore observationally equivalent.

Appendix C 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 $\mathcal{Y}_t$, and the transition matrix under $\mathbb{Q}$ is in companion form (22). The state-space representation of the model is

$$\mathcal{Y}_{t+1} = \mu^{\mathbb{P}}_{\mathcal{Y}} + A^{\mathbb{P}}_{\mathcal{Y}}\mathcal{Y}_t + \varepsilon^{\mathbb{P}}_{\mathcal{Y},t+1}, \tag{41}$$
$$b^o_t = B_0 + B_1\mathcal{Y}_t + v_t, \tag{42}$$

where b^o_t 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 $\varepsilon^{\mathbb{P}}_{\mathcal{Y},t+1}$. The covariance matrix of $\varepsilon^{\mathbb{P}}_{\mathcal{Y},t+1}$ is $\Sigma_{\mathcal{Y}}$.

The Kalman filter iteratively estimates the forward rates in $\mathcal{Y}_t$ based on the history of observed bond prices. Define $\hat{\mathcal{Y}}_t \equiv E(\mathcal{Y}_t|b^o_1, \ldots, b^o_{t-1})$ as the (prior) estimate of the state vector, and $\hat{\Sigma}_t \equiv E[(\mathcal{Y}_t - \hat{\mathcal{Y}}_t)(\mathcal{Y}_t - \hat{\mathcal{Y}}_t)']$ as the matrix measuring the uncertainty in the estimate. At every point in time the filter computes the innovation $a_t \equiv b^o_t - B_0 - B_1\hat{\mathcal{Y}}_t$, and uses it to update the current state, and to form next period's prior $\hat{\mathcal{Y}}_{t+1}$, according to the transition equation (41). 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.[42]

Suppose one knows the initial (multivariate Gaussian) distribution of the state vector, $\mathcal{Y}_1 \sim$

[42] A very good presentation of the Kalman filter is contained in Ljungqvist and Sargent (2012), p. 56.

$N(\hat{\mathcal{Y}}_1, \hat{\Sigma}_1)$, and observes a sample of bond prices $b_1^o, \ldots, b_T^o$. The filter equations are:

$$a_t = b_t^o - B_0 - B_1 \hat{\mathcal{Y}}_t, \tag{43}$$

$$\hat{K}_t = A_{\mathcal{Y}}^{\mathbb{P}} \hat{\Sigma}_t B_1' (B_1 \hat{\Sigma}_t B_1' + R)^{-1}, \tag{44}$$

$$\hat{\mathcal{Y}}_{t+1} = \mu_{\mathcal{Y}}^{\mathbb{P}} + A_{\mathcal{Y}}^{\mathbb{P}} \hat{\mathcal{Y}}_t, \tag{45}$$

$$\hat{\Sigma}_{t+1} = \Sigma_{\mathcal{Y}} + \hat{K}_t R \hat{K}_t' + (A_{\mathcal{Y}}^{\mathbb{P}} - \hat{K}_t B_1) \hat{\Sigma}_t (A_{\mathcal{Y}}^{\mathbb{P}} - \hat{K}_t B_1)'. \tag{46}$$

In the current application, the filter can be initiated with the first N_f 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 (44) and (46) 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 (43), and predicting the state (45).[43]

The likelihood of observing the data for a fixed set of model parameters is the same as the likelihood of observing a sequence of innovations a_t. For a Gaussian model with N_f factors, the log-likelihood function implied by the stationary filter is

$$\mathcal{L} = -\frac{TN_f}{2} \log 2\pi - \frac{T}{2} \log |\Omega| - \frac{1}{2} \sum_{t=1}^{T} a_t \Omega^{-1} a_t', \tag{47}$$

where Ω is the covariance matrix of innovations, $\Omega = R + B_1 \hat{\Sigma} B_1'$ (as is evident from (43)).

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 Estimation Under Observable Factors

This appendix discusses GDTSM estimation under the assumption that there exist exactly N_f 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

[43] Anderson et al. (1996) summarize the conditions under which the steady-state filter can be applied.

companion-form parametrization defined in Propositions 1 and 2 is

$$\mathcal{Y}_{t+1} = \mu_{\mathcal{Y}}^{\mathbb{P}} + A_{\mathcal{Y}}^{\mathbb{P}} \mathcal{Y}_t + \varepsilon_{\mathcal{Y},t+1}^{\mathbb{P}}, \tag{48}$$

$$b_t^o = B_0 + B_1 \mathcal{Y}_t + v_t. \tag{49}$$

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

Assumption 3 *For a N_f-factor model, there exist exactly N_f a priori known linear combinations of observed bond prices that uncover the factors, i.e., there exists a matrix W such that*

$$W v_t = 0, \tag{50}$$

Without loss of generality, it can be assumed that the rows of W have unit norm, and that they are linearly independent.[44] One can define the bond portfolios given by W as $\mathcal{P}_t \equiv W b_t^o$. In applications, W is usually chosen as the PCA loading matrix, corresponding to the scores of maximum unconditional variance, and the portfolios $\mathcal{P}_t$ correspond to the PCA factor scores.[45]

By Assumption 3, the observation equation (49) can be used to measure the factors in every given sample,

$$\mathcal{P}_t = W b_t^o = W B_0 + (W B_1) \mathcal{Y}_t. \tag{51}$$

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 3, 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

[44] 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.

