A New Perspective on Gaussian DTSMs

Scott Joslin[*], Kenneth J. Singleton[†] and Haoxiang Zhu[‡]

First version: November 9, 2008
This version: October 3, 2009

Abstract

We develop a canonical Gaussian dynamic term structure model (*GDTSM*) in which the pricing factors are observable portfolios of yields and, within this framework, offers several new perspectives on *GDTSM*s. When the pricing factors are measured without error our *GDTSM* has the feature that ordinary least squares gives the maximum likelihood estimates of the physical distribution of the pricing factors. It follows immediately that no-arbitrage pricing does not restrict the historical distribution of bond yields. Moreover, regardless of any constraints imposed on the risk neutral dynamics, we show that otherwise unrestricted *GDTSM*s induce the same physical distribution of bond yields. Additionally, using our canonical model, standard optimization algorithms converge to the global optimum of the likelihood function almost instantaneously, even in the presence of measurement error for all yields. These insights are extended to a class of *GDTSM*s with macro variables as pricing factors. Empirical examples and comparisons of the out-of-sample forecasting performance of various *GDTSM*s are also presented for U.S. Treasury yields.

[*]Sloan School of Management, MIT, sjoslin@mit.edu
[†]Graduate School of Business, Stanford University, singleton_ken@gsb.stanford.edu
[‡]Graduate School of Business, Stanford University, haoxiang.zhu@stanford.edu.

1 Introduction

The Gaussian dynamic term structure model (*GDTSM*) has been a "work horse" for studying the properties of interest rate risk,[1] because it accommodates flexible specifications of feedback and correlations among the risk factors and it yields closed-form solutions for zero-coupon bond prices and their associated likelihood functions. This flexibility has given rise to a variety of different econometric representations of the risk factors in *GDTSM*s, including the canonical specifications in Dai and Singleton (2000) (DS), Collin-Dufresne, Goldstein and Jones (2008) (CGJ), and Joslin (2007), as well as the model Christensen, Diebold and Rudebusch (2007,2009) referred to as the arbitrage-free, Nelson-Siegel ($AFNS$) model.

This paper shows that there is an all-encompassing, canonical *GDTSM*(N) in which the N pricing factors $\mathcal{P}_t$ are observable yields on portfolios of bonds with known weights, such as the first N principal components of yields. Moreover, the risk-neutral ($\mathbb{Q}$) distribution of bond yields is shown to be fully characterized by the long-run mean of the short rate $r^{\mathbb{Q}}_\infty$ and the N-vector of eigenvalues $\lambda^{\mathbb{Q}}$ of the $\mathbb{Q}$-drift of $\mathcal{P}_t$. Importantly, $(r^{\mathbb{Q}}_\infty, \lambda^{\mathbb{Q}})$ constitutes a low-dimensional and rotation-invariant (and thus economically meaningful) parameter space over which researchers typically have informed priors.

A striking insight from our canonical form is that, when $\mathcal{P}_t$ is priced without error, ordinary least squares gives the maximum likelihood (ML) estimates of the physical distribution of $\mathcal{P}_t$.[2] It follows immediately that, *within any canonical GDTSM*, no-arbitrage pricing does not restrict the $\mathbb{P}$ distribution of bond yields (see also Duffee (2009)). Additionally, *regardless of the constraints imposed on the risk neutral dynamics of* $\mathcal{P}$, we show that otherwise unrestricted *GDTSM*s imply the same physical distribution of bond yields. It follows, in particular, that forecasts produced by an $AFNS$ model are identical to those from an unrestricted VAR.

When constraints are imposed on the structure of risk premia, no-arbitrage restrictions may affect the conditional distribution of bond yields. One such constraint is that expected excess returns on bonds lie in a space of dimension smaller than N. Motivated by the descriptive analysis of Cochrane and Piazzesi (2005, 2008) and Duffee (2008), and using the methods developed in Joslin, Priebsch and Singleton (2009b), we explore the restriction that expected excess returns lie in a space of dimension $\mathcal{L}\,(< N)$ *without restricting a priori which of the N pricing factors $\mathcal{P}_t$ represent priced risks.* If $\mathcal{L} < N$, then there are necessarily

[1] The arbitrage-free Gaussian model dates back at least to Vasicek (1977) and the multi-factor versions by Langetieg (1980). The formal characterization of affine $DTSM$s was subsequently extended by Duffie and Kan (1996), Dai and Singleton (2000), and Duffee (2002), among others. See Dai and Singleton (2003) and Piazzesi (2006) for recent surveys.

[2] Due to the nature of identification and observational equivalence, the fact that this result holds for our canonical form implies that it must hold for *any* canonical form; thus this is a generic feature of such models.

restrictions linking the physical and risk-neutral drifts of $\mathcal{P}_t$. We show that, in this case, the forecasts of future yields implied by a *GDTSM* are different than those from an unrestricted VAR – arbitrage-free pricing matters for forecasting.

We exploit the structure of our canonical *GDTSM* to obtain near-instantaneous convergence of standard search algorithms to the global optimum of our likelihood function. Fast convergence is achieved regardless of the number or risk factors or bond yields used in estimation. When, instead of having N bond positions priced perfectly by the *GDTSM*, $\mathcal{P}_t$ measures the "true" pricing factors with error and Kalman filtering is used, the same algorithm convergences nearly as fast, because our initial guesses for the parameter remain very accurate. To put this computational advantage into perspective one need read no further than Duffee and Stanton (2007) and Duffee (2009) who highlight numerous computational challenges and multiple local optima associated with their likelihood functions. For example, Duffee reports that each optimization for his parameterization of a three-factor model takes about two days. For the *GDTSM*(3) models examined in this paper, fit with roughly three times the number of observations used by Duffee and using Kalman filtering to accommodate measurement errors, convergence to the global optimum of the likelihood function was typically achieved in about ten seconds.[3]

The rapid convergence to global optima using our canonical *GDTSM* makes it feasible to explore the rolling out-of-sample forecasts from different *GDTSM*s. For a variety of *GDTSM*s, with and without exact pricing (that is, with and without filtering) and with and without constraints imposed on the dimension $\mathcal{L}$ of risk premiums, we compare the out-of-sample forecasting performance relative to a benchmark unconstrained VAR.

We also investigate the implications of expanding the number of pricing factors $\mathcal{P}_t$ in settings where all of the bonds are priced with errors. Using data on U.S. Treasury zero-coupon bond yields we find, consistent with many prior studies, that three pricing factors leads to a reasonable fit over the range of maturities examined. When the dimension of $\mathcal{P}_t$ is expanded to four or five, the filtered values of the true fourth and fifth PCs bear only weak resemblance to their observed sample counterparts. We conclude that the four- and five-factor models examined in Cochrane and Piazzesi (2008) and Duffee (2008) may be over parameterized and that, when filtering, some caution should be taken in interpreting factors as the higher-order PCs that are relevant for pricing. They may, instead, represent some other constructs that the likelihood function attempts to match through filtering.

Finally, we show that many of our theoretical results extend to a class of macro-*GDTSM*s

[3] Convergence for Duffee's model, expressed in its equivalent representation within our canonical *GDTSM*(3) and using his data, was obtained in less time. We obtained exactly the same likelihood function and parameter values as reported in his paper (to five decimal places) using MATLAB on a laptop computer.

in which some of the pricing factors are observable macro variables such as inflation or output.

2 A Canonical *GDTSM* with *Observable* Risk Factors

In this section we develop our "JSZ" canonical representation of *GDTSM*s. Towards this end we start with a generic representation of a *GDTSM* in which the discrete-time[4] evolution of the risk factors (state vector) $X_t \in \mathbb{R}^N$ is governed by the following equations:

$$\Delta X_t = K_{0X}^{\mathbb{P}} + K_{1X}^{\mathbb{P}} X_{t-1} + \Sigma_X \epsilon_t^{\mathbb{P}}, \tag{1}$$

$$\Delta X_t = K_{0X}^{\mathbb{Q}} + K_{1X}^{\mathbb{Q}} X_{t-1} + \Sigma_X \epsilon_t^{\mathbb{Q}}, \tag{2}$$

$$r_t = \rho_{0X} + \rho_{1X} \cdot X_t, \tag{3}$$

where r_t is the one-period spot interest rate, $\Sigma_X \Sigma_X{}'$ is the conditional covariance matrix of X_t, and $\epsilon_t^{\mathbb{P}}, \epsilon_t^{\mathbb{Q}} \sim N(0, I_N)$. A canonical *GDTSM* is one that is maximally flexible in its parameterization of both the $\mathbb{Q}$ and $\mathbb{P}$ distributions of X_t, subject only to normalizations that ensure econometric identification. Before formally deriving our canonical *GDTSM* we briefly outline the basic idea. Variations of our canonical form, as well as some of its key implications for model specification and analysis, are discussed subsequently.

Suppose that N zero-coupon bond yields or N linear combinations of such yields, $\mathcal{P}_t$, are priced perfectly by the model (subsequently we relax this assumption). By a slight abuse of nomenclature we will refer to these linear combinations of yields as portfolios of yields. Applying invariant transformations[5] we show (i) the pricing factors X_t in (3) can be replaced by the observable $\mathcal{P}_t$, and (ii) the $\mathbb{Q}$ distribution of $\mathcal{P}_t$ can be fully characterized by the parameters $\Theta_{\mathcal{P}}^{\mathbb{Q}} \equiv (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$, where $r_\infty^{\mathbb{Q}}$ is the risk-neutral long-run mean of the short rate r, $\lambda^{\mathbb{Q}}$ is the vector of eigenvalues of $K_{1X}^{\mathbb{Q}}$, and $\Sigma_{\mathcal{P}} \Sigma_{\mathcal{P}}'$ is the covariance of innovations to the portfolios of yields. The price of all coupon bonds (as well as interest rate derivatives) are now defined in terms of these observable pricing factors through no arbitrage. Importantly, though the pricing factors are now observable, the underlying parameter space of the $\mathbb{Q}$ distribution of $\mathcal{P}$ is still fully characterized by $\Theta_{\mathcal{P}}^{\mathbb{Q}}$.

Moreover, the parameters of the $\mathbb{P}$ distribution of the (newly rotated and observable) state vector $\mathcal{P}_t$ are $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ along with $\Sigma_{\mathcal{P}}$. It turns out that, in the absence of restrictions linking $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ and $(K_{0\mathcal{P}}^{\mathbb{Q}}, K_{1\mathcal{P}}^{\mathbb{Q}})$, *ML* estimates of the former are those given by ordinary

[4] The choice of discrete-time is without loss of generality. All of our results apply to a continuous-time Gaussian model as well.

[5] Invariant transforms (Dai and Singleton (2000)) involve rotating, scaling and translating the state and parameter vectors to keep the short rate and bond prices unchanged (invariant), usually by a mapping $Y_t = AX_t + b$, where A is an invertible matrix. The transformed parameters are outlined in Appendix B.

least squares. Accordingly, computation of ML estimates effectively amounts to searching only over the $N+1$ parameters $(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}})$ and this can be accomplished very quickly. The remainder of this section flushes out these ideas. Unless otherwise noted, we assume that bond prices are Markov and all of the eigenvalues of $K^{\mathbb{Q}}_1$ are non-zero.[6]

The model-implied yield on a zero-coupon bond of maturity m is an affine function of the state X_t (Duffie and Kan (1996)):

$$y_{t,m} = A_m(\Theta^{\mathbb{Q}}_X) + B_m(\Theta^{\mathbb{Q}}_X) \cdot X_t, \tag{4}$$

where (A_m, B_m) satisfy well-known Riccati difference equations (see Appendix A for a summary), and $\Theta^{\mathbb{Q}}_X = (K^{\mathbb{Q}}_{0X}, K^{\mathbb{Q}}_{1X}, \Sigma_X, \rho_{0X}, \rho_{1X})$ is the vector of parameters from (2–3) relevant for pricing. We let $(m_1, m_2, \ldots, m_J)$ be the set of maturities (in years) of the bonds used in estimation of a *GDTSM*, $J \geq N$, and $y'_t = (y_{t,m_1}, \ldots, y_{t,m_J}) \in \mathbb{R}^J$ be the corresponding set of model-implied yields.

In general, (4) may be violated in the data due to market effects (e.g. bid-ask spreads or repo specials), violations of no arbitrage or measurement errors. We will collectively refer to all of these possibilities simply as measurement or pricing errors. To distinguish between model-implied and observed yields in the presence of pricing errors we let $y^o_{t,m}$ denote the counterpart of $y_{t,m}$ that is observed *with measurement error*.

For any full-rank, portfolio matrix $P \in \mathbb{R}^{N\times J}$ we let $\mathcal{P}_t \equiv P y_t$ denote the associated N-dimensional set of portfolios of yields, where the i^{th} portfolio puts weight $P_{i,j}$ on the yield for maturity m_j. Applying (4) we obtain

$$\mathcal{P}_t = A_P(\Theta^{\mathbb{Q}}_X) + B_P(\Theta^{\mathbb{Q}}_X)' X_t, \tag{5}$$

where $A_P = P[A_{m_1}, \ldots, A_{m_J}]'$ and $B_P = [B_{m_1}, \ldots, B_{m_J}]P'$. Note that $B_P(K^{\mathbb{Q}}_{1X}, \rho_1)$ depends only the subset $(K^{\mathbb{Q}}_{1X}, \rho_1)$ of $\Theta^{\mathbb{Q}}_X$ (see (A3) in Appendix A).

Initially, we assume that there exist portfolios for which the no arbitrage pricing relations hold exactly:

Case P: There are N portfolios of bond yields $\mathcal{P}_t$, constructed with weights P, that are priced perfectly by the *GDTSM*: $\mathcal{P}^o_t = \mathcal{P}_t$.

We refer to the case where each portfolio consists of a single bond, so that N yields are priced perfectly, as Case **Y**. We defer until Section 6 the case where all bonds are priced with errors and estimation is accomplished by Kalman filtering.

[6] By Markov we mean that the yield curve y_t is a Markov processes. In particular, we require that the distribution of y_{t+1} conditional on y_t is the same as the distribution of y_{t+1} conditional on $(y_t, y_{t-1}, \ldots, y_0)$. See Appendix E for relaxation of both of these assumptions.

We now state our main result for Case **P**:

Theorem 1. *Suppose that Case* **P** *holds for given fixed portfolio weights P, and the eigenvalues $\lambda^{\mathbb{Q}}$ of $K_{1X}^{\mathbb{Q}}$ are all nonzero. Then any canonical* GDTSM *is observationally equivalent to a unique* GDTSM *whose pricing factors $\mathcal{P}_t$ are the portfolios of yields $Py_t = Py_t^o$. Moreover, the $\mathbb{Q}$-distribution of $\mathcal{P}_t$ is uniquely determined by $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$. That is,*

$$\Delta \mathcal{P}_t = K_{0\mathcal{P}}^{\mathbb{P}} + K_{1\mathcal{P}}^{\mathbb{P}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \epsilon_t^{\mathbb{P}} \tag{6}$$

$$\Delta \mathcal{P}_t = K_{0\mathcal{P}}^{\mathbb{Q}} + K_{1\mathcal{P}}^{\mathbb{Q}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \epsilon_t^{\mathbb{Q}} \tag{7}$$

$$r_t = \rho_{0\mathcal{P}} + \rho_{1\mathcal{P}} \cdot \mathcal{P}_t, \tag{8}$$

is a canonical GDTSM*, where $K_{0\mathcal{P}}^{\mathbb{Q}}, K_{1\mathcal{P}}^{\mathbb{Q}}, \rho_{0\mathcal{P}}$ and $\rho_{1\mathcal{P}}$ are explicit functions of $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$. Our canonical form is parametrized by $\Theta^{\mathcal{P}} = (\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}})$.*

Remark. *The eigenvalues $\lambda^{\mathbb{Q}}$ may be not be distinct and may be complex. In the empirical analysis in Section 5, we explicitly consider these possibilities.*

We refer to the *GDTSM* in Theorem 1 as the JSZ canonical form parametrized by $\Theta^{\mathcal{P}}$. Before formally proving Theorem 1, we outline the main steps. First, we want to show that any *GDTSM* is observationally equivalent to a model where the states are the observed bond portfolios $\mathcal{P}_t$ (with corresponding weights P.) Thus, for $\mathcal{G} = \{(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_0, \rho_1, K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma)\}$, the set of all possible *GDTSM*s,[7] we want to show that every $\Theta \in \mathcal{G}$ is observationally equivalent to some $\Theta_{\mathcal{P}} \in \mathcal{G}_{\mathcal{P}}^P$, where

$$\mathcal{G}_{\mathcal{P}}^P = \{(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_0, \rho_1, K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma) : \text{the factors are portfolios with weights } P\}.$$

This first step is easily established: for any *GDTSM* with latent state X_t, $\mathcal{P}_t$ satisfies (5). Following DS, we can, by applying the change of variables outlined in Appendix B, compute the dynamics (under both $\mathbb{P}$ and $\mathbb{Q}$) of $\mathcal{P}_t$ and express r_t as an affine function of $\mathcal{P}_t$. The parameters after this change of variables give an observationally equivalent model where the states are the portfolios of yields.

