The Conditional Distribution of Bond Yields Implied by Gaussian Macro-Finance Term Structure Models

Scott Joslin* Anh Le† Kenneth J. Singleton‡

First draft: June, 2010
This draft: September 24, 2010

Abstract

This paper explores the impact of the no-arbitrage structure of canonical Gaussian macro-finance term structure models (*GMTSM*s) on maximum likelihood estimates of their implied conditional distributions of the macro risk factors and bond yields. We show that, for the typical yield curves and macro variables studied in this literature, the maximum likelihood estimate of a canonical *GMTSM*-implied joint distribution is virtually identical to its counterpart estimated from the economic-model-free vector-autoregression (VAR). The practical implication of this finding is that a canonical *GMTSM* does not offer any new insights into economic questions regarding the historical distribution of the macro risk factors and yields, over and above what one can learn from a VAR. In particular, the discipline of a canonical *GMTSM* is empirically inconsequential for out-of-sample forecasts of bond yields or macro factors, tests of the expectations theory of the term structure, or analyses of impulse response functions of bond yields and macro factors. Certain classes of constraints may break these irrelevancy results, and we discuss what is known from the literature about this possibility.

*MIT Sloan School of Management, `sjoslin@mit.edu`
†Kenan-Flagler Business School, University of North Carolina at Chapel Hill, `anh_le@unc.edu`
‡Graduate School of Business, Stanford University, and NBER, `kenneths@stanford.edu`

1 Introduction

Gaussian macro dynamic term structure models (*GMTSM*s) have been used to examine whether yield curve information improves forecasts of macro variables (Ang, Piazzesi, and Wei (2006)), and to quantify the proportion of bond yield variation attributable to macro risk factors (Bikbov and Chernov (2010)). Additionally, Ang, Dong, and Piazzesi (2007), Chernov and Mueller (2009), Chun (2009), and Smith and Taylor (2009) use *GMTSM*s to explore the structure of monetary policy (Taylor-style) rules implicit in the term structure of bond yields. Separately, Joslin, Priebsch, and Singleton (2010) quantify the effects of unspanned macro risks on forward term premiums within a *GMTSM*. In principle, by exploiting the increased econometric efficiency from enforcing no-arbitrage, a *GMTSM* can shed new light on the joint distribution of the macro factors and bond yields.

This paper explores the impact of the no-arbitrage structure of canonical Gaussian macro-finance term structure models (*GMTSM*s) on maximum likelihood (ML) estimates of the conditional distribution of the macro risk factors and bond yields. We show that, for the typical yield curves and macro variables studied in this literature, the ML estimate of the *GMTSM*-implied joint distribution is virtually identical to its counterpart estimated from the economic-model-free vector-autoregression (VAR). The practical implication of this finding is that a canonical *GMTSM* does not offer any new insights into economic questions regarding the historical (data-generating) distribution of macro variables and yields, over and above what one can learn from a VAR. In particular, the imposition of the no-arbitrage structure of a canonical *GMTSM* is empirically inconsequential for out-of-sample forecasts of bond yields or macro factors, tests of the expectations theory of the term structure, or analyses of impulse response functions of bond yields and macro factors. The same is true for the special case of Gaussian dynamic term structure models fit to yields alone (*GYTSM*s).

These irrelevancy results apply to the maximally flexible *canonical* forms of *GMTSM*s. Certain types of restrictions, when imposed in combination with the no-arbitrage restrictions of a *GMTSM*, may increase the efficiency of ML estimators relative to those of the unconstrained VAR. For instance, the *GMTSM* with constrained market prices of risk studied by Joslin, Priebsch, and Singleton (2010) has very different dynamic properties than an unconstrained VAR. On the other hand, focusing on conditional means, Joslin, Singleton, and Zhu (2010) provide examples where forecasts from constrained *GYTSM*s and their associated VARs are identical. Most studies of *GMTSM*s have left open the question of whether their particular choices of constraints materially improved the efficiency of their ML estimators. Our concluding section reviews what we know about this important question.

We begin our exploration of these issues with *GMTSM*s that incorporate macroeconomic risk factors directly into the specifications of the short-term interest rate and hence, through standard arbitrage-pricing relations, into the prices of bonds of all maturities. To derive our irrelevance results we develop a canonical form for the family of $\mathcal{N}$-factor *GMTSM*s in which $\mathcal{M}$ of the factors ($\mathcal{M} < \mathcal{N}$) are macro variables. Under the assumption that the $\mathcal{M}$ macro factors incrementally affect bond prices beyond the other $\mathcal{N} - \mathcal{M}$ factors, we show that every canonical *GMTSM* in this family is observationally equivalent to our canonical model in which the pricing factors are the $\mathcal{M}$ macro factors and the first $\mathcal{N} - \mathcal{M}$ principal components (PC) of bond yields, and the macro factors are affine functions of the first $\mathcal{N}$ PCs of bond yields. This theorem has several powerful implications for the properties of ML estimators of arbitrage-free *GMTSM*s with macro risk factors.[1]

First, if the yield-based risk factors are priced perfectly by the *GMTSM*, then optimal forecasts of the risk factors based on ML estimates of a canonical *GMTSM* are always identical to the forecasts produced by OLS estimates of the corresponding unconstrained VAR.[2] Though in the presence of measurement errors on yields these forecasts are no longer exactly equal, our theoretical results show that they will nearly coincide and, in fact, this is what we find for a variety of *GMTSM*s. Our analysis both complements and provides a theoretical context for Duffee (2009)'s finding that yield-based term structure models do not materially improve out-of-sample forecasts of bond yields. Moreover, it shows that any differences must arise entirely as a consequence of the modeler's choice of measurement-error distribution, and not because of any inherent features of a canonical *GMTSM*.

Turning to the conditional variances of the risk factors, theoretically, the diffusion invariance assumption of no arbitrage links the conditional second moments of the $\mathbb{P}$ and $\mathbb{Q}$ distributions of bond yields. Consequently ML estimates of second moments within a *GMTSM* are typically more efficient than those from a VAR. From the analytic first-order conditions for the ML estimator of the covariance of innovations derived subsequently, we show that the efficiency gains from ML estimation of a *GMTSM* are linked to the model-implied ratios of the sample means of the pricing errors to their variances. As these ratios approach zero, the ML estimators of the *GMTSM*-implied second moments of the risk factors approach their counterparts from OLS estimation of an unconstrained VAR. Though these "average-to-variance" ratios are rarely zero, for typical yield-curve data they are small. So the practical relevance of this second result is that the no-arbitrage restrictions inherent in a *GMTSM* are unlikely to lead to large efficiency gains for ML estimators of second moments of bond yields over those based on an unconstrained VAR.

Taken together, these results imply that the imposition of the no-arbitrage restrictions within a canonical *GMTSM* will typically give virtually the same ML estimates of the conditional $\mathbb{P}$-distribution of the risk factors as those produced by their unconstrained VAR model. This, in turn, implies that any canonical *GMTSM*-based results on relationships among the risk factors that are fully described in terms of the parameters of their $\mathbb{P}$ distribution will be identical to (or nearly identical to) what is obtained from a model-free VAR. This striking result is based on a mix of theoretical results that hold exactly for each sample of yields and on approximations that are accurate as the average-to-variance ratios of the model's pricing errors approach zero. Therefore, it is an empirical matter as to how close the *GMTSM*-implied distribution and the distribution implied by a factor model that does not impose no-arbitrage restrictions line up in practice.

Initially, we explore the empirical relevance of no-arbitrage restrictions for the joint

[1] This canonical form is the counterpart for *GMTSM*s of the form for *GYTSM*s developed by Joslin, Singleton, and Zhu (2010). As will be apparent subsequently, our analysis of *GMTSM*s also leads to several new observations about the empirical properties of models fit to yields alone.

[2] Joslin, Singleton, and Zhu (2010) show the analogous result for yield-only models. See Dai and Singleton (2003) and Piazzesi (2010) for recent surveys of these models.

distribution of macro and yield-based risk factors within a four-factor model with the factors being measures of output growth and inflation and the first two PCs (as in Bikbov and Chernov (2010)). Attention is then turned to two variants of three-factor *GMTSM*s: a model in which the risk factors are a measure of real output growth and the first two PCs of bond yields (in the spirit of Ang, Piazzesi, and Wei (2006)); and one in which the risk factors are output growth, inflation, and the first PC of bond yields (in the spirit of Ang and Piazzesi (2003)). There is a substantial deterioration in the accuracy in pricing individual bonds in the latter compared to the former models (root mean square errors in individual yields rise above 40 basis points). Nevertheless, for both models the conditional $\mathbb{P}$-distributions of the risk factors are nearly identical whether one imposes no arbitrage or applies filtering. These results illustrate a more general implication of our theoretical analysis: the near irrelevance of no-arbitrage restrictions does not rely on accurate pricing of individual bonds.

A different application of our results is to the literature assessing whether yield-based *GYTSM*s can replicate the substantial evidence against the expectations theory of the term structure found in U.S. data (e.g., Campbell and Shiller (1991)). The literature has shown that *GYTSM*-implied coefficients from Campbell-Shiller style regressions are close to their sample counterparts (e.g., Dai and Singleton (2002)). However our theoretical results suggest that it is not the dynamic term structure models per se that are resolving the expectations puzzles. Rather, if the *GYTSM*-implied conditional distribution of bond yields is nearly identical to its counterpart implied by an unconstrained factor model, then it must be the latter that resolves the puzzles. That is, our propositions suggest that the no-arbitrage structure of the *GYTSM*s has played effectively no role in prior resolutions of this puzzle. We document that this is in fact the case within a three-factor *GYTSM* of Treasury yields.

These illustrations presume that (i) the risk factors follows a first-order Markov process and (ii), in *GMTSM*s, the macro-factors are spanned in the sense that they can be expressed as linear combinations of the PCs of bond yields. While these assumptions are consistent with virtually all of the continuous-time and most the discrete-time analyses of *GYTSM*s, several implementations of *GMTSM*s have relaxed these assumptions. We show that our analysis is robust to these extensions in the sense that the estimates of the canonical no-arbitrage model remain nearly identical to those of the VAR. Of independent interest, we also find that, for our datasets, the empirical evidence supports multiple lags under the historical distribution $\mathbb{P}$, but a first-order Markov structure under the pricing measure $\mathbb{Q}$. Accordingly, we develop a new family of canonical *GMTSM*s with this lag structure.

To fix notation, suppose that a *GMTSM* is to be evaluated using a set of J yields y_t, with $J \geq \mathcal{N}$, where $\mathcal{N}$ is the number of pricing factors. We introduce a fixed, full-rank matrix of portfolio weights $W \in \mathbb{R}^{J\times J}$ and define the "portfolios" of yields $\mathcal{P}_t \equiv W y_t$ and, for any $G \leq J$, we let $\mathcal{P}_t^G$ and W^G denote the first G portfolios in $\mathcal{P}_t$ and their associated weights. The modeler's choice of W will determine which portfolios of yields enter the *GMTSM* as risk factors and which additional portfolios are used in estimation. Throughout our analysis y_t and y_t^o denote the model-implied and observed yields, respectively. In some cases we will allow corresponding entries in y_t and y_t^o, or in $\mathcal{P}_t$ and $\mathcal{P}_t^o$, to differ from each other owing to measurement or pricing errors arising from illiquidity, bid/ask bounce, and other market microstructure effects.

2 A Canonical *GMTSM*

A canonical model for a family of *GMTSM*s is one in which the $\mathbb{Q}$ distribution of the risk factors is maximally flexible, with a minimal set of *normalizations* imposed to ensure econometric identification. The normalizations typically chosen obscure the implications of the no-arbitrage structure of Gaussian models for the conditional $\mathbb{P}$ distribution of the risk factors X_t.[3] Joslin, Singleton, and Zhu (2010) (hereafter JSZ) propose a new canonical form for yield-only *GYTSM*s that clarifies the effect of no-arbitrage restrictions on the conditional mean of X_t. To explore the implications of no-arbitrage for the entire conditional distribution of X_t in models with macro factors, we derive an analogous canonical form for *GMTSM*s.

We focus on the following set of $\mathcal{N}$-factor *GMTSM*s with $\mathcal{M}$ macro risk factors:

Definition 1. *Fix the set of $\mathcal{M}$ macroeconomic variables M_t and suppose that an $\mathcal{N}$-factor model is non-degenerate in the sense that all $\mathcal{M}$ macro factors distinctly contribute to the pricing of bonds after accounting for the remaining $\mathcal{L}$ factors L_t.*[4] *$GMTSM_{\mathcal{N}}(\mathcal{M})$ is the set of all $\mathcal{N}$-factor* GMTSMs *that are invariant affine transformations*[5] *of* GMTSM*s in which r_t, the one-period interest rate, is an affine function of $\mathcal{M}$ macro (M_t) and $\mathcal{L}$ ($\equiv \mathcal{N} - \mathcal{M}$) latent ($L_t$) pricing factors,*[6]

$$r_t = \rho_0 + \rho_{\mathcal{M}} \cdot M_t + \rho_{\mathcal{L}} \cdot L_t; \tag{1}$$

and the risk factors $X_t' \equiv (M_t', L_t')$ follow the Gaussian processes

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

$$\Delta X_t = K_{0X}^{\mathbb{P}} + K_{1X}^{\mathbb{P}} X_{t-1} + \sqrt{\Sigma_X}\epsilon_t^{\mathbb{P}} \text{ under } \mathbb{P}, \tag{3}$$

where $\epsilon_t^{\mathbb{Q}}$ and $\epsilon_t^{\mathbb{Q}}$ are distributed as $N(0, I)$, and the eigenvalues of $K_{1X}^{\mathbb{Q}}$ are all non-zero.[7]

[3] See Dai and Singleton (2000) and Collin-Dufresne, Goldstein, and Jones (2008) for examples of two (observationally equivalent) canonical latent factor, yield-only models. Most of the extant literature on *GMTSM*s has adapted the set of normalizations used by Dai and Singleton (2000) to the case where there are both latent and observed pricing factors. Pericoli and Taboga (2008) formalize this adaptation, and Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007), and Bikbov and Chernov (2010), among others, are specific implementations. While it is straightforward to impose the Pericoli-Taboga normalization scheme, as with the Dai-Singleton scheme for *GYTSM*s, the implications of no-arbitrage for the conditional distribution of X_t are far from transparent. Furthermore, the Pericoli-Taboga scheme inherits the computational challenges associated with estimation when using the Dai-Singleton normalization scheme.

