On the Observational Equivalence of Kalman-Filter Estimates of Gaussian Macro-Term Structure and Unconstrained State-Space Models

Scott Joslin* Anh Le† Kenneth J. Singleton‡

First draft: June, 2010
This draft: February 15, 2011
PRELIMINARY AND INCOMPLETE

Abstract

This paper explores the impact of simultaneously enforcing the no-arbitrage structure of a Gaussian macro-finance term structure model (*MTSM*) and accommodating measurement errors on bond yield through filtering on the maximum likelihood estimates of the model-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 *MTSM*-implied joint distribution is nearly identical to its counterpart estimated from the economic-model-free state-space model (factor-VAR). The practical implication of this finding is that a canonical *MTSM* 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 factor-VAR. In particular, the discipline of a canonical *MTSM* is empirically inconsequential for analyses of impulse response functions of bond yields and macro factors or resolutions of the failure of the expectations theory of the term structure. 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 (*MTSM*s) typically feature three key ingredients: (i) a low-dimensional factor-structure in which the risk factors are both macroeconomic and yield-based variables; (ii) the assumption of no arbitrage opportunities in bond markets; and (iii) accommodation of measurement errors in bond markets owing to the presence of microstructure noise or errors introduced by the bootstrapping of zero-coupon yields. The low-dimensional factor structure is motivated by the observation that most of the variation in bond yields is explained by a small number of principal components (PCs).[1] The overlay of an arbitrage-free *MTSM* on the representations of the short-term rate brings information about the entire yield curve to bear on the links between macroeconomic shocks and bond yields, in a consistent structured way. Thirdly, with measurement errors on bond yields,[2] *MTSM*s are formulated as state-space models and estimation proceeds using filtering.

This paper takes the low-dimensional factor structure of bond yields and macro factors imposed in *MTSM*s as given and explores the implications of no-arbitrage and the use of filtering for the maximum likelihood (ML) estimator of the joint distribution of these variables. We derive sufficient and easily verified theoretical conditions for a canonical[3] *MTSM* and an unconstrained state-space model to lead to identical ML estimators of the joint distribution of the risk factors, even when all bonds are priced imperfectly by the *MTSM* and estimation proceeds by filtering. We proceed to show, using data on a variety of yield and macro risk factors, that these conditions are very nearly satisfied by the canonical versions of several prominent specifications of *MTSM*s. The practical implication of our analysis is that canonical *MTSM*s typically do not offer any new insights into economic questions regarding the historical distribution of macro variables and yields, over and above what one can learn from an economics-free state-space model (factor-VAR).

Among the most widely studied properties *MTSM*s are their implied impulse responses (IRs) of bond yields to shocks to output or inflation,[4] and their ability to reproduce the anomalous correlations between change in long-term bond yields and the slope of the yield curve relative to the expectations hypothesis.[5] Both of these features of *MTSM*s co-depend

[1] This has been widely documented for U.S. Treasury yields (e.g., Litterman and Scheinkman (1991)). Ang, Piazzesi, and Wei (2006) and Bikbov and Chernov (2010) are among the many studies of *MTSM*s that base their selection of a small number of risk factors (typically three or four) on similar PC evidence.

[2] A low-dimensional factor structure does not perfectly fit the term structure of yields. See Duffee (1996) for a discussion of measurement issues at the short end of the Treasury curve. In addition, the use of splines to extract zero-coupon yields from coupon yield curves and the differing degrees of liquidity of individual bonds along the yield curve introduce errors in the measurement of yields.

[3] A canonical model for a family of *MTSM*s is one in which maximally flexible (in the sense that each member of the family is represented) and which has a minimal set of normalizations imposed to ensure econometric identification.

[4] Examples include Ang and Piazzesi (2003) who examine the responses of bond yields to their macro risk factors; Bikbov and Chernov (2010) who quantify the proportion of bond yield variation attributable to macro risk factors; and Joslin, Priebsch, and Singleton (2010) who quantify the effects of unspanned macro risks on forward term premiums.

[5] The expectations puzzle (e.g., Campbell and Shiller (1991)) has been examined within Gaussian term structure models by Dai and Singleton (2002) and Kim and Orphanides (2005), among others.

on model-implied conditional means and variances of the risk factors. For impulse responses this codependence arises through the normalizations required to identify the innovations in (shocks to) the yields and macro variables. This codependence arise directly in evaluating expectations puzzles by definition of the *MTSM*-implied slope coefficients in projections of changes in bond yields onto the slope of the yield curve.

We show that the imposition of the structure of a canonical *MTSM* is empirically inconsequential for analyses of IR functions and inferences about the effects of macro factors on term premiums. More precisely we show that, when all bond yields are measured with errors, the joint distributions of bond yields and macro factors implied by a canonical *MTSM* and its unconstrained factor-VAR counterpart are virtually identical. Imposition of the over-identifying no-arbitrage restrictions has little impact on Kalman filter estimates of either the conditional mean or variance parameters of the the risk factors Z_t. This result is fully rotation invariant: any normalization of the latent or yield-based risk factors necessarily gives identical results.[6] Moreover when the non-macro pricing factors are normalized to be the theoretical low-order PCs of bond yields, then the model-implied joint distribution of the risk factors Z_t is virtually identical to the one implied by a standard unconstrained VAR model of the *observed* risk factors Z_t^o. That is, for studying the joint distribution of the low-order PCs and macro factors, neither the no-arbitrage structure of a *MTSM* nor filtering to accommodate measurement errors lead to new insights over what is learned from a standard VAR model estimated by OLS.

Illustrative of the practical implications of our theoretical results are the IRs in a *MTSM* that has two macro risk factors, representing real growth and inflation, and one latent risk factor (model $GM_3(g, \pi)$).[7] The no-arbitrage structure of $GM_3(g, \pi)$ implies over-identifying restrictions on the distribution of bond yields, and it is estimated using the Kalman filter to accommodate measurement errors in all of the bond yields. Nevertheless, the IRs of the first PC of bond yields ($PC1$) to a shock to CPI inflation implied by $GM_3(g, \pi)$ and by its corresponding factor-VAR ($FVAR$) are virtually indistinguishable (Figure 1).

At the core of our irrelevancy results is the proposition that the *MTSM*- and factor-VAR-based ML estimators of the conditional covariance matrix of the risk factors converge to each other (in a sense we make precise subsequently) as the average pricing errors on yield-based risk factors, relative to their standard deviations, approach zero. Without loss of generality, canonical *MTSM*s can be rotated so that any risk factors that are latent or represented by yields on specific bonds are replaced by low-order PCs of bond yields. So, essentially, what gives rise to the irrelevance of no-arbitrage restrictions or the presence of measurement errors on individual yields in *MTSM*s is the empirical regularity that the filtered low-order PCs from *MTSM*s tend to accurately replicate their sample counterparts. As we document through several examples, accurate matching of PCs does not require accurate pricing of individual bonds. In particular, this explains why our propositions apply with equal force to

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

[7]Full details of the data and estimation results are provided in Section 4.2.

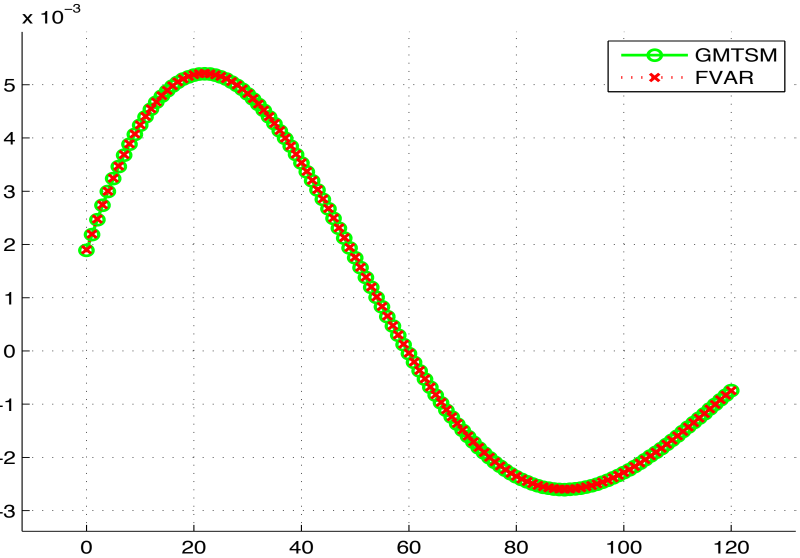


Figure 1: Impulse Responses of $PC1$ to a shock to CPI Inflation in Model $GM_3(g,\pi)$.

$GM_3(g,\pi)$ which implies large mean-squared pricing errors (RMSEs) on individual yields.

To derive our irrelevance results we develop a canonical form for the family of $\mathcal{N}$-factor $MTSM$s in which $\mathcal{M}$ of the factors ($\mathcal{M} < \mathcal{N}$) are macro variables, M_t. Under the assumption that M_t incrementally affects bond prices beyond the other $\mathcal{L} = \mathcal{N} - \mathcal{M}$ risk factors, we derive a canonical form for $MTSM$s in which the pricing factors are M_t and the first $\mathcal{L}$ principal components (PCs) of bond yields, $\mathcal{P}_t^{\mathcal{L}}$. This form is particularly revealing about the nature of the over-identifying restrictions implied by term structure models with macro risk factors, and of sufficient theoretical conditions for no-arbitrage restrictions to have no effect on the ML estimator of the historical distribution of the risk factors. When these conditions are (nearly) satisfied, a canonical $MTSM$ and a model-free factor-VAR will produce (nearly) identical estimates of any relationships among the risk factors that are fully describable in terms of the parameters of their historical ($\mathbb{P}$) distribution.

These propositions are a mix of theoretical results that hold exactly for each sample of yields and of approximations that are accurate as the average-to-variance ratios of a $MTSM$'s pricing errors approach zero. We initially explore their empirical relevance within three-factor $MTSM$s in which the risk factors are various combinations of output growth, inflation, and PCs of bond yields (in the spirit of Ang and Piazzesi (2003)). Attention is then turned to a $MTSM$ in which M_t is not spanned by bond yields and the question of whether the imposition of the structure of a $MTSM$ sheds new light on the failure of the expectations theory. Based on a variety of standard data sets, the conditional $\mathbb{P}$-distributions of the risk factors implied by these $MTSM$s and their unconstrained factor-VAR counterparts are nearly identical.

These illustrations presume that Z_t follows a first-order Markov process. Several implementations of $MTSM$s have allowed for higher-order lags. 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 factor-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 *MTSM*s with this asymmetric $\mathbb{P}/\mathbb{Q}$ lag structure.

This paper builds upon and complements several recent studies. Joslin, Singleton, and Zhu (2010) show theoretically for any canonical Gaussian model with latent or yield-based risk factors (*YTSM*) that if the $\mathcal{N}$ pricing factors are priced perfectly, then the ML estimator of the conditional mean of these factors is invariant to the imposition of no-arbitrage restrictions. Building on their results, Duffee (2011) shows through Monte Carlo analysis (again with perfectly priced risk factors in *YTSM*s) that estimates of the loadings in the mapping between risk factors and bond yields are also largely invariant to the imposition of no-arbitrage restrictions. Our focus is on the implications of no-arbitrage and filtering for the ML estimator of the *entire* conditional distribution of the risk factors and bond yields in *over-identified MTSM*s when *all yields are priced imperfectly* by the model.

Our irrelevancy results apply to the maximally flexible *canonical* forms of *MTSM*s. Certain types of restrictions, when imposed in combination with the no-arbitrage restrictions of a *MTSM*, may increase the efficiency of ML estimators relative to those of the unconstrained VAR. Most studies of *MTSM*s have left open the question of whether their particular formulations led to materially different estimates of historical distributions relative to those from a VAR.[8] In our concluding section we draw upon on analysis to assess what types of constraints might create such a wedge.

To fix notation, suppose that a *MTSM* is to be evaluated using a set of J yields $y_t = (y_t^{m_1}, \ldots, y_t^{m_J})'$ with maturities $(m_1, \ldots, m_J)$ and 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 $\mathcal{P}_t = W y_t$ and, for any $j \leq J$, we let $\mathcal{P}_t^j$ and W^j denote the first j portfolios and their associated weights. The modeler's choice of W will determine which portfolios of yields enter the *MTSM* as risk factors and which additional portfolios are used in estimation.

2 A Canonical *MTSM*

This section gives a heuristic construction of our canonical form; formal regularity conditions and a proof that our form is canonical are presented in Appendix A. Suppose that $\mathcal{M}$ macroeconomic variables M_t enter a *MTSM* as risk factors and that the one-period interest rate r_t is an affine function of M_t and an additional $\mathcal{L}$ pricing factors $\mathcal{P}_t^{\mathcal{L}}$,

$$r_t = \rho_{0Z} + \rho_{1M} \cdot M_t + \rho_{1\mathcal{P}} \cdot \mathcal{P}_t^{\mathcal{L}} \equiv \rho_0 + \rho_1 \cdot Z_t, \tag{1}$$

[8] In the context of *YTSM*s, Joslin, Singleton, and Zhu (2010) and Duffee (2011) explore empirically whether various constraints on the $\mathbb{P}$ distribution of the risk factors improve out-of-sample forecasts of these factors. They did not consider second moments, nor any properties of distributions that involve these moments.

where the risk factors are $Z_t \equiv (M_t', \mathcal{P}_t^{\mathcal{L}\prime})'$. Some treat $\mathcal{P}_t^{\mathcal{L}}$ in (1) as a set of $\mathcal{L}$ latent risk factors,[9] while others include portfolios of yields as risk factors.[10] Fixing M_t and the dimension $\mathcal{L}$ of $\mathcal{P}_t^{\mathcal{L}}$, these two theoretical formulations are observationally equivalent. In fact, as we show, we are free to rotate the entire vector Z_t to express bond prices in terms of $\mathcal{P}_t^{\mathcal{N}}$, the first $\mathcal{N} = \mathcal{M} + \mathcal{L}$ entries of the modeler's chosen portfolios of yields. This is an implication of affine pricing of $\mathcal{P}_t^{\mathcal{N}}$ in terms of Z_t. Accordingly, in characterizing a canonical form for the family of *MTSM*s with short-rate processes of the form (1), we are free to start with either interpretation of $\mathcal{P}_t^{\mathcal{L}}$ (latent or yield-based) and to use any of these rotations of the risk factors Z_t.

We select a rotation of Z_t and its associated risk-neutral ($\mathbb{Q}$) distribution so that our maximally flexible canonical form is particularly revealing about the joint distribution of Z_t and bond yields implied by *MTSM*s with $\mathcal{N}$ pricing factors and macro pricing factors M_t.

