Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs

Scott Joslin* Anh Le† Kenneth J. Singleton‡

First draft: June, 2010
This draft: March 11, 2011

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. For the typical yield curves and macro variables studied in this literature, the estimated joint distribution within a canonical *MTSM* is nearly identical to the estimate from an economic-model-free factor vector-autoregression (factor-VAR), even when measurement errors are large. It follows 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 is learned 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 empirical studies of term premiums. These results are rotation-invariant and, therefore, apply to many of the specifications of risk factors in the literature. In deriving these results we develop a new canonical form for *MTSM*s that is particularly revealing about the nature of the over-identifying restrictions implied by *MTSM*s relative to yield-based factor models.

*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 presence of measurement errors on yields for the Kalman filter estimator of the joint distribution of these variables. We derive sufficient and easily verified theoretical conditions for the Kalman filter estimator of this distribution within a canonical[3] *MTSM* to be (nearly) identical to the ordinary least-squared (OLS) estimator of an unconstrained factor-VAR. We show that these conditions are very nearly satisfied by the canonical versions of several prominent *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 factor-VAR.

Our theoretical propositions focus on the *entire conditional distribution* of the risk factors and bond yields in models where *all bond yields are measured with errors* and so filtering must be used in estimation. Both of these ingredients are essential for exploring what *MTSM*s teach us about say the impulse responses (IRs) of bond yields to shocks to output or inflation,[4] or about expectations puzzles in bond markets.[5] The theoretical propositions and empirical illustrations about the role of no-arbitrage restrictions in Joslin, Singleton, and

[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] Recent analyses of IRs within *MTSM*s 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.

Zhu (2010) (JSZ) and Duffee (2011a) are largely silent on these issues, because they focus on the conditional means (forecasts) of yield-based risk factors that are priced perfectly, or nearly perfectly, by their models and focus exclusively on models which maintain a good cross-sectional fit to the yield curve. In contrast, in this paper we allow all of the individual yields to be priced imperfectly, possibly with large errors, and then examine whether the imposition of the structure of a *MTSM* affects features of the risk factors that depend on *both* the conditional mean and variance parameters (as do IRs and term premiums).

A major reason that the answers to these questions cannot be inferred from prior work on Gaussian models with latent or yield-based risk factors (*YTSM*s) is that measurement errors on bond yields are a nontrivial consideration in *MTSM*s. Filtering often has little effect on ML estimators in *YTSM*s, in large part because the standard deviations of these errors are typically small (only a few basis points).[6] In contrast, pricing errors on individual bond yields in *MTSM*s are often much larger, exceeding 100 basis points in some prominent *MTSM*s. Accordingly, we provide sufficient conditions for the Kalman filter estimator of a *MTSM* and the OLS estimator of its factor-VAR counterpart to produce (nearly) identical conditional distributions of the risk factors when there are pricing errors of this magnitude. A key condition is that the ratio of the average pricing errors to their standard deviations for the yield-based risk factors be approximately zero. Historical and *MTSM*-implied low-order PCs track each other very closely, even though the pricing errors on individual bonds are at times large, and this is what drives our empirical findings of irrelevance. Our propositions also provide a theoretical underpinning for the findings in JSZ and Duffee (2011b) that higher-order PCs are not accurately priced in five-factor *YTSM*s.

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 are the macro variables M_t and the remaining $\mathcal{L} = \mathcal{N} - \mathcal{M}$ risk factors are the first $\mathcal{L}$ principal components (PCs) of bond yields, $\mathcal{P}_t^{\mathcal{L}}$. This form provides an organizing framework within which it is easy to determine whether a *MTSM* is econometrically identified. Moreover it leads directly to a formal characterization of the added flexibility of a *MTSM* (relative to an $\mathcal{N}$-factor model with no observed macro risk factors) in terms of a theoretical spanning condition of M_t by the first $\mathcal{N}$ PCs of yields.

Using this canonical form we show that our irrelevancy propositions are fully rotation invariant:[7] if our sufficient conditions are satisfied, then all choices of individual yields or PCs of yields as elements of $\mathcal{P}_t^{\mathcal{L}}$ necessarily result in identical (inconsequential) effects of no-arbitrage restrictions. Moreover when $\mathcal{P}_t^{\mathcal{L}}$ is normalized to be $\mathcal{L}$ low-order PCs, then the model-implied joint distribution of $Z_t' \equiv (M_t', \mathcal{P}_t^{\mathcal{L}\prime})$ is virtually identical to the one implied by a standard unconstrained VAR model of the *observed* risk factors Z_t^o.