[4] That is, we truly have an $\mathcal{N}$-factor model, one that does not reduce to an $\mathcal{N}^*$ ($< \mathcal{N}$) factor model owing to degeneracies in how the pricing factors affect bond yields. Formal conditions that guarantee the absence of such degenerate cases are discussed in Joslin (2006).

[5] See Dai and Singleton (2000) for the definition of invariant affine transformations. Such transformations lead to equivalent models in which the pricing factors $\tilde{\mathcal{P}}_t$ are obtained by applying affine transformations of the form $\tilde{\mathcal{P}}_t = C + D\mathcal{P}_t$, for nonsingular $\mathcal{N} \times \mathcal{N}$ matrix D.

[6] Studies that posit an r_t process of the form (1) include Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007), Bikbov and Chernov (2010), Chernov and Mueller (2009), and Smith and Taylor (2009).

[7] The case of a unit root in the $\mathbb{Q}$ process of the risk factors X_t is easily treated in a manner analogous to how JSZ treat a zero root in yield-only models. See their online supplement for details.

To construct our canonical form, we start by replacing the latent pricing factors L_t with $\mathcal{L}$ linear combinations of y_t. Specifically, for the first $\mathcal{N}$ entries $\mathcal{P}_t^{\mathcal{N}}$ of the modeler's chosen portfolios of yields $\mathcal{P}_t = W y_t$, the affine structure of a $GMTSM$ (Duffie and Kan (1996))) implies that

$$\mathcal{P}_t^{\mathcal{N}} = A_{\mathcal{P}}^{\mathcal{N}}(\Theta^{\mathbb{Q}}) + B_{\mathcal{P}}^{\mathcal{N}}(\Theta^{\mathbb{Q}}) X_t, \tag{4}$$

for a $\mathcal{N} \times 1$ vector $A_{\mathcal{P}}$ and $\mathcal{N} \times \mathcal{N}$ matrix $B_{\mathcal{P}}$ (both indexed by the choice of W) that depend only on the parameters $\Theta^{\mathbb{Q}}$ governing the $\mathbb{Q}$ distribution of X. Therefore, for any macro-$DTSM$ in $GMTSM_{\mathcal{N}}(\mathcal{M})$, the risk factors X_t can be replicated by portfolios of yields $\mathcal{P}_t^{\mathcal{N}}$ by inverting (4) to express X_t in terms of $\mathcal{P}_t^{\mathcal{N}}$:

$$\begin{pmatrix} M_t \\ L_t \end{pmatrix} = \begin{pmatrix} \gamma_0^W \\ \nu_0^W \end{pmatrix} + \begin{pmatrix} \gamma_1^W \\ \nu_1^W \end{pmatrix} \mathcal{P}_t^{\mathcal{N}}. \tag{5}$$

Next, we use the fact that all the macro-factors contribute to pricing to construct a canonical model for the family $GMTSM_{\mathcal{N}}(\mathcal{M})$ from a canonical model for the family of Gaussian $DTSM$s with pricing factors $\mathcal{P}_t^{\mathcal{N}}$. For this latter family we adopt the canonical form from Theorem 1 in JSZ which has $r_t = \rho_{0\mathcal{P}} + \rho_{1\mathcal{P}} \cdot \mathcal{P}_t^{\mathcal{N}}$ and

$$\Delta \mathcal{P}_t^{\mathcal{N}} = K_{0\mathcal{P}}^{\mathbb{Q}} + K_{1\mathcal{P}}^{\mathbb{Q}} \mathcal{P}_{t-1}^{\mathcal{N}} + \sqrt{\Sigma_{\mathcal{P}}} \epsilon_{\mathcal{P}t}^{\mathbb{Q}}, \tag{6}$$

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 the long-run $\mathbb{Q}$-mean of r_t, $r_\infty^{\mathbb{Q}}$, and the $\mathcal{N}$-vector of distinct, real eigenvalues of the feedback matrix $K_{1\mathcal{P}}^{\mathbb{Q}}$, $\lambda^{\mathbb{Q}}$.[8] Under this normalization, bond yields are fully characterized by the rotation-invariant risk-neutral parameters $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$.

Finally, we rotate this canonical form to an equivalent model in which the risk factors are M_t and $\mathcal{P}_t^{\mathcal{L}}$ (the first $\mathcal{L}$ entries of $\mathcal{P}^{\mathcal{N}}$) using the invariant transformation

$$X_t = \begin{pmatrix} M_t \\ \mathcal{P}_t^{\mathcal{L}} \end{pmatrix} = \begin{pmatrix} \gamma_0^W \\ 0 \end{pmatrix} + \begin{pmatrix} & \gamma_1^W \\ I_{\mathcal{L}} & 0_{\mathcal{L} \times \mathcal{N}} \end{pmatrix} \mathcal{P}_t^{\mathcal{N}}. \tag{7}$$

This rotation gives a canonical model for the family $GMTSM_{\mathcal{N}}(\mathcal{M})$ in which the pricing factors are $X_t' = (M_t', \mathcal{P}_t^{\mathcal{L}\prime})$ (based on the modeler's choice of W), the short rate satisfies

$$r_t = \rho_{0X} + \rho_{1X} \cdot X_t, \tag{8}$$

and, under $\mathbb{Q}$ and $\mathbb{P}$, X follows the processes (2) and (3). The risk-neutral parameters $(K_{0X}^{\mathbb{Q}}, K_{1X}^{\mathbb{Q}}, \rho_{0X}, \rho_{1X})$ are known functions of the parameters $\Theta^{\mathbb{Q}} = (\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_X, \gamma_0^W, \gamma_1^W)$, where Σ_X is any positive semi-definite matrix.[9]

The following theorem summarizes these derivations:

[8] Theorem 1 in JSZ allows for the more general case of $K_{1\mathcal{P}}^{\mathbb{Q}}$ being in ordered real Jordan form, which is relevant when the eigenvalues of $K_{1\mathcal{P}}^{\mathbb{Q}}$ are not distinct. We focus on the case of distinct eigenvalues to simplify our exposition.

[9] In the JSZ canonical model the matrix $\Sigma_{\mathcal{P}}$ is unconstrained. Therefore, after applying the rotation (7) we are free to treat Σ_X as the primitive volatility matrix.

Theorem 1. *Fix a full-rank portfolio matrix $W \in \mathbb{R}^{J\times J}$, and let $\mathcal{P}_t = Wy_t$. Any canonical form for the family of $\mathcal{N}$-factor models $GMTSM_{\mathcal{N}}(\mathcal{M})$ is observationally equivalent to a unique member of $GMTSM_{\mathcal{N}}(\mathcal{M})$ in which the first $\mathcal{M}$ components of the pricing factors X_t are the macro variables M_t, and the remaining $\mathcal{L}$ components of X_t are $\mathcal{P}_t^{\mathcal{L}}$; r_t is given by (8); M_t is related to $\mathcal{P}_t$ through*

$$M_t = \gamma_0^W + \gamma_1^W \mathcal{P}_t^{\mathcal{N}}, \tag{9}$$

for $\mathcal{M}\times 1$ vector γ_0^W and $\mathcal{M}\times\mathcal{N}$ matrix γ_1^W; and X_t follows the Gaussian $\mathbb{Q}$ and $\mathbb{P}$ processes (2) and (3), where $K_{0X}^{\mathbb{Q}}$, $K_{1X}^{\mathbb{Q}}$, ρ_{0X}, and ρ_{1X} are explicit functions of $\Theta^{\mathbb{Q}}$. For given W, our canonical form is parametrized by $\Theta^X = (\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, K_{0X}^{\mathbb{P}}, K_{1X}^{\mathbb{P}}, \Sigma_X, \gamma_0^W, \gamma_1^W)$.

Implicit in our proof of Theorem 1 is the proposition that first-order Markov versions of virtually all of the *GMTSM*s that have been explored in the literature are observationally equivalent to Gaussian term structure models in which the $\mathcal{N}$ pricing factors are $\mathcal{P}_t$. For the other short-rate process that has been examined, besides (1), has r_t being an affine function of the $\mathcal{M}$ macro factors M_t and a set of $\mathcal{L}$ linear combinations $\mathcal{P}_t^{\mathcal{L}}$ of bond yields:[10]

$$r_t = \rho_0^P + \rho_M^P \cdot M_t + \rho_{\mathcal{P}\mathcal{L}}^P \cdot \mathcal{P}_t^{\mathcal{L}}. \tag{10}$$

Models with this short-rate process are clearly also nested in our canonical form. Indeed, one can in general replace any given choice of yield factors $\mathcal{P}_t^{\mathcal{L}}$ in X_t with the first $\mathcal{L}$ entries of any other $\mathcal{P}$ constructed from a distinct full-rank W. That is, Theorem 1 delivers a large set of observationally equivalent canonical forms for *GMTSM*s, indexed by the set of full-rank portfolio matrices W. As we will illustrate through examples, a modeler's choice of W is often governed by the set of yields that are presumed to be priced perfectly by a *GMTSM* (if any) or by the set of yield portfolios used in model assessment (e.g., *PC*s of bond yields).

2.1 Model Flexibility and Econometric Identification

Intuitively, a *GMTSM* with pricing factors $(M_t, \mathcal{P}_t^{\mathcal{L}})$ offers more flexibility in fitting the joint distribution of bond yields than a pure latent factor model (one in which $\mathcal{N} = \mathcal{L}$), because the "rotation problem" of the risk factors is most severe in the latter setting. Characterizing this added flexibility is particularly straightforward within our canonical form. In the canonical form for a $\mathcal{N}$ latent-factor model, rotated into the JSZ normalization with pricing factors $\mathcal{P}_t^{\mathcal{N}}$, the underlying parameter set is $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}})$. The spanning property (9) of a *GMTSM* in which M_t constitutes $\mathcal{M}$ of the risk factors introduces the additional $\mathcal{M}(\mathcal{N}+1)$ free parameters γ_0^W and γ_1^W. This parameter accounting is generic: any canonical model for the family $GMTSM_{\mathcal{N}}(\mathcal{M})$, whether it be expressed with the partially latent factor set (M_t, L_t) or a fully observed factor set $(M_t, \mathcal{P}_t^{\mathcal{L}})$, effectively gains $\mathcal{M}(\mathcal{N}+1)$ free parameters relative to pure latent-factor Gaussian models. Of course this added flexibility in terms of parameter count is gained at a "cost:" the realizations of the latent risk factors must be related to the macro factors M_t through equation (9).

[10]An example of this second approach is Ang, Piazzesi, and Wei (2006) in which $\mathcal{P}_t^{\mathcal{L}}$ includes short- and long-term bond yields and the slope of the yield curve.

2.2 The Likelihood Function for Our Canonical *GMTSM*

In constructing the likelihood function of the data $\{M_t, \mathcal{P}_t\}$ used in estimation of a *GMTSM*,[11] it is instructive to distinguish between two cases: (i) the portfolios of yields comprising $\mathcal{P}^{\mathcal{L}}$ (that enter as risk factors) are priced perfectly by the *GMTSM* (that is, $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$); and (ii) the entire vector of observed yield portfolios $\mathcal{P}_t^o$ differs from its model-implied counterpart $\mathcal{P}_t$ by a vector of *i.i.d.* measurement errors. Consistent with the extant literature, the macro pricing factors M_t are presumed to be measured without error.

ML Estimation When The Risk Factors Are Measured Perfectly

Initially, we suppose that the first $\mathcal{L}$ entries of $\mathcal{P}_t$ are priced perfectly by the model ($\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$), and the last $J-\mathcal{L}$ entries, say $\mathcal{P}_t^e$, are priced up to the *i.i.d.* measurement errors $e_t = \mathcal{P}_t^{eo} - \mathcal{P}_t^e$, $e_t \sim N(0, \Sigma_e)$.[12] In this case, the joint (conditional) density of the data can be written as:

$$f^{\mathbb{P}}(X_t^o, \mathcal{P}_t^{eo}|X_{t-1}^o; \Theta^X, \Sigma_e) = f^{\mathbb{P}}(\mathcal{P}_t^{eo}|X_t^o; \Theta^X, \Sigma_e) \times f^{\mathbb{P}}(X_t^o|X_{t-1}^o; \Theta^X, \Sigma_e). \tag{11}$$

Since $X_t^o = X_t$, the conditional density $f^{\mathbb{P}}(X_t|X_{t-1}; \Theta)$ is given by

$$f^{\mathbb{P}}\left(X_t|X_{t-1}; K_{1X}^{\mathbb{P}}, K_{0X}^{\mathbb{P}}, \Sigma_X\right) = (2\pi)^{-\mathcal{N}/2}|\Sigma_X|^{-1/2} \exp\left(-\frac{1}{2} i_t' \Sigma_X i_t\right), \tag{12}$$

where $i_t = \Delta X_t - \left(K_{0X}^{\mathbb{P}} + K_{1X}^{\mathbb{P}} X_{t-1}\right)$, the time t innovation. Importantly, the only parameters from the set $\Theta^{\mathbb{Q}} = (\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_X, \gamma_0^W, \gamma_1^W)$ (those governing the risk-neutral distribution of X_t) that appear in (12) are those in Σ_X. On the other hand, the density of $\mathcal{P}_t^{eo}$ depends only on $\Theta^{\mathbb{Q}}$ and Σ_e and is

$$f^{\mathbb{P}}(\mathcal{P}_t^{eo}|X_t^o; \Theta^{\mathbb{Q}}, \Sigma_e) = (2\pi)^{-(J-\mathcal{L})/2}|\Sigma_e|^{-1/2} \exp\left(-\frac{1}{2} e_t(\Theta^{\mathbb{Q}})' \, \Sigma_e^{-1} \, e_t(\Theta^{\mathbb{Q}})\right). \tag{13}$$

Optimization of (11) proceeds from initial starting values of Θ^X as follows. Given $\Theta^{\mathbb{Q}}$ we compute $K_{0X}^{\mathbb{Q}}$, $K_{1X}^{\mathbb{Q}}$, ρ_{0X}, and ρ_{1X} using the known mappings from JSZ and the invariant transformation (7). These parameters are then used to compute bond yields as functions of the *observed* state X_t^o using standard affine pricing formulas as in (4). Finally, the model-implied yields along with the portfolio matrix W are used to compute the conditional density $f^{\mathbb{P}}(\mathcal{P}_t^{eo}|X_t; \Theta^{\mathbb{Q}}, \Sigma_e)$. Here, we compute the ML estimates which are conditional on the observed initial term structure and macro-variables. As will be discussed as part of our subsequent empirical analyses, reliable starting values are easily obtained for most of the parameters in Θ^X. We have found in practice that our search algorithm converges to the global optimum extremely quickly, usually in just a few seconds.