2.1 The Canonical Form

Consider a *MTSM* with risk factors Z_t and short rate as in (1), with Z_t following a Gaussian process under the risk-neutral distribution,

$$\Delta Z_t = K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}} Z_{t-1} + \sqrt{\Sigma}\epsilon_t^{\mathbb{Q}},\ \epsilon_t^{\mathbb{Q}} \sim N(0, I). \tag{2}$$

Absent arbitrage opportunities in this bond market, (1) and (2) imply affine pricing of bonds of all maturities (Duffie and Kan (1996)). The yield portfolios $\mathcal{P}_t$ can be expressed as

$$\mathcal{P}_t = A_{TS}(\Theta_{TS}^{\mathbb{Q}}) + B_{TS}(\Theta_{TS}^{\mathbb{Q}}) Z_t, \tag{3}$$

where the loadings (A_{TS}, B_{TS}) are known functions of the parameters $\Theta_{TS}^{\mathbb{Q}}$ governing the $\mathbb{Q}$ distribution of yields, and hereafter we use "TS" to denote features of a *MTSM*. A canonical version of this model is obtained by imposing normalizations that ensure that the only admissible rotation of Z_t that leaves the distribution of r_t unaffected is the identity matrix. To arrive at our canonical form we observe that from the first $\mathcal{N}$ entries of (3), Z_t, and hence all bond yields y_t, can be expressed as affine functions of $\mathcal{P}_t^{\mathcal{N}}$.[11] After rotating to a pricing model with risk factors $\mathcal{P}_t^{\mathcal{N}}$, we adopt the canonical form of Joslin, Singleton, and Zhu (2010) (JSZ). What is distinctive about their canonical form is that $\Theta_{TS}^{\mathbb{Q}}$ is fully characterized by the covariance matrix Σ and the rotation invariant (and hence economically interpretable) long-run $\mathbb{Q}$-mean of r_t, $r_\infty^{\mathbb{Q}} = E^{\mathbb{Q}}[r_t]$, and the $\mathcal{N}$-vector $\lambda^{\mathbb{Q}}$ of distinct real eigenvalues of the feedback matrix $K_{1Z}^{\mathbb{Q}}$.[12]

[9] Studies with this formulation include Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007), Bikbov and Chernov (2010), Chernov and Mueller (2009), and Smith and Taylor (2009).

[10] Examples include Ang, Piazzesi, and Wei (2006) and Jardet, Monfort, and Pegoraro (2010).

[11] This inversion presumes that the $\mathcal{N}$-factor *MTSM* 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. Formal regularity conditions are provided in Appendix A.

[12] Extensions to the more general case of $K_1^{\mathbb{Q}}$ being in ordered real Jordan form, or to a zero root in the $\mathbb{Q}$ process of Z_t, are straightforward along the lines of Theorem 1 in JSZ.

A second key implication of (3) is that, within any *MTSM* that includes M_t as pricing factors in (1), these macro factors must be spanned by $\mathcal{P}_t^{\mathcal{N}}$:

$$M_t = \gamma_0 + \gamma_1 \mathcal{P}_t^{\mathcal{N}}, \tag{4}$$

for some conformable γ_0 and γ_1 that implicitly depend on W. Using (4), we apply the rotation

$$Z_t = \begin{pmatrix} M_t \\ \mathcal{P}_t^{\mathcal{L}} \end{pmatrix} = \begin{pmatrix} \gamma_0 \\ 0 \end{pmatrix} + \begin{pmatrix} & \gamma_1 \\ I_{\mathcal{L}} & 0_{\mathcal{L}\times(\mathcal{N}-\mathcal{L})} \end{pmatrix} \mathcal{P}_t^{\mathcal{N}} \tag{5}$$

to the canonical form in terms to $\mathcal{P}_t^{\mathcal{N}}$ to obtain an equivalent model in which the risk factors are M_t and $\mathcal{P}_t^{\mathcal{L}}$, r_t satisfies (1), and Z_t follows the Gaussian $\mathbb{Q}$ process (2). Our specification is completed by assuming that, under the historical distribution $\mathbb{P}$, Z_t follows the process

$$\Delta Z_t = K_0^{\mathbb{P}} + K_1^{\mathbb{P}} Z_{t-1} + \sqrt{\Sigma}\epsilon_t^{\mathbb{P}},\ \ \epsilon_t^{\mathbb{P}} \sim N(0, I). \tag{6}$$

Summarizing, in our canonical form the first $\mathcal{M}$ components of the pricing factors Z_t are the macro variables M_t, and without loss of generality the risk factors are rotated so that the remaining $\mathcal{L}$ components of Z_t are the yield portfolios $\mathcal{P}_t^{\mathcal{L}}$; r_t is given by (1); M_t is related to $\mathcal{P}_t$ through (4); and Z_t follows the Gaussian $\mathbb{Q}$ and $\mathbb{P}$ processes (2) and (6). Moreover, for given W, the risk-neutral parameters $(\rho_0, \rho_1, K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}})$ are explicit functions of $\Theta_{TS}^{\mathbb{Q}} = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma)$.

Our canonical construction reveals the essential difference between term structure models based entirely on yield-based pricing factors $\mathcal{P}_t^{\mathcal{N}}$ and those that include macro risk factors. A *MTSM* 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. In the JSZ canonical form with pricing factors $\mathcal{P}_t^{\mathcal{N}}$, the underlying parameter set is $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, K_{0\mathcal{P}}^{\mathbb{P}}, K_{1\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}})$. A *MTSM* adds the spanning property (4) with its $\mathcal{M}(\mathcal{N}+1)$ free parameters. Thus, any canonical $\mathcal{N}$-factor *MTSM* with macro factors M_t gains $\mathcal{M}(\mathcal{N}+1)$ free parameters relative to pure latent-factor Gaussian models. Of course this added flexibility (by parameter count) of a *MTSM* is gained at a cost: the realizations of the yield-based risk factors must be related to the macro factors M_t through equation (4).

In taking the model to the data, we must accommodate the fact that the observed data, $\{M_t^o, \mathcal{P}_t^o\}$, will not be perfectly matched by a theoretical no-arbitrage model. Accordingly we suppose that the observed data are given by the theoretical values plus a mean-zero measurement error. Absent any guidance from economic theory, and consistent with the literature, we presume that the measurement errors are *i.i.d.* normal, thereby giving rise to a Kalman filtering problem.[13] The observation equation is simply (3) allowing for errors:

$$\mathcal{P}_t^o = A_{TS}(\Theta_{TS}^{\mathbb{Q}}) + B_{TS}(\Theta_{TS}^{\mathbb{Q}})\, Z_t + e_t, \quad e_t \sim N(0, \Sigma_e), \tag{7}$$

[13]This formulation subsumes the case of cross-sectionally uncorrelated pricing errors (Σ_e is diagonal) adopted by Ang, Dong, and Piazzesi (2007) and Bikbov and Chernov (2010), as well as the case where Σ_e is singular with the first $\mathcal{L}$ rows and columns of Σ_e equal to zero. In the latter case, $\mathcal{P}_t^{\mathcal{L}} = \mathcal{P}_t^{\mathcal{L}o}$.

and the state equation is (6). Consistent with the literature, we assume always that the observed macro factors M_t^o coincide with their theoretical counterparts M_t, though this assumption is easily relaxed. Together (6) and (7) comprise the state space representation of the *MTSM*. The full parameter set is $\Theta_{TS} = (\Theta_{TS}^{\mathbb{Q}}, K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma_e)$.

2.2 State-Space Formulations Under Alternative Hypotheses

Throughout our subsequent analysis we compare the *MTSM*s characterized by (6) and (7) to their "unconstrained alternatives". Since a *MTSM* involves multiple over-identifying restrictions, the relevant alternative model depends on which of these restrictions one is interested in relaxing. We find it useful to distinguish between the following three alternative formulations which we label by FV, TS^n, and FV^n.

The FV alternative follows Duffee (2011) and maintains the state equation (6), but generalizes the observation equation to

$$\mathcal{P}_t^o = A_{FV} + B_{FV}\, Z_t + e_t, \tag{8}$$

for conformable matrices A_{FV} and B_{FV}, with e_t normally distributed from the same family as the *MTSM*. The subscript "FV" is short-hand for the factor-VAR structure of (6) and (8). For identification we normalize the first $\mathcal{L}$ entries of A_{FV} to zero and the first $\mathcal{L}$ rows of B_{FV} to the corresponding standard basis vectors. Except for this, A_{FV} and B_{FV} are free from any restrictions.[14] The full parameter set is $\Theta_{FV} = (A_{FV}, B_{FV}, K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma_, \Sigma_e)$. Since all bonds are priced with errors, the FV model is estimated using the Kalman filter.

Special cases of models TS and FV that are also of interest arise when their respective error covariance matrices Σ_e have rank $J - \mathcal{L}$. In this case, $\mathcal{L}$ linear combinations of the yield portfolios $\mathcal{P}_t$ are priced perfectly by the model, along the lines of Chen and Scott (1993). The particular case we focus on is where the first $\mathcal{L}$ portfolios of yields $\mathcal{P}^{\mathcal{L}}$ are measured perfectly. We distinguish these special cases by the notation TS^n and FV^n (for *n*o pricing errors on the risk factors). The Kalman filtering problem then simplifies to conventional ML estimation. In particular, for the FV^n model, estimation conveniently reduces to two sets of OLS regressions: a VAR for the observed risk factors Z_t^o gives the parameters in (6),[15] and an OLS regression of $\mathcal{P}_t^o$ on Z_t^o recovers the parameters characterizing (7).

Relative to model TS, model FV relaxes the over-identifying restrictions implied by the assumption of no arbitrage, but maintains the low-dimensional factor structure of returns and the presumption of measurement errors on bond yields. Thus, in assessing whether these two models imply nearly identical joint distributions for (y_t, M_t), the focus is on whether the no arbitrage restrictions induce a difference. On the other hand, differences between the TS and TS^n models, which both maintain a similar no-arbitrage structure, should arise

[14] A subtle issue is that this is slightly over-identifying since it implies that a relationship of the form $\alpha + \beta \cdot \mathcal{P}_t^{\mathcal{L}} = 0$ cannot hold in the model. Certainly this would be rejected in the data for typical choices of W. However, the ODE theory implies this normalization is just-identifying in the no-arbitrage model.

[15] The ML estimators of $K_{0Z}^{\mathbb{P}}$ and $K_{1Z}^{\mathbb{P}}$ are the standard OLS estimators, and the ML estimator of Σ_Z is the usual sample covariance matrix based on the OLS residuals.

mainly out of the different treatments of measurement errors of the pricing factors. Finally, in moving from model TS to model FV^n one is relaxing both the no arbitrage restrictions and the presumption that the yield-based pricing factors $\mathcal{P}_t^{\mathcal{L}}$ are measured with errors ($\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$ in model FV^n), while again maintaining the low-dimensional factor structure.

3 Conditions for the (Near) Observational Equivalence of *MTSM*s and Factor-VARs

To derive sufficient conditions for the general agreement of Kalman filter estimators of models TS and FV, we fix a choice of W and derive (stronger) sufficient conditions for the Kalman filter estimators of the distribution of Z_t from model TS and FV to be (nearly) identical to those implied by the FV^n model. As it becomes clear subsequently, this indirect approach, as opposed to a direct comparison between models TS and FV, is justified by the ease in estimating model FV^n (two OLS regressions). As long as there exists one W^* such that these conditions are satisfied, it *must* mean that models TS and FV imply (nearly) identical distributions of Z_t *for all admissible portfolio matrices* W. This is true despite that the individual comparisons for the pairs (TS, FV^n) and (FV, FV^n) are rotation–dependent. Equally importantly, for such a W^*, everything that one can learn about the $\mathbb{P}$ distribution of this model's risk factors Z_t from a canonical $MTSM$ can be equally learned from analysis of the corresponding economics-free factor-VAR model FV^n.

The filtering problems in both models TS and FV is one of estimating the true values of $\mathcal{P}_t^{\mathcal{L}}$, the first $\mathcal{L}$ PCs of the bond yields y_t. Intuitively, a key condition for the Kalman filter estimates of models (TS, FV) to match the OLS estimates of model FV^n is that the filtered pricing factors equal their observed counterparts. However, this observation begs the more fundamental question of when this approximation holds. Additionally, this condition is not sufficient for the Kalman filter estimates of either the drift nor the volatility of Z_t to match the OLS estimates of model FV^n. The remainder of this section addresses these issues.

To fix the notation, we will denote the filtered and smoothed version of any random variable X_t by $X_t^f = E[X_t|\mathcal{F}_t]$ and $X_t^s = E[X_t|\mathcal{F}_T]$.

3.1 When do the filtered yields differ from the observed yields?

The filtered yields will agree closely with the observed yields when the filtered measurement errors are close to zero. This difference will depend on two quantities. First, it will depend on how large the measurement errors are for the yields. Second, it will depend on how accurately the yield portfolios can be forecasted based on current and lagged observeables, excluding the current yields themselves.

We can see the relationship as follows. Focusing on the first $\mathcal{L}$ yield portfolios, let us define the vector

$$I_t = (M_1, \mathcal{P}_1^o, \ldots, M_{t-1}, \mathcal{P}_{t-1}^o, M_t, \mathcal{P}_t^{-\mathcal{L}o}),$$

where $\mathcal{P}_t^{-\mathcal{L}o}$ are the observed yield portfolios at time t excluding the first $\mathcal{L}$ portfolios. The filtered measurement error, $e_t^{\mathcal{L}}$, is equal to

$$E[e_t^{\mathcal{L}}|I_t, \mathcal{P}_t^{\mathcal{L}o}].$$

Now, conditional on I_t, the measurement error and the observed yields are jointly normal so by the standard formula for updating of jointly normal variables[16] we have

$$E[e_t^{\mathcal{L}}|I_t, \mathcal{P}_t^{\mathcal{L}o}] = \Sigma_{e\mathcal{L}} S_t^{-1}(\mathcal{P}_t^{\mathcal{L}o} - E[\mathcal{P}_t^{\mathcal{L}o}|I_t]), \tag{9}$$

where $\Sigma_{e\mathcal{L}}$ is the covariance of the measurement errors, $e_t^{\mathcal{L}}$, and S_t is the conditional forecast variance of $\mathcal{P}_t^{\mathcal{L}o}$ based on I_t.

Intuitively, the conditional updating places more weight on the forecast the more accurate it is (i.e. the smaller S_t is) and more weight on the observed state the smaller the measurement errors. The difference in magnitude between the observed and filtered states will thus depend on the relative size of $\Sigma_{e\mathcal{L}}$ to S_t.

How small will the portfolio measurement errors be?

$\Sigma_{e\mathcal{L}}$ will reflect the size of the pricing errors within the model. Typical studies find mean square pricing errors for individual bond yields on the order of 10 basis points, though with fewer pricing factors the pricing errors can be substantially larger. However, $\sigma_{e\mathcal{L}}$ will be determined not only by the size of the pricing errors of the indivdual yields, but also by the choice of the weighting matrix W. In general, the weighting matrix W can allow for averaging of errors across individual yields, resulting $\Sigma_{c,\mathcal{L}}$ being smaller than the corresponding mean-square measurement errors of the individual yields. For example, if the errors are independent and we scale $\mathcal{P}_t$ to take the first row of W to give equal weights (corresponding to a level factor), the square root of the mean square pricing error should be reduced by a factor of $1/\sqrt{J}$ due to the cancelation of the signed errors across maturities.[17] This averaging effect means that even if individual bonds are priced with relatively large errors, Σ_e can still be relatively low due to a rich amount of cross-sectional information.

How large will the forecast error be?

S_t reflects the uncertainty about $\mathcal{P}_t^o$ given past realizations of the yield curve and macro variables and current information about the macro-variables and the higher order portfolios.

[16] When random vectors (X, Y) follow a multivariate normal distribution, $E[X|Y] = \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mu_Y)$, where μ_X and μ_Y are the mean of X and Y, Σ_Y is the variance of Y and Σ_{XY} is the covariance of X and Y. In the case of $X = e_t^{\mathcal{L}}$ and $Y = \mathcal{P}_t^{\mathcal{L}o}$, the covariance between X and Y is just the variance of the errors by independence.