[45] 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.

representation of the rotated model,

$$\mathcal{P}_{t+1} = \mu_{\mathcal{P}}^{\mathbb{P}} + A_{\mathcal{P}}^{\mathbb{P}}\mathcal{P}_t + \varepsilon_{\mathcal{P},t+1}^{\mathbb{P}}, \tag{52}$$

$$b_t^{\perp} = H_0 + H_1\mathcal{P}_t + v_t^{\perp}, \tag{53}$$

where the parameters are related to those of the original model through

$$\mu_{\mathcal{P}}^{\mathbb{P}} = WB_1\mu_{\mathcal{Y}}^{\mathbb{P}} + [I - (WB_1)A_{\mathcal{Y}}^{\mathbb{P}}(WB_1)^{-1}], \tag{54}$$

$$A_{\mathcal{P}}^{\mathbb{P}} = (WB_1)A_{\mathcal{Y}}^{\mathbb{P}}(WB_1)^{-1}, \tag{55}$$

$$H_0 = W^{\perp}B_0 - W^{\perp}B_1(WB_1)^{-1}WB_0, \tag{56}$$

$$H_1 = W^{\perp}B_1(WB_1)^{-1}. \tag{57}$$

The covariance matrices of the random terms in (52) and (53) are, respectively

$$\Sigma_{\mathcal{P}} \equiv (WB_1)\Sigma_{\mathcal{Y}}(WB_1)', \tag{58}$$

$$R^{\perp} \equiv W^{\perp}R(W^{\perp})'. \tag{59}$$

Since the factors in the new transition equation (52) 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 $\mu_{\mathcal{P}}^{\mathbb{P}}$, and $A_{\mathcal{P}}^{\mathbb{P}}$ coincide with their OLS estimates, and do not depend on the covariance matrix $\Sigma_{\mathcal{P}}$, 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 θ_μ and θ_A determine the conditional expectation in (48), θ_Σ is the vectorized triangular matrix in the Cholesky decomposition of $\Sigma_{\mathcal{Y}}$, μ and c parametrize the conditional $\mathbb{Q}$ dynamics of factors $\mathcal{Y}_t$, and θ_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 θ_μ, and θ_A.

Given the equivalence of the state-space representations (48)-(49), and (52)(53) under Assumption 3, the probability of observing a given sample of bond prices $\{b_t^o\}_{t\in(1,\ldots,T)}$, conditional on the parameters in Θ, can be factored into

$$\text{prob}(\{b_t^o\}|\Theta) = \text{prob}(\{\mathcal{P}_t\}|\Theta) \times \text{prob}(\{b_t^\perp\}|\{\mathcal{P}_t\};\Theta). \tag{60}$$

In light of (52), 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 2π part, is

$$\mathcal{L}_1 \equiv -\frac{(T-1)}{2}\log|\Sigma_{\mathcal{P}}| - \frac{1}{2}\sum_{t=1}^{T-1}(\mathcal{P}_{t+1} - \mu_{\mathcal{P}}^{\mathbb{P}} - A_{\mathcal{P}}^{\mathbb{P}}\mathcal{P}_t)\Sigma_{\mathcal{P}}^{-1}(\mathcal{P}_{t+1} - \mu_{\mathcal{P}}^{\mathbb{P}} - A_{\mathcal{P}}^{\mathbb{P}}\mathcal{P}_t)', \tag{61}$$

where $\Sigma_{\mathcal{P}}$ is given in (58). As noted above, $\mu_{\mathcal{P}}^{\mathbb{P}}$ and $A_{\mathcal{P}}^{\mathbb{P}}$ 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 $\Sigma_{\mathcal{P}}$, i.e., $\Theta_1 \equiv \{\theta_\Sigma, c\}$, as is evident from (58).

The other term in the factorization (60) is the probability of observing the sequence of T realizations of $v_t^\perp$, for given $\mathcal{P}_t$, and model parameters that determine H_0, H_1, and $R^\perp$, as indicated by (53). The log of this probability is (again ignoring the constant part)

$$\mathcal{L}_2 \equiv -\frac{T}{2}\log|R^\perp| - \frac{1}{2}\sum_{t=1}^{T}(b_t^\perp - H_0 - H_1\mathcal{P}_t)(R^\perp)^{-1}(b_t^\perp - H_0 - H_1\mathcal{P}_t)'. \tag{62}$$

Equations (56), (57), and (59) 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.[46] Under the companion-form parametrization, the latter only depends on c. To find B_0, one first needs to find the risk-neutral drift $\mu_{\mathcal{Y}}^{\mathbb{Q}}$, which according to Proposition 2 depends on parameters θ_Σ, and μ. 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 (61) and (62), and the corresponding parameter set is $\Theta_1 \cup \Theta_2$, which does not depend on θ_μ and θ_A.

[46] Consistent with Assumption 3, 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 `null`).