[11] The likelihood function for $\{M_t, y_t\}$ is identical up to the constant Jacobian of the known transformation between $\mathcal{P}_t$ and y_t, so all of our results carry over to the joint distribution of (M_t, y_t).

[12] In the literature on *GYTSM*s this assumption was introduced by Chen and Scott (1993) and it has been adopted in many empirical studies of latent-factor term structure models. It is also maintained in studies of *GMTSM*s with observable yield-based and macro risk factors.

ML Estimation When The Yield Risk Factors Are Measured Imperfectly

Several studies of *GMTSM*s have relaxed the exact pricing assumption $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$ by introducing *i.i.d.* measurement errors for the entire vector of yields y_t^o.[13] With the addition of measurement errors, ML estimation involves the use of the Kalman filter. To set up the Kalman filtering problem we start with a given set of portfolio weights $W \in \mathbb{R}^{J\times J}$ and our normalization based on Theorem 1 with theoretical pricing factors $X_t' = (M_t, \mathcal{P}_t^{\mathcal{L}\prime})$. From W and $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, \Sigma_X, \gamma_0^W, \gamma_1^W)$, we construct $(K_{0X}^{\mathbb{Q}}, K_{1X}^{\mathbb{Q}}, \rho_0, \rho_1)$. Based on the no-arbitrage pricing of bonds we then construct $A \in \mathbb{R}^J$ and $B \in \mathbb{R}^{J\times\mathcal{N}}$ with $\mathcal{P}_t = A + BX_t$, as in (4). The state equation is (3), with its parameters governed by Θ^X as described in Theorem 1, and the associated observation equation is

$$\mathcal{P}_t^o = A + BX_t + e_t, \tag{14}$$

where $e_t \sim N(0, \Sigma_e)$ are the measurement errors. Note that the dimension of e_t is now J, and not $J - \mathcal{L}$, because all J yields y_t^o are presumed to be measured with error. As before, we compute the ML estimates which are conditional on the observed initial term structure and macro-variables, which we assume is drawn from the steady-state distribution.

3 The Alternative Unconstrained Factor Model

Throughout our subsequent empirical analysis we will make frequent use of the "unconstrained alternatives" to the *GMTSM*s being investigated. The structure of the density $f^{\mathbb{P}}(X_t^o, \mathcal{P}_t^{eo}|X_{t-1})$ reveals the natural alternative model for assessing the impact of the no-arbitrage restrictions inherent in a *GMTSM* on the conditional distribution of X_t. The state-space representation of this unconstrained model reflects the presumption that bond yields have a low-dimensional factor structure, but it does not impose the restrictions implied by a no-arbitrage $DTSM$. Specifically, the state equation is

$$\Delta X_t = K_{0X} + K_{1X}X_{t-1} + \sqrt{\Sigma_X}\epsilon_t, \quad \epsilon_t \sim N(0, I_{\mathcal{N}}), \tag{15}$$

where $X_t' = (M_t', \mathcal{P}_t^{\mathcal{L}\prime})$, and the (yield) observation equation is

$$\mathcal{P}_t^o = H + GX_t + e_t, \quad e_t \sim N(0, \Sigma_e), \tag{16}$$

for conformable matrices H and G. We always assume the additional relationship that the macro-variables are observed without error ($M_t = M_t^o$), but this assumption is easily relaxed. The parameter set is $\Theta^{SS} = (K_{0X}, K_{1X}, \Sigma_X, H, G, \Sigma_e)$. When all bonds are priced with errors, Σ_e has rank $dim(\mathcal{P}_t) = J$ and the system (15–16) is estimated using the Kalman filter. If, instead, $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$, then X_t is fully observed and the first $\mathcal{L}$ entries of e_t are zero (so $\text{rank}(\Sigma_e) = J - \mathcal{L}$), and standard ML methods apply. In particular, when we assume that $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$ the maximum likelihood estimates of the matrices (H, G) are given by the OLS

[13] See, for example, Ang, Dong, and Piazzesi (2007) and Bikbov and Chernov (2010).

regression of (16). Henceforth, we will refer to the alternative unconstrained factor model (15–16) as the factor-VAR ($FVAR$).

Importantly, the unconstrained state-space representation is not the unconstrained VAR representation of the J yield portfolios $\mathcal{P}_t$:

$$\Delta \mathcal{P}_t^o = G_{0\mathcal{P}}^{\mathbb{P}} + G_{1\mathcal{P}}^{\mathbb{P}} \mathcal{P}_{t-1}^o + \sqrt{\Omega_{\mathcal{P}}} \epsilon_{\mathcal{P}t}^{\mathbb{P}}. \tag{17}$$

The model (17) relaxes both the no-arbitrage (and any over-identifying restrictions) enforced in the *GMTSM* *and* the assumed factor structure of bond yields (the dimension of X_t is less than the dimension of $\mathcal{P}_t^o$). Differences in the conditional distributions of $\mathcal{P}_t^o$ implied by a *GMTSM* and an unconstrained VAR model of $\mathcal{P}_t^o$ may arise from the VAR being over-parametrized relative to the unconstrained $FVAR$, the imposition of no-arbitrage restrictions within the *GMTSM*, or both.

4 No-Arbitrage and Conditional Means in *GMTSM*s

There has been considerable interest in the implications of no-arbitrage *GMTSM*s for out-of-sample forecasts of the macro factor M_t. For instance, Ang, Piazzesi, and Wei (2006) examine the forecasts of output in their *GMTSM*, and Chernov and Mueller (2009) examine forecasts of inflation within a similar model.

4.1 Forecasting X When $\mathcal{P}^{\mathcal{L}}$ is Measured Perfectly

Suppose that the risk factors $X_t = (M_t', \mathcal{P}_t^{\mathcal{L}\prime})'$ are measured without error, $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$. Then an immediate implication of Theorem 1 and (12) is:[14]

Corollary 1. *Assuming that $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$ and the conditions of Theorem 1 hold then, for any full-rank portfolio matrix W, optimal forecasts of future values of the risk factors X_t^o are invariant to the imposition of the no-arbitrage restrictions of a* GMTSM. *Equivalently, forecasts of X_{t+s}^o, $s > 0$, are identical to those from an unconstrained VAR model for X^o.*

For example, consider the pricing factors $X_t' = (g_t, r_t, y_t^n)$, similar to the set examined in Ang, Piazzesi, and Wei (2006), where y_t^n is the yield on a zero-coupon bond with maturity n and g_t is the output gap. Under the assumptions of Corollary 1, and presuming that $(r_t^o, y_t^{no}) = (r_t, y_t^n)$, forecasts of g_t from a canonical *GMTSM* estimated by the method of ML will be identical to those from an unrestricted VAR model of (g_t, r_t^o, y_t^{no}). This irrelevance of the no-arbitrage restrictions holds for any canonical model for the family $GMTSM_3(1)$ with $M_t = g_t$ and $X_t^o = X_t$.

4.2 Forecasting X When $\mathcal{P}^{\mathcal{L}}$ is Measured Imperfectly

To what extent is the relaxation of the assumption that $\mathcal{P}_t^{\mathcal{L}} = \mathcal{P}_t^{\mathcal{L}o}$ likely to change the ML estimates of the conditional mean of X_t? We now show that, provided the filtered model (be it

[14] This is the counterpart for *GMTSM*s of Proposition 3 in JSZ.

the no arbitrage model or the $FVAR$) accurately prices $\mathcal{P}_t$, the ML estimates of $(K^{\mathbb{P}}_{0X}, K^{\mathbb{P}}_{1X})$ will be quite close to their OLS counterparts. Now, as our likelihood function resides in the exponential family, the expectation maximization theory applies (see Dempster, Laird, and Rubin (1977) and Digalakis, Rohlicek, and Ostendorf (1993)) and it must be that at the ML estimates of the model we have[15]

$$\mathcal{E}_T \left(E[\partial_{(K^{\mathbb{P}}_{0X}, K^{\mathbb{P}}_{1X})} \log f^{\mathbb{P}}(X_t|X_{t-1})|\mathcal{F}_T] \right) = 0,$$

where $\mathcal{E}_T$ dentoes the sample average operator and $E[\cdot|\mathcal{F}_T]$ denotes the expectation conditioned on the entire sample of observables, $\{y_\tau, M_\tau\}_{\tau=0}^T$. As standard, we refer to these as the smoothed realizations of the variable. It follows that

$$[K^{\mathbb{P}}_{0X}, K^{\mathbb{P}}_{1X}] = \left(\mathcal{E}_T \left[E[Z_t Z_t'|\mathcal{F}_T]\right]\right)^{-1} \left(\mathcal{E}_T[E[Z_t' X_{t+1}|\mathcal{F}_T]]\right),$$

where $Z_t = [1, X_t]$. Notice that this is not quite the OLS estimate of $[K^{\mathbb{P}}_{0X}, K^{\mathbb{P}}_{1X}]$ obtained upon replacing the actual data with $E[X_t|\mathcal{F}_T]$. However,

$$E[X_t X_t'|\mathcal{F}_T] = \text{var}_T(X_t|\mathcal{F}_T) + E[X_t|\mathcal{F}_T]E[X_t|\mathcal{F}_T]'.$$

This equation (and the analogous extension to $E[X_t X_{t+1}'|\mathcal{F}_T]$) reveal that, provided $E[X_t|\mathcal{F}_T]$ is close to X_t^o (on average) and $\text{Var}_T(X_t|\mathcal{F}_T)$ is small, the OLS estimates of $[K^{\mathbb{P}}_{0X}, K^{\mathbb{P}}_{1X}]$ obtained by regressing X_t^o on its lag will produce similar estimates to those obtained by Kalman filtering. Since X_t is chosen to have elements which represent salient characteristics of the yield curve (such as level and slope), the condition $E[X_t|\mathcal{F}_T] \approx X_t^o$ is likely to hold even in the presence of large pricing errors for individual bonds. And we only need to match the time-series average of these characteristics of the yield curve. Additionally, it is very likely that $\text{Var}_T(X_t|\mathcal{F}_T)$ will be quite small since this represents the level of our uncertainty about, say, the level or slope of the yield curve conditioned on the entire time-series of yields and macro variables. Indeed, the current yield curve provides sufficient information to infer these broad characteristic of the yield curve.[16]

In summary, ML estimates of $(K^{\mathbb{P}}_{0X}, K^{\mathbb{P}}_{1X})$ from a canonical $GMTSM$ and a $FVAR$ are identical when $\mathcal{P}^{\mathcal{L}}_t = \mathcal{P}^{\mathcal{L}o}_t$; they are the OLS estimates from the $FVAR$. When $\mathcal{P}^{\mathcal{L}}_t \neq \mathcal{P}^{\mathcal{L}o}_t$ and the $GMTSM$ and $FVAR$ are estimated using Kalman filtering, the ML estimates of the conditional mean parameters in these models will differ from the OLS estimates when either (1) the sample average of $E[\mathcal{P}^{\mathcal{L}}_t|\mathcal{F}_T]$ differs greatly from the sample average of $\mathcal{P}^{\mathcal{L}o}_t$, or (2) there are characteristics of the yield curve that remain very uncertain even after observing the entire time series of yields.[17] Neither of these conditions arises in the illustrative $GMTSM$s examined in our subsequent empirical analysis: all four models– $GMTSM$ and

[15] For this argument, we assume a flat prior on (M_0, y_0). When one computes the ML estimate with a parametric distribution for (M_0, y_0) (or the ML estimate conditional on (M_0, y_0)), similar results will obtain.

[16] As JSZ show, when we increase the number of latent factors, both of these conditions start to fail.

[17] Note that for the time series of conditional estimates of the mean to agree, it must be that $E[\mathcal{P}^{\mathcal{L}}_t|\mathcal{F}_T] \approx \mathcal{P}^{\mathcal{L}o}_t$ for each t which is slightly stronger than the statement that these agree on average. However, for our estimated models, the filtered states agree with the observed states closely for each t.

$FVAR$, estimated with and without filtering– produce identical, or nearly identical, estimates of the conditional mean of X_t^o.

We emphasize that these conditions for the (near) irrelevance of no-arbitrage restrictions for forecasts of the risk factors do not require that the *GMTSM* accurately price each bond yield in y_t. Rather, it depends only on the accurate pricing of (small measurement errors on) the portfolios of yields $\mathcal{P}_t^{\mathcal{L}}$ that enter as risk factors.

4.3 Forecasting Bond Yields y When the Risk Factors Are Measured Imperfectly

These observations extend to forecasts of individual bond yields with one additional requirement. Specifically, the factor loadings A and B from OLS projections of y_t^o onto X_t^o need to be close to their model-based counterparts estimated using the Kalman filter. Now large $RMSE$s in the pricing of individual bonds might lead to large efficiency gains from ML estimation of the loadings within a *GMTSM*, in which case a *GMTSM* might produce sizable gains in forecast accuracy. This is an empirical question that we take up subsequently.

5 No-Arbitrage and Conditional Variances in *GMTSM*s

Whereas the parameters governing the conditional $\mathbb{P}$-mean of X_t enter only $f^{\mathbb{P}}(X_t^o|X_{t-1}^o)$ and not $f^{\mathbb{P}}(\mathcal{P}_t^{eo}|X_t^o)$, the matrix Σ_X enters both the $\mathbb{P}$ and $\mathbb{Q}$ representations of X by diffusion invariance. Therefore, even when the risk factors are assumed to be measured perfectly, the ML estimator of Σ_X will be more efficient than the standard sample estimator constructed from the residuals of the OLS estimator of the VAR model for X. In this section we show that such efficiency gains are likely to be inconsequential when the average pricing errors of the yield portfolios $\mathcal{P}_t$ are relatively small (in a sense we make precise). For the initial theoretical discussion of this issue we maintain the assumption that $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$.

5.1 Estimating Σ_X When $\mathcal{P}^{\mathcal{L}}$ is Measured Perfectly

Let $X_t = \Gamma_0 + \Gamma_1 \mathcal{P}_t^{\mathcal{N}}$ denote the transformation in (7). In Appendix A we show that

$$
\begin{aligned}
y_t &= A_X + B_X X_t, && (18)\\
A_X &= \alpha_{\mathcal{P}} r_\infty^{\mathbb{Q}} - B_X \Gamma_0 + \beta_X vec(\Sigma_X), && (19)
\end{aligned}
$$

where $\alpha_{\mathcal{P}}$ depends only on $\lambda^{\mathbb{Q}}$, and β_X and B_X are conformable matrices that depend only on $\lambda^{\mathbb{Q}}$ and Γ_1. The parameters $r_\infty^{\mathbb{Q}}$, Σ_X, and Γ_0 affect the intercept term A_X and not the loadings of the risk factors, B_X. Among the parameters governing (A_X, B_X), only Σ_X appears in the physical density. The first-order conditions with respect to Σ_X are built up as follows.