[17] Typically principal components are normalized so that the sum of the square of the weights is one. This condition also ensures the observational equivalence of Section 2 if one supposes that the errors are independent with equal variances. For ease of interpretation, it is convenient to rescale the principle component to more interpretable values such as setting the sum of the weights equal to one for the first principal component. This rescaling gives an observational equivalent model with the adjusted Σ_e (which will violate equal variances due to the rescaling).

Consider again the case where the first portfolio of $\mathcal{P}_t$ is associated with a level factor. First notice that $\mathcal{P}^o_t$ incorporates the measurement error which is independent of I_t, so that S_t will always be larger than Σ_e. Second, observe that the lagged value of $\mathcal{P}^o_{t-1}$ will contain a substantial amount of information about $\mathcal{P}^o_t$. Monthly volatilities of changes in yields are on the order of 20 to 30 basis points, depending on the sample period. To the extent that most of these changes are unpredictable relative to I_t, this should provide a guideline for the variance of $\mathcal{P}_t$ conditioned on lagged information. Furthermore, given the lagged information, incorporating the current macro-variables and current values of other yield portfolios are likely to have only a small effect on the conditional variance.

Discussion

This heuristic analysis allows us to see the approximate average magnitude of the difference between the filtered and observed states. For example, if there is a single state portfolio ($\mathcal{L} = 1$) which is a level factor with equal weights ($1/J$), then square root of the filtered mean squared measurement error is

$$\frac{\sigma_y}{J} \times \frac{\sigma_y}{\sigma_{f,t}} \tag{10}$$

where σ_y is the pricing error for a single yield, σ_f is the forecast error for time t based on I_t (i.e. $\sigma_t^2 = S_t$). The ratio of σ_y to $\sigma_{f,t}$ will always be less than one since the forecast is of the observed state portfolio which includes the independent measurement error. So if the pricing errors are independent and around 10 basis points, the forecast errors are on the order of 20 basis points, and there are $J = 10$ yields use in the estimation, the magnitude will be about half of a basis point.

Since the forecast error is normal distributed with covariance S_t, if $\Sigma_{e\mathcal{L}} S_t^{-\frac{1}{2}}$ is small, the difference between the filtered states and observed states will be small on average. This has a number of implication. First, since S_t is at least as large as $\Sigma_{e\mathcal{L}}$, it follows that as long as the measurement errors for the portfolios are small the filtered states and observed states will be close. This means for example for a level factor that increasing the number of yields used in the estimation is likely to reduce the measurement error for the level portfolio and increase the match between the observed level and the filtered level. Second, S_t will be much larger than $\Sigma_{e\mathcal{L}}$ when there is a lot of uncertainty in the forecasting $\mathcal{P}^{\mathcal{L}o}_t$ from the information in I_t. This forecasting uncertainty is likely to rise as the sampling frequency decreases. Thus we conclude that the observed and filtered states will agree when W is chosen so that (i) there is cancelation of measurement errors across maturities, (ii) more cross-sectional information is available, and (iii) there is a large amount of unpredictable variation in $\mathcal{P}^{\mathcal{L}o}_t$ relative to the information in I_t.

We stress that the magnitude of $\Sigma_{\mathcal{L}e} S_t^{-1}$ depends on the particular choice of the weighting matrix W. For some choices of W this product may be much smaller than for other choices. For example, if one chooses weighting matrices that select individual yields, there will be no cancelation of measurement errors so that $\Sigma_{\mathcal{L}e}$ will likely be substantially larger. However, as discussed in Section 2 , the choice of W only reflects the parameterization of the model. So although one can parameterize the model with W taken to be the identity, it may be that there exists an alternative W^* satisfying the above conditions so that filtering has little effect on the observed portfolios with respect to this W^*. Put differently, for a given model, some linear combinations of the observed yields may correspond closely to its filtered counterparts while other may not: the above conditions provide insights into which linear combination weighting will fall into each class.

These results also provide a context for interpreting previous work with large numbers of latent or yield-based risk factors. The reported large differences between the filtered and observed values of the high-order PCs in the five-factor, yield-only models studied by Duffee (2009) and Joslin, Singleton, and Zhu (2010) models could be attributed to the smaller variances of forecast errors associated with higher PCs. Under the typical assumption of *i.i.d.* measurement errors and normalized loadings, the cross-sectional uncertainty remains fixed across PCs. The sample standard deviations of the fourth and fifth PCs, about 19 and 13 basis points for our data, are much smaller than those for the first three PCs and more importantly, borderline the typical RMS of pricing errors. Since the forecast errors variances of the fourth and fifth PCs must be smaller than their respective unconditional variances, it is likely that the elements of $\Sigma_{\mathcal{L}e}\Omega_{\mathcal{L}t}^{-1}$ corresponding to these PCs are not small. Therefore, in this case, the Kalman filter will tend to emphasize measurement error reduction, through smoothing of the higher-order factors over their past innovations, over fitting the cross-section of yields.

3.2 ML Estimation of the Conditional Distribution of (M_t, y_t)

We now turn to characterize the conditions for computing the maximum likelihood estimates of the models. For either of the models TS or FV, the joint likelihood of the observed data, $\{M_t^o, y_t^o\}$, follow a multivariate normal distribution that can be computed efficiently by using the Kalman filter. From a theoretical perspective, we can think of building the likelihood of the data by by integrating the joint density $f_m^{\mathbb{P}}(\vec{Z} = z, \vec{\mathcal{P}}^o, \vec{M}^o; \Theta_m)$ over the missing data $\vec{Z}$:

$$f_m^{\mathbb{P}}(\vec{\mathcal{P}}^o, \vec{M}^o; \Theta_m) = \int_z f_m^{\mathbb{P}}(\vec{Z} = z, \vec{\mathcal{P}}^o, \vec{M}^o; \Theta_m) dz, \tag{11}$$

for $m = TS$ or FV, with $\vec{X}$ denoting the sequence of the entire sample: $\vec{X} = (X_1, X_2, \ldots, X_T)$. For ease of notation, we omit the subscript m from $f_m^{\mathbb{P}}$ and Θ_m in all expressions that apply to both the $MTSM$s and the factor-VARs. Taking the derivative of (11) with respect to Θ and setting this equal to zero, and dividing by the marginal density of $(\vec{\mathcal{P}}^o, \vec{M}^o)$, gives the first-order conditions[18]

$$0 = E\left[\partial_\Theta \log f^{\mathbb{P}}(\vec{Z}, \vec{\mathcal{P}}^o, \vec{M}^o; \hat{\Theta}) \middle| \mathcal{F}_T\right], \tag{12}$$

where T is the sample size and $\mathcal{F}_T$ is all of the observable information.[19]

[18] This relation arises in the literature on the "EM" algorithm (e.g., Dempster, Laird, and Rubin (1977)).

[19] In model FV^n with our choice of W, $Z_t = Z_t^o$ and (12) holds without the conditional expectation.

The density $\log f^{\mathbb{P}}(\vec{Z}, \vec{\mathcal{P}}^o, \vec{M}^o)$ in (12) is equal to

$$\sum_{t=1}^{T} \log f^{\mathbb{P}}\left(\mathcal{P}_t^o | Z_t; \Theta^{\mathbb{Q}}, \Sigma_e\right) + \sum_{t=1}^{T} \log f^{\mathbb{P}}\left(Z_t | Z_{t-1}; K_1^{\mathbb{P}}, K_0^{\mathbb{P}}, \Sigma\right). \tag{13}$$

This construction reveals two important and distinguishing properties of our canonical *MTSM*. The conditional distribution of the risk factors Z_t in (13) depends only on $(K_1^{\mathbb{P}}, K_0^{\mathbb{P}}, \Sigma)$, *and* $(K_1^{\mathbb{P}}, K_0^{\mathbb{P}})$ enter only $f^{\mathbb{P}}(Z_t|Z_{t-1})$ and not $f^{\mathbb{P}}(\mathcal{P}_t^o|Z_t)$. Moreover, both of these observations are equally true of model FV. This similarity between these null and alternative models is immediately apparent in our canonical form, while being largely obscured in say the more standard identification schemes of *MTSM*s based on Dai and Singleton (2000).

A key difference between models TS and FV is how Σ enters the two components of $f^{\mathbb{P}}$. The functional dependence of $f^{\mathbb{P}}(Z_t|Z_{t-1})$ on Σ is identical for these two models. However, owing to the diffusion invariance property of the no-arbitrage model, Σ only affects $f_{TS}^{\mathbb{P}}(\mathcal{P}_t^o|Z_t)$ and not $f_{FV}^{\mathbb{P}}(\mathcal{P}_t^o|Z_t)$. Nevertheless, for our canonical form, this difference turns out to be largely inconsequential for Kalman filter estimates of Σ.

Using the fact that $f(\mathcal{P}_t^o|Z_t)$ does not depend on $(K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$, the *ML* estimators of the conditional mean parameters $(K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$ satisfy

$$[\hat{K}_0^{\mathbb{P}}, \hat{K}_1^{\mathbb{P}}]' = \left(\left(\tilde{Z}'\tilde{Z}\right)^{\mathrm{s}}\right)^{-1}\left(\tilde{Z}'\Delta Z\right)^{\mathrm{s}}, \tag{14}$$

where $\tilde{Z}_t = [1, Z_t']'$, and Z and $\tilde{Z}$ are matrices with rows corresponding to Z_t and $\tilde{Z}_t$, respectively, for t ranging from 1 to T.

From (14) it is seen that a key ingredient for Kalman filter estimates of $(K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$ from models TS and FV to agree with each other and with those from model ZV is that $(\tilde{Z}_t\tilde{Z}_t')^{\mathrm{s}}$ be close to $\tilde{Z}_t^o\tilde{Z}_t^{o\prime}$, period-by-period. For (14) is *almost* the estimator of $[K_0^{\mathbb{P}}, K_1^{\mathbb{P}}]$ obtained from *OLS* estimation of a VAR on the smoothed risk factors Z_t^{s}. The difference between (14) and the VAR estimator run on Z_t^{s} lies in the fact that

$$(Z_t Z_t')^{\mathrm{s}} = \mathrm{Var}(Z_t|\mathcal{F}_T) + Z_t^{\mathrm{s}} Z_t^{\mathrm{s}\prime}. \tag{15}$$

This equation and the analogous extensions reveal that, provided the smoothed state is close to the observed state and $\mathrm{Var}(Z_t|\mathcal{F}_T)$ is small, the *ML* estimates of $(K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$ from model ZV will be similar to those obtained by Kalman filtering within a *MTSM*. Even in the presence of large pricing errors, the conditions derived in Section 3.1 show that the observed and smoothed states may agree closely depending on the amoutn of cross sectional information, the frequency of observations, and the amount of cancellation of measurement errors implicit in the choice of W.

Turning to estimation of Σ, in model FV there is no diffusion invariance and $f_{FV}^{\mathbb{P}}(\mathcal{P}_t^o|Z_t)$ does not depend on Σ. Therefore, the first-order conditions for maximizing the likelihood function depend only on $\log f_{FV}^{\mathbb{P}}(Z_t|Z_{t-1}; \hat{\Theta}_{FV})$. This leads to the first-order condition

$$E\left[vec\left((\hat{\Sigma}_{FV})^{-1} - (\hat{\Sigma}_{FV})^{-1}\hat{\Sigma}_{FV}^{u}(\hat{\Sigma}_{FV})^{-1}\right)\Big|\,\mathcal{F}_T\right] = 0, \tag{16}$$

where $\hat{\Sigma}^u_{FV}$ is the sample covariance matrix based on the residuals $\hat{i}^u_{FV,t} = \Delta Z_t - (\hat{K}^{\mathbb{P}}_{0FV} + \hat{K}^{\mathbb{P}}_{1FV} Z_{t-1})$. Both $\hat{\Sigma}^u_{FV}$ and $\hat{i}^u_{FV,t}$ are unobserved since they depend on the partially observed $\vec{Z}$. From (16), we obtain $\hat{\Sigma}_{FV} = (\Sigma^u_{FV})^{\mathrm{s}}$. Using the logic of our discussion of the conditional mean, as long as the estimated model FV accurately prices the risk factors, then $(\Sigma^u_{FV})^{\mathrm{s}}$ will be nearly identical to the OLS estimator of Σ from the VAR model ZV.

The ML estimator of Σ in model TS will in general be more efficient than in model ZV and this is true even when $\mathcal{P}^{\mathcal{L}o}_t = \mathcal{P}^{\mathcal{L}}_t$. The first-order conditions for Σ in model TS have an additional term since, in this model, the density $f^{\mathbb{P}}_{TS}(\mathcal{P}^o_t | Z_t; \Theta)$ also depends on Σ. Combining this term, derived in Appendix E as (71), with (16) gives

$$E\left[vec\left(\frac{1}{2}\left[(\hat{\Sigma}_{TS})^{-1} - (\hat{\Sigma}_{TS})^{-1}\hat{\Sigma}^u_{TS}(\hat{\Sigma}_{TS})^{-1}\right]\right) - \hat{\beta}'_Z(\hat{\Sigma}_{e,TS})^{-1}\frac{1}{T}\sum_t \hat{e}^u_{TS,t} \middle| \mathcal{F}_T\right] = 0,$$

where $\hat{\Sigma}^u_{TS}$ is the sample covariance of the residuals $\hat{i}^u_t = \Delta Z_t - (\hat{K}^{\mathbb{P}}_{0TS} + \hat{K}^{\mathbb{P}}_{1TS} Z_t)$, $\hat{\beta}_Z$ is a constant vector defined in Appendix E, and the unobserved pricing errors $\hat{e}^u_{TS,t}$ from (7) are evaluated at the ML estimators and depend on the partially observed $\vec{Z}$.

The following two conditions are sufficient for the Kalman filter estimators of Σ in models TS and FV to be approximately equal. First, we require that the risk factors be priced sufficiently accurately for

$$\hat{\Sigma}_{FV} = \left(\hat{\Sigma}^u_{FV}\right)^{\mathrm{s}} \approx \left(\hat{\Sigma}^u_Z\right)^{\mathrm{s}}. \tag{17}$$

To guarantee that the right hand side of (17) is close to the estimate of Σ in the $MTSM$, our second requirement is that the average-to-variance ratio $(\hat{\Sigma}_e)^{-1}(T^{-1}\sum \hat{e}^o_t)$ of pricing errors be close to zero, where $\hat{e}^o_t$ is computed from (7) evaluated at the ML estimates and using $\vec{Z}^o$. When both conditions are satisfied, $(\hat{\Sigma}_e)^{-1}(T^{-1}\sum \hat{e}^u_t)^{\mathrm{s}}$ will be close to zero as well, ensuring that $\hat{\Sigma}_{Z,TS} \approx (\hat{\Sigma}^u_{FV})^{\mathrm{s}}$. Moreover, when these conditions hold, the estimators from all three models TS, FV, and ZV will approximately agree with each other.

3.3 Discussion

The first-order conditions of the ML estimators do not set the sample mean of the pricing error $\hat{e}^u_{TS,t}$ to zero. However, it is easily verified that the first-order conditions with respect to the "constant terms" $(r^{\mathbb{Q}}_{\infty}, \gamma_0)$ set $\mathcal{M}+1$ linear combinations of the filtered means $(T^{-1}\sum \hat{e}^u_{TS,t})^{\mathrm{s}}$ to zero. So, effectively, the likelihood function has $\mathcal{M}+1$ degrees of freedom to use in making the mean-to-variance ratios close to zero.