Appendix E Tables

Table 1: Parameter values estimated by Maximum Likelihood (with Kalman filtering) for a three-factor model, using the full sample of Fama-Bliss bonds (June 1956 – July 2015).

A. Monthly Dynamics ($\mathbb{P}$)							
$\mu_{\mathcal{Y}}^{\mathbb{P}}$		$A_{\mathcal{Y}}^{\mathbb{P}}$				$\Sigma_{\mathcal{Y}}$	
0.0003	0.8802	0.1952	-0.0861	1e-5 $\times$	0.2043	0.1432	0.1009
0.0007	-0.0695	1.1786	-0.1211		0.1432	0.1242	0.1070
0.0012	-0.0999	0.3282	0.7563		0.1009	0.1070	0.1075
B. Annualized Dynamics ($\mathbb{P}$)							
$\mu_{\mathcal{Y}}^{\mathbb{P}}$		$A_{\mathcal{Y}}^{\mathbb{P}}$				$\Sigma_{\mathcal{Y}}$	
0.0039	0.0476	1.4826	-0.6447	1e-4 $\times$	0.1950	0.1544	0.1241
0.0071	-0.4559	2.1303	-0.7962		0.1544	0.1394	0.1244
0.0102	-0.5844	2.0422	-0.6007		0.1241	0.1244	0.1198
C. Annual Dynamics ($\mathbb{Q}$)							
$\mu_{\mathcal{Y}}^{\mathbb{Q}}$		$A_{\mathcal{Y}}^{\mathbb{Q}}$				$\Sigma_{\mathcal{Y}}$	
0.0000	0	1	0	1e-4 $\times$	0.1950	0.1544	0.1241
0.0000	0	0	1		0.1544	0.1394	0.1244
0.0074	-1.7034	3.8287	-1.2272		0.1241	0.1244	0.1198

D. Pricing Errors, and Noise Std

RMSE (b.p.)					Noise St.Dev. (b.p.)
15.3	12.6	19.4	10.8	9.5	19.3
MAPE (b.p.)					
11.8	9.2	14.2	7.0	6.8	

E. Eigenvalues

Monthly $\mathbb{P}$ Dynamics		
0.991	0.929	0.895
Annualized $\mathbb{P}$ Dynamics		
0.899	0.415	0.262
Annual $\mathbb{Q}$ Dynamics		
0.922	0.658	-2.807

Table 2: Parameter values estimated by Maximum Likelihood (with Kalman filtering) for a three-factor model, using the Cochrane and Piazzesi (2005) sample of Fama-Bliss bonds (January 1965 – December 2003).

A. Monthly Dynamics ($\mathbb{P}$)

$\mu_{\mathcal{Y}}^{\mathbb{P}}$	$A_{\mathcal{Y}}^{\mathbb{P}}$				$\Sigma_{\mathcal{Y}}$		
0.0017	0.8020	0.3022	-0.1354	1e-5 $\times$	0.4161	0.2841	0.2066
0.0015	-0.0730	1.0762	-0.0278		0.2841	0.2286	0.1777
0.0020	-0.2153	0.5527	0.6342		0.2066	0.1777	0.1586

B. Annualized Dynamics ($\mathbb{P}$)

$\mu_{\mathcal{Y}}^{\mathbb{P}}$	$A_{\mathcal{Y}}^{\mathbb{P}}$				$\Sigma_{\mathcal{Y}}$		
0.0157	-0.0340	0.8544	-0.0964	1e-4 $\times$	0.2946	0.2231	0.1791
0.0158	-0.4000	1.1158	0.0441		0.2231	0.1948	0.1714
0.0171	-0.5849	1.2286	0.1201		0.1791	0.1714	0.1618

C. Annual Dynamics ($\mathbb{Q}$)

$\mu_{\mathcal{Y}}^{\mathbb{Q}}$	$A_{\mathcal{Y}}^{\mathbb{Q}}$				$\Sigma_{\mathcal{Y}}$		
0.0000	0	1	0	1e-4 $\times$	0.2946	0.2231	0.1791
0.0000	0	0	1		0.2231	0.1948	0.1714
0.0072	-1.5063	3.6892	-1.2688		0.1791	0.1714	0.1618

D. Pricing Errors, and Noise St. Dev.

RMSE (b.p.)					Noise St.Dev. (b.p.)
14.7	12.6	18.6	10.5	9.0	18.7
MAPE (b.p.)					
11.2	9.4	13.3	6.6	6.6	

E. Eigenvalues

Monthly $\mathbb{P}$ Dynamics		
0.9814	0.9267	0.6042
Annualized $\mathbb{P}$ Dynamics		
0.798	0.401	0.002
Annual $\mathbb{Q}$ Dynamics		
0.9471	0.5707	-2.787

Table 3: Parameter values estimated by Maximum Likelihood (under the assumption of observable PCA factors) for a three-factor model, using the full sample of Fama-Bliss bonds (June 1956 – July 2015).