[7] More formally, we think of the set of *GDTSM*s as a set of stochastic processes for the yield curve rather than as a set of parameters governing the stochastic process of the yield curve. To see the correspondence, we define on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (with associated filtration $\{\mathcal{F}_t\}$) the processes $y : \Omega \times \mathbb{N} \to \mathbb{R}^{\mathbb{N}_+}$. Here $y_t^m(\omega)$ is the m-period yield at time t when the state is $\omega \in \Omega$. When our additional assumption that y is a Gaussian Markov process and no arbitrage is maintained (with risk premia at time t depending only on $\mathcal{F}_t$), these processes take the form of (1–3) and (4) for some parameters. In this way, we define a surjective map from the set of *GDTSM* parameters $(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_0, \rho_1, K_0^{\mathbb{P}}, K_1^{P}, \Sigma)$ to the set of *GDTSM* stochastic processes. With this association, two *GDTSM*s are observationally equivalent when the corresponding stochastic processes have the same finite-dimensional distributions.

Second, we establish uniqueness by showing that no two *GDTSMs* in $\mathcal{G}^P_{\mathcal{P}}$ are observationally equivalent. Clearly, if two *GDTSMs* are observationally equivalent and have the same observable factors, it must be that $(K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma)$ are the same. Intuitively, if the parameters $(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_0, \rho_1)$ are not the same, the price of some bonds would depend differently on the factors, a contradiction. In the second step, we formalize this intuition. Moreover, we show that for given $\lambda^{\mathbb{Q}}$ and $r^{\mathbb{Q}}_{\infty}$, there exists a unique $(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_0, \rho_1)$ consistent with no arbitrage and the states being the portfolios of yields $\mathcal{P}_t$. In the third and final step, we reparamatrize $\mathcal{G}^P_{\mathcal{P}}$ in terms of the free parameters $(\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_{\infty}, \Sigma_{\mathcal{P}})$.

In the second step of our proof of Theorem 1, we will use the following analogue of the canonical form in Joslin (2007), proved in Appendix C.

Proposition 1. *Every canonical* GDTSM *is observationally equivalent to the canonical* GDTSM *with* $r_t = r^{\mathbb{Q}}_{\infty} + \iota \cdot X_t$*,*

$$\Delta X_t = K^{\mathbb{Q}}_{1X} X_{t-1} + \Sigma_X \epsilon^{\mathbb{Q}}_t, \quad (9)$$

$$\Delta X_t = K^{\mathbb{P}}_{0X} + K^{\mathbb{P}}_{1X} X_{t-1} + \Sigma_X \epsilon^{\mathbb{P}}_t, \quad (10)$$

where Σ_X *is lower triangular (with positive diagonal),* $K^{\mathbb{Q}}_{1X}$ *is in ordered real Jordan form, and* $\epsilon^{\mathbb{Q}}_t, \epsilon^{\mathbb{P}}_t \sim N(0, I_N)$*.*

Here we specify the Jordan form with each eigenvalue associated with a single Jordan block (that is, each eigenvalue has geometric multiplicity of one.) Thus, when the eigenvalues are all real, $K^{\mathbb{Q}}_{1X}$ takes the form

$$K^{\mathbb{Q}}_{1X} = J(\lambda^{\mathbb{Q}}) \equiv \operatorname{diag}(J^{\mathbb{Q}}_1, J^{\mathbb{Q}}_2, \ldots, J^{\mathbb{Q}}_m), \quad \text{where each } J^{\mathbb{Q}}_i = \begin{pmatrix} \lambda^{\mathbb{Q}}_i & 1 & \cdots & 0 \\ 0 & \lambda^{\mathbb{Q}}_i & \cdots & 0 \\ \vdots & \vdots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda^{\mathbb{Q}}_i \end{pmatrix},$$

and where the blocks are in order of the eigenvalues. (See Appendix C for the real Jordan form when the eigenvalues are complex.) We refer to the set of Jordan canonical *GDTSMs* as $\mathcal{G}_J$ and it is parametrized by $\Theta^J = (\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_{\infty}, K^{\mathbb{P}}_{0X}, K^{\mathbb{P}}_{1X}, \Sigma_X)$.

Proof of Theorem 1: Having already established that we can rotate any model to one with $\mathcal{P}_t$ as the observed states, we proceed to prove the second step. Suppose that $\Theta_1, \Theta_2 \in \mathcal{G}^P_{\mathcal{P}}$ index two observationally equivalent canonical models. By the existence result in Proposition 1, each Θ_i is observationally equivalent to a *GDTSM*, Θ^J_i, which is in real ordered Jordan canonical form. Since

$$\mathcal{P}_t = A_P(\Theta_i^J) + B_P(\Theta_i)' X_{ti}^J, \tag{11}$$

where X_{ti}^J is the latent state for model Θ_i^J, it must be that

$$\Theta_i = A_P(\Theta_i^J) + B_P(\Theta_i^J)' \Theta_i^J. \tag{12}$$

Here, we use the notation that for a *GDTSM* with parameter vector Θ and state X_t, the observationally equivalent *GDTSM* with latent state $\hat{X}_t = C + DX_t$ has parameter vector $\hat{\Theta} = C + D\Theta$, as computed in Appendix B. Since observational equivalence is transitive, Θ_1^J is observationally equivalent to Θ_2^J; the uniqueness result in Proposition 1 implies that $\Theta_1^J = \Theta_2^J$. The equality in (12) then gives $\Theta_1 = \Theta_2$, which establishes our second step.

To establish the reparametrization in the third step, we focus on (11) and (12). The key is to show explicitly how given $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$ (from Θ_i^J) we can (i) choose the parameters $(K_{0J}^{\mathbb{P}}, K_{1J}^{\mathbb{P}}, \Sigma_J)$ to get any desired $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}})$ and (ii) construct the $(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_0, \rho_1)$ consistent the factors being $\mathcal{P}_t$. Details are provided in Appendix D.

For reference, we summarize the transformations computed in the last step as:

Proposition 2. *Any canonical* GDTSM *with* $\mathbb{Q}$ *parameters* $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$ *has the JSZ representation in Theorem 1 with*

$$K_{1\mathcal{P}}^{\mathbb{Q}} = BJ(\lambda^{\mathbb{Q}})B^{-1} \tag{13}$$

$$K_{0\mathcal{P}}^{\mathbb{Q}} = -K_{1\mathcal{P}}^{\mathbb{Q}} A \tag{14}$$

$$\rho_{1\mathcal{P}} = (B^{-1})' \iota \tag{15}$$

$$\rho_{0\mathcal{P}} = r_\infty^{\mathbb{Q}} - A \cdot \rho_1 \tag{16}$$

where $B = B_P(J(\lambda^{\mathbb{Q}}), \iota)'$, $A = A_P\left(0, J(\lambda^{\mathbb{Q}}), r_\infty^{\mathbb{Q}}, \iota, (B^{-1})' \Sigma_{\mathcal{P}} \Sigma_{\mathcal{P}}' B^{-1}\right)'$, *where* (A_P, B_P) *are defined in (5) and (A2–A3).*

3 $\mathbb{P}$-dynamics and Maximum Likelihood Estimation

Rather than defining latent states indirectly through a $\mathbb{Q}$-normalization, the JSZ normalization has instead prescribed observable yield portfolios $\mathcal{P}$ and parametrized their $\mathbb{Q}$ distribution in a maximally flexible way consistent with no arbitrage. A distinctive feature of our normalization is that, in estimation, there is an inherent separation between the parameters of the $\mathbb{P}$ and $\mathbb{Q}$

distributions of $\mathcal{P}_t$. In contrast, when the risk factors are latent, estimates of the parameters governing the $\mathbb{P}$ distribution necessarily depends on those of the $\mathbb{Q}$ distribution of the state, since the pricing model is required to either invert the model for the fitted states (when N bonds are priced perfectly) or to filter for the unobserved states (when all bonds are priced with errors). This section formalizes this "separation property" of the JSZ normalization.

By Theorem 1, we can, without loss of generality, use N portfolios of the yields y_t, $\mathcal{P}_t = \mathcal{P}_t^o \in \mathbb{R}^N$, as observed factors. Suppose that, in addition to $\mathcal{P}_t$, W of the individual bond yields, say $\mathcal{Y}_t \in \mathbb{R}^W$, are to be used in estimation and priced with measurement errors,[8] and that these errors $\mathcal{Y}_t^o - \mathcal{Y}_t$ are conditionally independent of lagged values of the state vector.[9] Then the conditional likelihood function (under $\mathbb{P}$) of the observed data $(\mathcal{P}_t, \mathcal{Y}_t^o)$ is

$$L(\mathcal{P}_t, \mathcal{Y}_t^o | y_{t-1}^o; \Theta) = L(\mathcal{Y}_t^o | \mathcal{P}_t; \lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}}) \times L(\mathcal{P}_t | \mathcal{P}_{t-1}; K_{1\mathcal{P}}^{\mathbb{P}}, K_{0\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}}). \tag{17}$$

Notice the convenient separation of parameters in the likelihood function. The conditional distribution of the yields priced with errors depends only on $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$ and not on $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$. In contrast, the conditional $\mathbb{P}$-density of the pricing factors $\mathcal{P}_t$ depends only on $(K_{1\mathcal{P}}^{\mathbb{P}}, K_{0\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}})$, and not on $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$. Using the assumption that $\mathcal{P}_t$ is conditionally Gaussian, the second term in (17) can be expressed as

$$L(\mathcal{P}_t | \mathcal{P}_{t-1}; K_{1\mathcal{P}}^{\mathbb{P}}, K_{0\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}}) = (2\pi)^{-N/2} |\Sigma_{\mathcal{P}}|^{-1} \exp\left(-\frac{1}{2} \| \Sigma_{\mathcal{P}}^{-1} \left(\mathcal{P}_t - E_{t-1}[\mathcal{P}_t] \right) \|_2^2 \right), \tag{18}$$

where $E_{t-1}[\mathcal{P}_t] = K_{0\mathcal{P}}^{\mathbb{P}} + (I + K_{1\mathcal{P}}^{\mathbb{P}})\mathcal{P}_{t-1}$. The parameters $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ that maximize the likelihood function L, namely

$$(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}) = \operatorname{argmax} \sum_{t=1}^{T} L(\mathcal{P}_t | \mathcal{P}_{t-1}; K_{1\mathcal{P}}^{\mathbb{P}}, K_{0\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}}) = \operatorname{argmin} \sum_{t=1}^{T} \| \Sigma_{\mathcal{P}}^{-1} \left(\mathcal{P}_t - E_{t-1}[\mathcal{P}_t] \right) \|_2^2, \tag{19}$$

are the sample ordinary least squares (OLS) estimates, independent of $\Sigma_{\mathcal{P}}$ (Zellner (1962)). Summarizing these observations:

Proposition 3. *Under Case* **P** *the ML estimates of the* $\mathbb{P}$ *parameters* $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ *are given by the OLS estimates of the conditional mean of* $\mathcal{P}_t$*.*

Absent constraints linking the $\mathbb{P}$ and $\mathbb{Q}$ dynamics, one can effectively separate the time-series properties ($\mathbb{P}$) of $\mathcal{P}_t$ from the cross-sectional constraints imposed by no arbitrage

[8]In the case that some of the portfolios have a single yield, we would delete this yield from those priced with error.

[9]They may be correlated, autocorrelated, non-normal, or have time-varying conditional moments depending on X_t. Typically, one assumes that $\mathcal{Y}_t^o - \mathcal{Y}_t \sim N(0, \Sigma_e)$, i.i.d.

($\mathbb{Q}$). The parameters of the $\mathbb{P}$ distribution thus can be estimated from time series alone, regardless of the cross-sectional restrictions. Furthermore, *independent* of $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$, the OLS estimates of $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ are by construction globally optimal. With $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ at hand, we use the sample conditional variance of $\mathcal{P}_t$, $\hat{\Sigma}_{\mathcal{P}}\hat{\Sigma}_{\mathcal{P}}'$, computed from the OLS innovations as the starting value for the population variance $\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}'$. Given $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$, this starting value for $\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}'$ is again by construction close to the global optimum. Therefore we have greatly reduced the number of parameters to be estimated. For instance, in a $GDTSM(3)$ model, the maximum number of parameters is 22 (3 for $\lambda^{\mathbb{Q}}$, 1 for $r_\infty^{\mathbb{Q}}$, 6 for $\Sigma_{\mathcal{P}}$, 3 for $K_{0\mathcal{P}}^{\mathbb{P}}$ and 9 for $K_{1\mathcal{P}}^{\mathbb{P}}$), but our normalization and estimation technique make it possible to search over only 4 parameters $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$. This underlies the remarkable improvement in estimation speed for the JSZ normalization over other canonical forms.

4 The $\mathbb{P}$-Distribution of Bond Yields Under Case P

Proposition 3 leads immediately to several striking properties of *GDTSM*s under Case **P**. We fix the N portfolios of zero coupon yields that are priced perfectly by the model so that the JSZ normalization has $\mathcal{P}_t^o = \mathcal{P}_t$ as the state variable of the model.

4.1 Irrelevance of No Arbitrage for Forecasting

No-arbitrage is irrelevant for forecasting the observable pricing factors $\mathcal{P}_t$. Under Case **P** the no-arbitrage feature of a $GDTSM$ has no effect on the ML estimates of $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$ and this, in turn, implies that forecasts of future values of $\mathcal{P}$ are identical to those from an unconstrained VAR(1) model for $\mathcal{P}_t$.[10]

This result sharpens Duffee (2009)'s finding that the restrictions on a VAR implied by an arbitrage-free $GDTSM$ cannot be rejected against the alternative of an unrestricted VAR. When forecasting the N portfolios of yields $\mathcal{P}_t$, Proposition 3 shows *theoretically* that a similar result *must* hold insofar as Case **P** is (approximately) valid.

The JSZ normalization makes these observations particularly transparent. In contrast, in the (observationally equivalent) specification in (1–3), portfolio yield forecasts are

$$\begin{aligned} E_t[\mathcal{P}_{t+1}] - \mathcal{P}_t &= B_P(\Theta^{\mathbb{Q}})\left(E_t[X_{t+1}] - X_t\right) = B_P(\Theta^{\mathbb{Q}})\left(K_{0X}^{\mathbb{P}} + K_{1X}^{\mathbb{P}} X_t\right) \\ &= B_P(\Theta^{\mathbb{Q}})\left(K_{0X}^{\mathbb{P}} + K_{1X}^{\mathbb{P}}(B_P(\Theta^{\mathbb{Q}})^{-1}\left(\mathcal{P}_t - A_P(\Theta^{\mathbb{Q}})\right)\right). \end{aligned} \tag{20}$$

[10] Note that, in principle, enforcing no-arbitrage restrictions may be relevant for the construction of forecast confidence intervals through the dependence on $\Sigma_{\mathcal{P}}$. However, empirically this effect is likely to be small.

Thus, with latent states, the portfolio forecasts are expressed in terms of both the $\mathbb{P}$ and $\mathbb{Q}$ parameters of the model. From (20) it is not obvious that OLS recovers the ML estimates of $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$. The JSZ normalization makes the implicit cancellations in (20) explicit.