These observations regarding the conditional distribution of the risk factors Z_t extend to individual bond yields with one additional requirement. Specifically, the factor loadings from OLS projections of y^o_t onto Z^o_t need to be close to their model-based counterparts estimated using the Kalman filter. Using our earlier logic, if $\mathcal{P}^{\mathcal{L}}_t$ is reasonably accurately priced, the

OLS loadings are likely to be close to those implied by model FV.[20] Nevertheless, large $RMSE$s in the pricing of individual bonds might lead to large efficiency gains from ML estimation of the loadings within a *MTSM*. This is an empirical question that we take up subsequently.

Importantly, the conditions we have derived for the near observational equivalence of the *MTSM*- and factor-VAR-implied Kalman filter estimators of the joint distribution of $(\mathcal{P}_t, M_t)$ apply even though the canonical *MTSM* implies over-identifying restrictions on this distribution and individual bond yields are priced imperfectly. Further intuition for our results comes from exploring two restrictive special cases: the yield-based risk factors $\mathcal{P}_t^{\mathcal{L}}$ are priced perfectly by the *MTSM* ($\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$) and, on top of this, the *MTSM* is just-identified in the sense that the restriction of no arbitrage is non-binding on the factor-VAR model for the risk factors. We discuss each of these in turn.

A stark version of our results is obtained directly under the assumption $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$. In this case, the relevant comparison is between models TS^n and FV^n where $\Sigma_{\mathcal{L}e}\Omega_{\mathcal{L}t}^{-1}$ is zero by construction. Hence, the ML estimates of the conditional mean parameters $(K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}})$ from TS^n exactly coincide with the OLS estimates for model FV^n, regardless of the magnitude of the mean-to-variance ratios of pricing errors.[21] Therefore, a sufficient condition for the conditional distribution of the risk factors Z_t in a *MTSM* to be fully invariant to the imposition of the no-arbitrage restrictions is that the average-to-variance ratio of the pricing errors $(\hat{\Sigma}_e)^{-1}(T^{-1}\sum \hat{e}_{TS,t})$ is zero. Owing to the Gaussian property, these invariance results extend to the *unconditional* distributions of $\{Z_t\}$ as well.

Insight into circumstances when the sample mean of $\hat{e}_{TS,t}$ is exactly equal to zero comes from Duffee (2011)'s analysis of a yield-based TS^n model where the number of yields used in estimation (J) is $\mathcal{N}+1$. In this case, the term structure model is just-identified,[22] and the mean of $\hat{e}_{TS,t}$ for the one imperfectly priced bond is zero. Thus, the ML estimators of the joint distribution of $\mathcal{P}_t^{\mathcal{N}}$ from the term structure model and the unrestricted factor-VAR are always identical to each other.

---

[20] To see this, first note that the loadings of y_t on Z_t are simply the loadings of $\mathcal{P}_t$ on Z_t, premultiplied by the inverse of W. Second note that, for the FV model, the loadings of $\mathcal{P}_t$ on Z_t are given by:

$$(\hat{A}_{FV}, \hat{B}_{FV}) = \left(\frac{1}{T}\sum_t \left[\mathcal{P}_t^o \left(\tilde{Z}_t'\right)^{\mathrm{s}}\right]\right)\left(\frac{1}{T}\sum_t \left[\left(\tilde{Z}_t\tilde{Z}_t'\right)^{\mathrm{s}}\right]\right)^{-1},$$

which should be close to the loadings from projecting $\mathcal{P}_t^o$ on Z_t^o if $\mathcal{P}_t^{\mathcal{L}o}$ is accurately priced.

[21] This is the counterpart for *MTSM*s of the irrelevancy result for conditional means derived in JSZ when $Z_t = \mathcal{P}_t^{\mathcal{N}}$.

[22] More precisely, what Duffee shows is that the parameter count in the factor-VAR and in his corresponding model TS are the same. This turns out *not* to be sufficient to leave model TS just-identified. Duffee gives examples of factor-VAR loadings B_{FV} that do not correspond to B_{TS}. His observation leads naturally to the question of whether $(K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}}, \Sigma)$ can differ from their counterparts in model FV. Our results show that this is not the case. Whenever $J = \mathcal{N}+1$, the first-order condition with respect to $r_\infty^{\mathbb{Q}}$ (unrestricted by the no-arbitrage model) guarantees that the mean of $\hat{e}_{TS,t}$ is zero. Importantly, this result holds for any given $\lambda^{\mathbb{Q}}$ and, therefore, does not depend on the extent to which the no-arbitrage model matches (A_{FV}, B_{FV}). Thus our results show that when $J = \mathcal{N}+1$ with $\mathcal{P}_t^{\mathcal{N}o} = \mathcal{P}_t^{\mathcal{N}}$, the ML estimators of $f(\mathcal{P}_t^{\mathcal{N}}|\mathcal{P}_{t-1}^{\mathcal{N}})$ from his models TS and FV are always identical to each other.

The same condition ($J = \mathcal{N}+1$) in an $\mathcal{N}$-factor *MTSM* with $\mathcal{M}$ macro factors guarantees that $\hat{\Sigma}_Z$ exactly agrees for models TS^n and FV^n when the yield risk factors $\mathcal{P}_t^{\mathcal{L}}$ are perfectly priced. As discussed above, our canonical *MTSM* reveals that there are $\mathcal{M}+1$ degrees of freedom available to force the mean of $\hat{e}_{TS,t}$ to zero. Therefore, if exactly $\mathcal{M}+1$ portfolios of yields are included with measurement errors in the *ML* estimation of a *MTSM*, the mean-to-variance ratios will be optimized at zeros.

The typical *MTSM* examined in the literature has $J \gg \mathcal{N}+1$ and, hence, the *MTSM* is an over-identified model of bond yields. Moreover, most of the recent literature has presumed that all bonds are priced imperfectly ($\mathcal{P}_t^{\mathcal{L}o} \neq \mathcal{P}_t^{\mathcal{L}}$). Our results show that much of the intuition from just-identified *MTSM*s will carry over to over-identified *MTSM*s whenever the *MTSM* accurately prices the yield-based factors $\mathcal{P}_t^{\mathcal{L}}$, and this may be true even when the *MTSM*-implied errors in pricing individual bonds are quite large.

The remainder of this paper explores the empirical relevance of the sufficient conditions just derived for the Kalman filter estimates of the distribution of (M_t, y_t) in our canonical *MTSM* and its associated model FV to (nearly) equal those implied by the *OLS* estimates of model FV^n.

4 Empirical (Near) Observational Equivalence of *MTSM*s and Factor-*VAR*s

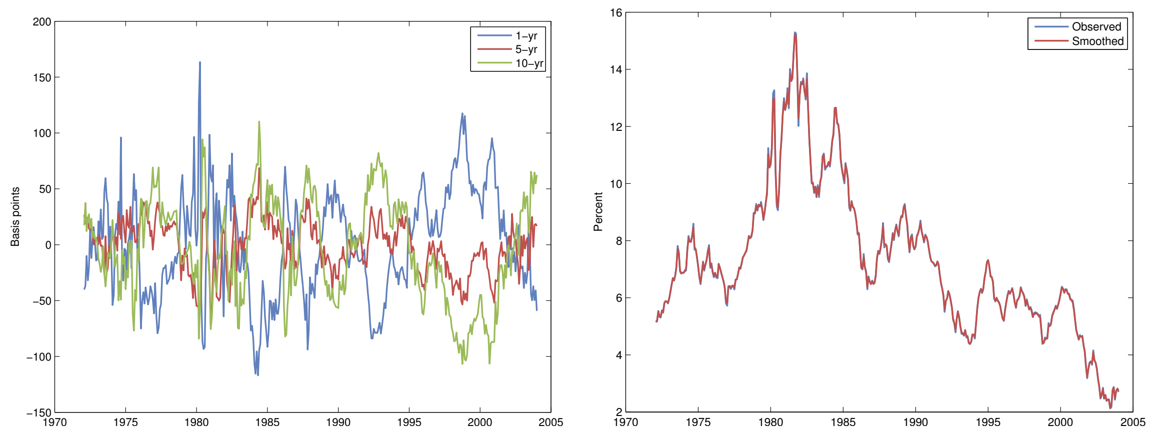
We now turn to assess the empirical relevance of the theory we developed in Section 3. We examine, step-by-step, to what extent the conditions hold that allow for the observational equivalence of *MTSM*s and factor-*VAR*s.

To examine the conditions, we first focus on a three-factor model in $M_t = (g_t, \pi_t)'$, where g_t is a measure of real output growth and π_t is a measure of inflation as in, for example, Ang, Dong, and Piazzesi (2007) and Smith and Taylor (2009). We follow Ang and Piazzesi (2003) and 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, 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.[23] We use the notation $GM_3(g, \pi)$ to denote this three-factor model with growth and inflation.

4.1 On the Need For Filtering *PC*s Within Our Canonical Models

A key aspect of our argument that the filtered versions of the first $\mathcal{L}$ portfolios agree with the observed portfolios was the cancellation effect of the errors across maturities. That is, even if individual yields are very noisy in the sense that filter error are large, it can be the case that portfolios are observed much more precisely.

[23] All of our results are qualitatively the same if we replace these measures of (g, π) by the help wanted index and *CPI* inflation as in Bikbov and Chernov (2010).



(a) Comparison of individual yields (b) Comparison of PC1

Figure 2: This figure compares observed yields with smoothed yields estimated from model $GM_3(g, \pi)$. Panel (a) plots the difference between observed yields and the smoothed versions from the model $(y_t^m)^{\text{s}}$. Panel (b) plots the observed $PC1$ and its smoothed version $PC1^{\text{s}}$.

Figure 2, Panel (a), plots the time series of the differences between observed yields, y_t^{mo}, and their smoothed counterparts $(y_t^m)^{\text{s}}$, for m=1, 5, and 10 years. These pricing errors are large, occasionally exceeding 100 basis points, so this model clearly has difficulty matching *individual yields*. The estimated standard deviation of the measurement errors is 43.1 basis points. The reason for this poor fit is that the macro-variables are only able to capture a small amount of the variation in the slope and curvature of the yield curve.

Although the individual yields are poorly fit by the model, the model provides an excellent fit the level of yields. Figure 2, Panel (b), plots the time series of the observed first principal component and its smoothed counterpart. Here, we scale the principal component so that the sum of the weights are one, which approximately gives equal weight to each maturity. The sample standard deviation of the difference between the observed and smoothed first PC is only 1.7 basis points; the standard deviation between the filtered and observed first PC is only 4.3 basis points.

These numbers are aligned with out the prediction of our theory in Section 3.1. The standard deviation of the forecast of the observed first PC based on lagged information and the current macro-variables and higher order principal component is 40.7 basis points. The sample standard deviation of changes in the first PC are 42.5 basis points, indicating that based on the information structure of $GM_3(g, \pi)$, very little of the changes are predictable consistent with the near-random walk behavior of the level of interest rates. The estimated standard deviation of the measurement error for the first PC is 12.5 basis points, approximately equal to the standard deviation of the individual yields (43.1 basis points) divided by the square root of the number of yields used in estimation ($J = 12$). According to (10), the estimated standard deviation of the difference between the observed and filtered first PC is 3.8 basis points – close to the sample value of 4.3 basis points.

	$K_{0Z}^{\mathbb{P}}$	$I + K_{1Z}^{\mathbb{P}}$		
	1	1	1	0.999
$\frac{TS}{FV}$	1	1	1	1
	1	0.999	1	1
	1.12	0.998	1.04	
$\frac{TS}{TS^n}$	0.999	0.999	1	1
	0.988	0.93	1	1
	1.12	0.998	1.04	1.02
$\frac{FV}{FV^n}$	0.999	0.999	1	1
	0.989	0.929	1	1

Table 1: Ratios of estimated $K_{0Z}^{\mathbb{P}}$ and $I + K_{1Z}^{\mathbb{P}}$ for the $GM_3(g, \pi)$ model. The first block compares the estimates for models TS and FV, the second block compares models TS and TS^n, and the third compares models FV and FV^n.

4.2 ML estimation of the conditional distribution

In Section 3.2, we observed several conditions for the ML estimators of the parameters governing the distribution of $(M_t, \mathcal{P}_t^{\mathcal{L}})$ to agree. These conditions were that first the smoothed lower order portfolios should agree closely with the observed lower order portfolios. Second, for the lower order portfolios there should be a low amount of uncertainty about their unobserved theoretical values. Finally, the time series average of the errors, relative to the error variances, should be low for the higher order portfolios.

We have just seen in Section 4.1, the first two conditions are satisfied in the estimation of the $GM_s(g, \pi)$ model. The final condition that the time series average of the measurement errors (relative to their variances) is small for all the yields also holds. Although Panel (a) of Figure 2 indicates that at times the errors for individual yields can be very large, visually we can see that the time series average

Given that all of the conditions are approximately satisfied, we should anticipate that the ML parameter estimates for $(K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma)$ should agree for the $MTSM$ and the FVAR with either filtering or by assuming that that PC1 is observed with no error. Table 1 displays the ratios and scaled differences of the conditional mean parameters from model $GM_3(g, \pi)$ and its associated $FVAR$, with and without filtering. Consistent with our theory, we see that the conditional mean parameters are virtually identical. Table 2 then compares the conditional variance parameters. Again, since the time series average errors (relative to their variances) of the yields are all near zero, the conditional variance parameters are nearly identical across the three specifications.

	Ratio			Scaled Difference		
	1.01	-	-	0.00359	-	-
$\frac{GM_3^{KF}}{FV_3^{KF}}$	0.987	1	-	-0.00649	0.000123	-
	0.998	1.01	1	-0.000901	0.00287	5.09e-05
	1.07	-	-	0.0335	-	-
$\frac{GM_3^{KF}}{GM_3}$	0.9	1	-	-0.0526	0.000953	-
	0.885	1.11	1.01	-0.0609	0.0503	0.00472
	1.07	-	-	0.0339	-	-
$\frac{FV_3^{KF}}{FV_3}$	0.898	1	-	-0.0535	0.000986	-
	0.885	1.11	1.01	-0.0612	0.0516	0.00474

Table 2: Ratios and scaled differences of conditional second moment parameters in Σ_Z for model $GM_3(g, \pi)$. Standard deviations are compared along the diagonals and correlations are compared on the off-diagonals.

4.3 Statistics of the distribution of (M_t, y_t)

Given that all of our conditions approximately hold for the $GM_3(g, \pi)$ model and both the conditional mean and conditional variance parameters are nearly identical, it follows that the distribution of the risk factors are the same across the specifications. This implies also that all statistics of the distribution, such as the impulse response functions will be nearly identical as well. This returns to our previous example in Figure 1, where we plotted the impulse response of $PC1$ to a shock to CPI Inflation in Model $GM_3(g, \pi)$. Both the term structure model and the factor-VAR have almost identical impulse responses for $PC1$ in response to CPI shocks, whether we impose no arbitrage or not.

5 Extensions

In this section, we explore a number of extensions and applications of our results. First, we turn our analysis to study the impact of no arbitrage to the study of the violations of the expectations theory of the term structure. Second, we consider the effect of relaxing the spanning assumption of the macro-variables by the yields as in Joslin, Priebsch, and Singleton (2010). Finally, we look at the effects of higher-order Markov processes for the yields and macro-variables.