A. Monthly Dynamics of PCA factors ($\mathbb{P}$)							
$\mu_{\mathcal{P}}^{\mathbb{P}}$	$A_{\mathcal{P}}^{\mathbb{P}}$				$\Sigma_{\mathcal{P}}$		
-0.0044	0.9931	-0.1056	-0.1467	1e-3 $\times$	0.6881	0.0004	0.0038
-0.0011	0.0005	0.9441	0.0665		0.0004	0.0159	0.0005
0.0010	-0.0001	0.0039	0.5966		0.0038	0.0005	0.0069
B. Annualized Dynamics of PCA factors ($\mathbb{P}$)							
$\mu_{\mathcal{P}}^{\mathbb{P}}$	$A_{\mathcal{P}}^{\mathbb{P}}$				$\Sigma_{\mathcal{P}}$		
-0.0486	0.9185	-0.9145	-0.4668	1e-2 $\times$	0.7691	-0.0042	0.0005
-0.0085	0.0038	0.5029	0.0947		-0.0042	0.0111	0.0003
0.0024	-0.0002	0.0058	0.0032		0.0005	0.0003	0.0011
C. Annual Dynamics of Forward-Rate Factors ($\mathbb{Q}$)							
$\mu_{\mathcal{Y}}^{\mathbb{Q}}$	$A_{\mathcal{Y}}^{\mathbb{Q}}$				$\Sigma_{\mathcal{Y}}$		
0.0001	0	1	0	1e-3 $\times$	0.1903	0.1574	0.1325
0.0002	0	0	1		0.1574	0.1507	0.1420
0.0340	-8.120	20.412	-11.731		0.1325	0.1420	0.1434

D. Pricing Errors, and Noise St. Dev.

RMSE (b.p.)					Noise St.Dev. (b.p.)
12.9	10.3	16.6	10.7	1.3	17.9
MAPE (b.p.)					
10.0	7.6	11.8	7.8	0.9	

E. Eigenvalues

Monthly $\mathbb{P}$ Dynamics		
0.992	0.946	0.596
Annualized $\mathbb{P}$ Dynamics		
0.910	0.512	0.002
Annual $\mathbb{Q}$ Dynamics		
0.906	0.673	-13.310

Table 4: Parameter values estimated by Maximum Likelihood (under the assumption of observable PCA factors) for a three-factor model, using the Cochrane and Piazzesi (2005) sample of Fama-Bliss bonds (January 1965 – December 2003).

A. Monthly Dynamics of PCA factors ($\mathbb{P}$)							
$\mu_{\mathcal{P}}^{\mathbb{P}}$	$A_{\mathcal{P}}^{\mathbb{P}}$				$\Sigma_{\mathcal{P}}$		
-0.0080	0.9869	-0.1165	-0.8163	1e-4 $\times$	0.9519	-0.0271	0.0047
-0.0009	0.0008	0.9457	0.0413		-0.0271	0.0195	0.0022
0.0008	-0.0002	0.0003	0.4450		0.0047	0.0022	0.0084
B. Annualized Dynamics of PCA factors ($\mathbb{P}$)							
$\mu_{\mathcal{P}}^{\mathbb{P}}$	$A_{\mathcal{P}}^{\mathbb{P}}$				$\Sigma_{\mathcal{P}}$		
-0.0956	0.8506	-0.9695	-1.3475	1e-2 $\times$	1.0367	-0.0298	-0.0002
-0.0079	0.0068	0.5076	0.0330		-0.0298	0.0137	0.0004
0.0014	-0.0003	0.0006	0.0005		-0.0002	0.0004	0.0010
C. Annual Dynamics of Forward-Rate Factors ($\mathbb{Q}$)							
$\mu_{\mathcal{Y}}^{\mathbb{Q}}$	$A_{\mathcal{Y}}^{\mathbb{Q}}$				$\Sigma_{\mathcal{Y}}$		
0.000	0	1	0	1e-3 $\times$	0.3060	0.2331	0.1894
0.000	0	0	1		0.2331	0.1984	0.1766
0.004	-3.285	8.296	-4.059		0.1894	0.1766	0.1676

D. Pricing Errors, and Noise Std

RMSE (b.p.)					Noise St.Dev. (b.p.)
14.0	11.2	17.3	8.6	0.6	18.6
MAPE (b.p.)					
10.9	8.4	11.8	6.0	0.5	

E. Eigenvalues

Monthly $\mathbb{P}$ Dynamics		
0.985	0.948	0.445
Annualized $\mathbb{P}$ Dynamics		
0.831	0.528	0.000
Annual $\mathbb{Q}$ Dynamics		
0.982	0.594	-5.635

Table 5: Root mean squared errors (RMSE; in basis points) for out-of-sample forecasts at various horizons in 300 simulated data sets, obtained by three different methods (ML with Kalman filtering, ML with observable factors, and random walk). The data were generated by a three-factor model with parameters as in Table 1. Every artificial panel of bonds contains 758 monthly observations on 5 annual maturities (as in the full Fama-Bliss data set), plus 60 months used to compute the forecast errors.