Initially we explore the empirical relevance of our propositions within a three-factor *MTSM*– model $GM_3(g, \pi)$– in which the risk factors are output growth, inflation, and the first

[6] This is documented in JSZ for estimates of the conditional mean parameters in *YTSM*s, and in Duffee (2011a) for the loadings that link the yield-based risk factors to the prices of individual bonds.

[7] See Dai and Singleton (2000) for the definition of invariant affine transformations. Such transformations lead to equivalent models in which the pricing factors $\tilde{\mathcal{P}}_t^{\mathcal{N}}$ 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.

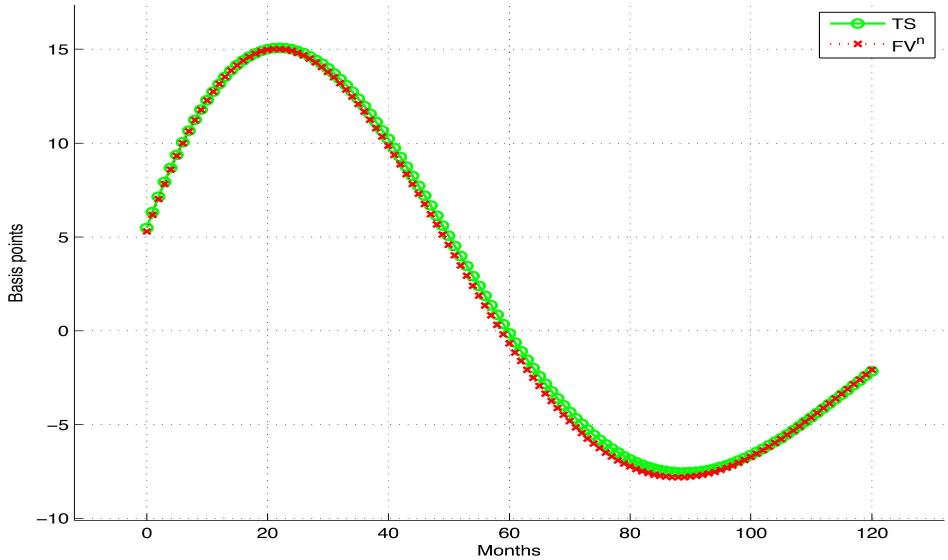


Figure 1: Impulse responses in basis points of $PC1$ to a shock to inflation in model $GM_3(g, \pi)$ (TS) and its corresponding factor-VAR (FV^n).

PC of bond yields ($PC1$).[8] The no-arbitrage structure of $GM_3(g, \pi)$ implies over-identifying restrictions on the distribution of bond yields, and its Kalman filter estimates imply root mean-squared pricing errors on the order of forty basis points (see Section 4.1). Nevertheless, the IRs of $PC1$ to a shock to CPI inflation implied by $GM_3(g, \pi)$ and by its corresponding factor-VAR (FV^n) are virtually indistinguishable (Figure 1). The analysis of $GM_3(g, \pi)$ is followed by an example of a four-factor model with $(g_t, \pi_t, PC1_t, PC2_t)$ as risk factors, and a model with unspanned macro risks in the sense of Joslin, Priebsch, and Singleton (2010).

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.

Certain types of restrictions, when imposed in combination with the no-arbitrage restrictions of a $MTSM$, may overturn our irrelevancy results and 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.[9] In our concluding section

[8] Full details of the data and estimation results are provided in Section 4.3. Unless otherwise noted, the loadings for $PC1$ are rescaled so that they add up to one.

[9] JSZ and Duffee (2011a) explore empirically whether various constraints on the $\mathbb{P}$ distribution of the risk

we draw upon our 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 $(m_1, \ldots, m_J)$ in periods and with $J \geq \mathcal{N}$, where $\mathcal{N}$ is the number of pricing factors. To be consistent with our empirical work, we fix the period length to be one month. 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. Throughout, we assume a flat prior on the initial observed data.

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 $\mathcal{P}_t^{\mathcal{L}}$,

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

where the risk factors are $Z_t = (M_t', \mathcal{P}_t^{\mathcal{L}\prime})'$. Some treat $\mathcal{P}_t^{\mathcal{L}}$ in (1) as a set of $\mathcal{L}$ latent risk factors,[10] while others include portfolios of yields as risk factors.[11] 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[12] 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,

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

factors in *YTSM*s improve out-of-sample forecasts of these factors. We look beyond their focus on conditional means and perfectly priced risk factors to the new issues that arise in *MTSM*s.