5.1 Resolutions of Expectations Puzzles by $YTSM$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

$$Proj\left[y_{t+1}^{n-1} - y_t^n | y_t^n - r_t\right] = \alpha_n + \phi_n \left(\frac{y_t^n - r_t}{n-1}\right), \tag{18}$$

where y_t^n denotes the n-period zero yield and $Proj[\cdot|\cdot]$ denotes linear least-squares projection. The ETTS implies that $\phi_n = 1$, for all maturities n. To see under what circumstances (18) holds it is instructive to compare this relationship to the general premium-adjusted expression

$$E^{\mathbb{P}}\left[y_{t+1}^{n-1} - y_t^n - (c_{t+1}^{n-1} - c_t^{n-1}) + \frac{p_t^{n-1}}{n-1} | y_t\right] = \left(\frac{y_t^n - r_t}{n-1}\right), \tag{19}$$

where

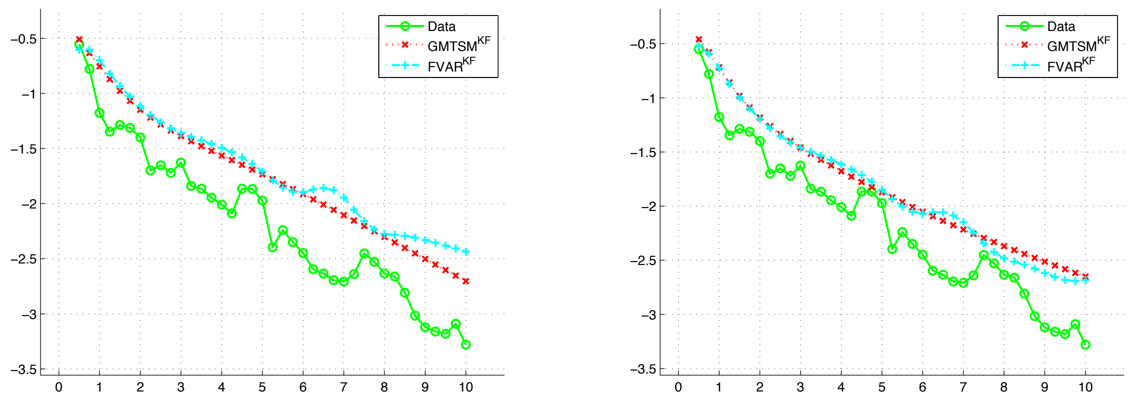
$$c_t^n \equiv y_t^n - \frac{1}{n}\sum_{i=0}^{n-1} E^{\mathbb{P}}[r_{t+1}|y_t] \text{ and } p_t^n \equiv f_t^n - E^{\mathbb{P}}[r_{t+n}|y_t] \tag{20}$$

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 *YTSM* is considered successful at explaining the failure of the ETTS if the term premiums it generates through time-varying market prices of risk reproduce (19) and, thereby, leads to a pattern in the model-implied ϕ_n^{GYTSM} that matches the ϕ_n in the sample.

Since the projections in (18) are restrictions on the conditional $\mathbb{P}$-distribution of y_t, our earlier analysis implies that, absent measurement errors on y_t^o, and if the model-implied loadings of yields on the risk-factors are close to the $FVAR$ estimates, 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 $\mathcal{P}_t$. In practice, bond yields are not priced perfectly by low-dimensional *YTSM*s. Likewise, in the presence of large pricing errors, risk loadings of individual yields may differ from their $FVAR$ counterpart. Therefore, 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 *YTSM*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), assuming that all yields are priced with errors. Using the covariances of the steady-state distribution of $\mathcal{P}_t^{\mathcal{N}}$ implied by these *YTSM*s evaluated at the ML estimates, we compute the projection 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.[24] The

[24] A practical problem that arises in computing the regression coefficients for the $FVAR$ model is that, from



(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.

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.[25] Consistent with the extant evidence, these low-dimensional $YTSM$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 $YTSM$ 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 (8) is a pattern of projection coefficients ϕ_n^{FVAR} that approximately matches those from the regression equations (18). The $YTSM$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 $YTSM$s and the $FVAR$'s produce almost identical conditional distributions of bond yields.[26] With regard to the ETTS, this is manifested in the nearly identical patterns of projection coefficients ϕ_n^{FVAR} and ϕ_n^{GYTSM} in Figure 3.

the twelve yields used in estimation of the $YTSM$ we cannot determine the loadings on the risk factors for all of the maturities. For those maturities not used in estimation, we obtain their loadings from cubic splines fitted through the loadings of the twelve maturities used in estimation. Very similar results are obtained by projecting all yields onto the risk factors using OLS regression and using these loadings to compute the $FVAR$-implied coefficients.

[25] 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 $YTSM$. The results for shorter and longer holding periods are qualitatively similar.

[26] 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.

5.2 Models with Unspanned Risk Factors

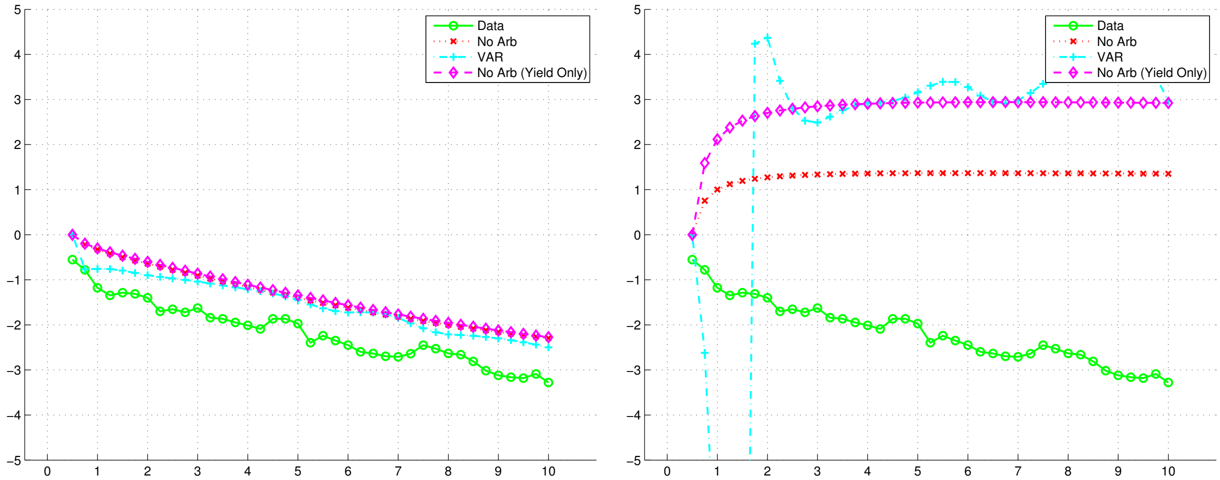
Up to this point, the models we have considered have the macro variables enter directly as risk factors determining interest rates, as is the case with the large majority of the extant literature on *MTSM*s. Joslin, Priebsch, and Singleton (2010) have developed a different class of models that allow for unspanned yield and macro risks– risks that cannot be replicated by linear combinations of bond yields.[27] Canonical versions of these models with unspanned risks share two important properties: (1) except for the volatility parameter (Σ), the $\mathbb{P}$-parameters are distinct from the $\mathbb{Q}$-parameters; and (2) Σ only affects yield levels and not the loadings of yields on the risk factors. Analogous to the spanned models, the implication of property (1) is: when one assumes the risk factors are observed without error, forecasts agree identically with the corresponding $FVAR$. Likewise, property (2) implies: the deviation of the ML estimate of Σ from its $FVAR$-counterpart is proportional to the average-to-variance ratio of the pricing errors. Therefore, so long as one considers canonical models with unspanned risk factors, the historical distribution of the yields and macro-factors estimated using either the $FVAR$ or the no-arbitrage model will be nearly identical.[28]

We verify these assertions by reconsidering the *MTSM*s explored above, but now reformulated as models with unspanned macro risks. Specifically, for the three-factor model $GM_3(g)$ with $g = REALPC$, we adopt a variant of the setup in Joslin, Priebsch, and Singleton (2010) in which there are two pricing factors ($\mathcal{P}_t^2$) and the full Z_t has forecasting power for excess returns. Similarly, for model $GM_3(g, \pi)$, the single pricing factor is $PC1$ and again the full state ($REALPC, INFPC, PC1$) determines risk premiums. Finally, for the four-factor model with $M_t' = (CPI, HELP)$, the pricing factors are $\mathcal{P}_t^2$ and ($\mathcal{P}_t^2, M_t$) determine risk premiums. In all of these cases, the estimated $K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}}$ and Σ from the *MTSM* are close to their $FVAR$ counterparts, whether we assume the pricing factors are measured without or with errors.

That the joint $\mathbb{P}$ distributions of the state Z_t implied by the *MTSM*s and their $FVAR$s are nearly identical is reflected in the model-implied projection coefficients (18) for evaluating the $ETTS$, as discussed in Section 5.1. Figure 4 plots these coefficients for the *MTSM*s and the corresponding yield-only models with the same number of factors underlying bond pricing (one factor for $GM_3(g, \pi)$ and two factors for $GM_3(g)$). All of the models with two pricing factors are successful in matching the failures of the expectations hypothesis, and the no-arbitrage and $FVAR$ models line up very closely. The projections for models $GM_3(g)$ and GY_2^{KF} are virtually on top of one another, indicating that the crucial component in matching the projections are the inclusion of PC2 (slope), and not the macro-information. This is underscored by considering the results for model $GM_3(g, \pi)$. Again, all the no-arbitrage and $FVAR$ versions produce similar patterns for the projection coefficients, consistent with our main arguments. However, in these models the projection coefficients are positive and

[27] For additional applications of their framework, see Wright (2009) and Barillas (2010). Duffee (2009) discusses a complementary model of unspanned risks in yield-only models.

[28] 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.



(a) Model $GM_3(g)$

(b) Model $GM_3(g, \pi)$

Figure 4: Regression coefficients ϕ_n implied by the sample Treasury yields, the models including two lags, $GM_3(g)^F$ and $GM_3(g,\pi)^F$, and their corresponding unconstrained factor-VAR model. All yields were allowed to be priced with errors. Additionally, the yield only models GY_1^{KF} and GY_2^{KF} are plotted in panels (a) and (b), respectively. The horizontal axis is in years.

increasing with maturity rather than negative and decreasing with maturity, as in the data. That is, without including PC2, none of the specifications are able to match the violations of the expectations hypothesis summarized by the sample projection coefficients.

5.3 Higher-Order VAR Models of Risk Factors

Up to this point we have focused on the class of $MTSM$s that presume that Z_t follows a first-order Markov process under $\mathbb{P}$ and $\mathbb{Q}$. We now show that our central arguments carry over to formulations based on higher-order VARs: the corresponding canonical models produce nearly identical estimates of the historical distribution of macro-factors and bond yields as their associated $FVAR$s.

We now consider the following extended family of models. For a fixed set of $\mathcal{M}$ macroeconomic variables M_t, we let $GMTSM_{\mathcal{N},p}(\mathcal{M})$ be the set of invariant affine transformations of non-degenerate $MTSM$s in which r_t is an affine function of $\mathcal{L}$ latent (L_t) and $\mathcal{M}$ macro (M_t) factors and the vector $Z_t^{\mathcal{L}} = (M_t', L_t')'$ follows the Gaussian processes (2) under $\mathbb{Q}$ and

$$Z_t^{\mathcal{L}} = \kappa_{0Z}^{\mathbb{P}} + \kappa_{1Z}^{\mathbb{P}} \overrightarrow{Z^{\mathcal{L}}}_{t-1,p} + \sqrt{\Omega_Z}\epsilon_t^{\mathbb{P}} \tag{21}$$

under $\mathbb{P}$, where, for any U_t, $\overrightarrow{U}_{t,p} \equiv (U_t', U_{t-1}', \ldots, U_{t-p+1}')'$.

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.[29] These studies adopt specifications of the market prices of risk that imply that the $\mathbb{Q}$ distribution of $Z^{\mathcal{L}}$ inherits the lag structure of the $VAR(p)$ under $\mathbb{P}$. The choice of $p > 1$ in (21) is often supported by descriptive evidence under the historical distribution. Left open by such evidence is the nature of the dependence of r_t and $Z_t^{\mathcal{L}}$ on lags of $Z_t^{\mathcal{L}}$ under the pricing distribution. As previously noted, within the family of reduced-form $MTSM$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.

Indirect guidance on the lag structure of the $\mathbb{Q}$ distribution of $(r_t, Z_t^{\mathcal{L}})$ is provided by the projections of yields onto lagged values of the state vector $Z_t^{\mathcal{L}}$. Starting with the null that $Z_t^{\mathcal{L}}$ follows the first order VAR under $\mathbb{Q}$, the information in $Z_t^{\mathcal{L}}$ is identical to that of $Z_t = (M_t, \mathcal{P}_t^{\mathcal{L}\prime})'$. Therefore, it is exactly equivalent for us to project yields onto lagged values of Z_t. If the regressions fit materially improves with additional lags of Z_t, this would constitute evidence against our null.[30] The standard deviations of the errors in the projections of yields onto $\overrightarrow{Z}^{o}_{t,q}$ ($RMSE$ 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 3. 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.[31] In all cases, the (AIC, BIC) model selection criteria select $q = 1$.

With these observations in mind, and supported by the evidence in Table 3, we proceed to explore asymmetric formulations of $MTSM$s in which Z_t follows a $VAR(p)$ under $\mathbb{P}$, and Z_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 (4). Further, the invariant transformation (5) gives an observationally equivalent model in which $Z_t' = (M_t', \mathcal{P}_t^{\mathcal{L}\prime})$, r_t is given by (1), and Z_t follows the process (2) under $\mathbb{Q}$, with the parameters governing (1) and (2) being explicit functions of $\Theta^{\mathbb{Q}}_{GMTSM} = (r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \gamma_0^W, \gamma_1^W, \Sigma)$.[32] These functions are identical to those in Section 2 and

[29] 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.

[30] 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.

[31] 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 60bp, that adding lags under $\mathbb{Q}$ improves the $RMSE$s a bit more, up to eight basis points.