A. RMSE (KF)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	16	25	34	48	61	70	80	82
2y	26	43	61	88	113	133	151	159
3y	34	58	85	122	160	191	217	233
4y	44	75	110	156	205	248	280	305
5y	52	89	131	183	244	296	335	369

B. RMSE (ML-OF)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	16	26	36	49	61	69	77	78
2y	26	43	62	88	113	131	146	151
3y	34	58	86	122	160	189	210	220
4y	44	75	110	156	205	245	271	288
5y	53	90	132	183	244	293	325	349

C. RMSE (RW)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	23	31	39	54	66	75	83	87
2y	31	45	63	89	114	134	150	160
3y	37	60	86	123	162	190	215	234
4y	46	75	110	156	208	247	278	306
5y	55	92	135	184	251	298	336	374

Table 6: Relative gains (in percents) for out-of-sample forecasts presented in Table 5. Each number is computed as the difference between two RMSEs corresponding to two different forecasting models, relative to the RMSE of the reference model.

A. RMSE, Gain of KF Over RW (%)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	32.6	18.9	12.2	10.2	8.2	6.0	4.6	5.1
2y	15.8	4.4	2.0	1.7	0.9	1.0	-0.5	0.5
3y	9.3	2.8	0.3	0.5	1.2	-0.4	-0.9	0.5
4y	3.6	0.9	-0.5	-0.1	1.5	-0.3	-0.9	0.5
5y	4.8	2.8	2.4	0.5	2.6	0.7	0.1	1.3

B. RMSE, Gain of ML-OF Over RW (%)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	29.8	17.1	8.9	9.3	8.2	7.9	7.9	10.4
2y	15.4	3.9	0.3	1.0	0.7	2.4	2.7	5.8
3y	9.3	2.9	-0.2	0.3	1.1	0.9	2.3	5.8
4y	3.8	0.6	-0.7	-0.4	1.6	0.9	2.6	5.9
5y	3.6	2.0	2.3	0.2	2.6	1.6	3.2	6.5

C. RMSE, Gain of KF Over ML-OF (%)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	4.0	2.1	3.6	1.1	0.0	-2.0	-3.6	-5.9
2y	0.5	0.6	1.7	0.7	0.3	-1.4	-3.3	-5.7
3y	0.0	-0.1	0.4	0.3	0.1	-1.3	-3.3	-5.6
4y	-0.3	0.3	0.2	0.3	-0.1	-1.2	-3.6	-5.8
5y	1.3	0.8	0.2	0.4	0.0	-1.0	-3.3	-5.6

Table 7: Mean absolute errors (MAE; in basis points) for out-of-sample forecasts at various horizons in 300 simulated data sets, obtained by three different methods (ML with Kalman filtering, ML with observable factors, and random walk). The data were generated by a three-factor model with parameters as in Table 1. Every artificial panel of bonds contains 758 monthly observations on 5 annual maturities (as in the full Fama-Bliss data set), plus 60 months used to compute the forecast errors.

A. MAE (KF)

	1m	3m	6m	1y	2y	3y	4y	5y
1y	13	20	27	39	50	56	62	65
2y	21	35	50	70	92	106	119	126
3y	27	46	69	97	130	153	171	183
4y	35	59	89	125	166	199	222	239
5y	41	70	106	146	198	238	265	290

B. MAE (ML-OF)

	1m	3m	6m	1y	2y	3y	4y	5y
1y	13	21	29	40	50	55	61	63
2y	21	34	50	71	93	105	116	121
3y	27	46	70	98	130	151	168	176
4y	35	58	90	125	167	197	217	229
5y	41	70	106	146	198	236	260	277

C. MAE (RW)

	1m	3m	6m	1y	2y	3y	4y	5y
1y	18	25	31	42	53	59	64	70
2y	24	36	51	70	92	107	119	129
3y	30	48	69	96	131	153	170	187
4y	36	58	88	122	168	200	219	247
5y	44	72	109	143	200	241	265	301

Table 8: Relative gains (in percents) for out-of-sample forecasts presented in Table 7. Each number is computed as the difference between two MAEs corresponding to two different forecasting models, relative to the MAE of the reference model.

A. MAE, Gain of KF Over RW (%)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	31.3	18.5	11.3	7.4	6.7	6.1	3.0	6.2
2y	14.2	4.3	3.2	0.2	0.5	0.7	0.0	2.1
3y	9.4	2.7	0.2	-1.1	0.7	-0.4	-0.7	2.1
4y	3.5	-0.4	-0.9	-2.3	1.3	0.3	-1.1	3.1
5y	6.9	2.1	3.3	-2.1	1.3	1.4	0.0	3.7

B. MAE, Gain of ML-OF Over RW (%)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	29.8	16.5	7.4	5.6	6.0	8.1	5.4	9.9
2y	14.0	4.7	1.5	-0.7	-0.2	1.9	2.1	6.0
3y	9.3	2.9	-0.6	-1.6	0.3	0.8	1.3	6.1
4y	3.0	-0.1	-1.4	-2.4	1.1	1.3	1.1	7.1
5y	5.2	1.7	2.8	-2.3	1.3	2.2	2.2	7.9

C. MAE, Gain of KF Over ML-OF (%)								
	1m	3m	6m	1y	2y	3y	4y	5y
1y	2.2	2.4	4.2	1.8	0.7	-2.2	-2.5	-4.1
2y	0.1	-0.4	1.8	0.9	0.7	-1.2	-2.2	-4.2
3y	0.1	-0.3	0.8	0.5	0.3	-1.2	-2.0	-4.3
4y	0.5	-0.3	0.5	0.1	0.2	-1.0	-2.3	-4.3
5y	1.8	0.4	0.5	0.2	0.0	-0.9	-2.2	-4.6

Appendix F   Figures

Figure 1: Forward rates in the Fama-Bliss data set (June 1956 – July 2015). The first sub-graph displays the historical values, the second shows the model-implied values obtained from a three-factor model with Kalman filtering (KF), and the third plots the fitted values from the model estimated by Maximum Likelihood under the assumption of observable factors (ML-OF).