Remark. *For Case* **P** *only the portfolios of yields $\mathcal{P}$ are assumed to be priced perfectly by the* GDTSM*, and this irrelevance result applies only to forecasting $\mathcal{P}$.*

Elaborating, if for instance $\mathcal{P}$ represents the first N PCs of the collection of J observed bond yields y_t^o, then each individual yield y_{t,m_j}^o $(j = 1, \cdots, J)$ will be priced with error, and the imposition of no-arbitrage restrictions may improve its forecast.[11] The primary source of this improvement within a *GDTSM* comes from the parsimonious representations of the y_{t,m_j} as affine functions of the PCs $\mathcal{P}_t$. Thus, if a *GDTSM* is (approximately) correctly specified, then forecasting y_t^o within this *GDTSM* should lead to improved forecasts relative to the (possibly) much more heavily parameterized unconstrained VAR representation of y_t^o. This improvement seems especially likely in a term structure setting, because the variance of $(y_{t,m_j}^o - y_{t,m_j})$ is typically small relative to the variance of y_{t,m_j}.[12]

4.2 Irrelevance of $\mathbb{Q}$-restrictions for Forecasting

Restrictions that only affect the $\mathbb{Q}$-parameters $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$ (as well as $\Sigma_{\mathcal{P}}$) are irrelevant for forecasting the bond-portfolio yields $\mathcal{P}_t$. Again the latent-factor representation offers little insight into the fact that this result holds since (20) suggests that the risk-neutral parameters enter into $E_t[\mathcal{P}_{t+1}]$. However, as the JSZ normalization shows, absent restrictions *across* the $\mathbb{P}$ and $\mathbb{Q}$ parameters of the model, these $\mathbb{Q}$-restrictions must affect $(K_{0X}^{\mathbb{P}}, K_{1X}^{\mathbb{P}})$ in a manner that "cancels" their impact on $E_t[\mathcal{P}_{t+1}]$.

An immediate implication of this second observation is that *forecasts of $\mathcal{P}$ using an arbitrage-free Nelson-Siegel (AFNS) model are equivalent to forecasts based on an unconstrained VAR(1) representation of $\mathcal{P}$.* To see this, we show that the AFNS model of Christensen, Diebold and Rudebusch (2007) is an invariant transformation of a special case of the JSZ normalization (and indeed of the DS normalization) with the additional constraint that $\lambda^{\mathbb{Q}} = (0, \lambda, \lambda)$ and $r_\infty^{\mathbb{Q}} = 0$.[13] The AFNS(3) model with latent state vector $X_t = (X_t^1, X_t^2, X_t^3)'$

[11] Of course if, for given j, y_{t,m_j} is one of the observable pricing factors in the model, selected by the portfolio matrix P (i.e., $y_{t,m_j}^o = y_{t,m_j}$), then our irrelevance result applies to y_{t,m_j}.

[12] Nevertheless, when forecasting y_{t,m_j}^o within a *GDTSM*, one might want to take account of the conditional distribution of the error $(y_{t,m_j}^o - y_{t,m_j})$ in constructing an optimal forecast.

[13] Christensen, Diebold and Rudebusch (2007) also impose zero drift of the non-stationary factor under $\mathbb{Q}$.

has a feedback matrix $K^{\mathbb{Q}}_{1X}$ of the form

$$K^{\mathbb{Q}}_{1X} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & -\lambda \end{pmatrix}, \tag{21}$$

and the short rate depends only on the first two latent pricing factors: $r_t = X^1_t + X^2_t$. This model is obtained by starting with the Jordan form underlying the JSZ normalization chosen to be the three-factor case with $r_t = r^{\mathbb{Q}}_{\infty} + \iota' Y_t$ and $K^{\mathbb{Q}}_{1Y}$ having two equal eigenvalues:

$$\Delta Y_t = \begin{pmatrix} -\alpha & 0 & 0 \\ 0 & -\lambda & 1 \\ 0 & 0 & -\lambda \end{pmatrix} Y_{t-1} + \Sigma_Y \epsilon^{\mathbb{Q}}_t. \tag{22}$$

Applying the invariant transform

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1/\lambda \end{pmatrix}$$

to Y gives $X_t = BY_t$, $\Sigma_X = B\Sigma_Y$, $r_t = r^{\mathbb{Q}}_{\infty} + (1,1,0)X_t$, and

$$\Delta X_t = \begin{pmatrix} -\alpha & 0 & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & -\lambda \end{pmatrix} X_{t-1} + \Sigma_X \epsilon^{\mathbb{Q}}_t. \tag{23}$$

Therefore, the AFNS model is the constrained special case of the JSZ normalization with $\lambda^{\mathbb{Q}} = (0, \lambda, \lambda)$ and $r^{\mathbb{Q}}_{\infty} = 0$. Proposition 3 implies that these restrictions do not affect the ML estimates of $K^{\mathbb{P}}_{0\mathcal{P}}$ and $K^{\mathbb{P}}_{1\mathcal{P}}$ and, hence, they *cannot* improve the forecasts of $\mathcal{P}$ relative to an unconstrained VAR(1). It follows that the forecast gains that Christensen, Diebold and Rudebusch (2007) attribute to the structure of their AFNS pricing model must in fact have been a consequence of restrictions that they imposed directly on the $\mathbb{P}$-distribution of bond yields – the no-arbitrage restrictions implicit in the AFNS model played no role in their forecasts of the first three PCs of bond yields.

4.3 Constraints on the Dimensionality of Excess Returns

Central to the preceding irrelevance results is the absence of restrictions across the parameters of the $\mathbb{P}$ and $\mathbb{Q}$ distributions of $\mathcal{P}_t$. Such constraints would arise in practice if, for instance, the *GDTSM*-implied expected excess returns on bonds of different maturities line in a space of dimension $\mathcal{L}$ less than $\dim(\mathcal{P}_t) = N$. Cochrane and Piazzesi (2005, 2008) conclude that $\mathcal{L} = 1$ when conditioning risk premiums only on yield curve information. Joslin, Priebsch and Singleton (2009b) present evidence that, when expected excess returns are conditioned on $\mathcal{P}_t$, inflation, and output growth, $\mathcal{L}$ is at least two. We explore the relevance for forecasting bond yields of imposing the constraint $\mathcal{L} = 1$ within *GDTSM*s that condition risk premiums on the pricing factors $\mathcal{P}$. When this constraint is (approximately) valid, improved forecasts of y_t could well arise from the associated reduction in the dimensionality of the parameter space.

To interpret this constraint, note from Cox and Huang (1989) and Joslin, Priebsch and Singleton (2009b) that one-period, expected excess returns on portfolios of bonds with payoffs that track the pricing factors $\mathcal{P}_t$, say $xr\mathcal{P}_t$, are given by the components of

$$xr\mathcal{P}_t = K_{0\mathcal{P}}^{\mathbb{P}} - K_{0\mathcal{P}}^{\mathbb{Q}} + (K_{1\mathcal{P}}^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{Q}})\mathcal{P}_t. \tag{24}$$

That, the i^{th} component of $(K_{1\mathcal{P}}^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{Q}})\mathcal{P}_t$ is the source of variation in the risk premium on the bond portfolio representing pure exposure to innovations in the i^{th} *PC*. Therefore, the constraint that the one-period expected excess returns on bond portfolios are driven by the same linear combination of the pricing factors $\mathcal{P}$ ($\mathcal{L} = 1$) amounts to the constraint that the rank of $A_{RRP} = K_{1\mathcal{P}}^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{Q}}$ is one.[14]

The reduced rank risk premium *GDTSM*s can be estimated through a concentration of the likelihood in the same spirit as (17). Given $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$, the ML estimates of $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ can be computed as follows. First, compute (α, β) from the regression

$$\mathcal{P}_{t+1} - (K_{0\mathcal{P}}^{\mathbb{Q}} + K_{1\mathcal{P}}^{\mathbb{Q}}\mathcal{P}_t) = \alpha + \beta\mathcal{P}_t + \epsilon_t^P, \tag{25}$$

where we fix $\Sigma_{\mathcal{P}}$ and impose the constraint that β has rank 1. The ML estimates of this constrained regression can be computed in closed form. For given $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$, the ML estimates of the $\mathbb{P}$ parameters are then given by

$$K_{0\mathcal{P}}^{\mathbb{P}} = K_{0\mathcal{P}}^{\mathbb{Q}} + \hat{\alpha}, \quad K_{1\mathcal{P}}^{\mathbb{P}} = K_{1\mathcal{P}}^{\mathbb{Q}} + \hat{\beta}. \tag{26}$$

[14] Alternatively, we could restrict the rank of $[K_{0\mathcal{P}}^{\mathbb{P}} - K_{0\mathcal{P}}^{\mathbb{Q}}, K_{1\mathcal{P}}^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{Q}}]$ to one. This would enforce the stronger restriction that only one linear combination of the factors has non-zero expected excess return.

In comparison to the setting underlying Proposition 3, reduced rank risk premia enforce constraints across the parameters of the $\mathbb{P}$ and $\mathbb{Q}$ distributions. Consequently, the ML estimates of the $\mathbb{P}$ parameters are no longer given by their OLS counterparts. This, in turn, means that the implications of Proposition 3 discussed in Section 4.2 will, in general, no longer apply. Under the reduced-rank restrictions any further assumptions on the $\mathbb{Q}$-parameters (such as the constraints of the AFNS) will directly affect the estimated $\mathbb{P}$ parameters as there is a link between the cross-section and time-series dynamics of yields. We explore the empirical implications of these observations in Section 5.

4.4 Comparing the JSZ Normalization to Other Canonical Models

The normalizations adopted by DS and Joslin (2007) preserve the latent factor structure in (9 - 10), in contrast to the rotation to observable pricing factors in the JSZ normalization. To our knowledge the only other normalization that has an "observable" state vector is the one explored by Collin-Dufresne, Goldstein and Jones (2008) (CGJ). All three of these canonical models– DS, Joslin, and CGJ– are observationally equivalent.[15]

In the CGJ setup the state vector X_t is completely defined by r_t and its first $N-1$ moments under $\mathbb{Q}$:

$$X_t = (r_t, \mu_{1t}, \mu_{2t}, \ldots, \mu_{N-1,t})', \tag{27}$$

where

$$\mu_{1t} = \frac{1}{dt} E^{\mathbb{Q}}(dr_t),\ \ \mu_{k+1,t} = \frac{1}{dt} E^{\mathbb{Q}}(d\mu_{kt}),\ \ k = 1, \ldots, N-2. \tag{28}$$

Under $\mathbb{Q}$, X_t follows

$$dX_t = (K^{\mathbb{Q}}_{0,CGJ} + K^{\mathbb{Q}}_{1,CGJ} X_t)dt + \Sigma_X dW_t, \tag{29}$$

where Σ_X is lower triangular and $K^{\mathbb{Q}}_{0,CGJ} = (0, 0, \ldots, 0, \gamma)'$. By construction, the matrix $K^{\mathbb{Q}}_{1,CGJ}$ is the companion matrix factorization of the feedback matrix $K^{\mathbb{Q}}_{1X}$ in (9).

[15] Different choices of normalizations, associated with different, unique matrix factorizations of the feedback matrix $K^{\mathbb{Q}}_{1X}$, give rise to observationally equivalent models, though models with different structure to their parameter sets. The JSZ normalization is based on the real Jordan factorization used in Proposition 1. CJG adopt the companion factorization. For any monic polynomial $p(x) = x^n - \mu_{n-1}x^{n-1} - \cdots - \mu_1 x - \mu_0$, the companion matrix is

$$C(p) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_{n-1} \end{pmatrix}.$$

Given any matrix K, its monic characteristic polynomial is unique, and the matrix K is similar to its companion matrix $C(p(K))$.

The sense in which X_t is observable in the CGJ normalization is quite different than in the JSZ normalization, and these differences may have practical relevance. First, it will not always be convenient to assume that the one-period short-rate r_t is observable. Duffee (1996) highlights various liquidity and "money-market" effects that might distort yields on short-term bond relative what is implied by a *GDTSM*. The *true* short rate– the one that implicitly underlies the pricing of long-term bonds– will not literally be observable absent an explicit model of these money-market effects. Second, actions by monetary authorities might necessitate the inclusion of additional risk factors or jumps in these factors when explicitly including short rates in the analysis of a $DTSM$ (Piazzesi (2005)). Within the JSZ normalization one is free to define the portfolio matrix P so as to focus on segments of the yield curve away from the very short end, while preserving fully observable $\mathcal{P}$.

More subtly, the construction of the state vector in the CGJ normalization requires the parameters of the $\mathbb{Q}$ distribution. Therefore, any change in the implementation of a *GDTSM* that changes the implied $\mathbb{Q}$ parameters will necessarily change the observed pricing factors under the CGJ normalization. Fitting the same model to two overlapping sample periods could, for example, give rise to different values of the observed state variables during the overlapping period. In contrast, under the JSZ normalization we are led to identical values of $\mathcal{P}$ for all overlapping sample periods.

Full separation of the $\mathbb{P}$ and $\mathbb{Q}$ sides of the unrestricted model appears to be a unique feature of the JSZ normalization. It is this separation that clarifies the role of no-arbitrage restrictions in *GDTSM*s, and gives rise to the enormous computational advantages of our normalization relative to the DS, Joslin, and CGJ canonical models.

5 Empirical Results

We estimate the three-factor *GDTSM*s summarized in Table 1 by ML using the JSZ canonical form and the methods outlined in Section 4.[16] The data are end-of-month, Constant Maturity Treasury (CMT) yields from release Fed H.15 over the period from January 1990 to December 2007 (216 observations). The maturities considered are 6 months, and 1-, 2-, 3-, 5-, 7- and 10-years. These coupon yields are then used to bootstrap a zero coupon curve assuming constant forward rates between maturities. Within Case **P**, we consider several subcases. With distinct real eigenvalues, we assume the first three *PC*s are priced without error (RPC); or the 0.5-, 2-, and 10-year zero coupon yields are priced without error (RY). Additionally, we estimate models that price the first three *PC*s of the zero curve exactly under the constraints

[16] $\bar{\lambda}_i^{\mathbb{Q}}$ denotes the complex conjugate of the i^{th} element of $\lambda^{\mathbb{Q}}$. Also, we defer discussion of case RKF, in which all yields are priced with error, until Section 6.

Model Name	Specification
RPC	Real $\lambda^{\mathbb{Q}\prime} = (\lambda_1^{\mathbb{Q}}, \lambda_2^{\mathbb{Q}}, \lambda_3^{\mathbb{Q}})$, PC1, PC2, PC3 priced exactly
RY	Real $\lambda^{\mathbb{Q}\prime} = (\lambda_1^{\mathbb{Q}}, \lambda_2^{\mathbb{Q}}, \lambda_3^{\mathbb{Q}})$, 0.5-, 2-, and 10-year zeros priced exactly
CPC	Complex $\lambda^{\mathbb{Q}\prime} = (\lambda_1^{\mathbb{Q}}, \lambda_2^{\mathbb{Q}}, \bar{\lambda}_2^{\mathbb{Q}})$, PC1, PC2, PC3 priced exactly
JPC	Real repeated $\lambda^{\mathbb{Q}\prime} = (\lambda_1^{\mathbb{Q}}, \lambda_2^{\mathbb{Q}}, \lambda_2^{\mathbb{Q}})$, PC1, PC2, PC3 priced exactly
RPC_1	RPC and rank 1 risk premia
RKF	Real distinct $\lambda^{\mathbb{Q}}$, and all yields are priced with error
RCMT	Real $\lambda^{\mathbb{Q}\prime} = (\lambda_1^{\mathbb{Q}}, \lambda_2^{\mathbb{Q}}, \lambda_3^{\mathbb{Q}})$, 0.5-, 2-, and 10-year CMTs priced exactly

Table 1: Summary of Model Specifications

of repeated eigenvalues (JPC) and complex eigenvalues (CPC). Model JPC imposes the eigenvalue constraint of the AFNS model examined by Christensen, Diebold and Rudebusch (2009). Finally, a subscript of "1" indicates the case of reduced-rank risk premiums ($\mathcal{L} = 1$) with the one-period expected excess returns being perfectly correlated across bonds. In all cases, except as noted, the measurement errors are assumed to be independent and normally distributed with the same standard deviation across maturities.