[32] 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 $MTSM$ in which the risk factors $Z_t^{\mathcal{L}}$ are latent and satisfy

$$Z_t^{\mathcal{L}} = \kappa^{\mathbb{Q}}_{0Z} + \kappa^{\mathbb{Q}}_{1Z} \overrightarrow{Z}^{\mathcal{L}}_{t-1,q} + \sqrt{\Omega_Z}\epsilon_t^{\mathbb{Q}},$$

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

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	10	15	17	10	10	18
GM_3	$REALPC$	6	9	14	16	10	10	18
GM_3	$REALPC$	12	9	14	16	9	9	17
GM_3	$REALPC, INFPC$	1	60	45	22	25	36	47
GM_3	$REALPC, INFPC$	6	57	42	20	24	34	45
GM_3	$REALPC, INFPC$	12	53	37	18	22	31	40
GM_4	$HELP, CPI$	1	8	15	17	10	10	18
GM_4	$HELP, CPI$	6	7	14	16	9	10	17
GM_4	$HELP, CPI$	12	7	14	15	9	9	16
GM_4	$REALPC, INFPC$	1	9	15	17	10	10	18
GM_4	$REALPC, INFPC$	6	8	14	16	9	10	17
GM_4	$REALPC, INFPC$	12	8	13	15	9	9	16

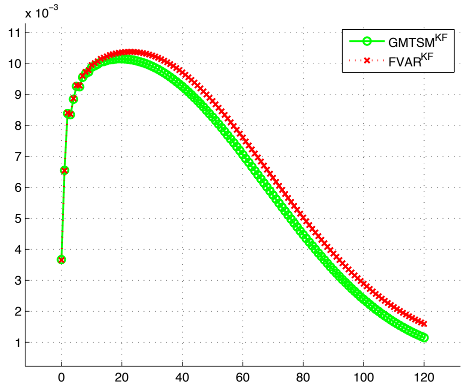
Table 3: Root-mean-squared fitting errors, measured in basis points, from projections of bond yields onto current and lagged values of the risk factors Z_t^o. For given q, the conditioning information is $(Z_t, Z_{t-1}, \ldots, Z_{t-q+1})$.

provided in Appendix A. Additionally, the $\mathbb{P}$-dynamics of Z_t is unrestricted:

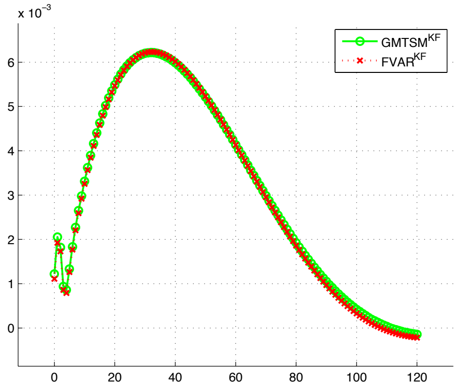
$$Z_t = K_{0Z}^{\mathbb{P}} + K_{1Z}^{\mathbb{P}} \overrightarrow{Z}_{t-1,p} + \sqrt{\Sigma}\epsilon_t^{\mathbb{P}}. \tag{22}$$

Our previous argument extends to as follows. For fixed W, any member of 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 $\mathcal{P}_t^{\mathcal{L}}$; r_t is given by (1); M_t is related to $\mathcal{P}_t^{\mathcal{N}}$ according to (4); the risk factors follow the Gaussian process (2) under $\mathbb{Q}$ and (22) under $\mathbb{P}$, where $K_{0Z}^{\mathbb{Q}}$, $K_{1Z}^{\mathbb{Q}}$, Σ, ρ_0, and ρ_1 are explicit functions of $\Theta_{GMTSM}^{\mathbb{Q}}$. For given W, our canonical form is parametrized by $\Theta_Z = (\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma, \gamma_0^W, \gamma_1^W)$.

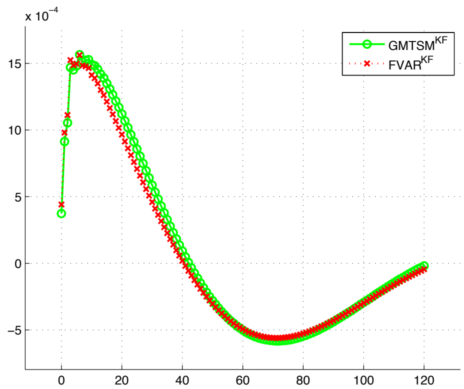
Our main analysis comparing the properties of no arbitrage models with their associated $FVAR$s follows through for this generalized $MTSM$ with lags. The source of this robustness of our analysis is again the separation of the parameter space: except for Σ, the $\mathbb{P}$-parameters are distinct from the $\mathbb{Q}$-parameters. Therefore, when Z_t is measured without errors, ML estimates of the $\mathbb{P}$-mean parameters can be obtained directly by fitting Z_t to a $VAR(p)$. Moreover, with or without lags, Σ 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 must be proportional to the average-to-variance ratio of errors defined earlier, so earlier invariance results regarding the ML estimate of conditional variances continue to hold. Provided that the average pricing errors are small relative to their standard deviations, one should expect the model-implied conditional $\mathbb{P}$-distribution of Z_t to be nearly identical to its counterpart implied by an unconstrained $FVAR$ of yields.



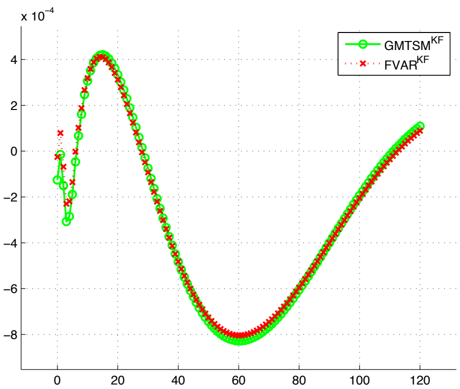
(a) Response of $PC1$ to $HELP$



(b) Response of $PC1$ to CPI



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



(d) Response of $PC2$ to CPI

Figure 5: 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 $(CPI, HELP, PC1, PC2)$.

To examine this issue empirically we introduce $VAR(p)$ processes under $\mathbb{P}$ to models $GM_3(REALPC)$ and $GM_4(HELP, CPI)$. For each of these models, we first fit Z_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 (7). 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 $p = 7$ ($p = 4$) in generalizing GM_3 (GM_4). We denote these models by $GM_{3,7}$ and $GM_{4,4}$, respectively, where the second subscripts refer to the number of lags in each model.

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.[33] Figure 5 displays the impulse responses implied by model $GM_{4,4}^{KF}$ which depend on both $K_{1Z}^{\mathbb{P}}$ and Σ.[34] As before, the imposition of no arbitrage is virtually inconsequential for how shocks to macro factors impact the yield curve.[35] This finding is robust to choices of which macro variables are included. For example, a very similar result is obtained from model $GM_{3,7}$.

6 Concluding Remarks

We have shown theoretically and documented empirically that the no-arbitrage restrictions of canonical *MTSM*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 *MTSM*s in the literature. This finding is robust to whether a subset of the bond yields are priced perfectly by a *MTSM*, or all yields are measured with error and filtering is used in ML estimation.

Of course this finding does not imply that *YTSM*s or *MTSM*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 *MTSM* will lead to more efficient estimates of the $\mathbb{P}$ distribution of yields relative to those of the factor model (6)-(8).

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 $K_{1Z}^{\mathbb{P}}$ imposed by Christensen, Diebold, and Rudebusch (2009) in their analysis of *YTSM*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 ($\mathcal{M} = 2, \mathcal{L} = 1$) *MTSM* 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 *MTSM*s (and *YTSM*s) carry over to non-canonical models.

More generally, the results on higher order Markov term structure models combined with the propositions in JSZ imply that one widely imposed class of restrictions will not break our irrelevancy results. Specifically, constraints on the market price of risk that amount to

[33]Given their similarity to our previous findings, we omit these tables.

[34]The higher-order lag structure alters somewhat the individual responses. The choppy behavior over short horizons for some of the responses in Figure 5 is also evident in the impulse responses reported in Ang and Piazzesi (2003) for their *MTSM* with lags.

[35]The difference in the responses of $PC1$ to innovations in $HELP$ is roughly 5bp. Since the loadings of yields on $PC1$ are about 0.3, this translates to a difference in yields' responses of about 1.5bp.

zero restrictions on the lag structure of Z_t under the pricing measure $\mathbb{Q}$ do not affect the factorization of the likelihood function under our normalization scheme. Therefore, they do not impinge on the parameters of the matrix $K^{\mathbb{P}}_{1Z}$, whence their affects on the entire conditional $\mathbb{P}$ distribution of Z_t are likely to be negligible.

For instance, the *MTSM* with constrained market prices of risk studied by Joslin, Priebsch, and Singleton (2010) has very different dynamic properties than an unconstrained VAR. Similarly, Duffee (2011) obtained improved out-of-sample forecasts of bond yields in a *YTSM* with one of the risk factors constrained to follow a random walk. On the other hand, focusing on conditional means, Joslin, Singleton, and Zhu (2010) provide examples where forecasts from constrained *YTSM*s and their associated VARs are identical.

Similarly, unit-root or cointegration-type restrictions imposed directly on the $\mathbb{P}$ distribution of the risk factors (JSZ, Duffee (2011), Jardet, Monfort, and Pegoraro (2010)) also seem unlikely to induce large differences between the conditional $\mathbb{P}$ distributions implied by a *MTSM* and the corresponding $FVAR$ with the same restrictions imposed. JSZ, for instance, obtained very similar conditional mean parameters under cointegration from a *YTSM* and a $FVAR$ estimated using US Treasury data.

On the other hand, Joslin, Priebsch, and Singleton (2010) found that the constraints across the $\mathbb{P}$ and $\mathbb{Q}$ parameters selected by Schwarz (1978)'s Bayesian information criteria led to substantial differences 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.[36] 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 *MTSM* 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.

[36] 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).

A A Canonical Form for *MTSM*s

Our objective is to show that each $GMTSM_{\mathcal{N}}(\mathcal{M})$ where

$$r_t = \rho_{0Z}^{\mathcal{L}} + \rho_{1Z}^{\mathcal{L}} \cdot Z_t^{\mathcal{L}} \tag{23}$$

with the risk factors $Z_t^{\mathcal{L}} \equiv (M_t', L_t')'$ following the Gaussian processes

$$\Delta Z_t^{\mathcal{L}} = \kappa_{0Z}^{\mathbb{Q}} + \kappa_{1Z}^{\mathbb{Q}} Z_{t-1}^{\mathcal{L}} + \sqrt{\Omega_Z}\epsilon_t^{\mathbb{Q}} \text{ under } \mathbb{Q} \text{ and} \tag{24}$$
$$\Delta Z_t^{\mathcal{L}} = \kappa_{0Z}^{\mathbb{P}} + \kappa_{1Z}^{\mathbb{P}} Z_{t-1}^{\mathcal{L}} + \sqrt{\Omega_Z}\epsilon_t^{\mathbb{P}} \text{ under } \mathbb{P}, \tag{25}$$

is observationally equivalent to a *unique* member of $GMTSM_{\mathcal{N}}(\mathcal{M})$ in which $Z_t = (M_t', \mathcal{P}_t^{\mathcal{L}\prime})'$ with some $\mathcal{L}$ yield portfolios $\mathcal{P}_t^{\mathcal{L}}$:

$$r_t = \rho_{0Z} + \rho_{1Z} \cdot Z_t, \tag{26}$$
$$\Delta Z_t = K_{0Z}^{\mathbb{Q}} + K_{1Z}^{\mathbb{Q}} Z_{t-1} + \sqrt{\Sigma}\epsilon_t^{\mathbb{Q}} \text{ under } \mathbb{Q} \text{ and} \tag{27}$$
$$\Delta Z_t = K_{0Z}^{\mathbb{P}} + K_{1Z}^{\mathbb{P}} Z_{t-1} + \sqrt{\Sigma}\epsilon_t^{\mathbb{P}} \text{ under } \mathbb{P} \tag{28}$$

where $(\rho_{0Z}, \rho_{1Z}, K_{0Z}^{\mathbb{Q}}, K_{1Z}^{\mathbb{Q}})$ are explicit functions of some underlying parameter set $\Theta_{GMTSM}^{\mathbb{Q}} = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0^W, \gamma_1^W, \Sigma)$ to be described. We will make precise the sense in which $\Theta_Z = (\Theta_{GMTSM}^{\mathbb{Q}}, K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}})$ uniquely characterizes the latter $GMTSM_{\mathcal{N}}(\mathcal{M})$.

Observational Equivalence

Assuming, for ease of exposition, that $\kappa_{1Z}^{\mathbb{Q}}$ has nonzero, real and distinct eigenvalues with the standard eigendecomposition:[37] $\kappa_{1Z}^{\mathbb{Q}} = A^{\mathbb{Q}} diag(\lambda^{\mathbb{Q}}) A^{\mathbb{Q}^{-1}}$, we follow Joslin (2006) by adopting the rotation:

$$X_t = \mathcal{V}^{-1}\left(Z_t^{\mathcal{L}} + (\kappa_{1Z}^{\mathbb{Q}})^{-1}\kappa_{0Z}^{\mathbb{Q}}\right) \text{ where } \mathcal{V} = A^{\mathbb{Q}} diag((\rho_{1Z}^{\mathcal{L}})' A^{\mathbb{Q}})^{-1} \tag{29}$$

to arrive at the following $\mathbb{Q}$ specification:

$$r_t = r_\infty^{\mathbb{Q}} + \iota \cdot X_t, \text{ and } \Delta X_t = diag(\lambda^{\mathbb{Q}}) X_{t-1} + \sqrt{\Sigma_X}\epsilon_t^{\mathbb{Q}} \tag{30}$$

where $\lambda^{\mathbb{Q}}$ is ordered and ι denotes a vector of ones and

$$r_\infty^{\mathbb{Q}} = \rho_{0Z}^{\mathcal{L}} + (\rho_{1Z}^{\mathcal{L}})'(\kappa_{1Z}^{\mathbb{Q}})^{-1}\kappa_{0Z}^{\mathbb{Q}} \text{ and } \Omega_Z = \mathcal{V}\Sigma_X\mathcal{V}'.$$

From (30), the $J \times 1$ vector of yields y_t is affine in X_t:

$$y_t = A_X(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) + B_X(\lambda^{\mathbb{Q}}) X_t \tag{31}$$

[37] See *JSZ* for detailed treatments of cases with complex, repeated or zero eigenvalues.

with A_X, B_X obtained from standard recursions. Following JSZ, we fix a full-rank loadings matrix $W \in \mathbb{R}^{J\times J}$, and let $\mathcal{P}_t = W y_t$. Focusing on the first $\mathcal{N}$ portfolios $\mathcal{P}_t^{\mathcal{N}}$, we have:

$$\mathcal{P}_t^{\mathcal{N}} = W^{\mathcal{N}} A_X + W^{\mathcal{N}} B_X X_t. \tag{32}$$

Based on (29) and (32), there is a linear mapping between M_t and $\mathcal{P}_t^{\mathcal{N}}$:

$$M_t = \gamma_0^W + \gamma_1^W \mathcal{P}_t^{\mathcal{N}} \tag{33}$$

where

$$\gamma_1^W = \mathcal{V}_{\mathcal{M}}(W^{\mathcal{N}} B_X)^{-1} \text{ and } \gamma_0^W = -\gamma_1^W W^{\mathcal{N}} A_X - A_{\mathcal{M}}^{\mathbb{Q}}(\lambda^{\mathbb{Q}})^{-1} A^{\mathbb{Q}^{-1}} \kappa_{0Z}^{\mathbb{Q}}, \tag{34}$$

and $\mathcal{V}_{\mathcal{M}}$, $A_{\mathcal{M}}^{\mathbb{Q}}$ denote the first $\mathcal{M}$ rows of $\mathcal{V}$, $A^{\mathbb{Q}}$, respectively. This allows us to write:

$$Z_t = \Gamma_0 + \Gamma_1 \mathcal{P}_t^{\mathcal{N}} = \Gamma_0 + \Gamma_1(W^{\mathcal{N}} A_X + W^{\mathcal{N}} B_X X_t) = \mathcal{U}_0 + \mathcal{U}_1^{-1} X_t \tag{35}$$

where

$$\Gamma_0 = (\gamma_0^{W\prime}, 0_{\mathcal{L}}')', \;\; \Gamma_1 = \begin{pmatrix} \gamma_1^W \\ I_{\mathcal{L}}, 0_{\mathcal{L}\times\mathcal{M}} \end{pmatrix}, \;\; \mathcal{U}_0 = \Gamma_0 + \Gamma_1 W^{\mathcal{N}} A_X, \text{ and } \mathcal{U}_1 = (\Gamma_1 W^{\mathcal{N}} B_X)^{-1}.$$

Combining (30) and (35), the $\mathbb{Q}$-specification of Z_t is:

$$r_t = \rho_{0Z} + \rho_{1Z} \cdot Z_t \text{ and } \Delta Z_t = K_{0Z}^{\mathbb{Q}} + K_{1Z}^{\mathbb{Q}} Z_{t-1} + \sqrt{\Sigma}\epsilon_t^{\mathbb{Q}} \tag{36}$$

where

$$\rho_{1Z} = (\mathcal{U}_1)'\iota \text{ and } \rho_{0Z} = r_\infty^{\mathbb{Q}} - \rho_{1Z} \cdot \mathcal{U}_0,$$
$$K_{1Z}^{\mathbb{Q}} = \mathcal{U}_1^{-1}\lambda^{\mathbb{Q}}\mathcal{U}_1, \;\; K_{0Z}^{\mathbb{Q}} = -K_{1Z}^{\mathbb{Q}}\mathcal{U}_0 \text{ (and } \Sigma_X = \mathcal{U}_1\Sigma\mathcal{U}_1').$$

Based on (29) and (35), there must be a linear mapping between Z_t and $Z_t^{\mathcal{L}}$. It follows that the $\mathbb{P}$-dynamics of Z_t must be Gaussian as in (28).