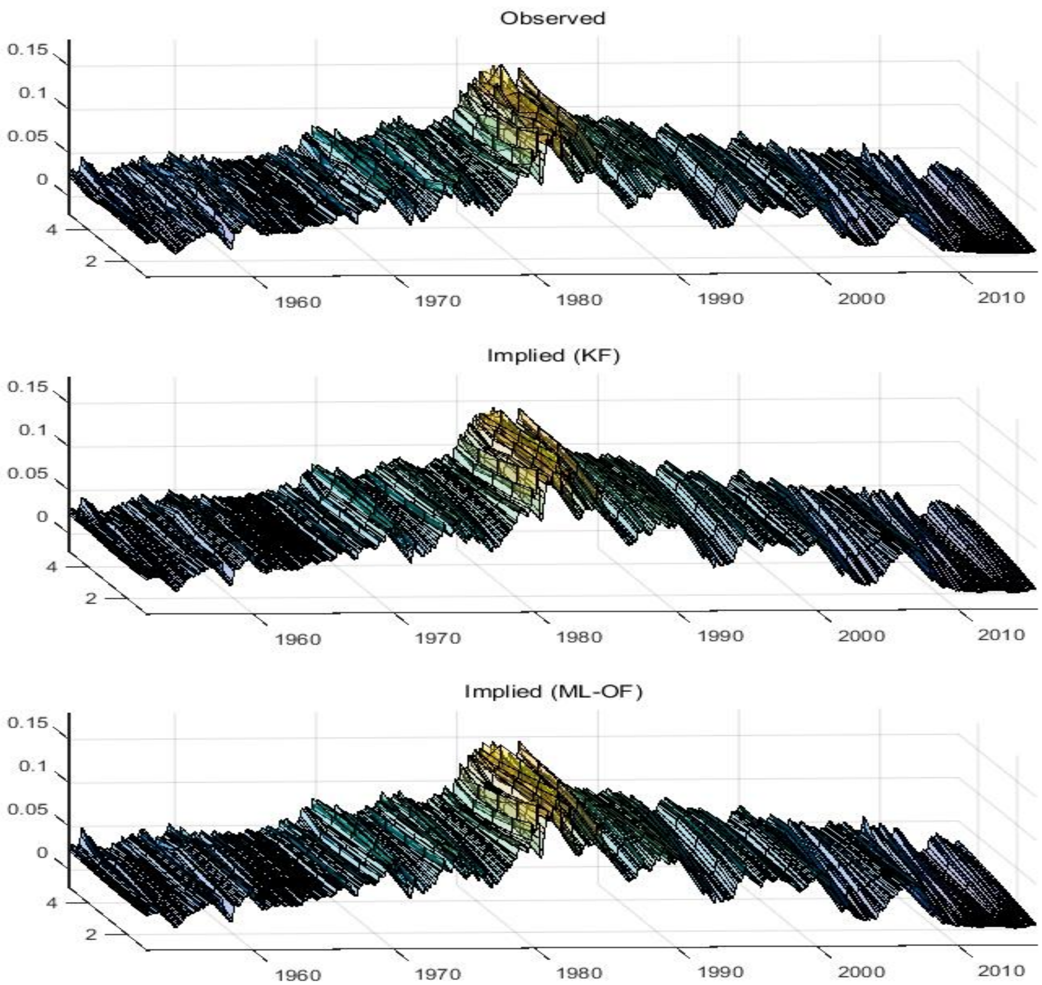


Figure 2: Factor values under the canonical companion form. The solid lines correspond to the actual rates (measured with errors). The dashed lines are obtained by the Kalman-filter. The dotted lines were found by inverting the model-implied relationship between the canonical factors and the observed PCA scores. I shift the series by 50 b.p. for greater clarity.