First,

$$
\mathcal{E}_T[\partial \log f^{\mathbb{P}}(X_t^o|X_{t-1}^o)/\partial \Sigma_X] = -\frac{1}{2}\left[\Sigma_X^{-1} - \Sigma_X^{-1}\Sigma_{XT}^{OLS}\Sigma_X^{-1}\right], \quad (20)
$$

where Σ_{XT}^{OLS} is the innovation covariance matrix of X_t estimated from the OLS residuals of a VAR. Second, letting W^e denote the $(J-\mathcal{L}) \times J$ matrix of weights that generate the yield portfolios $\mathcal{P}_t^e$ that are priced with errors, the pricing errors are

$$\begin{aligned} e_t =& W^e(y_t^o - y_t) \\ =& W^e(y_t^o - \alpha_{\mathcal{P}} r_\infty^{\mathbb{Q}} + B_X\Gamma_0 - \beta_X vec\,(\Sigma_X) - B_X X_t)\,. \end{aligned} \tag{21}$$

It follows that

$$\mathcal{E}_T[\partial \log f^{\mathbb{P}}(\mathcal{P}_t^{eo}|X_t^o)/\partial vec(\Sigma_X)] = \beta_X'(W^e)'\Sigma_e^{-1}\mathcal{E}_T[e_t]. \tag{22}$$

Combining these terms gives:

$$vec\left(\frac{1}{2}\left[(\Sigma_{XT}^{ML})^{-1} - (\Sigma_{XT}^{ML})^{-1}\Sigma_{XT}^{OLS}(\Sigma_{XT}^{ML})^{-1}\right]\right) = \beta_{XT}^{ML\prime}(W^e)'(\Sigma_{eT}^{ML})^{-1}\mathcal{E}_T[e_t] \tag{23}$$

where the superscript "ML" indicates empirical ML estimates. Though the first order condition (23) does not have a closed form solution, this equation is easily linearized. Since Σ_{XT}^{OLS} will likely be close to the maximizer of the likelihood (for fixed $(\lambda^{\mathbb{Q}}, \gamma_1^W, r_\infty^{\mathbb{Q}}, \Sigma_e)$) this linearization and Newton updating should provide near instantaneous convergence. Therefore, we can effectively view (23) as a closed form expression for Σ_{XT}^{ML} in terms of $(\lambda^{\mathbb{Q}}, \gamma_1^W, r_\infty^{\mathbb{Q}}, \Sigma_e)$. This leads to a concentrated likelihood function and, thereby, further simplifies the estimation of both our canonical $GMTSM$ and the canonical $GYTSM$ proposed by JSZ.

5.2 No-Arbitrage Restrictions and Efficient Estimation of Σ_X

Our characterization of Σ_X in terms of $(\lambda^{\mathbb{Q}}, \gamma_1^W, r_\infty^{\mathbb{Q}}, \Sigma_e)$ reveals a simple sufficient condition for the ML and OLS residual-based estimators of Σ_X to coincide. Specifically, from (23) it follows that when the average observation errors for all maturities is zero, $\mathcal{E}_T[e_t] = 0$, then $\Sigma_{XT}^{ML} = \Sigma_{XT}^{OLS}$. Furthermore, holding all else constant, the smaller is the average-to-variance ratio of errors $(\Sigma_{eT}^{ML})^{-1}\mathcal{E}_T[e_t]$, the less Σ_{XT}^{ML} will be impacted by no arbitrage restrictions. In practice, this ratio is typically small compared to the first derivative of $\partial f^{\mathbb{P}}(X_t^o|X_{t-1}^o)/\partial\Sigma_X$ with respect to Σ_X, evaluated at Σ_{XT}^{OLS}. Thus, accommodating the right hand side of (23) only requires a small deviation of Σ_X from Σ_{XT}^{OLS}. Therefore, we anticipate that the OLS estimate of Σ_X will typically be very close to the ML estimate from the $GMTSM$.

Within standard yield-only models, Duffee (2009) shows that when only one yield is priced with error $(J - \mathcal{N} = 1)$, a $GYTSM$ is just-identified and $\mathcal{E}_T[e_t] = 0$. The construction of our canonical $GMTSM$ reveals that the $\mathcal{M}+1$ free parameters $(r_\infty^{\mathbb{Q}}, \Gamma_0)$ are available to force $\mathcal{E}_T[e_t]$ to zero through their first-order conditions. This is a consequence of the fact that these parameters do not assist in fitting the time-series properties of X_t, nor do they impact the cross-section of yields through the risk-factor loadings. It follows that if exactly $\mathcal{M}+1$ portfolios of yields are included with measurement errors in the ML estimation, then Σ_X^{ML} will coincide with Σ_X^{OLS}. Indeed, if fewer than $\mathcal{M}+1$ yield portfolios are included in estimation then the parameters $(r_\infty^{\mathbb{Q}}, \Gamma_0)$ are not fully econometrically identified.

Combining the preceding results on Σ_X with our results in Section 4 leads to:

Corollary 2. *Under the assumptions of Theorem 1 and assuming that $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$, if the average-to-variance ratio of errors $(\Sigma_{eT}^{ML})^{-1}\mathcal{E}_T[e_t]$ is zero, then the conditional distribution of the risk factors X_t is fully invariant to the imposition of the no-arbitrage restrictions of a* GMTSM*. Equivalently, any functions of the moments of the conditional or unconditional distributions of $\{X_t\}$ are unaffected by the no-arbitrage structure of a* GMTSM*.*

5.3 Estimating Σ_X When $\mathcal{P}^{\mathcal{L}}$ Measured Imperfectly

To what extent is the relaxation of the assumption that $\mathcal{P}_t^{\mathcal{L}} = \mathcal{P}_t^{\mathcal{L}o}$ likely to change Σ_{XT}^{ML}? Invoking the EM theory once again, at the ML estimates of the model, we have:

$$\mathcal{E}_T\left(E[\partial_{\Sigma_X}\log f^{\mathbb{P}}(X_t|X_{t-1}) + \partial_{\Sigma_X}\log f^{\mathbb{P}}(\mathcal{P}_t^e|X_t)|\mathcal{F}_T]\right) = 0,$$

For the FVAR model where there is no diffusion invariance, this leads to

$$E\left[vec\left(\frac{1}{2}\left[(\Sigma_X)^{-1} - (\Sigma_X)^{-1}\Sigma_X^{OLS}(\Sigma_X)^{-1}\right]\right)\middle|\,\mathcal{F}_T\right] = 0, \tag{24}$$

where $\Sigma_X^{OLS} = E_T[i_t^{ML}(i_t^{ML})']$, where $i_t^{ML} = \Delta X_t - (K_{0X}^{ML} + K_{1X}^{ML}X_{t-1})$. As such, for an $FVAR$, we obtain

$$\Sigma_{XT}^{ML} = E\left[\Sigma_X^{OLS}\,\middle|\,\mathcal{F}_T\right].$$

Using logic of Section 4.2, as long as the estimated $FVAR$ accurately prices the risk factors, then we must have $E\left[\Sigma_X^{OLS}\,\middle|\,\mathcal{F}_T\right] \approx \Sigma_{XT}^{OLS}$, the conditional variance of the risk factors obtained from OLS regressions.

For a *GMTSM*, the first order condition of Σ_X is

$$E\left[vec\left(\frac{1}{2}\left[(\Sigma_X)^{-1} - (\Sigma_X)^{-1}\Sigma_X^{OLS}(\Sigma_X)^{-1}\right]\right) - \beta'(W^e)'(\Sigma_e)^{-1}\mathcal{E}_T[e_t]\middle|\,\mathcal{F}_T\right] = 0. \tag{25}$$

Compared to the corresponding $FVAR$, there is one extra term coming from the density of pricing errors. Nevertheless, as long as the mean-to-variance ratio of (smoothed) errors is relatively small and the risk factors are priced accurately, we should expect for a *GMTSM* too that $\Sigma_{XT}^{ML} \approx \Sigma_{XT}^{OLS}$.

Taken together, these observations suggest that even when the risk factors are measured imperfectly, the estimates of the conditional variances parameter Σ_X can be largely unaffected by the imposition of no-arbitrage.

The remainder of this paper explores the practical relevance of these observations. We begin our empirical exploration with a four-factor model that shares the basic structure of models recently examined in the literature. The joint distributions of bond yields implied by various three-factor *GMTSM*s, and the links between these models and "Taylor rules," are examined in Section 7.

6 Empirical Exploration of the Joint Distribution of Macro and Yield Factors in a $\boldsymbol{GMTSM}_4(2)$ Model

We first examine model GM_4 with risk factors $(PC1_t, PC2_t, CPI_t, HELP_t)$, where $HELP$ is the help wanted index and CPI is the inflation rate, as in Bikbov and Chernov (2010). For further comparability to previous studies, we use monthly unsmoothed Fama-Bliss zero yields over the sample period 1972 through 2003. The specific maturities included are three- and six-months, and one through ten years.[18] The rotation of $\mathcal{P}_t^2$ to the first two PCs of these bond yields uses sample estimates of the PC loadings.

The unconstrained alternative is the relevant four-factor state-space model described in Section 3 in which bond yields are related to the risk factors according to (16). In all models, the macro factors are assumed to be observed without errors. Additionally, for the yield factors we consider two cases: (i) the yield portfolios $\mathcal{P}_t^2$ are priced without errors with the remaining 10 yield portfolios priced with errors distributed $N(0, \sigma_e^2 I_{10})$, and (ii) all twelve yields are priced with $i.i.d.$ uncorrelated errors with common variance, $(y_t^o - y_t) \sim N(0, \sigma_e^2 I_{12})$. These measurement assumptions are applied equally to the corresponding versions of $FVAR$. So, for instance, if all yields are priced with errors by the $GMTSM$, then we include measurement errors for all twelve portfolios $\mathcal{P}_t^o$ in (16).

When all of the bond yields y_t are allowed to be measured with errors ($\mathcal{P}_{jt} \neq \mathcal{P}_{jt}^o$, for all $j = 1, \ldots, J$), then Theorem 1 implies that the normalization of the non-macro factors to be the first two PCs of bond yields is without loss of generality. On the other hand, when it is assumed that $\mathcal{P}_t^{\mathcal{L}} = \mathcal{P}_t^{\mathcal{L}o}$, then the ML estimates of a $GMTSM$ are not invariant to the choice of portfolio weights W. When all yields are measured imperfectly we use the Kalman filter in estimation and identify the associated models by a superscript "KF."

6.1 Estimates of the Conditional Mean of X_t

Table 1 displays the ratios and scaled differences of the conditional mean parameters from model GM_4 and its associated $FVAR$, with and without filtering (measurement errors on $\mathcal{P}_t$). Consider the first block with ratios of estimates from models GM_4^{KF} and $FVAR_4^{KF}$. Most of the ratios are between 0.99 and 1 and almost all of the scaled differences are zero to three decimal places. The near equality of the parameter estimates translates into virtually identical forecasts of X_t^o from these models across all forecast horizons that we examined.[19]

Now the root-mean-squared differences ($RMSE$s) between $(PC1_t^o, PC2_t^o)$ and the output of the Kalman filter, $(PC1_t^{KF}, PC2_t^{KF})$, are only $(1.7, 4.6)$ basis points.[20] Drawing upon our

[18] The Fama-Bliss data extending out to ten years ends in 2003. We started our sample in 1972, instead of in 1970 as in Bikbov and Chernov (2010), because data on yields for the maturities between five and ten years are sparse before 1972.

[19] We examined forecasts for horizons out to six months and they are graphically indistinguishable, with root mean square difference around 2 or 3 basis points. These particular results corroborate the results on forecasts using $GYTSM$s in Duffee (2009).

[20] To put these numbers in perspective, if we scale the loadings on $PC1$ so that they sum to one, then the $RMSE$s for $PC1$ are 0.5 basis points.

	Ratio					Scaled Difference				
	$K^{\mathbb{P}}_{0X}$	$I + K^{\mathbb{P}}_{1X}$				$K^{\mathbb{P}}_{0X}$	$I + K^{\mathbb{P}}_{1X}$			
$\frac{GM_4^{KF}}{FVAR_4^{KF}}$	0.999	1	0.998	0.999	0.999	-0.05	0.0033	0.08	-0.06	-0.07
	0.998	0.997	1	0.996	0.998	0.09	0.14	0.01	-0.20	-0.11
	0.999	0.995	1	1	0.999	-0.06	0.24	0.09	-0.0022	-0.04
	1	0.998	1	0.999	1	-0.0019	0.10	-0.11	0.06	0.00027
$\frac{GM_4^{KF}}{GM_4}$	1.02	0.999	1.01	1.01	1.01	8.0	-0.028	-0.50	0.43	0.00451
	1.19	1.09	0.986	1.19	1.08	-8.9	-4.50	-0.73	8.50	4.00
	0.993	1.01	0.984	1	1.01	-0.36	-0.58	-0.79	0.01	0.28
	0.975	0.926	0.963	1.09	0.999	1.24	3.85	1.88	-4.26	-0.03
$\frac{FVAR_4^{KF}}{FVAR_4}$	1.02	1	1.01	1.01	1.01	0.75	-0.02	-0.42	0.37	0.39
	1.19	1.09	0.986	1.18	1.08	-8.78	-4.32	-0.72	8.27	3.85
	0.992	1.01	0.986	1	1	-0.42	-0.33	-0.70	0.0078	0.24
	0.979	0.924	0.965	1.09	0.999	1.05	3.95	1.78	-4.20	-0.03

Table 1: Ratios and scaled differences of conditional mean parameters $K^{\mathbb{P}}_{0X}$ and $I + K^{\mathbb{P}}_{1X}$. The first block compares the estimates for models GM_4^{KF} and $FVAR_4^{KF}$, the second block compares the estimates with and without filtering within model GM_4, and the third compares estimates with and without filtering within $FVAR_4$. The scaled difference between a and b is computed as: $(a - b)/(|a| + |b|)$.

analysis in Section 4, it should therefore be equally true that the estimated mean parameters from models GM_4^{KF} and GM_4 are similar. The second block of results in Table 1 shows that this is in fact the case. Additionally, we know that forecasts of the risk factors from models GM_4 and $FVAR_4$ are identical, in and out of sample. The last block in Table 1 closes the loop among all four of the models considered and shows that, with our without measurement errors, the estimated mean parameters in the $FVAR$s are nearly the same. In sum, the imposition of no-arbitrage restrictions have essentially no impact on forecasts of the risk factors in our four-factor *GMTSM*, and this is true whether or not one assumes that bond prices are measured with errors.