To summarize, the $GMTSM_{\mathcal{N}}(\mathcal{M})$ with mixed macro-latent risk factors $Z_t^{\mathcal{L}}$, described by (23), (24), and (25), is observationally equivalent to one with observable mixed macro-yield-portfolio risk factors Z_t, characterized by (26), (27), and (28). The *primitive* parameter set is $\Theta_Z = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0^W, \gamma_1^W, \Sigma, K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}})$. The mappings between $(\rho_{0Z}, \rho_{1Z}, K_{0Z}^{\mathbb{Q}}, K_{1Z}^{\mathbb{Q}})$ and $\Theta_{GMTSM}^{\mathbb{Q}} = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0^W, \gamma_1^W, \Sigma)$ are:

$$\rho_{1Z} = (\mathcal{U}_1)'\iota, \;\; \rho_0 = r_\infty^{\mathbb{Q}} - \rho_{1Z} \cdot \mathcal{U}_0, \;\; K_{1Z}^{\mathbb{Q}} = \mathcal{U}_1^{-1}\lambda^{\mathbb{Q}}\mathcal{U}_1, \;\; K_{0Z}^{\mathbb{Q}} = -K_{1Z}^{\mathbb{Q}}\mathcal{U}_0 \tag{37}$$

where

$$\mathcal{U}_1 = (\Gamma_1 W^{\mathcal{N}} B_X(\lambda^{\mathbb{Q}}))^{-1}, \;\; \mathcal{U}_0 = \Gamma_0 + \Gamma_1 W^{\mathcal{N}} A_X(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \mathcal{U}_1\Sigma\mathcal{U}_1'), \text{ and}$$
$$\Gamma_0 = (\gamma_0^{W\prime}, 0_{\mathcal{L}}')', \;\; \Gamma_1 = \begin{pmatrix} \gamma_1^W \\ I_{\mathcal{L}}, 0_{\mathcal{L}\times\mathcal{M}} \end{pmatrix}.$$

Uniqueness

Consider two parameter sets, Θ_Z and $\tilde{\Theta}_Z$, that give rise to two observationally equivalent $GMTSM_{\mathcal{N}}(\mathcal{M})$'s with risk factors Z_t. Since Z_t is observable, the parameters, $\Sigma, K^{\mathbb{P}}_{0Z}, K^{\mathbb{P}}_{1Z}$, describing the $\mathbb{P}$-dynamics of Z_t must be identical. Additionally, based on (33), the following identity must hold state by state:

$$M_t \equiv \gamma_0^W + \gamma_1^W \mathcal{P}_t^{\mathcal{N}} \equiv \tilde{\gamma}_0^W + \tilde{\gamma}_1^W \mathcal{P}_t^{\mathcal{N}}. \tag{38}$$

Since W is full rank, hence $\mathcal{P}_t^{\mathcal{N}}$ are linearly independent, it follows that:

$$\gamma_0^W = \tilde{\gamma}_0^W \text{ and } \gamma_1^W = \tilde{\gamma}_1^W. \tag{39}$$

Finally, writing the term structure with $\mathcal{P}_t^{\mathcal{N}}$ as risk factors:

$$y_t = A_X + B_X(W^{\mathcal{N}}B_X)^{-1}(P_t^{\mathcal{N}} - W^{\mathcal{N}}A_X), \tag{40}$$

it follows that

$$B_X(W^{\mathcal{N}}B_X)^{-1} = \tilde{B}_X(W^{\mathcal{N}}\tilde{B}_X)^{-1}, and \tag{41}$$

$$(I_J - B_X(W^{\mathcal{N}}B_X)^{-1}W^{\mathcal{N}})A_X = (I_J - \tilde{B}_X(W^{\mathcal{N}}\tilde{B}_X)^{-1}W^{\mathcal{N}})\tilde{A}_X. \tag{42}$$

Now (41) is equivalent to:

$$diag(\frac{1-\lambda_i^n}{1-\lambda_i})(W^{\mathcal{N}}B_X)^{-1} = diag(\frac{1-\tilde{\lambda}_i^n}{1-\tilde{\lambda}_i})(W^{\mathcal{N}}\tilde{B}_X)^{-1} \tag{43}$$

for every horizon n. As long as both $W^{\mathcal{N}}B_X$ and $W^{\mathcal{N}}\tilde{B}_X$ are full rank, it must follow that $\lambda_i^{\mathbb{Q}} \equiv \tilde{\lambda}_i^{\mathbb{Q}}$ for all i's.

Turning to (42), we note that

$$A_X = \iota\, r_\infty^{\mathbb{Q}} + \beta_X vec(\Sigma_X) \tag{44}$$

where β_X is a function of $\lambda^{\mathbb{Q}}$, and thus must be the same for both Θ_Z and $\tilde{\Theta}_Z$. Likewise, $\Sigma_X = \mathcal{U}_1\Sigma\mathcal{U}_1'$, dependent only on $(\gamma_1^W, \lambda^{\mathbb{Q}}, \Sigma)$, must be the same for both parameter sets. It follows that $r_\infty^{\mathbb{Q}} = \tilde{r}_\infty^{\mathbb{Q}}$. Therefore, $\Theta_Z \equiv \tilde{\Theta}_Z$.

Regularity Conditions

First, we assume that the diagonal elements of $\lambda^{\mathbb{Q}}$ are non-zero, real and distinct. These can be easily relaxed - see *JSZ* for detailed treatments. Second, we assume that the $GMTSM_{\mathcal{N}}(\mathcal{M})$'s are non-degenerate in the sense that there is no transformation such that the effective number of risk factors is less than $\mathcal{N}$. For this, the requirement is that all elements of $(\rho_{1Z}^{\mathcal{L}})'A^{\mathbb{Q}}$ are non-zero. In terms of the parameters of our canonical form, none of the eigenvectors of the risk-neutral feedback matrix $K_{1Z}^{\mathbb{Q}}$ is orthogonal to the loadings vector ρ_{1Z} of the short rate. Finally, to maintain valid transformations between alternative choices of risk factors, we require that the matrices $W^{\mathcal{N}}B_X$ and Γ_1 be full rank. These are conditions on $(\lambda^{\mathbb{Q}}, W)$ and γ_1^W, respectively.

The following theorem summarizes the above derivations:

Theorem 1. *Fix a full-rank portfolio matrix $W \in \mathbb{R}^{J\times J}$, and let $\mathcal{P}_t = W y_t$. 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 Z_t are the macro variables M_t, and the remaining $\mathcal{L}$ components of Z_t are $\mathcal{P}^{\mathcal{L}}_t$; r_t is given by (26); M_t is related to $\mathcal{P}_t$ through*

$$M_t = \gamma_0^W + \gamma_1^W \mathcal{P}^{\mathcal{N}}_t, \tag{45}$$

for $\mathcal{M}\times 1$ vector γ_0^W and $\mathcal{M}\times\mathcal{N}$ matrix γ_1^W; and Z_t follows the Gaussian $\mathbb{Q}$ and $\mathbb{P}$ processes (27), and (28), where $K^{\mathbb{Q}}_{0Z}$, $K^{\mathbb{Q}}_{1Z}$, ρ_{0Z}, and ρ_{1Z} are explicit functions of $\Theta^{\mathbb{Q}}_{GMTSM} = (r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \gamma_0^W, \gamma_1^W, \Sigma)$, given by (37). For given W, our canonical form is parametrized by $\Theta_Z = (\Theta^{\mathbb{Q}}_{GMTSM}, K^{\mathbb{P}}_{0Z}, K^{\mathbb{P}}_{1Z})$.

B Filtering Invariance

It is standard to write the Kalman filtering equation as:

$$E[\mathcal{P}^{\mathcal{L}}_t|\mathcal{F}_t] = E[\mathcal{P}^{\mathcal{L}}_t|\mathcal{F}_{t-1}] + cov(\mathcal{P}^{\mathcal{L}}_t, U^o_t|\mathcal{F}_{t-1}) var(U^o_t|\mathcal{F}_{t-1})^{-1}(U^o_t - E[U^o_t|\mathcal{F}_{t-1}]) \tag{46}$$

where $U^o_t = (\mathcal{P}^{\mathcal{L},o\prime}_t, D^{o\prime}_t)'$. To evaluate the Kalman gain term, we write:

$$\begin{aligned} cov(\mathcal{P}^{\mathcal{L}}_t, U^o_t|\mathcal{F}_{t-1}) var(U^o_t|\mathcal{F}_{t-1})^{-1} =& cov(\mathcal{P}^{\mathcal{L},o}_t, U^o_t|\mathcal{F}_{t-1}) var(U^o_t|\mathcal{F}_{t-1})^{-1} \\ & - cov(e_{\mathcal{L},t}, U^o_t|\mathcal{F}_{t-1}) var(U^o_t|\mathcal{F}_{t-1})^{-1} \\ =& (I_{\mathcal{L}}, 0_{\mathcal{L},J+\mathcal{M}-\mathcal{L}}) - (\Sigma_{\mathcal{L}e}, 0_{\mathcal{L},J+\mathcal{M}-\mathcal{L}}) var(U^o_t|\mathcal{F}_{t-1})^{-1}. \end{aligned}$$

Applying block inversion to $var(U^o_t|\mathcal{F}_{t-1})$, collecting the first $\mathcal{L}$ rows corresponding to $\mathcal{P}^{\mathcal{L},o}$, and substitute back the Kalman gain term into (46), we can write:

$$\begin{aligned} E[\mathcal{P}^{\mathcal{L}}_t|\mathcal{F}_t] =& \mathcal{P}^{\mathcal{L},o}_t \underbrace{-\Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t}}_{K_{\mathcal{L},t}}(\mathcal{P}^{\mathcal{L},o}_t - E[\mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1}]) \\ & \underbrace{-\Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t} cov(\mathcal{P}^{\mathcal{L},o}_t, D^o_t|\mathcal{F}_{t-1}) var(D^o_t|\mathcal{F}_{t-1})^{-1}}_{K_{D,t}}(D^o_t - E[D^o_t|\mathcal{F}_{t-1}]) \end{aligned} \tag{47}$$

where

$$\Omega_{\mathcal{L}t} = var(\mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1}) - cov(\mathcal{P}^{\mathcal{L},o}_t, D^o_t|\mathcal{F}_{t-1}) var(D^o_t|\mathcal{F}_{t-1})^{-1} cov(D^o_t, \mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1}). \tag{48}$$

From this, the conditional variances are:

$$var(K_{\mathcal{L},t}\mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1}) = \Sigma_{\mathcal{L},e}\Omega^{-1}_{\mathcal{L}t} var(\mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1})\Omega^{-1}_{\mathcal{L}t}\Sigma_{\mathcal{L},e}, \tag{49}$$

$$\begin{aligned} var(K_{D,t}D^o_t|\mathcal{F}_{t-1}) =& \Sigma_{\mathcal{L},e}\Omega^{-1}_{\mathcal{L}t} cov(\mathcal{P}^{\mathcal{L},o}_t, D^o_t|\mathcal{F}_{t-1}) var(D^o_t|\mathcal{F}_{t-1})^{-1} cov(D^o_t, \mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1})\Omega^{-1}_{\mathcal{L}t}\Sigma_{\mathcal{L},e} \\ =& \Sigma_{\mathcal{L},e}\Omega^{-1}_{\mathcal{L}t}(var(\mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1}) - \Omega_{\mathcal{L}t})\Omega^{-1}_{\mathcal{L}t}\Sigma_{\mathcal{L},e}. \end{aligned} \tag{50}$$

It is obvious that $var(K_{D,t}D^o_t|\mathcal{F}_{t-1})$ is strictly smaller than $var(K_{\mathcal{L},t}\mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1})$. Finally, since the conditional mean of $\mathcal{P}^{\mathcal{L},o}_t - E[\mathcal{P}^{\mathcal{L},o}_t|\mathcal{F}_{t-1}]$ is zero, its unconditional variance is simply the unconditional mean of its conditional variance.

C Speed of Convergence to Steady States

Consider the following *generic* state space system:

$$Z_{t+1} = K_0 + K_1 Z_t + \sqrt{\Sigma}\epsilon_{t+1}, \tag{51}$$

$$Z^o_{t+1} = Z_{t+1} + e_{Z,t+1}, \tag{52}$$

$$Y^o_{t+1} = A + BZ_{t+1} + e_{Y,t+1} \tag{53}$$

where $e_{Z,t}$ and $e_{Y,t}$ are independent and $e_{Z,t} \sim N(0, \Sigma_{Ze})$ and $e_{Y,t} \sim N(0, \Sigma_{Ye})$. Let Σ_{t+1}, Ω_{t+1} denote $var(Z_{t+1}|\mathcal{F}_t)$ and $var(Z^o_{t+1}|Y^o_{t+1}, \mathcal{F}_t)$ respectively. It is standard to show that Σ_{t+1} follows the recursion:

$$\Sigma_{t+1} = \Sigma + K_1(\Sigma_t - \Sigma_t \tilde{B}'(\tilde{B}\Sigma_t \tilde{B}' + \Sigma_e)^{-1}\tilde{B}\Sigma_t)K_1' \tag{54}$$

where Σ_e is the variance matrix of $(e'_{Z,t}, e'_{Y,t})'$ and $\tilde{B}' = (I, B')$. We will first show that when $\Sigma_e \Omega_t^{-1}$ is small then Σ_t, and therefore the Kalman gain matrix, will approach its steady state values rapidly. Next, we will show what this condition translates to under our particular setup.

Standard linear algebra allows us to express the term between K_1 and K_1' in (54) as:

$$\Sigma_{Ze} - (\Sigma_{Ze}, 0)\left(\tilde{B}\Sigma_t\tilde{B}' + \Sigma_e\right)^{-1}\begin{pmatrix} \Sigma_{Ze} \\ 0 \end{pmatrix}. \tag{55}$$

Now consider a small variation in Σ_t of $\partial\Sigma_t$, the corresponding change in Σ_{t+1} (the Fréchet derivative) will be:

$$\partial\Sigma_{t+1} = \Phi\partial\Sigma_t\Phi' \text{ with } \Phi = K_1(\Sigma_{Ze}, 0)\left(\tilde{B}\Sigma_t\tilde{B}' + \Sigma_e\right)^{-1}\begin{pmatrix} I \\ B \end{pmatrix}. \tag{56}$$

Now replace $\left(\tilde{B}\Sigma_t\tilde{B}' + \Sigma_e\right)$ by $var\left(\begin{matrix} Z^o_t \\ Y^o_t \end{matrix} |\mathcal{F}_{t-1}\right)$ and apply block-wise inversion to this matrix, we have:

$$\Phi = K_1\Sigma_{Ze}\Omega_t^{-1}(I - \Sigma_t B'(B\Sigma_t B' + \Sigma_{Ye})^{-1}B). \tag{57}$$

As a result, as $\Sigma_{Ze}\Omega_t^{-1}$ approaches zeros, so do the eigenvalues of Φ. Since the recursion (54) can be written approximately as:

$$vec(\Sigma_{t+1} - \bar{\Sigma}) \approx (\Phi \otimes \Phi)vec(\Sigma_t - \bar{\Sigma}) \tag{58}$$

where $\bar{\Sigma}$ denotes the steady state value of Σ_t, small eigenvalues of Φ (and hence $\Phi\otimes\Phi$) should induce fast convergence to the steady state.