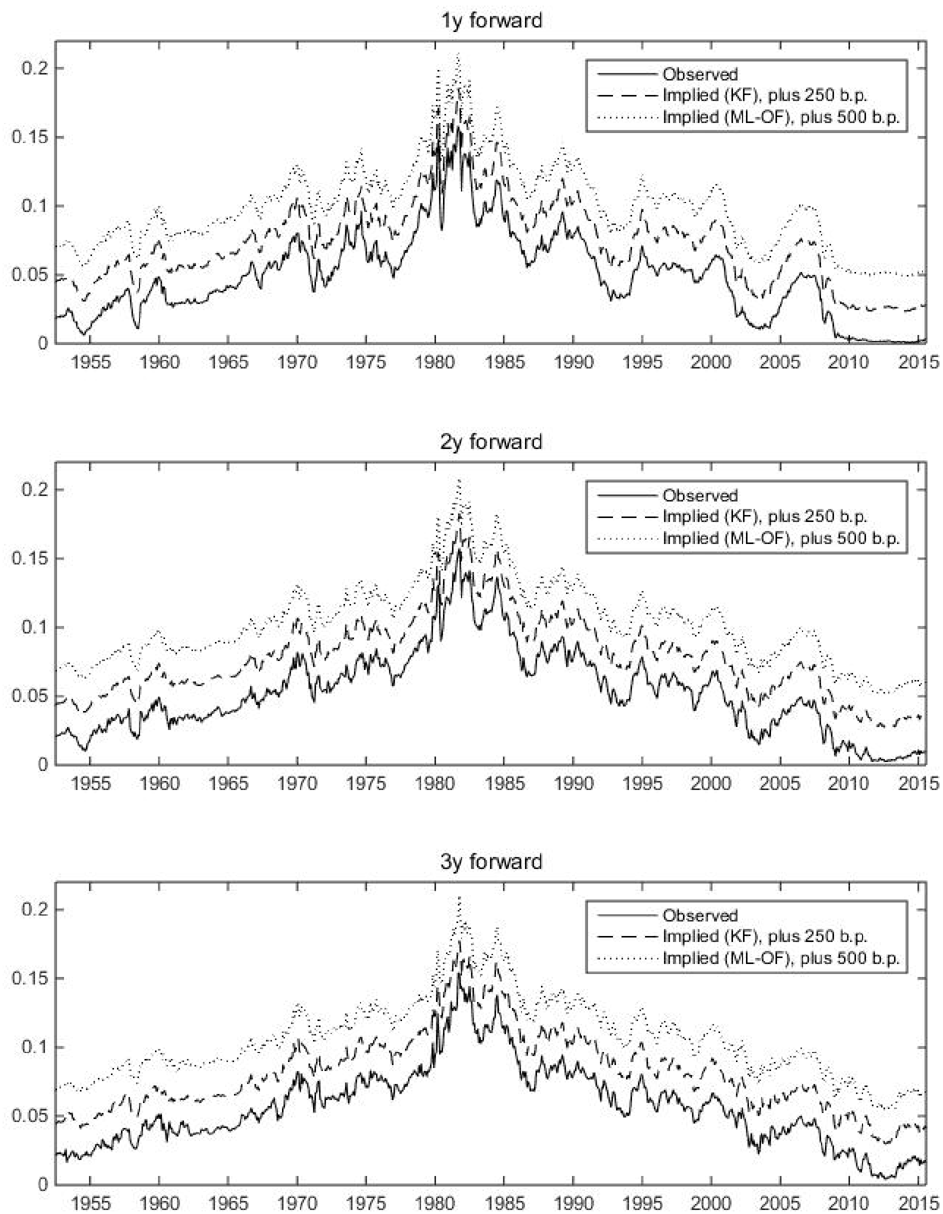
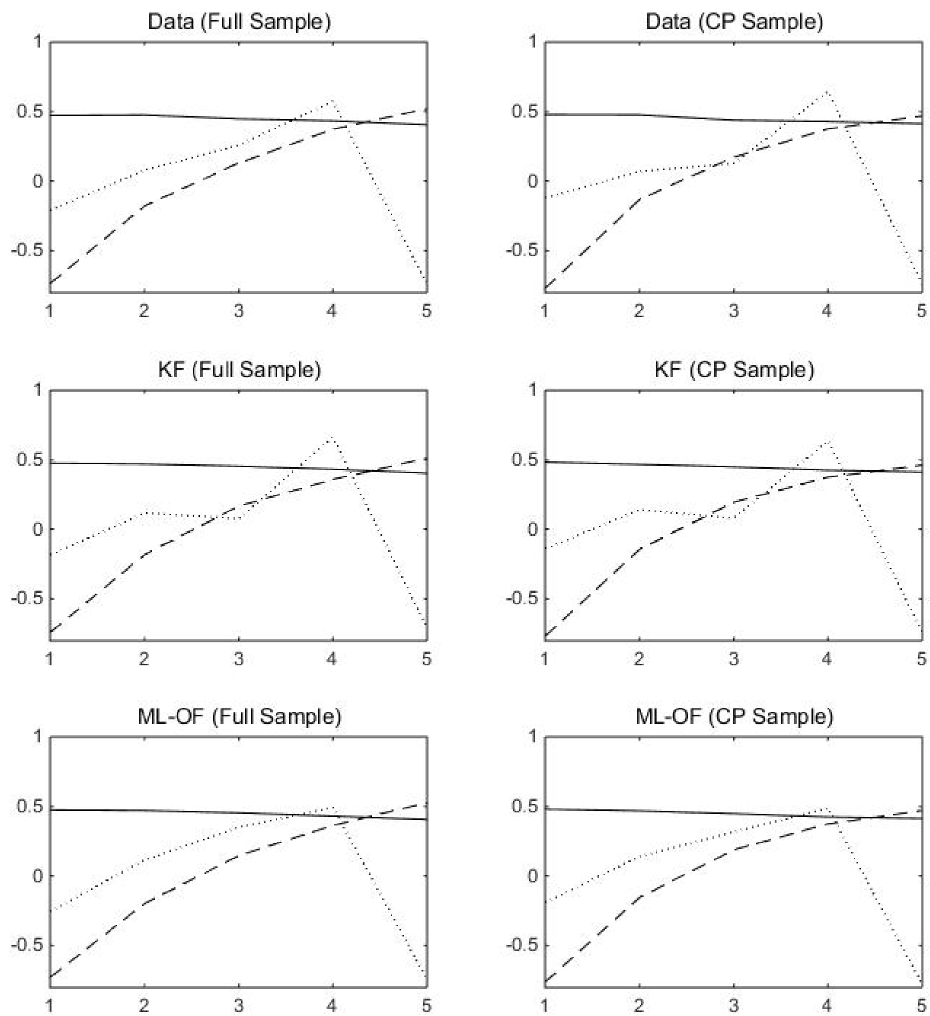