6.2 Estimates of the Conditional Second Moments of X_t

Comparisons of the ML estimates of the conditional second moments are displayed in Table 2. Ratios and differences between standard deviations are reported along the diagonals, and their counterparts for correlations among the innovations are reported in the off-diagonal entries. Recall from Section 5 that even when all of the risk factors are measured perfectly there may be efficiency gains to estimating Σ_X within a *GMTSM*. The first block compares models GM_4 and GM_4^{KF} to their respective $FVAR$s (without and with filtering). The no-arbitrage structure of model GM_4 has negligible impact on the ML estimates of Σ_X, both when all

(a) Ratio of Model *GMTSM* to Model *FVAR*

	Ratio				Scaled Difference(%)			
$\frac{GM_4}{FVAR_4}$	1	-	-	-	4.0e-05	-	-	-
	0.975	1	-	-	-0.0053	9.3e-06	-	-
	1.01	0.978	1	-	0.0008	-0.0008	-6.6e-06	-
	1	0.996	1	1	-0.0001	-0.0008	-2.9e-05	3.9e-06
$\frac{GM_4^{KF}}{FVAR_4^{KF}}$	1	-	-	-	3.0e-05	-	-	-
	0.983	1	-	-	-0.0039	-8.2e-07	-	-
	1.01	0.966	1	-	0.0010	-0.0010	-6.6e-06	-
	1	1	1	1	-3.9e-06	-8.2e-05	0.0002	-4.1e-06

(b) *GMTSM*s With and Without Filtering

	Ratio				Difference			
$\frac{GM_4^{KF}}{GM_4}$	1.01	-	-	-	0.0001	-	-	-
	0.874	1.08	-	-	-0.0292	0.0003	-	-
	0.978	1.24	1	-	-0.0022	0.0067	1.8e-05	-
	0.979	0.870	0.987	1	-0.0059	-0.0306	-0.0010	8.3e-05
$\frac{FVAR_4^{KF}}{FVAR_4}$	1.01	-	-	-	0.0001	-	-	-
	0.882	1.07	-	-	-0.0278	0.0003	-	-
	0.980	1.23	1	-	-0.0020	0.0066	1.8e-05	-
	0.980	0.873	0.990	1	-0.0057	-0.0299	-0.0008	7.5e-05

Table 2: Ratios and differences of conditional second moment parameters. Standard deviations are compared along the diagonals and correlations are compared on the off-diagonals.

bonds are measured with error and with the first two PCs are priced perfectly.

Turning to comparisons across filtered and unfiltered versions of the *GMTSM* and $FVAR$ in the lower block of Table 2, again the ratios are very close to one. The only notable exceptions are the $(1, 2)$, $(3, 2)$ and $(4, 2)$ entries that represent the conditional correlations between $PC2$ and other risk factors. These differences do *not* arise as consequence of imposing no-arbitrage restrictions– the comparisons in the lower block of Table 2 vary only whether we apply the Kalman filter and fix whether no arbitrage is imposed. Rather, they arise because, given the assumed distributions of the measurement errors for $\mathcal{P}_t^2$, even small differences between the observed and filtered $PC2$ translate into different estimates of the conditional correlations with and without filtering. The cautionary implication of this finding is that the choice of distribution for the measurement errors $\mathcal{P}_t^{2o} - \mathcal{P}_t^2$ may affect ML estimates of the conditional correlations of the risk factors within *GMTSM*s.

Comparably close estimates for both the conditional mean and variance parameters are obtained from models that substitute the first PC of the help wanted index, unemployment, the growth rate of employment, and the growth rate of industrial production ($REALPC$) for $HELP$, or that use the first PC of various inflation indices in place of CPI, as in Ang and Piazzesi (2003). Accordingly, we are comfortable concluding that the no-arbitrage restrictions of canonical *GMTSM*s produce nearly identical ML estimates of the conditional moments of the risk factors to those obtained from the corresponding unrestricted $FVAR$s.

6.3 No-Arbitrage Restrictions and Impulse Response Functions

Though these estimates of conditional moments are nearly the same, they are not exactly identical. Therefore we proceed to compare the model-implied impulse response functions. By construction these response functions are built up from functions of both the first and second moments, and they embody a temporal dimension as well.

Figure 1 displays the responses of $PC1$ and $PC2$ to innovations in $HELP$ and CPI implied by the ML estimates of model GM_4^{KF} and its corresponding $FVAR_4^{KF}$ model (in both models all bonds are priced with errors). These responses are based on the ordering $(CPI, HELP, PC1, PC2)$ in order to give the macro factors priority. It is immediately evident that their responses are essentially indistinguishable; comparable matching is obtained for other orderings of X_t and for the responses to other risk factors. This verifies that no-arbitrage restrictions add no incremental insight into the patterns of these responses over what is obtained from the unconstrained $FVAR_4$, even in the presence of pricing errors on yields.

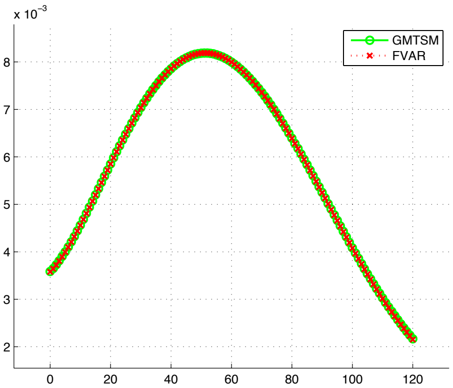
7 *GMTSM*$_3(\mathcal{M})$s With Output Growth and Inflation

Building upon the extensive literature on yield-only *GYTSM*s, much of the literature on *GMTSM*s has focused on three-factor models in which $M_t = g_t$ or $M_t = (g_t, \pi_t)'$, where g_t is a measure of real output growth and π_t is a measure of inflation. In the former case $X_t' = (\mathcal{P}_t^{2\prime}, g_t)$ (e.g., Ang, Piazzesi, and Wei (2006)), and in the later case $X_t' = (\mathcal{P}_t^1, g_t, \pi_t)$ (e.g., Ang, Dong, and Piazzesi (2007), Smith and Taylor (2009)). By examining these formulations in parallel we are able to explore how, for a given number of risk factors (three), changing the mix of macro and yield-based factors affects the impact of no-arbitrage restrictions on the conditional distribution of X_t. To differentiate between the three-factor *GMTSM*s with $M_t = g_t$ and $M_t = (g_t, \pi_t)'$ we use the notation $GM_3(g)$ and $GM_3(g, \pi)$, respectively.

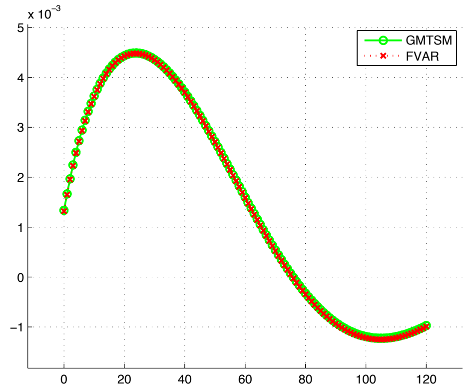
Prior to undertaking this comparison, we briefly revisit the economic interpretation of the loadings on the risk factors in the expression (1) for the short rate r_t. Ang, Dong, and Piazzesi (2007), Bikbov and Chernov (2010), and Smith and Taylor (2009), among others, interpret their versions of the short-rate equation (1) as a Taylor-style policy rule. That is, with $M_t' = (g_t, \pi_t)$ and $\mathcal{L} = 1$,

$$r_t = r^* + \alpha\pi_t + \beta g_t + L_t, \tag{26}$$

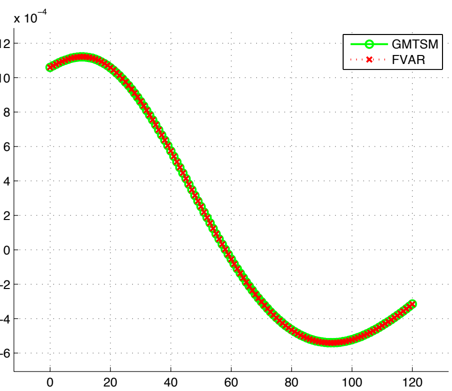
and typical formulations of policy rules that have central banks setting the short rate according to targets for inflation and the output gap are similar in form. Just as in standard structural formulations of a Taylor rule, there is also a latent shock L_t to the short rate in (26). However,



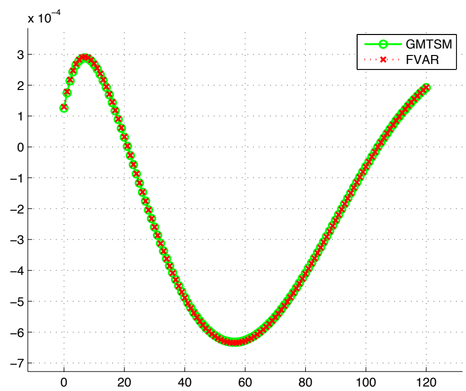
(a) Response of $PC1$ to $HELP$



(b) Response of $PC1$ to CPI



(c) Response of $PC2$ to $HELP$



(d) Response of $PC2$ to CPI

Figure 1: Impulse responses of $PC1$ and $PC2$ to innovations in $HELP$ and CPI based on ML estimates of model GM_4^{KF}. The risk factors are ordered as $(CPI, HELP, PC1, PC2)$. The horizontal axis is in months.

without imposing additional economic structure on a *GMTSM*, the parameters (α, β) in (26) are not meaningfully interpretable as the reaction coefficients of a central bank. This is why we can express ρ_{0X} and ρ_{1X} in (8) as explicit functions of $\Theta^{\mathbb{Q}}$ in our canonical form without loss of economic generality.

To see this in the simplest, benchmark case we focus on the family $GMTSM_3(\mathcal{M})$ with short-rate given by (26). As in the *GMTSM*s of Ang, Dong, and Piazzesi (2007) and Chernov and Mueller (2009), for example, the latent shock L_t is allowed to be correlated with M_t.[21] Within this setting, an immediate implication of Theorem 1 is that, for a given full-rank portfolio matrix W, we can construct a canonical model for $GMTSM_3(\mathcal{M})$ in which r_t is

[21] Cochrane (2007) argues that it is an inherent feature of new-Keynesian models that the policy shock "jumps" in response to changes in inflation or output gaps.

	Model $GM_3(g,\pi)$				Model $GM_3(g)$			
	$K_{0X}^{\mathbb{P}}$	$I+K_{1X}^{\mathbb{P}}$			$K_{0X}^{\mathbb{P}}$	$I+K_{1X}^{\mathbb{P}}$		
	1	1	1	0.997	1.08	1	1.35	0.961
$\frac{GM_3^{KF}}{FVAR_3^{KF}}$	0.999	1	1	1	0.985	1.12	0.999	0.996
	1.01	0.965	1	1	0.997	1	0.998	1
	1.1	0.998	1.03	1.01	0.962	1	0.744	1.03
$\frac{GM_3^{KF}}{GM_3}$	0.999	0.999	1	1	1.17	0.97	0.99	1.05
	0.987	1.04	0.999	1	1	0.999	1	1
	1.11	0.998	1.03	1.01	1.04	1	1	0.985
$\frac{FVAR_3^{KF}}{FVAR_3}$	0.999	0.998	1	1	1.15	1.09	0.989	1.05
	0.998	1.01	1	1	0.998	1	0.998	1

Table 3: Ratios of estimated $K_{0X}^{\mathbb{P}}$ (first column) and $I + K_{1X}^{\mathbb{P}}$ (the next 3) for models $GM_3(g,\pi)$ and $GM_3(g)$. The first block compares the estimates for $GMTSM_3$ and $FVAR_3$ with filtering, the second block compares the estimates of each $GMTSM$ with and without filtering, and the third compares estimates with and without filtering within the two $FVAR_3$s.

given by

$$r_t = r^{W*} + \beta_\pi^W \pi_t + \beta_g^W g_t + \beta_{\mathcal{P}}^W \mathcal{P}_t^1. \tag{27}$$

One choice of W, the choice we will adopt for our subsequent empirical analysis, places the first PC (the "level" factor) in place of L_t in (26)– sets $\mathcal{P}_t^1 = PC1_t$. Another choice of W sets $\mathcal{P}_t^1 = PC2_t$ (the "slope" factor), etc. All of these choices give rise to theoretically equivalent representations of the short-term rate and, through the no-arbitrage term structure model, identical bond yields at all maturities. Absent the imposition of additional economic structure, there appears to be no basis for interpreting any one of the infinity of equivalent rotations (27) as *the* Taylor rule of a structural model.

To explore the properties of the conditional distribution of the risk factors in model $GM_3(g)$ we use the first PC of the help wanted index, unemployment, the growth rate of employment, and the growth rate of industrial production ($REALPC$) as our measure of g, and for model $GM_3(g,\pi)$ we use $REALPC$ for g and the first PC of measures of inflation based on the CPI, the PPI of finished goods, and the spot market commodity prices ($INFPC$) for π. These are the output and inflation measures used by Ang and Piazzesi (2003).

Table 3 displays the ratios of the parameters of the conditional means of X_t for the models $GM_3(g)$ and $GM_3(g,\pi)$. Once again these ratios are very close to one across all pairs of models thereby establishing that, holding fixed the content of the three risk factors, imposition of no-arbitrage restrictions has only small effects on the forecasts of the risk factors.

A notable difference between models $GM_3(g)$ and $GM_3(g,\pi)$ is that standard deviations of the errors in pricing individual bonds are much larger in the latter model. For bond with maturities $(1yr, 3yr, 5yr, 7yr, 10yr)$, these measures of mispricing (in basis points) are $(12.1, 14.5, 9.7, 9.6, 17.3)$ for model $GM_3(g)$ and $(46.6, 14.9, 17.8, 28.6, 39.9)$ in model $GM_3(g, \pi)$.[22] There is a sizable deterioration in fit with three-factor *GMTSM*s that include both inflation and output growth as risk factors, relative to models in which the only macro factor is output growth. The reason is that the PCs of bond yields capture the variation in bond yields much more accurately than the macro risk factors, so model $GM_3(g, \pi)$ is handicapped in terms of this measure of fit relative to model $GM_3(g)$.[23]

An interesting aspect of the results in Table 3 is that the (near) irrelevance of no-arbitrage restrictions for forecasts of X_t arises regardless of how well the models price the individual bonds and, in light of the magnitudes of these mispricings, regardless of whether $\mathcal{P}_t^{\mathcal{L}}$ is assumed to be measured with error. These findings illustrate an important point that we emphasized in Section 4. Namely, for the no-arbitrage restrictions to be (nearly) irrelevant for forecasting the risk factors in a *GMTSM*, it is not necessary that it accurately price the individual bonds used in estimation. All that is required is that the yield factors $\mathcal{P}_t^{\mathcal{L}}$ are accurately priced, and this is in fact the case in both models $GM_3(g)$ and $GM_3(g, \pi)$.