An alternative measurement error structure arises when one supposes that coupon bonds are priced without error. In this case, portfolios of zero bond yields will necessarily incorporate measurement error. To that end, we consider

Case C: N coupon bonds are priced exactly and W coupon bonds are priced with normally distributed errors in the *GDTSM*.

In implementing Case **C** with coupon-bond data, one can still select N portfolios of zero coupon yields and construct the rotation where these portfolios comprise the state vector. Even though such yields may not be observed, this rotation is still valuable because the portfolios of model-implied zero yields $\mathcal{P}_t$ can be approximated from the observed data. For example, one could bootstrap or spline an approximate zero coupon yield curve from the observed coupon bond prices and form an approximation of $\mathcal{P}_t$, call it $\mathcal{P}_t^a$. Importantly, the projection of $\mathcal{P}_t^a$ onto its own lag will recover reliable starting values for $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$. However, because coupon bond yields are nonlinear functions of $\mathcal{P}$, the irrelevance propositions discussed in Section 4 do not apply to Case **C**. In our empirical implementation we consider the case of the 0.5-, 2-, and 10-year CMT yields priced without error, and the 1-, 3-, 5-, 7-year par coupon yields priced with errors (RCMT).

In order to facilitate comparison of the estimates across models with different pricing factors, all of our results are presented in terms of the implied $\mathbb{Q}$ distribution of the first

Model	Parameter Estimate			
	$\lambda_1^{\mathbb{Q}}$	$\lambda_2^{\mathbb{Q}}$	$\lambda_3^{\mathbb{Q}}/\mathrm{im}(\lambda_2^{Q})$	$r_\infty^{\mathbb{Q}}$
RPC	-0.0024	-0.0482	-0.0712	8.67
	(0.00032)	(0.00472)	(0.00756)	(0.457)
RY	-0.00196	-0.0404	-0.0896	9.44
	(0.000214)	(0.00156)	(0.00415)	(0.504)
RKF	-0.00245	-0.0472	-0.0739	8.45
	(0.000567)	(0.00724)	(0.0125)	(0.678)
RCMT	-0.00178	-0.0372	-0.103	11.2
	(7e-005)	(0.000819)	(0.0029)	(0.346)
JPC	-0.00225	-0.0582	-0.0582	8.94
	(0.000233)	(0.000707)	(0.000707)	(0.378)
CPC	-0.00225	-0.0582	-1.56e-005	8.94
	(0.000318)	(0.00154)	(0.668)	(0.518)
RPC_1	-0.00241	-0.0479	-0.0717	8.68
	(0.000319)	(0.00455)	(0.00739)	(0.458)
RY_1	-0.00197	-0.0404	-0.0899	9.44
	(0.000213)	(0.00155)	(0.00414)	(0.504)
RCMT_1	-0.00178	-0.0371	-0.103	11.2
	(6.92e-005)	(0.000828)	(0.003)	(0.345)
JPC_1	-0.00225	-0.0583	-0.0583	8.97
	(0.000233)	(0.000705)	(0.000705)	(0.383)

Table 2: ML estimates of the risk-neutral parameters of the model-implied principal components, with $r_\infty^{\mathbb{Q}}$ normalized to percent per annum (by multiplying by 12×100). Large-sample standard errors are given in parentheses.

three PCs of the zero yields.[17] Table 2 shows that these parameters are largely invariant to: (i) assumptions about the distribution of measurement errors, (ii) restrictions on the $\mathbb{Q}$-dynamics through restrictions on $\lambda^{\mathbb{Q}}$, and (iii) restrictions on the relation between the $\mathbb{Q}$- and $\mathbb{P}$-dynamics through the reduced rank assumption. The only mild exception is that model RCMT has a higher $r_\infty^{\mathbb{Q}}$, which is compensated for by slightly lower $\lambda_1^{\mathbb{Q}}$ and $\lambda_2^{\mathbb{Q}}$. The close alignment of results shows that the cross-section of bond yields provides a rich information set from which to extract the four relevant $\mathbb{Q}$-parameters, $r_\infty^{\mathbb{Q}}$ and $\lambda^{\mathbb{Q}}$. Imposing the reduced-rank risk premium constraint $\mathcal{L} = 1$ leads to measurable differences in estimates across corresponding models, particularly for some of the elements of $K_{1\mathcal{P}}^{\mathbb{P}}$.

[17] That is, under Case **Y** or when the CMT yields are priced perfectly by the *GDTSM*, after estimation, we impose the JSZ normalization based on the PCs of zero yields as the state variables.

Another notable feature of these estimates is that the results for model CPC are the same as those for model JPC. This is because, in the limit, as the complex part of the eigenvalues approach zero, the complex model approaches the Jordan model (see Appendix C). Thus we see that, for our data set, complex eigenvalues are not preferred over real eigenvalues.

Tables 3 and 4 present the parameters of the $\mathbb{P}$ distribution of $\mathcal{P}$. The final row presents parameters from a VAR (with no pricing involved) of the PCs. Table 4 reveals that initializing $\Sigma_{\mathcal{P}}$ using OLS residuals leads to very accurate starting values. By way of contrast, if we had instead used the DS canonical form, an accurate initialization of Σ_X would require a reliable initial value for $K_1^{\mathbb{Q}}$. We avoid this interplay between the values of Σ_X and $K_1^{\mathbb{Q}}$ in the JSZ canonical form by applying no-arbitrage constraints to determine $K_{1\mathcal{P}}^{\mathbb{Q}}$ independently of $\Sigma_{\mathcal{P}}$.

Across all specifications, the parameters are very comparable. Partly this is a consequence of Proposition 3: whether $\lambda^{\mathbb{Q}}$ is comprised of distinct real eigenvalues (RPC), complex eigenvalues (CPC), or repeated eigenvalues (JPC), the estimates of $K_{1\mathcal{P}}^{\mathbb{P}}$ and $K_{0\mathcal{P}}^{\mathbb{P}}$ are equal to each other and to the OLS estimates. However, stepping beyond this proposition, when we change whether it is PCs or individual yields (e.g., RPC versus RY) that are priced perfectly by the $GDTSM$ under Case **P**, the parameters of the corresponding $\mathbb{P}$ distributions remain very similar. This remains true even when we impose the constraint that $\mathcal{L} = 1$, so that expected excess holding period returns over one period are perfectly correlated.

Model	Parameter Estimate											
	$K^{\mathbb{P}}_{1,11}$	$K^{\mathbb{P}}_{1,12}$	$K^{\mathbb{P}}_{1,13}$	$K^{\mathbb{P}}_{1,21}$	$K^{\mathbb{P}}_{1,22}$	$K^{\mathbb{P}}_{1,23}$	$K^{\mathbb{P}}_{1,31}$	$K^{\mathbb{P}}_{1,32}$	$K^{\mathbb{P}}_{1,33}$	$\theta^{\mathbb{P}}_1$	$\theta^{\mathbb{P}}_2$	$\theta^{\mathbb{P}}_3$
RPC	-0.25	0.16	5.2	0.032	-0.32	4.2	-0.03	-0.028	-1.8	-0.11	0.025	0.0063
	(0.16)	(0.55)	(2.8)	(0.055)	(0.24)	(1.2)	(0.022)	(0.087)	(0.46)	(0.028)	(0.0077)	(0.00035)
RY	-0.25	0.11	5.5	0.037	-0.31	4.1	-0.03	-0.034	-1.8	-0.11	0.026	0.0061
	(0.15)	(0.55)	(2.7)	(0.055)	(0.23)	(1.2)	(0.023)	(0.091)	(0.47)	(0.027)	(0.0077)	(0.00035)
RKF	-0.12	0.33	6.7	0.0078	-0.35	4.7	-0.021	-0.007	-1.2	-0.14	0.026	0.0063
	(0.13)	(0.52)	(2.9)	(0.052)	(0.22)	(1.2)	(0.018)	(0.075)	(0.42)	(0.029)	(0.0055)	(0.00029)
RCMT	-0.25	0.11	5.7	0.037	-0.32	4.1	-0.031	-0.032	-1.8	-0.11	0.026	0.0062
	(0.15)	(0.55)	(2.6)	(0.056)	(0.23)	(1)	(0.02)	(0.071)	(0.43)	(0.044)	(0.0093)	(0.00052)
JPC	-0.25	0.16	5.2	0.032	-0.32	4.2	-0.03	-0.028	-1.8	-0.11	0.025	0.0063
	(0.15)	(0.55)	(2.8)	(0.055)	(0.24)	(1.2)	(0.022)	(0.086)	(0.46)	(0.027)	(0.0077)	(0.00035)
CPC	-0.25	0.16	5.2	0.032	-0.32	4.2	-0.03	-0.028	-1.8	-0.11	0.025	0.0063
	(0.16)	(0.55)	(2.8)	(0.055)	(0.24)	(1.2)	(0.022)	(0.087)	(0.46)	(0.028)	(0.0077)	(0.00035)
RPC_1	-0.24	-0.17	7.4	0.031	-0.14	3.3	-0.025	-0.061	-1.5	-0.11	0.025	0.0063
	(0.15)	(0.55)	(2.8)	(0.056)	(0.25)	(1.3)	(0.022)	(0.087)	(0.46)	(0.037)	(0.012)	(0.0004)
RY_1	-0.24	-0.14	7.3	0.038	-0.17	3.3	-0.026	-0.056	-1.6	-0.11	0.026	0.0061
	(0.15)	(0.55)	(2.7)	(0.055)	(0.23)	(1.2)	(0.023)	(0.091)	(0.47)	(0.031)	(0.011)	(0.00037)
$RCMT_1$	-0.25	-0.11	7.1	0.042	-0.18	3.3	-0.029	-0.045	-1.7	-0.11	0.025	0.0062
	(0.15)	(0.55)	(2.6)	(0.057)	(0.23)	(1.1)	(0.02)	(0.072)	(0.42)	(0.04)	(0.013)	(0.0005)
JPC_1	-0.23	-0.18	7.4	0.03	-0.14	3.3	-0.025	-0.064	-1.5	-0.11	0.025	0.0063
	(0.15)	(0.55)	(2.7)	(0.056)	(0.25)	(1.2)	(0.022)	(0.087)	(0.46)	(0.038)	(0.012)	(0.0004)

Table 3: ML estimates of the physical parameters of the model-implied principal components, with $K^{\mathbb{P}}_1$ annualized (by multiplying by 12). Large-sample standard errors are given in parentheses.

Model	Parameter Estimate					
	σ_1	σ_2	σ_3	ρ_{12}	ρ_{13}	ρ_{23}
RPC	2.22	0.732	0.301	-0.576	0.579	-0.416
	(0.128)	(0.0335)	(0.0153)	(0.0416)	(0.0493)	(0.0624)
RY	2.21	0.721	0.316	-0.571	0.566	-0.388
	(0.126)	(0.0343)	(0.0168)	(0.0422)	(0.0508)	(0.0638)
RKF	2.21	0.837	0.313	-0.603	0.725	-0.631
	(0.127)	(0.0423)	(0.0205)	(0.044)	(0.0493)	(0.0668)
RCMT	2.23	0.73	0.316	-0.591	0.541	-0.362
	(0.423)	(0.0215)	(0.0278)	(0.0325)	(0.108)	(0.0504)
JPC	2.22	0.732	0.301	-0.576	0.578	-0.415
	(0.126)	(0.0335)	(0.0152)	(0.0414)	(0.0493)	(0.0618)
CPC	2.22	0.732	0.301	-0.576	0.578	-0.415
	(0.128)	(0.0335)	(0.0153)	(0.0416)	(0.0493)	(0.0624)
RPC_1	2.22	0.733	0.301	-0.579	0.58	-0.418
	(0.129)	(0.0338)	(0.0153)	(0.0416)	(0.0493)	(0.0626)
RY_1	2.21	0.723	0.316	-0.573	0.567	-0.389
	(0.127)	(0.0346)	(0.0168)	(0.0422)	(0.0509)	(0.064)
$RCMT_1$	2.23	0.731	0.316	-0.593	0.541	-0.362
	(0.424)	(0.0215)	(0.0278)	(0.0324)	(0.108)	(0.0507)
JPC_1	2.22	0.733	0.301	-0.579	0.58	-0.418
	(0.127)	(0.0338)	(0.0153)	(0.0414)	(0.0494)	(0.0618)

Table 4: ML estimates of the conditional covariance of the model-implied principal components, with volatility estimates $\sigma_1, \sigma_2, \sigma_3$ normalized to percent per annum (by multiplying by $100 \times \sqrt{12}$). Large-sample standard errors are given in parentheses.

5.1 Statistical Inference Within the JSZ Canonical Form

There are two null hypotheses that are of particular interest given our observations in Section 4. The first test addresses the algebraic multiplicity of eigenvalues in the $A_0(3)$ model. We showed earlier that the AFNS model of Christensen, Diebold and Rudebusch (2007) is equivalent to the JSZ canonical form with three extra constraints, including a repeated eigenvalue of $K_1^{\mathbb{Q}}$. To assess the validity of the null hypothesis $\lambda_2^{\mathbb{Q}} = \lambda_3^{\mathbb{Q}}$, under the JSZ normalization, we perform a Likelihood Ratio (LR) test against the alternative that $\lambda^{\mathbb{Q}}$ is unconstrained. With this one linear constraint, the LR test statistic has an asymptotic χ^2 distribution with 1 degree of freedom, $\chi^2(1)$.

The second test of interest is the dimensionality of the one-period risk premium which, as discussed in Section 4.3, is captured by the rank of $A_{RRP} = K_{1\mathcal{P}}^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{Q}}$. To impose the constraint that $\mathcal{L} = 1$ we start with the singular value decomposition of A_{RRP}, UDV', where U and V are unitary matrices and D is diagonal with the diagonal sorted in decreasing order. The null hypothesis of interest– that A_{RRP} has rank 1– is therefore imposed by setting D_{22} and D_{33} to zero. To translate this representation into constraints on the parameter space, note that, for an N-factor $GDTSM$ with $\mathcal{L} = 1$,

$$DV'\mathcal{P}_t = D_{11} \sum_{j=1}^{N} V_{j1} \mathcal{P}_{jt}. \tag{30}$$

Therefore, the expected excess returns $xr\mathcal{P}_t$ (see Section 4.3) are given by

$$xr\mathcal{P}_t = \left(K_{0\mathcal{P}}^{\mathbb{P}} - K_{0\mathcal{P}}^{\mathbb{Q}}\right) + U_{\bullet 1} \left(D_{11} \sum_{j=1}^{N} V_{j1} \mathcal{P}_{jt} \right). \tag{31}$$

The second term on the right-hand side of (31) expresses the time-varying components of $xr\mathcal{P}_t$ in terms of a common linear combination $V'_{\bullet 1}\mathcal{P}_t$ of the pricing factors. All of the parameters in (31) are econometrically identified by virtue of the facts that $V'_{\bullet 1}V_{\bullet 1} = 1$ (which identifies D_{11}) and $U'_{\bullet 1}U_{\bullet 1}$ (which identifies the weights on $D_{11}V'_{\bullet 1}\mathcal{P}_t$). Furthermore, given N, (31) implies $(N-1)^2$ cross-equation restrictions on the parameters of the conditional expectation $xr\mathcal{P}_t$. In our case, $N = 3$, so there are 4 cross-equation restrictions.

Tests for the equality of two eigenvalues are reported on top panel of Table 5, where a leading J means that model was estimated under the constraint that $\lambda_2^{\mathbb{Q}} = \lambda_3^{\mathbb{Q}}$ (consistent with the specifications of $AFNS$ models). In the PC-based models this null hypothesis is not rejected, while for the yield-based models it is rejected at conventional significant levels. To interpret this difference across choices of risk factors, we note from Table 2 that the estimated

$H_0 : \lambda_2^{\mathbb{Q}} = \lambda_3^{\mathbb{Q}}$

H_0	$\log L_0$	H_a	$\log L_a$	LR stats $\chi^2(1)$	p-value
JPC	58.6178	RPC	58.6193	0.645	0.422
JPC_1	58.6132	RPC_1	58.6149	0.731	0.393
JY	38.1679	RY	38.1863	7.912	0.005
JY_1	38.1638	RY_1	38.1830	8.256	0.004
JRCMT	39.0123	RCMT	39.0414	12.513	0.000

$H_0 : \text{rank}\left(K_{1\mathcal{P}}^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{Q}}\right) = 1$

H_0	$\log L_0$	H_a	$\log L_a$	LR stats $\chi^2(4)$	p-value
RPC_1	58.6149	RPC	58.6193	1.892	0.756
JPC_1	58.6132	JPC	58.6178	1.978	0.740
RY_1	38.183	RY	38.1863	1.419	0.841
JY_1	38.1638	JY	38.1679	1.763	0.779
RCMT_1	39.0387	RCMT	39.0414	1.161	0.884

Table 5: The top panel reports tests equality of two eigenvalues, and bottom panel reports tests for rank-1 risk premium. All likelihood-ratio statistics are computed as $LR = -2(T-1)(\log L_0 - \log L_a)$, where $T = 216$ is sample size and $\log L_0$ and $\log L_a$ are the log-likelihoods under the null and alternative, respectively.