Figure 3: Principal component factor loadings, estimated in the whole Fama-Bliss sample (June 1956 – July 2015; left), and in the sample of Cochrane and Piazzesi (2005) (January 1965 – December 2003; right). The first row shows the results of empirical PCA. The middle row displays the loadings implied by the model estimated by ML, with Kalman filtering. The bottom row contains the loadings with respect to the empirical PCA factors, but consistent a no-arbitrage model (estimated by ML under the assumption of observable factors).



References

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

Aït-Sahalia, Y. and Kimmel, R. L. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions. *Journal of Financial Economics*, 98(1):113–144.

Anderson, E. W., McGrattan, E. R., Hansen, L. P., and Sargent, T. J. (1996). Mechanics of forming and estimating dynamic linear economies. *Handbook of computational economics*, 1:171–252.

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

Atiyah, M. F. and Macdonald, I. G. (1969). *Introduction to commutative algebra*, volume 2. Addison-Wesley Reading.

Babbs, S. H. and Nowman, B. K. (1999). Kalman filtering of generalized vasicek term structure models. *Journal of Financial and Quantitative Analysis*, 34(1).

Backus, D. K. and Zin, S. E. (1994). Reverse engineering the yield curve. Technical report, National Bureau of Economic Research.

Birkhoff, G. and Mac Lane, S. (1966). *A survey of modern algebra*. Universities Press.

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

Chen, R.-R. and Scott, L. (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *The Journal of Fixed Income*, 3(3):14–31.

Cheridito, P., Filipović, D., and Kimmel, R. L. (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics*, 83(1):123–170.

Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics*, 164(1):4–20.

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

Cochrane, J. H. and Piazzesi, M. (2009). Decomposing the yield curve. *SSRN Working Paper Series*.

Collin-Dufresne, P., Goldstein, R. S., and Jones, C. S. (2008). Identification of maximal affine term structure models. *The Journal of Finance*, 63(2):743–795.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, pages 385–407.

Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance*, 55(5):1943–1978.

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

Duarte, J. (2004). Evaluating an alternative risk preference in affine term structure models. *Review of Financial Studies*, 17(2):379–404.

Duffee, G. (2012). Forecasting interest rates. *Handbook of economic forecasting*, 2:385–426.

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

Duffee, G. R. (2011a). Forecasting with the term structure: The role of no-arbitrage restrictions. *Johns Hopkins University, Department of Economics Working Paper*.

Duffee, G. R. (2011b). Information in (and not in) the term structure. *Review of Financial Studies*, 24(9):2895–2934.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, 6(4):379–406.

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

Filipović, D., Larsson, M., and Trolle, A. B. (2015). Linear-rational term structure models. *Swiss Finance Institute Research Paper*, 14-15.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, 54(8):2291–2304.

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

Joslin, S., Le, A., and Singleton, K. J. (2013). Gaussian macro-finance term structure models with lags. *Journal of Financial Econometrics*, 11(4):581–609.

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

Joslin, S., Singleton, K., and Zhu, H. (2010). Supplement to *A New Perspective on Gaussian DTSMs*. In *Sloan School working paper*. MIT.

Joslin, S., Singleton, K. J., and Zhu, H. (2011). A new perspective on gaussian dynamic term structure models. *Review of Financial Studies*, 24(3):926–970.

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147.

Le, A., Singleton, K. J., and Dai, Q. (2010). Discrete-time affine$^{\mathbb{Q}}$ term structure models with generalized market prices of risk. *Review of Financial Studies*, 23(5):2184–2227.

Leippold, M. and Wu, L. (2002). Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis*, 37(02):271–295.

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

Ljungqvist, L. and Sargent, T. J. (2012). *Recursive macroeconomic theory*. MIT press.

McCulloch, J. H. (1975). The tax-adjusted yield curve. *The Journal of Finance*, 30(3):811–830.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, pages 473–489.

Radwański, J. F. (2010). Understanding bond risk premia uncovered by the term structure. *Vienna Graduate School of Finance Working Paper*.

Svensson, L. E. (1995). Estimating forward interest rates with the extended nelson & siegel method. *Sveriges Riksbank Quarterly Review*, 3(1):13–26.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, 57(298):348–368.