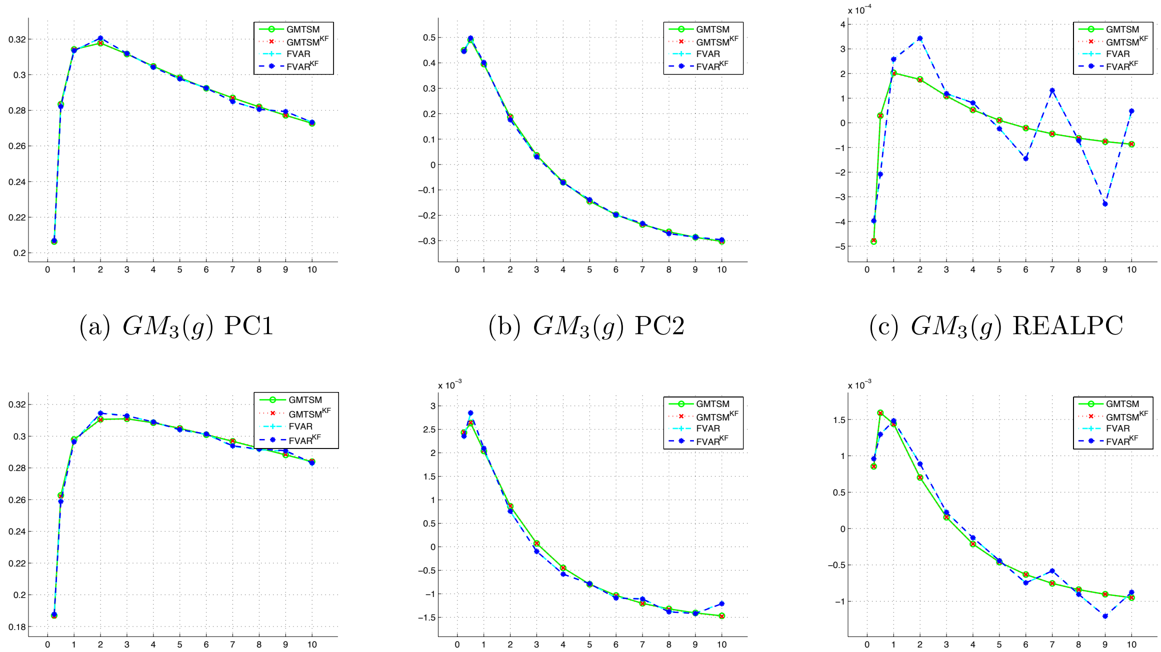
The (near) irrelevance is also true for the conditional second moments of the risk factors. The ratios and differences corresponding to those reported in Table 2 take on very similar values though, in most cases, they are closer to unity (zero) for the ratios (differences) in the three-factor models. Thus, we are again led to the conclusion that no-arbitrage restrictions have little impact on the estimates of the conditional variances of the risk factors. Further, combining these results for the first and second moments, it follows that the conditional distributions of the risk factors implied by these GM_3s and $FVAR_3$s are nearly identical.

One aspect of *GMTSM*s that we have not explored up to this point is how much their no-arbitrage structure affects the conditional distributions of the individual bond yields. Three key ingredients are required for the conditional distribution of bond yields to be (nearly) invariant to the imposition of the no-artitrage restrictions of a canonical *GMTSM*. We have already verified two of these: OLS estimates of the conditional mean of the factors $(\mathcal{P}_t^{\mathcal{L}o}, M_t)$ and their ML counterparts within the *GMTSM* are nearly identical, and the average pricing errors for the yield portfolios $\mathcal{P}_t^{eo}$ are small relative to their standard deviations. The third ingredient is that no-arbitrage restrictions must not substantially affect the ML estimators of the factor loadings (A_X, B_X) in (18).

Figure 2 displays the loadings on all three factors for both models $GM_3(g)$ (top frames) and $GM_3(g, \pi)$ (bottom frames), estimated with and without measurement errors (filtering) on the yield factors. In addition, in each frame we display the corresponding loadings from the associated $FVAR_3$, again estimated with and without filtering on the yield factors. The loadings on $PC1$ and $PC2$ in model $GM_3(g)$ are virtually on top of each other for all four factor representations. Similar results were obtained by Duffee (2009) in the context of three-factor, yield-only *GYTSM*s. On the other hand, for the macro factor $REALPC$ the loadings

[22] Precisely, we project the individual bond yields onto the pricing factors in these models and a constant, and compute the standard deviation of the projection errors.

[23] Pursuing this logic, neither *GMTSM* fits the individual bond yields as well as a three-factor *GYTSM* with the risk factors normalized to be the first three PCs of bond yields. For this model the corresponding measures of mispricing are $(10.4, 6.6, 7.9, 9.3, 12.5)$ basis points.



(a) $GM_3(g)$ PC1 (b) $GM_3(g)$ PC2 (c) $GM_3(g)$ REALPC

(d) $GM_3(g,\pi)$ PC1 (e) $GM_3(g,\pi)$ INFPC (f) $GM_3(g,\pi)$ REALPC

Figure 2: Loadings on the risk factors for models $GM_3(g,\pi)$ and $GM_3(g)$, estimated with and without filtering, and for their associated $FVAR_3$s. The horizontal axis is in years.

for model $GM_3(g)$, estimated with and without filtering, and for model $FVAR_3$, again with and without filtering, are virtually identical, but the patterns across the no-arbitrage and unconstrained $FVAR$ models are somewhat different. However, notice that the loadings on $REALPC$ are small, much smaller than those on $(PC1, PC2)$.[24] As such, the differences in loadings in Figure 2 are inconsequential: the correlations between the six-month ahead forecasts of the corresponding yields on bonds of all maturities considered across models $GM_3(g)$ and $FVAR_3$ are virtually one. In other words, the no-arbitrage restrictions of the canonical $GM_3(g)$ has no affect on forecasts of yields on individual bonds, whether or not the models are estimated accounting for measurement errors.

The same is true for the results from model $GM_3(g,\pi)$. We emphasize again that this irrelevance is obtained even though the standard deviations of the pricing errors in model $GM_3(g,\pi)$ are large both in absolute value and relative to those in model $GM_3(g)$. Accurate pricing of individual bond yields is not a necessary condition for the no-arbitrage structure of a $GMTSM$ to be (nearly) irrelevant for forecasts of individual bond yields.

[24]This is the case even when we scale $PC1$, $PC2$ and $REALPC$ so that they have the same standard deviations.

8 Resolutions of Expectations Puzzles by *GYTSM*s

One of the most widely studied features of the joint distribution of bond yields is the failure of the expectations theory of the term structure (ETTS). According to the ETTS changes in long-term bond yields should move one-to-one with changes in the slope of the yield curve, as long rates are hypothesized to differ from the average of expected future short rates by no more than a constant term premium. Instead, the evidence from US Treasury bond markets suggests that long-term bond yields tend to fall when the slope of the yield curve steepens (e.g., Campbell and Shiller (1991)). Dai and Singleton (2002) and Kim and Orphanides (2005), among others, have shown that the risk premiums inherent in Gaussian term structure models are capable of rationalizing the "puzzling" failure of the expectations theory.

At issue are the coefficients ϕ_n in the projections

$$Proj\left[y_{t+1}^{n-1} - y_t^n | y_t\right] = \alpha_n + \phi_n \left(\frac{y_t^n - r_t}{n-1}\right), \tag{28}$$

where $Proj[\cdot|y_t]$ denotes linear least-squares projection. The ETTS implies that $\phi_n = 1$, for all maturities n. To see under what circumstances (28) holds it is instructive to compare this relationship to the general premium-adjusted expression

$$E^{\mathbb{P}}\left[y_{t+1}^{n-1} - y_t^n - (c_{t+1}^{n-1} - c_t^{n-1}) + \frac{p_t^{n-1}}{n-1} | y_t\right] = \left(\frac{y_t^n - r_t}{n-1}\right), \tag{29}$$

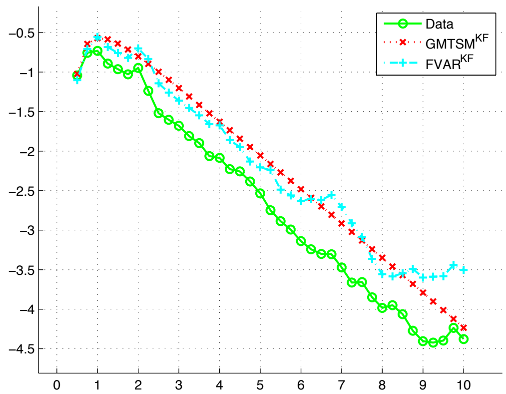
where

$$c_t^n \equiv y_t^n - \frac{1}{n}\sum_{i=0}^{n-1} E^{\mathbb{P}}[r_{t+1}|y_t] \text{ and } p_t^n \equiv f_t^n - E^{\mathbb{P}}[r_{t+n}|y_t] \tag{30}$$

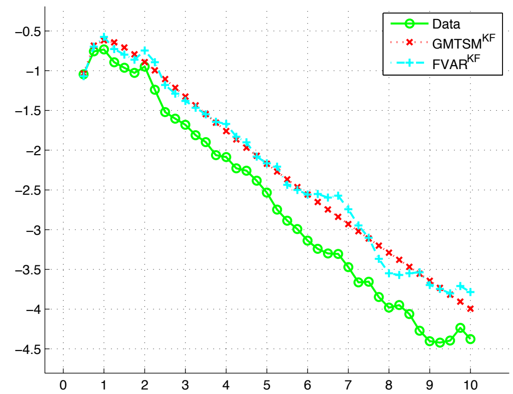
are the yield and forward term premiums, respectively, and f_t^n denotes the forward rate for one-period loans commencing at date $t+n$ (e.g., Dai and Singleton (2002)). A *GYTSM* is considered successful at explaining the failure of the ETTS if the term premiums it generates, through time-varying market prices of risk, reproduce (29). Put differently, a *GYTSM* explains the failure of the ETTS if the model-implied population projection coefficients ϕ_n^{GYTSM} replicate the pattern in the data.

Since the projections in (28) are restrictions on the conditional $\mathbb{P}$-distribution of y_t, Corollary 2 implies that, absent measurement errors on y_t^o, nothing is learned about the failure of the ETTS over and above what one learns from fitting an unconstrained factor-model for y_t based on a $FVAR$ model for $\mathcal{P}_t$. Now, in practice, bond yields are not priced perfectly by low-dimensional *GYTSM*s so, as with the previous illustrations, there is some scope for no-arbitrage restrictions to improve the efficiency of estimators of the ϕ_n.

To investigate the role of no-arbitrage restrictions for testing this hypothesis we estimated three- and four-factor *GYTSM*s (models GY_3 and GY_4, respectively) using the JSZ normalization with the risk factors rotated to be the first $\mathcal{N}$ PCs of bond yields (with $\mathcal{N} = 3$ or 4), and assuming that $\mathcal{P}_t^{\mathcal{N}o} = \mathcal{P}_t^{\mathcal{N}}$. Using the covariances of the steady-state distribution of $\mathcal{P}_t^{\mathcal{N}}$ implied by these *GYTSM*s evaluated at the ML estimates, we compute the projection



(a) Model GY_3^{KF}



(b) Model GY_4^{KF}

Figure 3: Regression coefficients ϕ_n implied by the sample Treasury yields, the GY_3^{KF}, GY_4^{KF} models and their corresponding unconstrained $FVAR$ models. All yields were allowed to be priced with errors. The horizontal axis is in years.

coefficients ϕ_n^{GYTSM}. For comparison we computed the $FVAR$-implied projection coefficients ϕ_n^{FVAR}, now with all of the risk factors being model-implied PCs of the bond yields.[25] The data are again the unsmoothed Fama-Bliss zero yields on US Treasury bonds for the period January, 1972 through December, 2003.

The results for the case of a three-month holding period (that is, the short-term positions are rolled every three months) are displayed in Figure 3.[26] Consistent with the extant evidence, these low-dimensional $GYTSM$s do resolve the expectations puzzle: the implied ϕ_n^{GYTSM} track the estimated ϕ_n from the data quite closely. More to the point of our analysis, the reason that the $GYTSM$ is successful at resolving the ETTS puzzle is because the unconstrained factor model resolves this puzzle. In other words, inherent in the reduced-form factor structure (16) is a pattern of projection coefficients ϕ_n^{FVAR} that approximately matches those from the regression equations (28). The $GYTSM$s also match the regression slopes because their no-arbitrage restrictions are empirically irrelevant for this feature of the conditional $\mathbb{P}$ distribution of bond yields– the $GYTSM$s and the $FVAR$'s produce almost identical conditional distributions of bond yields.[27] With regard to the ETTS, this is manifested in

[25]A practical problem that arises in computing the regression coefficients for the $FVAR$ model is that, from the twelve yields used in estimation of the $GYTSM$ we cannot determine the loadings on the risk factors for all of the maturities. For those maturities that were not used in estimation we projected the yields onto the risk factors using OLS regression and used these loading to compute the $FVAR$-implied coefficients.

[26]We focus on the case of three-month holding periods, because this is the shortest maturity Treasury bond that was used in estimation of the $GYTSM$. The results for shorter and longer holding periods are qualitatively similar.

[27]Implicit in this finding is a very close similarity between the cross-sectional patterns of factor loadings produced by OLS projections of bond yields onto the risk factors (PCs of bond yields) and the loadings produced by the arbitrage-free term structure models.

the nearly identical patterns of projection coefficients ϕ_n^{FVAR} and ϕ_n^{GYTSM} in Figure 3.

9 Extensions

Up to this point we have focused on the class of *GMTSM*s that presume that (i) X_t follows a first-order Markov process under $\mathbb{P}$ and $\mathbb{Q}$ and (ii) the macro-factors are spanned in the sense that they incrementally affect pricing after controlling for the remaining factors. Several studies of *GMTSM*s have relaxed these assumptions by assuming that either that the data-generating process for X_t is a higher-order VAR under both $\mathbb{P}$ and $\mathbb{Q}$ or that there are components of macro-economic risks that are unspanned by the yield curve. We now show that our central arguments follow in these instances: canonical models of either of these types produce nearly identical estimates of the historical distribution of macro-factors and bond yields as the associated factor vector autoregression.