$|\lambda_2^{\mathbb{Q}} - \lambda_3^{\mathbb{Q}}|$ is larger in model RY than in model RPC, with most of this difference being attributable to the larger value of $|\lambda_3^{\mathbb{Q}}|$ in model RY. The eigenvalue $\lambda_3^{\mathbb{Q}}$ governs the relatively high-frequency $\mathbb{Q}$-variation in yields and, thus, is particularly relevant for the behavior of the short end of the yield curve. Introducing the six-month yield directly as a pricing factor over weights the short end of the yield curve relative to having the PCs as pricing factors, as the latter are portfolios of yields along the entire maturity spectrum.

In the bottom panel we report tests of the reduced-rank, risk premium hypothesis that $\mathcal{L} = 1$. Under all model specifications this hypothesis cannot be rejected. This finding is consistent with the conclusions reached by Cochrane and Piazzesi (2005), though they effectively considered models with $N = 5$ as they examined $PC1$ through $PC5$.

5.2 Small-sample standard errors

Another feature of our normalization is that it facilitates the computation of smal-sample standard errors that can be compared to the asymptotic BHHH standard errors presented in the previous tables. The method which we used to bootstrap the standard errors is as

Parameter	Estimate	BHHH S.E.	Bootstrap S.E.
$K^{\mathbb{P}}_{1,11}$	-0.2543	(0.1562)	(0.2697)
$K^{\mathbb{P}}_{1,12}$	0.1595	(0.5495)	(0.8466)
$K^{\mathbb{P}}_{1,13}$	5.235	(2.795)	(2.956)
$K^{\mathbb{P}}_{1,21}$	0.03235	(0.05546)	(0.1025)
$K^{\mathbb{P}}_{1,22}$	-0.3153	(0.2421)	(0.3151)
$K^{\mathbb{P}}_{1,23}$	4.239	(1.243)	(1.222)
$K^{\mathbb{P}}_{1,31}$	-0.03047	(0.0225)	(0.04171)
$K^{\mathbb{P}}_{1,32}$	-0.02772	(0.08712)	(0.1344)
$K^{\mathbb{P}}_{1,33}$	-1.755	(0.4617)	(0.5039)
$\theta^{\mathbb{P}}_1$	-0.1109	(0.02762)	(0.02211)
$\theta^{\mathbb{P}}_2$	0.02539	(0.007709)	(0.006329)
$\theta^{\mathbb{P}}_3$	0.00631	(0.0003512)	(0.0002908)
$\lambda^{\mathbb{Q}}_1$	-0.002401	(0.0003234)	(0.0004967)
$\lambda^{\mathbb{Q}}_2$	-0.04817	(0.004761)	(0.006375)
$\lambda^{\mathbb{Q}}_3$	-0.07123	(0.007619)	(0.009381)
$r^{\mathbb{Q}}_\infty$	8.672	(0.4614)	(0.7907)
σ_1	0.02215	(0.001278)	(0.001038)
σ_2	0.008949	(0.0004206)	(0.0004213)
σ_3	0.003725	(0.0001632)	(0.0001753)
ρ_{12}	-0.5758	(0.04157)	(0.04592)
ρ_{13}	0.5789	(0.04925)	(0.04758)
ρ_{23}	-0.4156	(0.06245)	(0.05731)

Table 6: This table presents the standard errors of the parameter estimates computed both by the BHHH method and using a bootstrap method.

follows: we randomly choose a data $t \in \{1, 2, \ldots 216\}$ and initialize the state as the value of $\mathcal{P}$ on this date. Then, using the maximum likelihood estimate of the parameters, we simulate a path of the term structure for the sample size of 216 months and estimate the model based on this simulated data. These steps are repeated 1000 times.

Table 6 presents the results for the model RPC. The BHHH standard errors tend to overstate the precision with which we measure the effect of the level PC on the conditional means of the PCs ($K^{\mathbb{P}}_{1,11}, K^{\mathbb{P}}_{1,21}, K^{\mathbb{P}}_{1,31}$) by a factor of about two. Additionally, the precision with which we estimate the $\mathbb{Q}$ parameters is overstated by the BHHH method by a factor of about 50%. Overall though, the BHHH standard errors line up rather well with the bootsrapped standard errors.

5.3 Out-of-Sample Forecasting Results

An interesting question at this juncture is whether differences in parameter estimates translate into differences in the out-of-sample forecasting performance of these *GDTSM*s. We compute rolling re-estimation of each model using data from months $t = 1, \ldots, T$ ($T = 61, \ldots, 215$) and use the model to predict, out of sample, the changes in the principal components over the next 1-, 3-, 6-, and 12-month periods. As a benchmark, we use the corresponding forecasts from an unconstrained VAR. As we noted in Section 4, *theoretically* the forecasts of $\mathcal{P}_t$ are the same across all models that assume these PCs are priced without error and that differ only in the constraints they impose on the $\mathbb{Q}$ distribution of $\mathcal{P}_t$. In particular, with $\mathcal{L} = 3$, whether we assume distinct real eigenvalues, complex eigenvalues or repeated eigenvalues (as in the AFNS model), the forecasts of $\mathcal{P}_t$ are all *exactly* the same as those from an unconstrained VAR. This explains the rows of zeros in Table 7.

On the other hand, with the constraint $\mathcal{L} = 1$ imposed, models RPC_1 and JPC_1 have different predictions (though only slightly). This is because the differences under $\mathbb{Q}$ implied by the repeated root assumption now propagate to the $\mathbb{P}$-dynamics through the restriction relating the $\mathbb{P}$- and $\mathbb{Q}$-drifts.

Of greater quantitive significance, we see also that the reduced rank risk premia models offer superior out-of-sample predictions for all specifications relative to their unconstrained counterparts. In the event that the reduced rank assumption is true in the data, this is what one would anticipate due to improved accuracy of estimation.[18]

In concluding this section we comment briefly on the computational efficiency obtained using the the JSZ normalization. The only parameters that need to be estimated are $(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$ since, as discussed in Section 4, $(K^{\mathbb{P}}_{0,\mathcal{P}}, K^{\mathbb{P}}_{1\mathcal{P}})$ are determined by concentrating the likelihood and $(K^{\mathbb{Q}}_{0,\mathcal{P}}, K^{\mathbb{Q}}_{1,\mathcal{P}})$ are determined by no-arbitrage.[19] The models were estimated using sequential quadratic programming as implemented in Matlab's `fmincon`. Estimation under Case **P** using an informed guess of the $\mathbb{Q}$-eigenvalues took approximately 1.2 seconds.[20] We also estimated the models using multiple random seeds both for $(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}})$ and $\Sigma_{\mathcal{P}}$ in ranges consistent with a flat non-informative prior ($r^{\mathbb{Q}}_{\infty}$ was distributed uniformly on $[-.10, .30]$, for example). In this case, 100 random seeds were first initialized before applying the gradient search for the local optima to improve the starting value. These randomized searches converged in a slightly longer time of approximately 2.2 seconds. Furthermore, 99%+ of the

[18] Another interesting issue that we do not explore is whether, even if the assumption that $\mathcal{L} = 1$ is *false*, imposing this constraint improves forecasts by overcoming well-known small sample biases in estimation.

[19] The standard deviation of the pricing errors, σ_{pricing}, can be concentrated out as well, both when $\mathcal{L}$ equals 1 and 3.

[20] The computations were performed using a single-threaded application on a 2.4GHZ Intel Q6600 processor.

Forecast Error Reduction (%)	PC1				PC2				PC3			
	1m	3m	6m	12m	1m	3m	6m	12m	1m	3m	6m	12m
RPC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RY	-0.3	-0.5	-0.8	-0.7	0.2	0.4	0.4	0.0	0.1	0.8	1.3	0.8
RKF	0.9	3.0	5.9	12.9	-1.7	-4.7	-7.7	-10.0	1.2	3.3	7.7	10.6
JPC	-0.0	-0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.0
CPC	-0.0	-0.0	0.0	-0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RPC_1	-2.1	-4.5	-6.4	-7.4	-2.0	-3.9	-3.9	-1.6	-1.5	-2.7	-2.5	0.3
RY_1	-2.2	-4.8	-7.2	-8.8	-1.9	-3.9	-3.9	-1.8	-1.6	-2.7	-2.5	-1.0
JPC_1	-2.4	-4.8	-6.9	-8.4	-1.9	-3.7	-4.2	-2.8	-1.5	-2.6	-1.9	0.7

Table 7: This table presents the improvement in out-of-sample forecast accuracy relative to the forecasts from a VAR(1),

$$\sqrt{\frac{1}{T-59}\sum_{t=60}^{T}(\Delta PCi_{t+1} - E_t[\Delta PCi_{t+1}])^2},$$

where the expectation, E_t, is computed using the model estimated with data from time $\tau = 1, \ldots, t$.

searches converged to the same likelihood value (to within the tolerance) with very similar parameter estimates.[21] These computational advantages become even more important in the case where all yields are measured with error, which we consider next.

5.4 No Arbitrage and Restrictions on the $\mathbb{P}$-distribution of Yields

In Section 4, we demonstrated that neither the imposition of no arbitrage nor restrictions on the $\mathbb{Q}$-dynamics have any effect on the maximum likelihood estimates of $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$. However, restrictions on risk premia, such as the reduced-rank assumption, link $\mathbb{P}$ and $\mathbb{Q}$ and interact with no arbitrage to affect estimates of $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$. We now complete this discussion by examining whether no arbitrage affects the distribution of bond yields when one also imposes stand-alone restrictions on the $\mathbb{P}$-distribution of yields. Examples of such restrictions are that the yield portfolios are cointegrated or that the conditional mean of each portfolio yield does not depend on the other portfolio yields.[22] These restrictions focus entirely on the $\mathbb{P}$-distribution of yields with no reference to the $\mathbb{Q}$-distribution either directly or indirectly through risk premia. As such, one could impose such restrictions without reference to a no arbitrage model.

In these examples OLS no longer recovers the ML estimates of the parameters; rather, to obtain efficient estimates given $\Sigma_{\mathcal{P}}$, one must implement GLS. Let $(K_0^{c*}(\Sigma_{\mathcal{P}}), K_1^{c*}(\Sigma_{\mathcal{P}}))$ denote the GLS estimates of $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ given $\Sigma_{\mathcal{P}}$:

$$(K_0^{c*}(\Sigma_{\mathcal{P}}), K_1^{c*}(\Sigma_{\mathcal{P}})) = \underset{K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}}{\arg\max} \sum_{t=1}^{T} L(\mathcal{P}_t^o | \mathcal{P}_{t-1}^o; K_{1\mathcal{P}}^{\mathbb{P}}, K_{0\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}}), \tag{32}$$

where the arg max is taken over $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ satisfying the appropriate restriction on the $\mathbb{P}$-dynamics. In the presence of such restrictions, there is a non-degenerate dependence of (K_0^{c*}, K_1^{c*}) on $\Sigma_{\mathcal{P}}$. This dependence means that no arbitrage (which links $\Sigma_{\mathcal{P}}$ across $\mathbb{P}$ and $\mathbb{Q}$) affects the maximum likelihood estimates of $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$.

To explore the magnitudes of these effects, we estimate (1) a model with $K_{1\mathcal{P}}^{\mathbb{P}}$ constrained to be diagonal and (2) a model in which the $\mathcal{P}_t$ are cointegrated (with one unit root and no trend). The data again are as described in Section 5. In both cases, we assume risk premia have full rank and the $\mathbb{Q}$-distribution of yields is unconstrained. When the constraint of diagonal $K_{1\mathcal{P}}^{\mathbb{P}}$ is imposed (Table 8), no arbitrage has almost no effect on the parameters.[23]

[21] An exception here is the Jordan form where typically there were two local extrema with either the smaller or the larger eigenvalue repeated.

[22] See Campbell and Shiller (1991) (among others) for empirical evidence on cointegration among bond yields. Diebold and Li (2006) adopt an assumption very similar to the second example.

[23] The average log-likelihood (across t) for the no arbitrage model was 58.6193, while for the diagonal-

With No Arbitrage				Without No Arbitrage			
$K_{0\mathcal{P}}^{\mathbb{P}}$	$K_{1\mathcal{P}}^{\mathbb{P}}$			$K_{0\mathcal{P}}^{\mathbb{P}}$	$K_{1\mathcal{P}}^{\mathbb{P}}$		
-0.0129	-0.1511			-0.01288	-0.1509		
(0.01941)	(0.1362)						
0.00747		-0.2839		0.007613		-0.2889	
(0.006478)		(0.2055)					
0.01315			-1.991	0.01288			-1.949
(0.002929)			(0.4247)				

Table 8: This table presents the conditional mean parameters for the model with $K_{1\mathcal{P}}^{\mathbb{P}}$ constrained to be diagonal. The left panel imposed no arbitrage and uses yield data for all maturities. The right panel does not use no arbitrage and simply computes the estimates of a VAR of $\mathcal{P}_t$ with $K_{1\mathcal{P}}^{\mathbb{P}}$ constrained to be diagonal through GLS.

With No Arbitrage				Without No Arbitrage			
$K_{0\mathcal{P}}^{\mathbb{P}}$	$K_{1\mathcal{P}}^{\mathbb{P}}$			$K_{0\mathcal{P}}^{\mathbb{P}}$	$K_{1\mathcal{P}}^{\mathbb{P}}$		
-0.06204	-0.2627	0.00618	5.19	-0.06677	-0.2401	0.2661	5.291
-0.02014	0.04517	-0.0798	4.307	-0.01721	0.05194	-0.1684	4.317
0.00707	-0.02705	0.03503	-1.737	0.007129	-0.01835	0.06318	-1.707

Table 9: This table presents the conditional mean parameters for the model with cointegration with no trend and one unit root imposed. The left panel imposed no arbitrage and uses yield data for all maturities. The right panel does not use no arbitrage and simply computes the estimates of a VAR of $\mathcal{P}_t$ with cointegration imposed so that $[K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}]$ has rank 2.

When cointegration is imposed (Table 9), the no arbitrage assumption has a somewhat larger effect, but again the differences are generally small.

6 Observable Factors with Measurement Errors

Up to this point we have assumed that N portfolios of yields are priced perfectly by the *GDTSM*. We turn next to the case where all of the zero-coupon yields used in estimation equal their *GDTSM*-implied values plus measurement errors. Under the assumption that the measurement errors are jointly normal, this is a Kalman filtering problem.

Case F: The yields on $J(\geq N)$ zero-coupon bonds equal their *GDTSM*-implied values plus mean zero, normally distributed errors, $y_t^o - y_t$.

A number of researchers (see, e.g., Duffee and Stanton (2007) and Duffee (2009)) have

constrained model it was 58.5185. The corresponding likelihood ratio test statistic is 43.7369, far exceeding the 99% rejection region of 16.8 indicating a very strong rejection of this constraint.

emphasized the computational challenges of estimation under Case **F**. Under the normalization of DS, a researcher must estimate $(K_{1X}^{\mathbb{Q}}, K_{0X}^{\mathbb{P}}, K_{1X}^{\mathbb{Q}}, \rho_0, \rho_1)$, where $K_{1X}^{\mathbb{Q}}$ is lower triangular. In this parametrization, a researcher would likely have a diffuse prior on all of the parameters. Moreover, the states of the model depend on the parameters, so they too are unknown. We now show that our JSZ canonical representation extends to the setting of Case **F** and demonstrate its benefits both for interpretation and estimation of *GDTSM*s.