Now to apply this to our setup, since we assume that M_t is perfectly observed, the $\mathcal{M}$ rows and columns of Σ_e corresponding to M_t are zeros. Applying block inversion to Ω_t and collect the $\mathcal{L} \times \mathcal{L}$ block corresponding to the yield portfolios $\mathcal{P}^{\mathcal{L}}_t$, it follows that we need $\Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t}$ to be small.

D Conditional Mean Parameters

We are going to show that the filtered version of equation (14):

$$[\hat{K}^{\mathbb{P}}_{0Z}, I + \hat{K}^{\mathbb{P}}_{1Z}]' = \left(\frac{1}{T}\sum_t Z^f_{t+1}, \frac{1}{T}\sum_t (Z_{t+1}Z'_t)^f \right) \begin{pmatrix} 1 & \frac{1}{T}\sum_t Z^{f\prime}_t \\ \frac{1}{T}\sum_t Z^f_t & \frac{1}{T}\sum_t (Z_t Z'_t)^f \end{pmatrix}^{-1}, \quad (59)$$

with some mild assumptions, will deliver estimates that are close to the OLS estimates when $\Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t}$ is small. Assuming further that $(Z_t Z'_t)^s$ and $(Z_{t+1}Z'_t)^s$ are close to their filtered counterparts, it follows that the smoothed version of (14) will also give approximately the OLS estimates of $K^{\mathbb{P}}_{0Z}$, and $\hat{K}^{\mathbb{P}}_{1Z}$. We turn to this assumption at the end and show why it is likely to hold.

As shown in Appendix C, when $\Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t}$ is small, convergence to steady state will be fast. As such we can treat $\Omega_{\mathcal{L}t}$ (and Ω_t) as if it were a constant matrix. Denoting $\Omega_t = var(Z^o_t | \mathcal{P}^{-\mathcal{L}o}_t, \mathcal{F}_{t-1})$ with $\mathcal{P}^{-\mathcal{L}o}_t$ being the $J - \mathcal{L}$ higher order PCs, post-multiplying both terms on the right hand side of (59) by $\begin{pmatrix} 1 & 0 \\ 0 & \Omega^{-1}_t \end{pmatrix}$ we have:

$$\left(\frac{1}{T}\sum_t Z^f_{t+1}, \frac{1}{T}\sum_t (Z_{t+1}Z'_t)^f \Omega^{-1}_t \right) \begin{pmatrix} 1 & \frac{1}{T}\sum_t Z^{f\prime}_t \Omega^{-1}_t \\ \frac{1}{T}\sum_t Z^f_t & \frac{1}{T}\sum_t (Z_t Z'_t)^f \Omega^{-1}_t \end{pmatrix}^{-1}. \quad (60)$$

Now,

$$\begin{aligned} (Z_t Z'_t)^f \Omega^{-1}_t =& var(Z_t|\mathcal{F}_t)\Omega^{-1}_t + Z^f_t (Z^f_t)'\Omega^{-1}_t \\ =& var(Z_t|\mathcal{F}_t)\Omega^{-1}_t + Z^o_t (Z^o_t)'\Omega^{-1}_t \end{aligned} \quad (61)$$

where the second line follows by using the result of Appendix B. Using block inversion we can show that the non-zero block of the first term is:

$$\Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t} - \Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t}\Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t}$$

which by our assumption must be close to zeros. Therefore we can replace the $\frac{1}{T}\sum_t (Z_t Z'_t)^f \Omega^{-1}_t$ term in (60) by $\frac{1}{T}\sum_t Z^o_t Z^{o\prime}_t \Omega^{-1}_t$. Using a similar argument, the $\frac{1}{T}\sum_t (Z_{t+1}Z'_t)^f \Omega^{-1}_t$ term can also be replaced by $\frac{1}{T}\sum_t Z^o_{t+1} Z^{o\prime}_t \Omega^{-1}_t$. Using the result of Appendix B, we can replace all Z^f_t in (60) by its observed counter-part:

$$\left(\frac{1}{T}\sum_t Z^o_{t+1}, \frac{1}{T}\sum_t Z^o_{t+1} Z^{o\prime}_t \Omega^{-1}_t \right) \begin{pmatrix} 1 & \frac{1}{T}\sum_t Z^{o\prime}_t \Omega^{-1}_t \\ \frac{1}{T}\sum_t Z^o_t & \frac{1}{T}\sum_t Z^o_t Z^{o\prime}_t \Omega^{-1}_t \end{pmatrix}^{-1}. \quad (62)$$

Additionally, assuming that $var_T(Z^o_t) var(Z^o_t | \mathcal{P}^{-\mathcal{L}o}_t, \mathcal{F}_{t-1})^{-1}$ is non-degenerate relative to $\Sigma_{\mathcal{L}e}\Omega^{-1}_{\mathcal{L}t}$, then all Ω_t's cancel out and (62) reduces to the the familiar OLS estimates.

E Conditional Variance Parameters

To be integrated:

$$f(\mathcal{P}_t^o|Z_t;\Theta^{\mathbb{Q}},\Sigma_e) = (2\pi)^{-J/2}|\Sigma_e|^{-1/2}\exp\left(-\frac{1}{2}e_t'\,\Sigma_e^{-1}\,e_t\right), \tag{63}$$

where e_t is given by (7) for the $MTSM$ and (8) for the factor-VAR. The conditional density $f(Z_t|Z_{t-1})$ is given by

$$f\left(Z_t|Z_{t-1};K_{1Z}^{\mathbb{P}},K_{0Z}^{\mathbb{P}},\Sigma\right) = (2\pi)^{-\mathcal{N}/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}i_t'\Sigma^{-1}i_t\right), \tag{64}$$

where $i_t = \Delta Z_t - \left(K_{0Z}^{\mathbb{P}} + K_{1Z}^{\mathbb{P}}Z_{t-1}\right)$, the time t innovation.

The term structure corresponding to our canonical form with the observable risk factors Z_t can be obtained by substituting (35) into (40):

$$y_t \;=\; A_X + B_X(W^{\mathcal{N}}B_X)^{-1}(\Gamma_1^{-1}(Z_t-\Gamma_0) - W^{\mathcal{N}}A_X). \tag{65}$$

From this we can write

$$\begin{aligned}
\mathcal{P}_t &= A_{TSM} + WB_X\mathcal{U}_1Z_t, \text{ where} && (66)\\
A_{TSM} &= \mathcal{G}\gamma_r + \beta_Z vec(\Sigma), && (67)\\
\mathcal{G} &= W\left((I_J - B_X(W^{\mathcal{N}}B_X)^{-1}W^{\mathcal{N}})\iota, B_X\mathcal{U}_{1,\mathcal{M}}\right), && (68)\\
\beta_Z &= W(I_J - B_X(W^{\mathcal{N}}B_X)^{-1}W^{\mathcal{N}})\beta_X(\mathcal{U}_1\otimes\mathcal{U}_1), && (69)
\end{aligned}$$

$\gamma_r' = (r_\infty^{\mathbb{Q}}, \gamma_0^{W\prime})$, and $\mathcal{U}_{1,\mathcal{M}}$ denotes the first $\mathcal{M}$ columns of $\mathcal{U}_1$. Importantly, $\mathcal{G}$ and $\mathcal{T}$ are only dependent on $\lambda^{\mathbb{Q}}$ and γ_1^W. Therefore, from (7), the errors in pricing $\mathcal{P}_t$ are given by

$$e_t = \mathcal{P}_t^o - \mathcal{G}\gamma_r - \beta_Z vec(\Sigma) - B_{GMTSM}Z_t. \tag{70}$$

****Integrate in this additional term takes the form

$$E\left[\mathcal{E}_T[\partial\log f_{TSM}(\mathcal{P}_t^o|Z_t^o;\hat{\Theta}_{TSM})/\partial vec(\Sigma)]\middle|\,\mathcal{F}_T\right] = \hat{\beta}_Z'(\hat{\Sigma}_e)^{-1}E\left[\mathcal{E}_T[\hat{e}_t^u]|\,\mathcal{F}_T\right], \tag{71}$$

where the unobserved pricing errors $\hat{e}_t^u$ from (7) are evaluated at the ML estimators and depend on the partially observed $\vec{Z}$.

F Concentration of the Likelihood

In this section, we show how to concentrate out the likelihood when $\mathcal{P}_t^{\mathcal{L}} = \mathcal{P}_t^{\mathcal{L},o}$. Since in this case Σ_e is singular, to simplify exposition and avoid the use of Moore-Penrose pseudo inverse, we will treat Σ_e as the covariance matrix of pricing errors for the remaining $J-\mathcal{L}$ portfolios priced with errors; and Σ_e is non-singular again. We denote the weighting matrix associated with the $J - \mathcal{L}$ portfolios priced with errors by W_e. Ignoring constants, the log likelihood of the data is:

$$-\frac{1}{2}\left(\log|\Sigma_e| + \log|\Sigma| + \mathcal{E}_T\left[e_t'\Sigma_e^{-1}e_t + i_t'\Sigma^{-1}i_t\right]\right) \tag{72}$$

where

$$\begin{aligned} e_t &= W_e y_t^o - \mathcal{G}_e\gamma_r - \beta_{Z,e}vec(\Sigma) - B_{GMTSM,e}Z_t, & (73)\\ i_t &= \Delta Z_t - \hat{K}_{0Z,OLS}^{\mathbb{P}} - \hat{K}_{1Z,OLS}^{\mathbb{P}}Z_{t-1}, & (74)\end{aligned}$$

where the subscript e indexes the rows of $\mathcal{G}$, β_Z, and B_{GMTSM} that correspond to the yields portfolios priced with errors. From the first order condition with respect to γ_r, we have:

$$\gamma_r = \underbrace{(\hat{\mathcal{G}}_e'(\hat{\Sigma}_e)^{-1}\hat{\mathcal{G}}_e)^{-1}\hat{\mathcal{G}}_e'(\hat{\Sigma}_e)^{-1}}_{\mathcal{H}}\mathcal{E}_T\left[W_e y_t^o - \beta_{Z,e}vec(\Sigma) - B_{GMTSM,e}Z_t\right]. \tag{75}$$

Turning to Σ, the first order condition is:

$$vec\left(\frac{1}{2}\left[(\hat{\Sigma}_Z)^{-1} - (\hat{\Sigma}_Z)^{-1}\hat{\Sigma}_{Z,OLS}(\hat{\Sigma}_Z)^{-1}\right]\right) - \hat{\beta}_{Z,e}'(\hat{\Sigma}_e)^{-1}\mathcal{E}_T[\hat{e}_t] = 0. \tag{76}$$

Given the expression for γ_r from (75), we observe that $\hat{\beta}_{Z,e}'(\hat{\Sigma}_e)^{-1}\mathcal{E}_T[\hat{e}_t]$ is linear in $vec(\Sigma)$:

$$\hat{\beta}_{Z,e}'(\hat{\Sigma}_e)^{-1}\mathcal{E}_T[\hat{e}_t] = \mathcal{H}_0 + \mathcal{H}_1 vec(\hat{\Sigma}) \tag{77}$$

where $\mathcal{H}_0 = \hat{\beta}_{Z,e}'(\hat{\Sigma}_e)^{-1}(I_{J-\mathcal{L}} - \mathcal{G}_e\mathcal{H})\mathcal{E}_T[W_e y_t^o - B_{GMTSM,e}Z_t]$ and $\mathcal{H}_1 = -\hat{\beta}_{Z,e}'(\hat{\Sigma}_e)^{-1}(I_{J-\mathcal{L}} - \mathcal{G}_e\mathcal{H})\beta_{Z,e}$. As such, we can write (76) as:

$$-(\hat{\Sigma}_Z)^{-1} + (\hat{\Sigma}_Z)^{-1}\hat{\Sigma}_{Z,OLS}(\hat{\Sigma}_Z)^{-1} + 2\,vec^{-1}(\mathcal{H}_0 + \mathcal{H}_1 vec(\hat{\Sigma})) = 0. \tag{78}$$

Let's denote the left hand side by $F(\Sigma)$. We obtain the Fréchet derivative of F with respect to Σ:

$$DF(\Sigma)(\epsilon) = \left[\Sigma^{-1}\epsilon\Sigma^{-1} - \Sigma^{-1}\epsilon\Sigma^{-1}\Sigma_{Z,OLS}\Sigma^{-1} - \Sigma^{-1}\Sigma_{Z,OLS}\Sigma^{-1}\epsilon\Sigma^{-1}\right] + 2\text{vec}^{-1}\left(\mathcal{H}_1\text{vec}\left(\epsilon\right)\right). \tag{79}$$

This gives us the linearization $F(\Sigma+\epsilon) = F(\Sigma) + DF(\Sigma)(\epsilon) + o(\|\epsilon\|_2)$ and provides the Newton update equation $\Sigma_{Z,n+1} = \Sigma_{Z,n} + \epsilon_n$ where ϵ_n solves $F(\Sigma_{Z,n}) + DF(\Sigma_{Z,n})(\epsilon_n) = 0$. This is easily solved using the vec operation:

$$\text{vec}\left(\epsilon_n\right) = E^{-1}\text{vec}\left(F(\Sigma_{Z,n})\right) \tag{80}$$

where

$$E = \left[\left(\Sigma_{Z,n}^{-1}\otimes\Sigma_{Z,n}^{-1}\right) - \left(\Sigma_{Z,n}^{-1}\otimes\Sigma_{Z,n}^{-1}\Sigma_{Z,OLS}\Sigma_{Z,n}^{-1}\right) - \left(\Sigma_{Z,n}^{-1}\Sigma_{Z,OLS}\Sigma_{Z,n}^{-1}\otimes\Sigma_{Z,n}^{-1}\right)\right] + 2\mathcal{H}_1. \tag{81}$$

Since $\hat{\Sigma}$ should be close to $\hat{\Sigma}_{Z,OLS}$, using this algorithm with $\Sigma_{Z,0} = \hat{\Sigma}_{Z,OLS}$ should provide near-instantaneous convergence. Finally, we note that depending on the assumed structure of Σ_e, this parameter can also be analytically concentrated out. For example, if we let $\Sigma_e = I\sigma_e^2$, an assumption we maintain throughout our empirical implementations, σ_e can also be concentrated out.

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.

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

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.

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.

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

Duffee, G., 2009, "Information In (and Not In) The Term Structure," Discussion paper, Working Paper, Johns Hopkins University.

Duffee, G., 2011, "Forecasting with the Term Structure: the Role of No-Arbitrage," Discussion paper, Working Paper, Johns Hopkins University.

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

Hamilton, J., and J. Wu, 2010, "Identification and Estimation of Affine Term Structure Models," Discussion paper, University of California, San Diego.

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.

Litterman, R., and J. Scheinkman, 1991, "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1, 54–61.

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, forthcoming, *American Economic Review.*