9.1 Higher-Order VAR Models of Risk Factors

Consider the extended family of models:

Definition 2. *Fix the set of $\mathcal{M}$ macroeconomic variables M_t. $GMTSM_{\mathcal{N},p}(\mathcal{M})$ is the set of invariant affine transformations of non-degenerate* GMTSMs *in which r_t is an affine function of $\mathcal{L}$ latent (L_t) and $\mathcal{M}$ macro (M_t) factors and the vector $X_t = (M_t', L_t')'$ follows the Gaussian processes (2) under $\mathbb{Q}$ and*

$$\begin{pmatrix} M_t \\ L_t \end{pmatrix} = \begin{pmatrix} K_{0M}^{\mathbb{P}} \\ K_{0L}^{\mathbb{P}} \end{pmatrix} + \begin{pmatrix} K_{1MM}^{\mathbb{P}} & K_{1ML}^{\mathbb{P}} \\ K_{1LM}^{\mathbb{P}} & K_{1LL}^{\mathbb{P}} \end{pmatrix} \begin{pmatrix} \overrightarrow{M}_{t-1,p} \\ \overrightarrow{L}_{t-1,p} \end{pmatrix} + \sqrt{\Sigma_X}\epsilon_t^{\mathbb{P}} \tag{31}$$

under $\mathbb{P}$, where, for any Z_t, $\overrightarrow{Z}_{t,p} \equiv (Z_t, Z_{t-1}, \ldots, Z_{t-p+1})$.

With regard to the $\mathbb{P}$ distribution of the risk factors, this formulation nests many prior macro-finance term structure models with lags, including Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007), and Jardet, Monfort, and Pegoraro (2010). It does not, however, nest their representations under the pricing distribution.[28] These studies adopt specifications of the market prices of risk that imply that the $\mathbb{Q}$ distribution of X inherits the lag structure of the $VAR(p)$ under $\mathbb{P}$. Descriptive evidence supporting higher-order lags in *GMTSM*s can be based on the application of model selection criteria under the historical distribution. Left open by such descriptive evidence supporting $p > 1$ in (31) is the nature of the dependence of r_t and X_t on lags of X_t under the pricing distribution. As previously noted, within the family of reduced-form *GMTSM*s (with or without lags), neither the risk factors nor their weights have structural interpretations, so economic theory, as well, is silent on this issue.

[28] It is straightforward to extend our theoretical results to allow for $\mathbb{Q}$-dependence on lags of macro variables - the setups of Ang and Piazzesi (2003) and Ang, Dong, and Piazzesi (2007). Guided by the evidence below, we omit these additional lags from our analysis.

Model	M_t	History (q)	$y_t^{0.5yr}$	y_t^{1yr}	y_t^{2yr}	y_t^{5yr}	y_t^{7yr}	y_t^{10yr}
GM_3	$REALPC$	1	11	12	14	10	10	17
GM_3	$REALPC$	6	11	12	14	9	9	17
GM_3	$REALPC$	12	11	11	13	9	9	16
GM_3	$REALPC, INFPC$	1	54	47	26	18	29	40
GM_3	$REALPC, INFPC$	6	52	44	24	17	27	39
GM_3	$REALPC, INFPC$	12	49	40	21	16	26	35
GM_4	$HELP, CPI$	1	10	12	14	10	10	16
GM_4	$HELP, CPI$	6	9	12	13	9	9	15
GM_4	$HELP, CPI$	12	9	11	12	9	9	14
GM_4	$REALPC, INFPC$	1	10	12	14	10	9	16
GM_4	$REALPC, INFPC$	6	10	12	13	9	9	16
GM_4	$REALPC, INFPC$	12	9	11	12	9	9	15

Table 4: Root-mean-squared fitting errors, measured in basis points, from projections of bond yields onto current and lagged values of the risk factors X_t^o. For given q, the conditioning information is $(X_t, X_{t-1}, \ldots, X_{t-q+1})$.

Indirect guidance as to the lag structure of the $\mathbb{Q}$ distribution of (r_t, X_t) is provided by the link between the order of the VAR under $\mathbb{Q}$ (say q) and the effective number of risk factors ($q\mathcal{N}$). Consider, for example, the models GM_3 and GM_4. If the order of the VAR representation of X_t under $\mathbb{Q}$ is q, then bond yields will depend on the history $\overrightarrow{X}_{t,q}$. The quantitative relevance of extending a *GMTSM* to values of q larger than one can be determined by projecting yields onto current and $q-1$ lags of the state vector X_t^o in these models.[29] The standard deviations of the errors in these projections ($RMSE$ in basis points), for $q = 1, 6, 12$ (one year in our monthly data) and various compositions of macro risk factors M_t, are presented in Table 4. Clearly, for three of the four cases, the improvements in fitting these bond yields from allowing for $q > 1$ are tiny, at most one or two basis points.[30] In all cases, the (AIC, BIC) model selection criteria select $q = 1$.

With these observations in mind, and supported by the evidence in Table 4, we proceed to explore asymmetric formulations of *GMTSM*s in which X_t follows a $VAR(p)$ under $\mathbb{P}$, and X_t follows a $VAR(1)$ under $\mathbb{Q}$. Since the $\mathbb{Q}$ distributions for models in $GMTSM_{\mathcal{N},p}(\mathcal{M})$ and $GMTSM_{\mathcal{N}}(\mathcal{M})$ are identical, the risk factors (M_t, L_t) can once again be expressed as an affine function of $(M_t, \mathcal{P}_t^{\mathcal{L}})$ as in (5). Further, the invariant transformation (7) gives an observationally equivalent model in which $X_t' = (M_t', \mathcal{P}_t^{\mathcal{L}\prime})$, r_t is given by (8), and X_t follows

[29] Strictly speaking, these projections speak directly to variants of models that assume $\mathcal{P}_t^{2o} = \mathcal{P}_t^2$. However, we have seen that the filtered PCs are nearly identical to $\mathcal{P}_t^{2o}$ in the variants with measurement errors ($\mathcal{P}_t^{2o} \neq \mathcal{P}_t^2$), so the following observations are relevant for both cases.

[30] The only exception is model $GM_3(g, \pi)$ with state vector $(PC1, REALPC, INFPC)$. Here the fit is so poor, with $RMSE$s as large as 54bp, that adding lags under $\mathbb{Q}$ improves the $RMSE$s a bit more, up to five basis points.

the process (2) under $\mathbb{Q}$, with the parameters governing (8) and (2) being explicit functions of $\Theta^{\mathbb{Q}} = (\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_{\infty}, \Sigma_X, \gamma_0^W, \gamma_1^W)$.[31] Our previous argument extends to give

Theorem 2. *Fix a full-rank portfolio matrix $W \in \mathbb{R}^{J\times J}$, and let $\mathcal{P}_t = Wy_t$. Any canonical form for the family of models $GMTSM_{\mathcal{N},p}(\mathcal{M})$ is observationally equivalent to a unique member of $GMTSM_{\mathcal{N},p}(\mathcal{M})$ in which the first $\mathcal{M}$ components of the pricing factors are the macro variables M_t, and the remaining $\mathcal{L}$ components are $\mathcal{P}^{\mathcal{L}}_t$; r_t is given by (8); M_t is related to $\mathcal{P}^{\mathcal{N}}_t$ according to (9); the risk factors follow the Gaussian process (2) under $\mathbb{Q}$ and (31) under $\mathbb{P}$, where $K_0^{\mathbb{Q}}$, $K_1^{\mathbb{Q}}$, Σ_X, ρ_{0X}, and ρ_{1X} are explicit functions of $\Theta^{\mathbb{Q}}$. For given W, our canonical form is parametrized by $\Theta^X = (\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_{\infty}, K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma_X, \gamma_0^W, \gamma_1^W)$.*

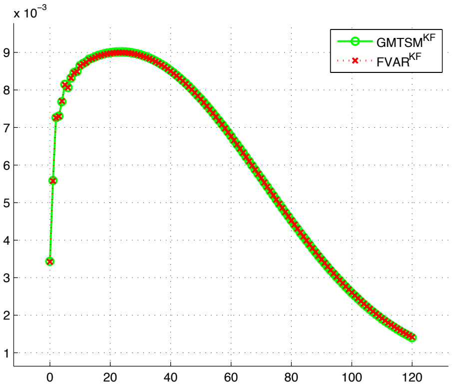
Our main analysis comparing the properties of no arbitrage models with their associated $FVAR$s follows through for this generalized *GMTSM* with lags. The source of this robustness of our analysis is again the separation of the parameter space: except for Σ_X, the $\mathbb{P}$-parameters are distinct from the $\mathbb{Q}$-parameters. Therefore, when X_t is measured without errors, ML estimates of the $\mathbb{P}$-mean parameters can be obtained directly by fitting X_t to a $VAR(p)$. Moreover, with or without lags, Σ_X only affects yield levels and not their factor loadings. Consequently, the first-order derivative of the density of the pricing errors with respect to Σ_X must be proportional to the average-to-variance ratio of errors defined earlier. Therefore, corollaries 1 and 2 continue to hold. Thus, provided that the average pricing errors are small relative to their standard deviations, one should expect the model-implied conditional $\mathbb{P}$-distribution of X_t to be nearly identical to its counterpart implied by an unconstrained factor-VAR model of bond yields. We proceed to explore our more parsimonious specification with the goal of examining whether a richer lag structure under $\mathbb{P}$ impacts our earlier finding that the conditional distributions of bond yields are largely invariant to the imposition of the no-arbitrage structure of a *GMTSM*.

To examine this issue empirically we generalize models $GM_3(REALPC)$ and $GM_4(HELP, CPI)$ to $VAR(p)$ processes under $\mathbb{P}$. For each of these models, we first fit X_t to a $VAR(p)$ over a wide range of p's and choose the optimal number of lags according to their BIC and AIC scores. For model GM_3, the optimal number of lags according to the BIC (AIC) is 2 (5). For model GM_4, the optimal number of lags according to the BIC (AIC) is 1 (4). Given our objective of examining the sensitivity of the conditional distribution of bond yields to high-order lag structures we choose $p = 5$ ($p = 4$) in generalizing GM_3 (GM_4). We denote these models by $GM_{3,5}$ and $GM_{4,4}$, respectively, where the second subscripts refer to the number of lags in each model.

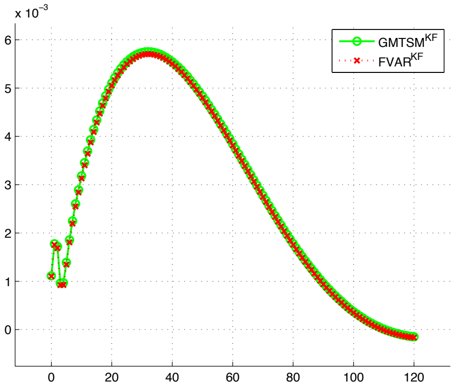
[31] The lag structure of the most flexible models with lags precludes the direct application of the normalization strategies in JSZ or in our Theorem 1. That is, if one starts with a *GMTSM* in which the risk factors X_t are latent and satisfy

$$X_t = K^{\mathbb{Q}}_{0X} + \sum_{i=1}^{q} K^{\mathbb{Q}}_{iX} X_{t-i} + \sqrt{\Sigma_X}\epsilon^{\mathbb{Q}}_t,$$

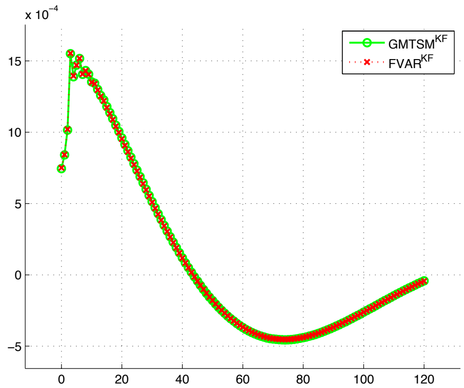
then in general it is not possible to find a portfolio matrix W such that $\mathcal{P}^{\mathcal{N}}_t$ can be substituted for X_t in this expression. Instead, premultiplying by the first $\mathcal{N}$ rows of W and inverting the lag polynomial, we can express X_t as an infinite-order distributed lag of $\mathcal{P}^{\mathcal{N}}_t$, and when this expression is substituted into the above VAR the resulting time-series model for $\mathcal{P}^{\mathcal{N}}_t$ inherits this infinite order.



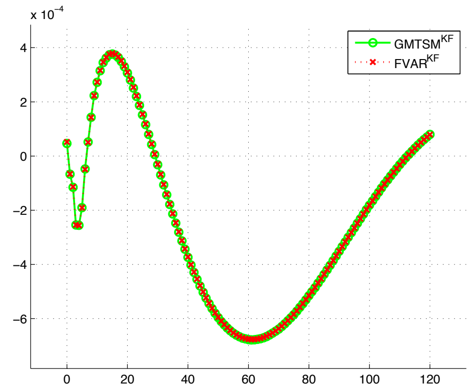
(a) Response of $PC1$ to $HELP$



(b) Response of $PC1$ to CPI



(c) Response of $PC2$ to $HELP$



(d) Response of $PC2$ to CPI

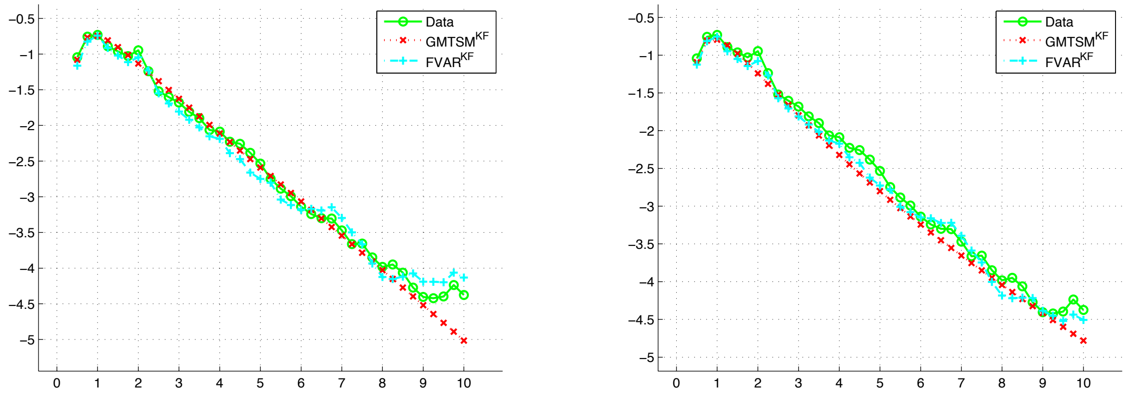
Figure 4: Impulse responses of $PC1$ and $PC2$ to innovations in $HELP$ and CPI based on ML estimates of model $GM_{4,4}^{KF}$. The horizontal axis is months, and the risk factors are ordered as $(CPI, HELP, PC1, PC2)$.