Theorem 1 shows that any *GDTSM* is observationally equivalent to a model where the latent states are a given set of portfolios of yields, purged of measurement errors. In Case **P**, when the portfolios are assumed to be observed without measurement errors, this means the states are simply these portfolios of yields. In Case **F** we can maintain the interpretation that the latent states are portfolios of yields with known portfolio matrix P, though now constructed with the model-implied (measurement-error free) yields y_t. Equivalently, under Case **F**, one can view $\mathcal{P}_t = Py_t$ as the "true" values of the pricing factors and view $\mathcal{P}_t^o = Py_t^o$ as its observed counterpart.[24]

To set up the Kalman filtering problem for Case **F** we start with a given set of portfolio weights $P \in \mathbb{R}^{J\times N}$. From P and $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$, we construct $(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_0, \rho_1)$ as prescribed in Proposition 2. From the no arbitrage relation (A2-A3) we then construct $A \in \mathbb{R}^J$ and $B \in \mathbb{R}^{J\times N}$ with $y_t = A + B\mathcal{P}_t$ and thus the relations

$$\Delta\mathcal{P}_t = K_{0\mathcal{P}}^{\mathbb{P}} + K_{1\mathcal{P}}^{\mathbb{P}}\mathcal{P}_t + \Sigma_{\mathcal{P}}\epsilon_t^{\mathbb{P}} \ , \tag{33}$$

$$y_t^o = A + B\mathcal{P}_t + \Sigma_Y\epsilon_t^{\mathrm{m}}, \tag{34}$$

where $\epsilon_t^{\mathbb{P}} \sim N(0, I_N)$ and $\epsilon_t^{\mathrm{m}} \sim N(0, I_M)$ are the measurement errors. Researchers have considered several parameterizations of the volatility matrix Σ_Y for ϵ_t^{m}. In our subsequent empirical examples we examine the cases of independent (diagonal Σ_Y) errors with distinct or common volatilities. These relations give the usual observation and state equations of the Kalman filter and they fully characterize the conditional distribution of the yield curve in terms of rotation-invariant parameters.

The computational benefits from using the JSZ normalization in this Case **F** arise, in part, from the observation that the least-squares projection of $\mathcal{P}_t^o$ onto $\mathcal{P}_{t-1}^o$ will nearly recover the *ML* estimates of $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$ to the extent that $\mathcal{P}_t^o \approx \mathcal{P}_t$ (and we can choose portfolios, such as the principal components, to make these errors small).[25] Additionally, although not exact, we have nearly concentrated the likelihood in that the optimal $\mathbb{P}$ parameters will typically

[24] In fact, an equivalent characterization of the JSZ normalization is that, for a given portfolio matrix P, $A_P(\Theta^{\mathbb{Q}}) = 0$ and $B_P(\Theta^{\mathbb{Q}}) = I_N$.

[25] This approximation can be verified empirically by comparing $\mathcal{P}_t^o$ to $E_t^{\mathbb{P}}[\mathcal{P}_t]$ or $E_T^{\mathbb{P}}[\mathcal{P}_t]$.

have weak dependence on the $\mathbb{Q}$ parameters owing to the fact that, as the $\mathbb{Q}$ parameters vary, the filtered states largely do not change.[26]

With the JSZ normalization, the parameter estimates are directly comparable across distributional assumptions on the measurement errors. That is, in analogy to Section 4, by fixing the yield portfolios, both measured with and without error, the $\mathbb{P}$ parameters are now directly comparable *regardless of the* $\mathbb{Q}$ *structure*. The parameters are also directly comparable across sample periods. When the $\mathbb{P}$-parameters are defined indirectly through a $\mathbb{Q}$-normalization, such comparisons will in general not be possible.

6.1 Empirical Implication

To illustrate Case **F** we estimate model RKF in which all J zero-coupon bonds used in estimation are priced with errors, and the eigenvalues of $K_1^{\mathbb{Q}}$ are all real. From Table 2 it is seen that the estimates of the $\mathbb{Q}$ parameters for model RKF are similar to those for models RPC and RY that are fit with N portfolios of yields priced exactly by the $GDTSM(3)$. Similarly, from Table 3 and Table 4 we see that the $\mathbb{P}$ parameters also generally match up across the models with and without filtering. An exception is the $\mathbb{P}$ distribution of $PC3$: when filtering, the volatility of $PC3$ is reduced by about 10%, and $PC3$ has a larger effect on the conditional mean of PC1 and PC2 (higher $K_{1,13}^{\mathbb{P}}, K_{1,23}^{\mathbb{P}}$). That is, $PC3$ both becomes a bit smoother and the model attributes a slightly greater affect of $PC3$ on forecasts of changes in the level and slope of the yield curve. For out-of-sample forecasts using model RKF, Table 7 shows that $PC1$ is better predicted by a simple VAR, while $PC2$ is predicted better than a VAR (though the differences are modest).

Also of interest in the presence of filtering are comparisons of the model-implied PCs with their corresponding sample estimates that, by assumption, are contaminated by measurement errors. Figure 1 plots the time series of the PCs computed from data against those from models RCMT, RY and RKF. For model RKF we plot the model-implied filtered $PCi_t^f = E_t[PCi_t]$. For all three models, the PCi^o are nearly identical to their model-implied counterparts. This is not surprising: if the model is accurately pricing the cross section of bonds, then it is almost a necessity that it will accurately match level, slope, and curvature. $PC3^f$ deviates slightly from $PC3^o$, and this is the source of the small differences seen in Figure 1.

A quite different picture emerges when we increase the number of pricing factors to four or five using the JSZ normalization under Case **F**. For $i = 1, 2, 3$, PCi^f lines up well with PCi^o, as before. However from Figure 2 it is seen that $(PC4^f, PC5^f)$ appears to be a smoothed

[26]This is in contrast to, for example, the rotation of DS where, as the lower triangular $K_1^{\mathbb{Q}}$ is changed, the latent states vary as well. Thus, necessarily, so do the optimal $\mathbb{P}$ parameters given the specified $\mathbb{Q}$ parameters.

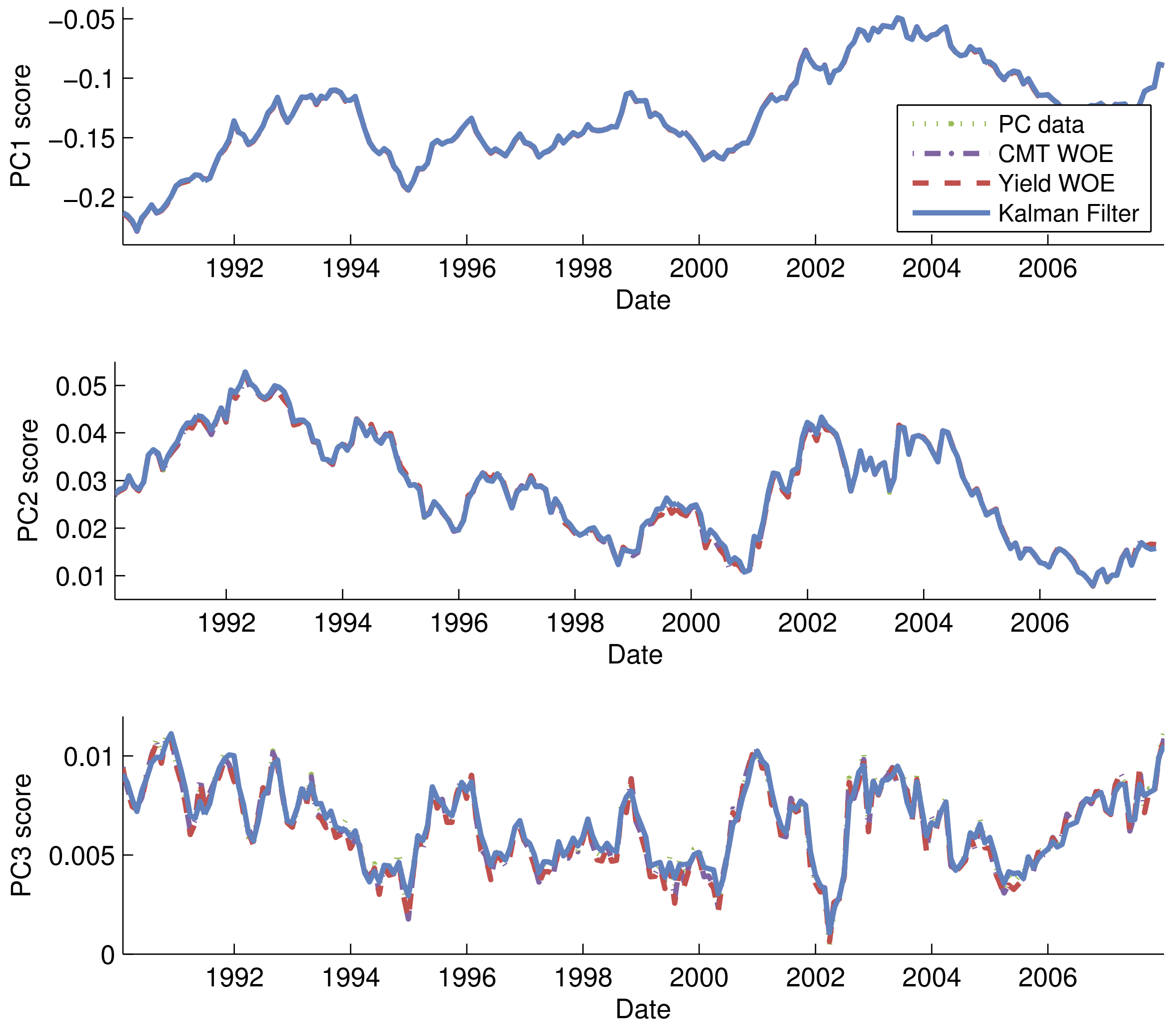

Figure 1: Model-Implied and Sample PCs: This figure plots the PCs implied by models RCMT, RY, and RKF against the estimated PCs from the data.

version of $(PC4^o, PC5^o)$, with the differences being substantial during some periods. To interpret these patterns we note the likelihood function, through the Kalman filter, attempts to match both the cross-sectional pricing relationships and the time series variation in excess returns. The higher order $PC4$ and $PC5$ only have small impacts on pricing since a three factor model already prices the cross-section of bonds well, but they do contain information about time-variation in expected returns.[27]

[27] Cochrane and Piazzesi (2005, 2008) find that a portfolio of smoothed forward rates, that is correlated with $PC4$, predicts bond returns. Joslin, Priebsch and Singleton (2009b) find that smoothed growth in industrial production, which is also correlated with $PC4$, is an important determinant of excess returns for

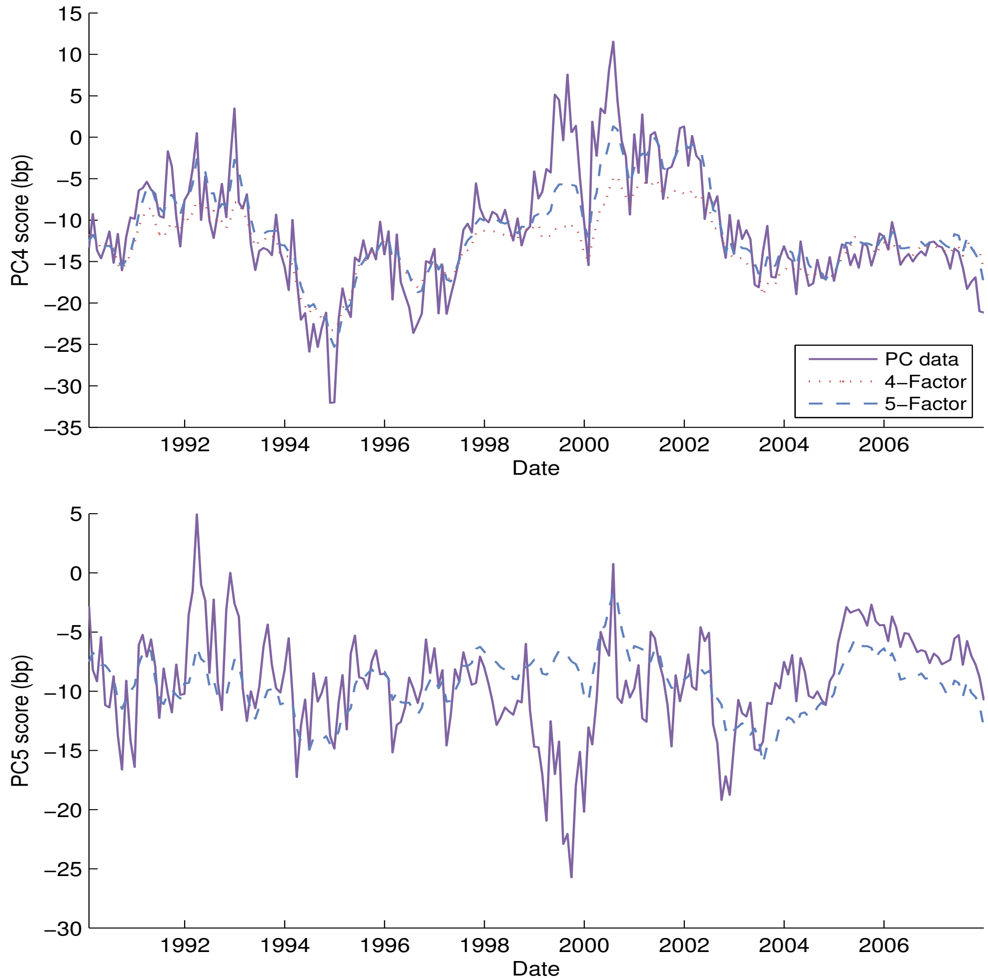

Figure 2: This figure plots the model implied and sample principal components for the fourth and fifth PCs when all PCs are assumed to be priced with normally distributed errors.

Further insight into how ML addresses this dual objective is revealed by the estimated half-lives of the pricing factors under $\mathbb{Q}$ (computed from the estimated $\lambda^{\mathbb{Q}}$). In the five-factor $GDTSM$ the $\mathbb{Q}$ half-lives of $\mathcal{P}_t$ are (in years) $(265, 16.2, 14.6, 0.52, 0.44)$, whereas they are $(24, 1.17, 0.78)$ in the three-factor model. Thus, $GDTSM(5)$ has three pricing factors with long $\mathbb{Q}$ half-lives, compared to only one in $GDTSM(3)$. Moreover, the half-life of $PC1$ in $GDTSM(5)$ is many times longer than its counterpart in $GDTSM(3)$. One consequence of these differences is that the properties of the model-implied long-term bonds (with maturities beyond 15 years) are wholly implausible in $GDTSM(5)$ (e.g., the thirty-year rate reaches values of -50%). In contrast, yields on long-term bonds take reasonable values in $GDTSM(3)$.

level and slope portfolios.

We interpret this evidence as being symptomatic of over-fitting, having too many pricing factors. Models with a lower dimensional $\mathcal{P}$ (say three) that allow higher order PCs to predict excess returns (and hence risk premiums) are easily developed along the lines of the modeling strategy in Joslin, Priebsch and Singleton (2009b). At a minimum, for our data set, the filtered higher-order pricing factors should not be interpreted literally as the PCs of the yield curve in models with $N > 3$.

Does the accommodation of filtering substantially increase the computational complexity of estimation using the JSZ normalization? The parameters $(K_{0,\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ and σ_{pricing} are now included as part of the parameter search. As we argued for $\Sigma_{\mathcal{P}}$ in the Case **RP** case, we obtain very accurate starting points for $(K_{0,\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ *irrespective of any inaccuracies in* $(r_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$. The additional cost of computing the Kalman filter as well as the lack of concentration of the likelihood function results in estimation times of approximately 10.4 seconds and, as without filtering, virtually all local optima are identical to within set tolerances. Using the results of the RP estimation as a starting point for the RF estimation decreased the estimation time to approximately 8.7 seconds. Thus, under the JSZ normalization, the estimation remains very fast even when all yields are priced with errors.

7 Macro *GDTSM*s

We now extend our analysis to a class of N-factor macro-dynamic term structure models in which the observable pricing factors $\mathcal{P}_t$ are comprised of K macroeconomic variables M_t and $N - K$ portfolios of bond yields $\mathcal{P}_{yt}$. Examples of this setup include Ang, Piazzesi and Wei (2003) ($\mathcal{P}_{yt}' = (y_{t,0.25y}^o, y_{t,5y}^o - y_{t,0.25y}^o)$ and M_t is the growth rate of real GDP), Garcia and Lugar (2009) (same $\mathcal{P}_{yt}$ and M_t includes the return on the market portfolio for equities, the inflation rate, and the growth rate of aggregate real consumption), and Jardet, Monfort and Pegoraro (2009) ($\mathcal{P}_{yt}' = (y_{t,0.25y}^o, y_{t,10y}^o)$ and M_t is the logarithm of real GDP). These studies assume that two yields are priced perfectly by their *GDTSM*s, and that the K macro variables are measured without error. Moreover, their models imply that M_t is spanned by any K linearly independent portfolios of the model-implied yields y_t. This follows directly from the observations that y_t is an affine function of $\mathcal{P}_t$, M_t is part of $\mathcal{P}_t$, and so the pricing model can be inverted to express M_t as an affine function of y_t.

Consistent with these studies we consider the case:

Case MSP: The K macro variables M_t and $N - K$ portfolios of bond yields $\mathcal{P}_{yt}$ are measured without error by the macro-*GDTSM*. Additionally, the macro variables are spanned by K portfolios of the model-implied yields y_t.

Note that Case **MSP** does not presume that M_t is spanned by the observed yields y_t^o, but rather only by model-implied yields.[28] However we view the spirit of models falling under Case **MSP** as capturing their macroeconomic content through their choices of $\mathcal{P}$, with the pricing errors $y_t^o - y_t$ being true measurement errors or unmodeled market microstructure effects. Models in which M_t is not spanned by y_t are conceptually very different in that the unspanned components of M_t are central economic factors underlying variation in risk premiums in bond markets. Joslin, Priebsch and Singleton (2009b) develop a family of macro-$DTSM$s with unspanned (by y_t) macro variables which they and Wright (2009) use to study macro-financial linkages in bond markets.

Before formally extending the JSZ normalization to Case **MSP**, we illustrate the mapping between the yield-based JSZ normalization and a canonical macro-$DTSM$ with M_t being part of $\mathcal{P}_t$. Suppose that a macro-$DTSM$ has as its pricing factors $\mathcal{P}_t' = (y^o_{t,0.5y}, y^o_{t,2y}, \pi_t)$, where π_t is the inflation rate, $y^o_{t,0.5y} = y_{t,0.5y}$, and $y^o_{t,2y} = y_{t,2yr}$. Under Case **MSP**, there is a unique mapping between a yield-based JSZ normalization and the macro-$DTSM$ with state $\mathcal{P}_t$. To see this we first invoke Theorem 1 to obtain a canonical $GDTSM$ with pricing factors $\hat{\mathcal{P}}_t' = (y^o_{t,0.5y}, y^o_{t,2y}, y_{t,10y})$ (equivalently, $\hat{\mathcal{P}}_t' = (y_{t,0.5y}, y_{t,2y}, y_{t,10y})$). Next, the spanning condition of **MSP** implies that inflation is an affine function of $\hat{P}$,

$$\pi_t = \gamma_0 + \gamma_1 \cdot \hat{\mathcal{P}}_t. \tag{35}$$

Thus, applying the invariant rotation

$$\mathcal{P}_t \equiv \begin{pmatrix} y_{t,0.5y} \\ y_{t,2y} \\ \pi_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \gamma_0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \gamma_{11} & \gamma_{12} & \gamma_{13} \end{pmatrix} \hat{\mathcal{P}}_t \tag{36}$$

to the canonical model with pricing factors $\hat{\mathcal{P}}$, we obtain the equivalent canonical model

$$\begin{aligned} \Delta \mathcal{P}_t &= K^{\mathbb{Q}}_{0\mathcal{P}} + K^{\mathbb{Q}}_{1\mathcal{P}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \epsilon^{\mathbb{Q}}_t \\ \Delta \mathcal{P}_t &= K^{\mathbb{P}}_{0\mathcal{P}} + K^{\mathbb{P}}_{1\mathcal{P}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \epsilon^{\mathbb{P}}_t \\ r_t &= \rho_{0\mathcal{P}} + \rho_{1\mathcal{P}} \cdot \mathcal{P}_t. \end{aligned}$$

Together with (35), these equations define explicit mappings between the rotation-invariant parameters $(\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_\infty, \Sigma_{\mathcal{P}}, \gamma_0, \gamma_1)$ and $(K^{\mathbb{Q}}_{0\mathcal{P}}, K^{\mathbb{Q}}_{1\mathcal{P}}, \rho_{0\mathcal{P}}, \rho_{1\mathcal{P}})$.

[28] In the studies cited above, $y^o_{t,m} = y_{t,m}$ for some of the maturities (those in $\mathcal{P}_{yt}$, e.g., $y^o_{t,0.25y} = y_{t,0.25y}$ and $y^o_{t,5y} = y_{t,5y}$). So M_t is spanned by linear combinations of $\mathcal{P}_{yt}$ and the subset of model-implied yields y_t not included in $\mathcal{P}_{yt}$.

In general, since $\hat{\mathcal{P}}_t$ is not observed ($y_{t,10y}$, the model-implied yield, may differ from $y^o_{t,10y}$, the observed yield), projecting π_t onto $\hat{\mathcal{P}}_t$ does not exactly recover the parameters (γ_0, γ_1) that enter macro-$DTSM$s. Nevertheless, the OLS projection of π_t onto $(y^o_{t,0.5y}, y^o_{t,2y}, y^o_{t,10y})$ should provide reliable starting values as the model implies that the only source of residual variation is the measurement error $y^o_{t,10y} - y_{t,10y}$.

7.1 The JSZ Normalization for macro-*GDTSM*s Under MSP

Theorem 2. *Under MSP, for any fixed yield portfolio weights $P \in \mathbb{R}^{N\times J}$, any canonical macro-*GDTSM *is observationally equivalent to a unique macro-*GDTSM *in which (1) the first $(N-K)$ components of $\mathcal{P}_t$ are the first $(N-K)$ yield portfolios $\mathcal{P}_{yt}$, and (2) the remaining K components of $\mathcal{P}_t$ are the macro variables M_t; and*

$$
\begin{aligned}
\Delta \mathcal{P}_t &= K^{\mathbb{Q}}_{0\mathcal{P}} + K^{\mathbb{Q}}_{1\mathcal{P}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \epsilon^{\mathbb{Q}}_t && (37)\\
\Delta \mathcal{P}_t &= K^{\mathbb{P}}_{0\mathcal{P}} + K^{\mathbb{P}}_{1\mathcal{P}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \epsilon^{\mathbb{P}}_t && (38)\\
r_t &= \rho_0 + \rho_1 \cdot \mathcal{P}_t && (39)\\
M_t &= \gamma_0 + \gamma_1'(P y_t), && (40)
\end{aligned}
$$

where $K^{\mathbb{Q}}_{0\mathcal{P}}, K^{\mathbb{Q}}_{1\mathcal{P}}, \rho_{0\mathcal{P}}$ and $\rho_{1\mathcal{P}}$ are explicit functions of $(\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_\infty, \Sigma_{\mathcal{P}}, \gamma_0, \gamma_1)$. For arbitrary P, our canonical form is parametrized by $\Theta^{\mathcal{P}} = (\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_\infty, K^{\mathbb{P}}_{0\mathcal{P}}, K^{\mathbb{P}}_{1\mathcal{P}}, \Sigma_{\mathcal{P}}, \gamma_0, \gamma_1)$.

For brevity, we omit the proof which is very similar to the proof of Theorem 1. The key results of Section 3 carry over to Case **MSP**.

Proposition 4. *In Case* **MSP***, the ML estimates of the $\mathbb{P}$ parameters $(K^{\mathbb{P}}_{0\mathcal{P}}, K^{\mathbb{P}}_{1\mathcal{P}})$ are given by the OLS estimates of the conditional mean of $\mathcal{P}_t$.*

The proof of this proposition is the same as before, since we can write the likelihood as

$$
L(\mathcal{P}_t, \mathcal{Y}^o_t | y^o_{t-1}; \Theta) = L(\mathcal{Y}^o_t | \mathcal{P}_t; K^{\mathbb{Q}}_{0\mathcal{P}}, K^{\mathbb{Q}}_{1\mathcal{P}}, r^{\mathbb{Q}}_\infty, \Sigma_{\mathcal{P}}, \gamma_0, \gamma_1) \times L(\mathcal{P}_t | \mathcal{P}_{t-1}; K^{\mathbb{P}}_{1\mathcal{P}}, K^{\mathbb{P}}_{0\mathcal{P}}, \Sigma_{\mathcal{P}}), \quad (41)
$$

where $\mathcal{P}_t$ now incorporates both yield and macro information.

7.2 On the Estimation of Macro-$DTSM$s in Case MSP

Proposition 4 implies that ML estimates of the $\mathbb{P}$ parameters in (38) are the OLS estimates. ML estimates of the remaining parameters $(\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_\infty, \Sigma_{\mathcal{P}}, \gamma_0, \gamma_1)$ are obtained by maximizing the joint likelihood (41). Starting values that are typically close to the global optima for the parameters $(\Sigma_{\mathcal{P}}, \gamma_0, \gamma_1)$ are available from OLS regressions and, as in the preceding yield-based analysis, one often has informed priors over the parameters $(\lambda^{\mathbb{Q}}, r_{\infty}^{\mathbb{Q}})$. Thus, again, ML estimation effectively involves searching over $N+1$ parameters in an N-factor model.

Instead of proceeding with ML estimation, Ang, Piazzesi and Wei (2003) propose a two-step approach. First, they estimate the parameters of (38) by OLS, and they use the fitted residuals to estimate $\Sigma_{\mathcal{P}}$. Then they estimate the parameters governing their market prices of risk– equivalently the parameters $(K_{0\mathcal{P}}^{\mathbb{Q}}, K_{1\mathcal{P}}^{\mathbb{Q}})$ of (37)– by minimizing the sum of squared differences between the historical and model-implied bond yields. Garcia and Lugar (2009) and Jardet, Monfort and Pegoraro (2009) adopt the same estimation strategy.

As these authors note, this two-step approach to estimation is consistent but inefficient. However, an immediate implication of Proposition 4 is that *their first-step estimates of* $(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$ *are in fact the* ML *estimates of the conditional* $\mathbb{P}$*-mean of* $\mathcal{P}$. It is their estimators of $(\Sigma_{\mathcal{P}}, K_{0\mathcal{P}}^{\mathbb{Q}}, K_{1\mathcal{P}}^{\mathbb{Q}})$, those governing the risk premiums, that are inefficient.

Furthermore note that, in the second stage of their estimation strategy, these authors are searching over the $N(N+1)$ parameters governing their market prices of risk.[29] As they parameterized their models, one typically has diffuse priors about the appropriate values for these parameters. In contrast, within the JSZ normalization, fully efficient ML estimates can be obtained by solving what turns out to be a search problem over an effectively much lower dimensional parameter space $(N+1)$ and, thereby, standard optimization algorithms converge extremely quickly to the global optimum of the likelihood function.

Analogously to Section 4, restrictions that affect only the $\mathbb{Q}$ distribution of $\mathcal{P}$ have no affect on the $\mathbb{P}$ distribution of $\mathcal{P}$. This irrelevance of $\mathbb{Q}$ restrictions is broken by constraints on the dimensionality of the model-implied risk premiums.

Finally, by construction, in the two-step estimation procedure adopted in previous studies, the forecasts of future values of $\mathcal{P}$ based on the past history of $\mathcal{P}_t$ are necessarily the same as those from an unconstrained VAR. What Proposition 4 adds is the insight that, even if these authors had computed the fully efficient ML estimates of their model, the same irrelevance result applies. In particular, in Case **MSP**, the out-of-sample forecasts of the macro variables M_t based on past $\mathcal{P}$ are invariant to the imposition of no-arbitrage restrictions.

[29]The number of parameters governing the market prices of risk is smaller in the equilibrium version of the models examined in Garcia and Lugar (2009), since they are tied directly to preference parameters. Also, Jardet, Monfort and Pegoraro (2009) allow for additional lags of $\mathcal{P}$ in their VAR representation of the state.

8 Conclusion

We derive a new canonical form for Gaussian dynamic term structure models. This canonical form allows for (essentially) arbitrary observable portfolios of zero-coupon yields to serve as the state variable. This allows us to characterize the properties of a *GDTSM* in terms of salient observables rather than latent states. Additionally, the risk-neutral distribution is parsimoniously characterized by the eigenvalues, $\lambda^{\mathbb{Q}}$, of the drift matrix and the long run mean of the short rate, $r^{\mathbb{Q}}_{\infty}$. Our canonical form reveals that simple OLS regression gives the maximum likelihood estimates of the parameters governing the physical distribution of bond yields. This results remains true even if restrictions are imposed on the risk-neutral distribution of the economy. An immediate implication of this result is that constraints such as imposing the arbitrage-free Nelson Siegel model or imposing complex $\mathbb{Q}$-eigenvalues are irrelevant for forecasting bond yields. However, when one imposes structure on risk premia, such as the reduced-rank risk premium of Joslin, Priebsch and Singleton (2009b), a wedge from the unconstrained OLS estimates arises. Our canonical form allows us to easily overcome the challenge of empirical estimation of *GDTSM*s in the case of filtering. We also extend our canonical form to macro-*GDTSM*s to obtain a number of analogous results. Taken together, our results shed new light on estimation and interpretation of *GDTSM*s, and the effects of different specifications of the risk premiums and the risk-neutral distribution of the state on the dynamic properties of the yield curve and macro-economic variables.

Appendices

A Bond Pricing in *GDTSM*s

Under (1–3), the price of an m-year zero-coupon bond is given by

$$D_{t,m} = E_t^{\mathbb{Q}}[e^{-\sum_{i=0}^{m-1} r_{t+i}}] = e^{\mathcal{A}_m + \mathcal{B}_m \cdot X_t}, \tag{A1}$$

where $(\mathcal{A}_m, \mathcal{B}_m)$ solve the first-order difference equations

$$\mathcal{A}_{m+1} - \mathcal{A}_m = K_0^{\mathbb{Q}'} \mathcal{B}_m + \frac{1}{2} \mathcal{B}_m' H_0 \mathcal{B}_m - \rho_0 \tag{A2}$$

$$\mathcal{B}_{m+1} - \mathcal{B}_m = K_1^{\mathbb{Q}'} \mathcal{B}_m - \rho_1 \tag{A3}$$

subject to the initial conditions $\mathcal{A}_0 = 0, \mathcal{B}_0 = 0$. See, for example, Dai and Singleton (2003). The loadings for the corresponding bond yield are $A_m = -\mathcal{A}_m/m$ and $B_m = -\mathcal{B}_m/m$.

B Invariant Transformations of *GDTSM*s

As in DS, given the *GDTSM* with parameters as in (1–3) and latent state X_t, if we may apply the invariant transformation $\hat{X}_t = C + DX_t$. We then have an observationally equivalent *GDTSM* with latent state $\hat{X}_t$ and parameters given by

$$K_{0\hat{X}}^{\mathbb{Q}} = (DK_{0X}^{\mathbb{Q}} - DK_{1X}^{\mathbb{Q}} D^{-1} C, \tag{A4}$$

$$K_{1\hat{X}}^{\mathbb{Q}} = DK_{1X}^{\mathbb{Q}} D^{-1}, \tag{A5}$$

$$\rho_{0\hat{X}} = \rho_{0X} - \rho_{1X}' D^{-1} C \tag{A6}$$

$$\rho_{1\hat{X}} = (D^{-1})' \rho_{1X}, \tag{A7}$$

$$K_{0\hat{X}}^{\mathbb{P}} = (DK_{0X}^{\mathbb{P}} - DK_{1X}^{\mathbb{P}} D^{-1} C, \tag{A8}$$

$$K_{1\hat{X}}^{\mathbb{P}} = DK_{1X}^{\mathbb{P}} D^{-1}, \tag{A9}$$

$$H_{0\hat{X}} = D' H_{0X} D \tag{A10}$$

Given a parameter vector Θ, we denote the parameter vector of $\hat{X}_t$ as $C + D\Theta$.