As in models with $p = 1$, the no-arbitrage and $FVAR$ models give very similar estimates for the parameters governing the conditional mean and conditional covariance of the risk factors, even when allowing for measurement errors on all yields.[32] Figure 4 displays the impulse responses implied by model $GM_{4,4}^{KF}$ which depend on both $K_1^{\mathbb{P}}$ and Σ_X.[33] As before, the imposition of no arbitrage is virtually inconsequential for how shocks to macro factors impact the yield curve. This finding is robust to choices of which macro variables are included. For example, a very similar result is obtained from model $GM_{3,5}$.

Finally, we take the two $GYTSM$s considered earlier (GY_3 and GY_4) and generalize their $\mathbb{P}$

[32] For brevity we do not report these ratios owing to the large number of parameters.

[33] The higher-order lag structure alters somewhat the individual responses (compare Figures 1 and 4). The choppy behavior over short horizons for some of the responses in Figure 4 is also evident in the impulse responses reported in Ang and Piazzesi (2003) for their $GMTSM$ with lags.



(a) Model $GY_{3,2}^{KF}$ (b) Model $GY_{4,2}^{KF}$

Figure 5: Regression coefficients ϕ_n implied by the sample Treasury yields, the models including two lags, $GY_{3,2}^{F}$ and $GY_{4,2}^{F}$, and their corresponding unconstrained factor-VAR model. All yields were allowed to be priced with errors. The horizontal axis is in years.

dynamics to a $VAR(p)$, models $GY_{3,2}$ and $GY_{4,2}$. Using the same model selection procedure, the optimal lags for both models is two. In Figure 5, we compare the Campbell-Shiller regression coefficients ϕ_n implied by the two models with those implied by the corresponding $FVAR$'s. Comparing Figure 5 to its counterpart without lags, Figure 3, it is seen that the additional lag helps the two models match the empirical coefficients. More to the point of our analysis, what remains intact is the similarity between the ϕ_n's implied by $GY_{3,2}^{KF}$ and $GY_{4,2}^{KF}$ and their corresponding $FVAR$'s. The no-arbitrage restrictions are inconsequential for the success of these $GYTSM$s in resolving the expectations puzzle.

9.2 Models with Unspanned Risk Factors

Up to this point, the models we have considered have the macro variables enter directly as risk factors determining the short-term interest rates. This is consistent with the large majority of the extant literature which has incorporated macro risks in this manner. Building upon the framework in Joslin, Priebsch, and Singleton (2010), a different class of models that allow for unspanned yield and macro risks– risks that cannot be replicated by linear combinations of bond yields– has recently been explored (for example, Wright (2009) and Barillas (2010)). Canonical versions of these models with unspanned risks also share two important properties: (1) except for the volatility parameter (Σ_X), the $\mathbb{P}$-parameters are distinct from the $\mathbb{Q}$-parameters; and (2) Σ_X only affects yields levels and not loadings of yields on risk factors. The implication of property (1) is analogous to Corollary 1: when one assumes the risk factors are observed without error, forecasts agree identically with the corresponding $FVAR$. Likewise, property (2) is analogous to Corollary 2: the deviation of ML estimate of Σ_X from its $FVAR$-counterpart is proportional to the average-to-variance ratio of pricing errors. Therefore, so long as one considers canonical models with unspanned risk factors, the historical distribution of the yields and macro-factors estimating using either the FVAR or the no-arbitrage model will be nearly identical.[34]

10 Concluding Remarks

We have shown theoretically and documented empirically that the no-arbitrage restrictions of canonical *GMTSM*s have essentially no impact on the ML estimates of the joint conditional distribution of the macro and yield-based risk factors, including models that nest some of the most widely studied *GMTSM*s in the literature. This finding is robust to whether a subset of the bond yields are priced perfectly by a *GMTSM*, or all yields are measured with error and filtering is used in ML estimation.

Of course this finding does not imply that *GYTSM*s or *GMTSM*s are of little value for understanding the risk profiles of portfolios of bonds. Our entire analysis has been conducted within canonical forms that offer maximal flexibility in fitting both the conditional $\mathbb{P}$ and $\mathbb{Q}$ distributions of the risk factors. Restrictions on risk premiums in bond markets typically amount to constraints across these distributions, and such constraints cannot be explored outside of a term structure model that (implicitly or explicitly) links the $\mathbb{P}$ and $\mathbb{Q}$ distributions of yields. Moreover, the presence of constraints on risk premiums will in general imply that ML estimation of a *GMTSM* will lead to more efficient estimates of the $\mathbb{P}$ distribution of yields relative to those of the factor model (15)-(16).

Whether such efficiency gains will be sizable is an empirical question and will likely depend on the nature of the constraints imposed. JSZ found that the constraints on the feedback matrix $K_{1X}^{\mathbb{P}}$ imposed by Christensen, Diebold, and Rudebusch (2009) in their analysis of *GYTSM*s had small effects on out-of-sample forecasts. Further, Ang, Dong, and Piazzesi (2007) found that impulse response functions implied by their three-factor $(\mathcal{M} = 2, \mathcal{L} = 1)$ *GMTSM* that imposes various zero restrictions on lag coefficients and the parameters governing the market prices of risk were nearly identical to those computed from their corresponding unrestricted VAR. Both of these studies illustrate cases where our propositions on the near irrelevance of no-arbitrage restrictions in *GMTSM*s (and *GYTSM*s) carry over to non-canonical models.

Similarly, unit-root or cointegration-type restrictions imposed directly on the $\mathbb{P}$ distribution of the risk factors (JSZ, Duffee (2009), Jardet, Monfort, and Pegoraro (2010)) also seem unlikely to induce large differences between the conditional $\mathbb{P}$ distributions implied by a *GMTSM* and the corresponding $FVAR$ with the same restrictions imposed. JSZ, for instance, obtained very similar conditional mean parameters under cointegration from a *GYTSM* and a $FVAR$ estimated using US Treasury data.

On the other hand, Joslin, Priebsch, and Singleton (2010) found that the constraints selected by Schwarz (1978)'s Bayesian information criteria led to substantial differences

[34] In the case that yields or macro variables are forecastable by variables not in their joint span, this applies only to the comparison of the no arbitrage model and the FVAR which are estimated by Kalman filtering. This is because in this case the assumption that $\mathcal{P}_t = \mathcal{P}_t^o$ cannot hold by construction.

between the selected model and both its canonical and unconstrained VAR counterparts. Thus, undertaking a systematic model-selection exercise may point to constraints that break the irrelevance results documented here.[35] Similarly, the constraint that expected excess returns lie in a lower than $\mathcal{N}$-dimensional space (Cochrane and Piazzesi (2005), JSZ), which effectively amounts to constraining the market prices of risk, might also have material effects.

From what we know so far, evaluating how one's choice of constraints on a $GMTSM$ affects the model-implied historical distribution of bond yields and macro variables, relative to the distribution from a VAR, seems likely to be an informative exercise.

[35] A typical strategy for achieving parsimony in dynamic term structure models is to first estimate a canonical (or nearly canonical) model, to then set to zero the parameters of the market price of risk that are small relative to their estimated standard errors, and finally to re-estimate this constrained model. See, for examples, Dai and Singleton (2000), Ang and Piazzesi (2003), and Bikbov and Chernov (2010).

References

Ang, A., S. Dong, and M. Piazzesi, 2007, "No-Arbitrage Taylor Rules," Discussion paper, Columbia University.

Ang, A., and M. Piazzesi, 2003, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745–787.

Ang, A., M. Piazzesi, and M. Wei, 2006, "What Does the Yield Curve Tell us About GDP Growth?," *Journal of Econometrics*, 131, 359–403.

Barillas, F., 2010, "Can we Exploit Predictability in Bond Markets?," Discussion paper, New York University.

Bikbov, R., and M. Chernov, 2010, "No-Arbitrage Macroeconomic Determinants of the Yield Curve," *Journal of Econometrics*, forthcoming.

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

Chen, R., and L. Scott, 1993, "Maximum Likelihood Estimation For a Multifactor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income*, 3, 14–31.

Chernov, M., and P. Mueller, 2009, "The Term Structure of Inflation Expectations," Discussion paper, London Business School.

Christensen, J. H., F. X. Diebold, and G. D. Rudebusch, 2009, "An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model," *The Econometrics Journal*, 12, 33–64.

Chun, A., 2009, "Expectations, Bond Yields and Monetary Policy," Discussion paper, Copenhagen Business School.

Cochrane, J., 2007, "Identification with Taylor Rules: A Critical Examination," Discussion paper, University of Chicago.

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

Collin-Dufresne, P., R. Goldstein, and C. Jones, 2008, "Identification of Maximal Affine Term Structure Models," *Journal of Finance*, LXIII, 743–795.

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

Dai, Q., and K. Singleton, 2002, "Expectations Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics*, 63, 415–441.

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

Dempster, A., N. Laird, and D. Rubin, 1977, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1–38.

Digalakis, V., J. Rohlicek, and M. Ostendorf, 1993, "ML Estimation of a Stochastic Linear System with the EM Algorithm and Its Application to Speech Recognition," *IEEE Transactions on Speech and Audio Processing*, 1, 431–442.

Duffee, G., 2009, "Forecasting with the Term Structure: the Role of No-Arbitrage," Discussion paper, Working Paper, UC Berkeley.

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

Jardet, C., A. Monfort, and F. Pegoraro, 2010, "No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth," Discussion paper, Banque de France.

Joslin, S., 2006, "Can Unspanned Stochastic Volatility Models Explain the Cross Section of Bond Volatilities?," Discussion paper, MIT.

Joslin, S., M. Priebsch, and K. Singleton, 2010, "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," Discussion paper, Stanford University.

Joslin, S., K. Singleton, and H. Zhu, 2010, "A New Perspective on Gaussian DTSMs," Discussion paper, forthcoming, *Review of Financial Studies*.

Kim, D., and A. Orphanides, 2005, "Term Structure Estimation with Survey Data on Interest Rate Forecasts," Discussion paper, Federal Reserve Board.

Pericoli, M., and M. Taboga, 2008, "Canonical Term-Structure Models with Observable Factors and the Dynamics of Bond Risk Premia," *Journal of Money, Credit and Banking*, 40, 1471–1488.

Piazzesi, M., 2010, "Affine Term Structure Models," in Y. Ait-Sahalia, and L. Hansen (ed.), *Handbook of Financial Econometrics* . chap. 12, pp. 691–766, Elsevier B.V.

Schwarz, G., 1978, "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Smith, J., and J. Taylor, 2009, "The Term Structure of Policy Rules," *Journal of Monetary Economics*, 56, 907–919.

Wright, J., 2009, "Term Premiums and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset," Discussion paper, Johns Hopkins University.

A Derivations for Section 5

First, starting from the canonical $\mathbb{Q}$ specification of JSZ:

$$Z_{t+1} = diag(\lambda^{\mathbb{Q}})Z_t + \sqrt{\Sigma_Z}\epsilon_{t+1} \tag{32}$$

the yield curve is given by:

$$y_t = A_Z + B_Z Z_t \tag{33}$$

where B_Z only dependent on $\lambda^{\mathbb{Q}}$, and:

$$A_Z = \vec{1}r_\infty^{\mathbb{Q}} + \beta_Z vec(\Sigma_Z) \tag{34}$$

where β_Z is only dependent on $\lambda^{\mathbb{Q}}$.

Second, rotating the risk factors to $\mathcal{P}_t^{\mathcal{N}} = W^{\mathcal{N}} y_t$ gives rise to the following yield curve:

$$y_t = A_{\mathcal{P}} + B_{\mathcal{P}}\mathcal{P}_t^{\mathcal{N}} \tag{35}$$

where $B_{\mathcal{P}}$ is still only dependent on $\lambda^{\mathbb{Q}}$, and:

$$A_{\mathcal{P}} = (I - B_{\mathcal{P}}W^{\mathcal{N}})A_Z = (I - B_{\mathcal{P}}W^{\mathcal{N}})(\vec{1}r_\infty^{\mathbb{Q}} + \beta_Z vec(\Sigma_Z)). \tag{36}$$

In addition, by diffusion invariance, we must have:

$$(W^{\mathcal{N}}B_Z)\Sigma_Z(W^{\mathcal{N}}B_Z)' = \Sigma_{\mathcal{P}}, \tag{37}$$

which, after taking vec of both sides and inverting, gives:

$$vec(\Sigma_Z) = ((W^{\mathcal{N}}B_Z) \otimes (W^{\mathcal{N}}B_Z))^{-1} vec(\Sigma_{\mathcal{P}}). \tag{38}$$

From (36) and (38), we can write:

$$A_{\mathcal{P}} = \alpha_{\mathcal{P}} r_\infty^{\mathbb{Q}} + \beta_{\mathcal{P}} vec(\Sigma_{\mathcal{P}}). \tag{39}$$

Third, rotating the risk factors further to $X_t = \Gamma_0 + \Gamma_1 \mathcal{P}_t^{\mathcal{N}}$ gives rise to the following yield curve:

$$y_t = A_X + B_X X_t \tag{40}$$

where

$$B_X = B_{\mathcal{P}}\Gamma_1^{-1}, \tag{41}$$

$$A_X = A_{\mathcal{P}} - B_{\mathcal{P}}\Gamma_1^{-1}\Gamma_0 = \alpha_{\mathcal{P}} r_\infty^{\mathbb{Q}} + \beta_{\mathcal{P}} vec(\Sigma_{\mathcal{P}}) - B_X\Gamma_0. \tag{42}$$

In addition, we need:

$$\Sigma_X = \Gamma_1 \Sigma_{\mathcal{P}} \Gamma_1', \tag{43}$$

which, similar to before, gives: $vec(\Sigma_X) = (\Gamma_1 \otimes \Gamma_1) vec(\Sigma_{\mathcal{P}})$. Therefore we can write:

$$A_X = \alpha_{\mathcal{P}} r_\infty^{\mathbb{Q}} - B_X\Gamma_0 + \beta_X vec(\Sigma_X) \tag{44}$$

where $\alpha_{\mathcal{P}}$ is only dependent on $\lambda^{\mathbb{Q}}$, β_X and B_X are only dependent on $\lambda^{\mathbb{Q}}$ and Γ_1.