C Proofs

C.1 Proof of Proposition 1

We require a slight variation of the standard Jordan canonical form of a square matrix which maintains all real entries and bears a similar relation to the real Schur decomposition and the Schur decomposition.

Definition 1. *We refer to the* **real ordered Jordan form** *of a square matrix $A \in \mathbb{R}^{n\times n}$ with eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ as*

$$A = J(\lambda) \equiv diag(J_1, J_2, \ldots, J_m),$$

and if λ_i is real

$$J_i = \begin{pmatrix} \lambda_i & 1 & \cdots & 0 \\ 0 & \lambda_i & \cdots & 0 \\ \vdots & \vdots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_i \end{pmatrix},$$

or if λ_i is strictly complex

$$J_i = \begin{pmatrix} R & I_2 & \cdots & 0 \\ 0 & R & \cdots & 0 \\ \vdots & \vdots & \ddots & I_2 \\ 0 & \cdots & 0 & R \end{pmatrix} \text{ with } R = \begin{pmatrix} real(\lambda_i) & -|imag(\lambda_i)| \\ |imagl(\lambda_i)| & real(\lambda_i) \end{pmatrix}$$

and where the blocks are in order of the eigenvalues.

We will be concerned only with the case that each eigenvalue has geometric multiplicty of one, and hence each eigenvalue is associated with only one block.

Proof of Proposition 1: We first prove the existence. Given θ, $\theta' = \theta + \left(K_1^{\mathbb{Q}}\right)^{-1} K_0^{\mathbb{Q}}$ has $(K_0^{\mathbb{Q}})' = 0$ (and under the assumption of non-zero eigenvalues, $K_1^{\mathbb{Q}}$ is invertible).[30] Thus without loss of generality $K_0^{\mathbb{Q}} = 0$. By standard linear algebra, there exists U so that $UK_1^{Q}U^{-1}$ is in the standard Jordan normal form. By Lemma 1 of the supplement to this paper (see Joslin, Singleton and Zhu (2009a)), we can further transform to have the real ordered form of Defintion 1. We consider separately the cases of real and imaginary Jordan blocks. Note that by Joslin (2007), each eigenvalue has geometric multiplicty one and this is associated with only one block.

[30] We use the notation $A\theta + b$ for some $A \in \mathbb{R}^{n\times n}$ (invertible) and $b \in \mathbb{R}^n$ to denote the action on the parameter space when the variable X_t is replaced by $AX_t + b$: $A\theta + b \equiv (\rho_0 - \rho_1^{'} A^{-1}b, (A^{-1})^{'}\rho_1, AK_0^{\mathbb{P}} - AK_1^{\mathbb{P}}A^{-1}b, AK_1^{\mathbb{P}}A^{-1}, AK_0^{\mathbb{Q}} - AK_1^{\mathbb{Q}}A^{-1}b, AK_1^{\mathbb{Q}}A^{-1}, A\Sigma)$.

(i) J_i corresponds to real eigenvalues with algebraic multiplicity k (k could be 1). Then J_i is $k \times k$ matrix

$$J_i = \begin{pmatrix} \lambda_i & 1 & \cdots & 0 \\ 0 & \lambda_i & \cdots & 0 \\ \vdots & \vdots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_i \end{pmatrix}$$

Write $\rho_{1i} = (\rho_{1i}^{(1)}, \dots, \rho_{1i}^{(k)})$. Note that $\rho_{1i}^{(1)} \neq 0$, for otherwise we can do without state variable $Y_{ti}^{(1)}$, contradicting our assumption of an N-factor model. Then one can check that $BJ_iB^{-1} = J_i$ if and only if B has the form

$$B = \begin{pmatrix} B_1 & B_2 & \cdots & B_k \\ 0 & B_1 & \cdots & B_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_1 \end{pmatrix}. \tag{A11}$$

The matrix

$$B_i = \begin{pmatrix} \rho_{1i}^{(1)} & \rho_{1i}^{(2)} - \rho_{1i}^{(1)} & \cdots & \rho_{1i}^{(k)} - \rho_{1i}^{(k-1)} \\ 0 & \rho_{1i}^{(1)} & \cdots & \rho_{1i}^{(k-1)} - \rho_{1i}^{(k-2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{1i}^{(1)} \end{pmatrix}$$

thus maintains the real Jordan structure and has $(B_i^{-1})'\rho_{1i} = \iota$.

(ii) J_i corresponds to complex eigenvalues with multiplicity k, which means J_i is $2k \times 2k$, as complex eigenvalues come in pairs. In this case

$$J_i = \begin{pmatrix} R & I_2 & \cdots & 0 \\ 0 & R & \cdots & 0 \\ \vdots & \vdots & \ddots & I_2 \\ 0 & \cdots & 0 & R \end{pmatrix}$$

with

$$R = \begin{pmatrix} \beta & -\mu \\ \mu & \beta \end{pmatrix}$$

The proof is very analogous to the real case, as the individual steps are the same but require lemmas to verify the intuitive steps hold with (2×2) block matrices replacing scalars. The details are available on request from the authors.

The uniqueness, follows from the uniqueness of an ordered Jordan decomposition and the fact that (1) the Jordan decomposition is maintained only by a block matrix where B has form (A11) (or () for the complex case) and (2) the only such B that satisifies $B^{'}\iota = \iota$ is $B = I$. Furthermore, if $a \neq 0$, and $\theta \in \Theta_{JSZ}$, $K_0^{\mathbb{Q}}(\theta + a) \neq \vec{0}$. This establishes the uniqueness and proves the result.

D Details of Step 3 in the Proof of Theorem 1

We have established that every *GDTSM* is observationally equivalent to a Jordan normalized model and the transformation relating the two models is found by computing the associated portfolio loadings:

$$\mathcal{G}_{\mathcal{P}}^{P} = \{A_P(\Theta^J) + B_P(\Theta^J)^{'}\Theta^J : \Theta^J \in \mathcal{G}_J\}. \tag{A12}$$

Observe that since $\rho_1^J = \iota$, $B_P(\Theta^J)$ depends only on $\lambda^{\mathbb{Q}}$; let us denote $B_{\lambda^{\mathbb{Q}}} \equiv B_P(\Theta^J)$. Similarly, let us denote $A_{\lambda^{\mathbb{Q}},\rho_0,\Sigma} \equiv A_P(\Theta^J)$. Since, for any $\lambda^{\mathbb{Q}}$, the map $s_{\lambda^{\mathbb{Q}}}(\Sigma) = B_{\lambda^{\mathbb{Q}}}^{-1}\Sigma$ is a bijection[31], we can reparametrize the conditional volatility by

$$\mathcal{G}_{\mathcal{P}}^{P} = \{A_{\Theta^J} + B_{\Theta^J}\Theta^J : \Theta^J = (0, J(\lambda^{\mathbb{Q}}), r_\infty^{\mathbb{Q}}, \iota, K_{0J}^{\mathbb{P}}, K_{1J}^{\mathbb{P}}, s_{\lambda^{\mathbb{Q}}}(\Sigma_{\mathcal{P}}))\}. \tag{A13}$$

Here we use $\Sigma_{\mathcal{P}}$ to denote the parameterization since, for $\Theta^J = (0, J(\lambda^{\mathbb{Q}}), r_\infty^{\mathbb{Q}}, \iota, K_{0J}^{\mathbb{P}}, K_{1J}^{P}, B_{\lambda^{\mathbb{Q}}}^{-1}\Sigma_{\mathcal{P}})$, the transformed model $A_{\Theta^J} + B_{\Theta^J}\Theta^J$ (which has $\mathcal{P}_t$ as the factors since it is in $\mathcal{G}_{\mathcal{P}}$) has innovation volatility of $B_{\lambda^{\mathbb{Q}}}B_{\lambda^{\mathbb{Q}}}^{-1}\Sigma_{\mathcal{P}} = \Sigma_{\mathcal{P}}$.

Define the bijective map k on $\mathbb{R}^N \times \mathbb{R}^{N\times N}$ by:

$$k_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}}(K_0, K_1) = \left(B_{\lambda^{\mathbb{Q}}}K_0 - B_{\lambda^{\mathbb{Q}}}K_1B_{\lambda^{\mathbb{Q}}}^{-1}A_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}},\ B_{\lambda^{\mathbb{Q}}}K_1B_{\lambda^{\mathbb{Q}}}^{-1}\right). \tag{A14}$$

The function k maps (K_0, K_1) under the change of variables $X_t \mapsto A_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}} + B_{\lambda^{\mathbb{Q}}}X_t$. Using k, we further reparametrize $\mathcal{G}_{\mathcal{P}}^{P}$ by

$$\mathcal{G}_{\mathcal{P}}^{P} = \{A_{\Theta^J} + B_{\Theta^J}\Theta^J : \Theta^J = (0, J(\lambda^{\mathbb{Q}}), r_\infty^{\mathbb{Q}}, \iota, k_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}}^{-1}(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}), s_{\lambda^{\mathbb{Q}}}(\Sigma_{\mathcal{P}}))\}. \tag{A15}$$

This gives our desired reparameterization of $\mathcal{G}_{\mathcal{P}}^{P}$ by $\Theta_{JSZ} = (\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}})$. This is because, for $\Theta^J = \left(0, J(\lambda^{\mathbb{Q}}), r_\infty^{\mathbb{Q}}, \iota, k_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}}^{-1}(K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}), s_{\lambda^{\mathbb{Q}}}(\Sigma_{\mathcal{P}})\right)$,

$$\begin{aligned} \Theta^{\mathcal{P}} &= A_{\Theta^J} + B_{\Theta^J}\Theta^J \\ &= \left(k_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}}(0, J(\lambda^{\mathbb{Q}})),\ r_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}}(r_\infty^{\mathbb{Q}}, \iota),\ K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}}\right), \end{aligned} \tag{A16}$$

[31]For simplicity, we denote the Cholesky factorization, Σ, but we have in mind the covariance $\Sigma\Sigma^{'}$.

where $r_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}}$ maps (ρ_0,ρ_1) under the change of variables $X_t \mapsto A_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}} + B_{\lambda^{\mathbb{Q}}}X_t$:

$$r_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}}(\rho_0,\rho_1) = \left(\rho_0 - \rho_1^{'} B_{\lambda^{\mathbb{Q}}} A_{\lambda^{\mathbb{Q}},r_\infty^{\mathbb{Q}},\Sigma_{\mathcal{P}}},\ B_{\lambda^{\mathbb{Q}}}^{'}\rho_1\right). \tag{A17}$$

E Fully Flexible GDTSM

In the body of this paper, we assume that (1) the eigenvalues under $\mathbb{Q}$ are non-zero and (2) bond yields follow a Markov process. The first assumption is purely for simplicity. When there is a zero eigenvalue, the models take the same form except that the non-stationarity now allows one to include a deterministic time-trend in the short rate within the fully flexible model. This might seem to introduce a new equivalence class of models. In fact, however, by letting $\lambda_1 \to 0^-$ and $r_\infty^{\mathbb{Q}} = -\alpha/\lambda_1$ (which will go to ∞ as $\lambda_1 \to 0^-$), we see that the 0 eigenvalue model with time trend α can be viewed as the limit of a strictly negative eigenvalue model. With these generalizations, our framework now captures all possible cases of positive, negative, zero, or complex eigenvalues with arbitrary geometric multiplicities.

The non-Markovian case does introduce new issues. For example, we could suppose that the yield curve is 3-dimensional (i.e., there are 3 factors that are relevant for pricing so that $N^{\mathbb{Q}} = 3$), but that there are additional factors that are relevant for predicting future bond prices (i.e. $N > 3$). Joslin, Priebsch and Singleton (2009b) make such an assumption. However, they also assume that, although the yield curve alone does not follow a Markov process, when the yield curve is augmented with macro variables the joint process is Markov. More generally, one could augment the *GDTSM* given by (1–3) with macro variables M_t through the expression $M_t = \delta_0 + \delta_1' X_t$. For a non-degenerate model where the maximal rank of $[B_P, \delta_1]$ is less than the dimension of X_t, there will be some factor which predicts future yields and/or macro variables but that is not determined by the current y_t^o and M_t. For example, we may have $N^{\mathbb{Q}} = 3$ while $N = 4$ so that there exists a latent factor which predicts future bond returns but is not determined by current bond yields. In the absence of identification through macro variables, such a factor must be filtered. An analog of our main normalization continues to apply, though we no longer have informed priors over all of the parameters. We plan to analyze such models in more detail in future research.

References

Ang, A., M. Piazzesi, and M. Wei, "What Does the Yield Curve Tell us About GDP Growth?," Working Paper, Columbia University 2003.

Campbell, J. and R. Shiller, "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*, 1991, *58*, 495–514.

Christensen, Jen H.E., Francis X. Diebold, and Glenn D. Rudebusch, "The Affine Arbitrage-Free Class of Nelson Siegel Term Structure Models," *Working Paper, Federal Reserve Bank of San Francisco*, 2007.

_ , _ , and _ , "An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model," *The Econometrics Journal*, 2009.

Cochrane, J. and M. Piazzesi, "Bond Risk Premia," *American Economic Review*, 2005, *95*, 138–160.

_ and _ , "Decomposing the Yield Curve," Technical Report, Stanford University 2008.

Collin-Dufresne, P., R. Goldstein, and C. Jones, "Identification of Maximal Affine Term Structure Models," Technical Report, forthcoming, *Journal of Finance* 2008.

Cox, John C. and Chifu Huang, "Optimum Consumption and Portfolio Policies When Asset Prices Follow a Diffusion Process," *Journal of Economic Theory*, 1989, *49*, 33–83.

Dai, Q. and K. Singleton, "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 2000, *55*, 1943–1978.

_ and _ , "Term Structure Dynamics in Theory and Reality," *Review of Financial Studies*, 2003, *16*, 631–678.

Diebold, F. and C. Li, "Forecasting the term structure of government bond yields," *Journal of Econometrics*, 2006, *130*, 337–364.

Duffee, G., "Idiosyncratic Variation in Treasury Bill Yields," *Journal of Finance*, 1996, *51*, 527–552.

_ , "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 2002, *57*, 405–443.

_ , "Information In (and Not In) The Term Structure," Technical Report, Working Paper, Johns Hopkins University 2008.

_ , "Forecasting with the Term Structure: the Role of No-Arbitrage," *Working Paper, UC Berkeley*, 2009.

— **and R. Stanton**, “Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation,” *Journal of Financial Econometrics*, 2007, *6* (1), 108–142.

Duffie, D. and R. Kan, “A Yield-Factor Model of Interest Rates,” *Mathematical Finance*, 1996, *6*, 379–406.

Garcia, R. and R Lugar, “Risk Aversion, Intertemporal Substitution, and the Term Structure of Interest Rates,” Technical Report, Cirano Working Paper 2009s-20 2009.

Jardet, C., A. Monfort, and F. Pegoraro, “No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premiums and GDP Growth,” Technical Report, Banque de France 2009.

Joslin, S., “Pricing and Hedging Volatility in Fixed Income Markets,” Technical Report, Working Paper, MIT 2007.

— **, K. Singleton, and H. Zhu**, “Supplement to ‘A New Perspective on Gaussian *DTSM*s’,” Technical Report, Sloan School, MIT 2009.

— **, M. Priebsch, and K. Singleton**, “Risk Premium Accounting in Macro-Dynamic Term Structure Models,” Technical Report, Stanford University 2009.

Langetieg, T., “A Multivariate Model of the Term Structure,” *Journal of Finance*, 1980, *35*, 71–97.

Piazzesi, M., “Bond Yields and the Federal Reserve,” *Journal of Political Economy*, 2005, *113*, 311–344.

— , “Affine Term Structure Models,” 2006.

Vasicek, O., “An Equilibrium Characterization of the Term Structure,” *Journal of Financial Economics*, 1977, *5*, 177–188.

Wright, J., “Empirical Evidence on Term Premiums from an International Panel Dataset,” Technical Report, Johns Hopkins University 2009.

Zellner, A., “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias,” *Journal of the American Statistical Association*, 1962, *57*, 348–368.