[10]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).

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

[12]Throughout, we will refer to an affine transformation of the state variable as in Dai and Singleton (2000) as a rotation. See Appendix C for details of these transformations.

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

$$\mathcal{P}_t = A_{TS} + B_{TS} Z_t, \tag{3}$$

where the loadings (A_{TS}, B_{TS}) are known functions of the parameters $(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_0, \rho_1)$ governing the risk neutral distribution of yields, and hereafter "TS" denotes 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 $\mathcal{P}_t^{\mathcal{N}}$.[13] After rotating to a pricing model with risk factors $\mathcal{P}_t^{\mathcal{N}}$, we adopt the canonical form of JSZ. What is distinctive about their canonical form is that the risk-neutral distribution of $\mathcal{P}_t^{\mathcal{N}}$ 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 $K_1^{\mathbb{Q}}$.[14]

A 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 "state yield portfolios" $\mathcal{P}_t^{\mathcal{L}}$ (the first $\mathcal{L}$ components of $\mathcal{P}_t^{\mathcal{N}}$); 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 $\Theta_{TS}^{\mathbb{Q}} \equiv (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma)$.

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

[13] 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.

[14] 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.

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 $(\lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}}, K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma)$. 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 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 yield portfolios $\mathcal{P}_t^o$ are equal to their 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.[15] The observation equation is then (3) adjusted for these errors:

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

and the state equation is (6). Here we consider (A_{TS}, B_{TS}) as functions of the parameters $\Theta_{TS}^{\mathbb{Q}}$ of our normalization. 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 (2011a) 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

[15] 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, $\mathcal{P}_t^{\mathcal{L}} = \mathcal{P}_t^{\mathcal{L}o}$.

restrictions.[16] 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 state yield portfolios are measured perfectly. We distinguish these special cases by the notation TS^n and FV^n (for *no* pricing errors on the risk factors).

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 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 state yield portfolios are measured without errors ($\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$ in model FV^n), while again maintaining the low-dimensional factor structure.

2.3 Discussion

A key feature of our normalization is that it imposes "pricing consistency" in the sense that the state yield portfolios recovered from the pricing equation (3) always agree with their theoretical values. Ang, Piazzesi, and Wei (2006) and Jardet, Monfort, and Pegoraro (2010) enforce pricing consistency by minimizing sums of squared pricing errors subject to a consistency constraint. Their approach requires that their state yield portfolios are priced perfectly by the *MTSM*, and their two-step estimation strategy is asymptotically inefficient. In this section we show our choice of canonical form automatically enforces pricing consistency even when all bonds are priced imperfectly by the *MTSM* and, accordingly, Kalman filter estimators are fully efficient.

Equally importantly, our canonical forms for the TS and FV models are invariant with respect to the modeler's choice of W. That is, all admissible choices of W– e.g., choices that set the state yield factors to individual yields or to low-order PCs of bond yields– lead to exactly the same Kalman filter estimates of the parameters of the joint distribution of (y_t^o, M_t^o). In fact, so long as one enforces the model-implied spanning condition (4), representations of model TS in which the risk factors are all yield portfolios (e.g., $Z_t = \mathcal{P}_t^{\mathcal{N}}$) or the mix $(M_t, \mathcal{P}_t^{\mathcal{L}})$ of macro and yield-based factors lead to identical fitted moments of (y_t^o, M_t^o) regardless of the choice of admissible W.

The remainder of this section discusses each of these points in turn.

[16] 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.

Pricing Consistency

To illustrate the consistency issue, consider the *MTSM* with a single macro variable ($\mathcal{M} = 1$), and two pricing factor ($\mathcal{L} = 2$) with W chosen so that the two state yield portfolios are the short rate and the two-year (twenty-four month) rate: $Z_t = (M_t, r_t, y_t^{24})$. Pricing consistency requires that when one computes the loadings for the two-year yield from (3) by solving the recurrence relation given in Appendix B, it must be that the intercept is 0 and the loadings on Z_t are $(0, 0, 1)$. The two-year rate, up to convexity, is the average of expected future short rates. Since our model is Gaussian, the convexity term is constant. Thus, for a monthly sampling frequency, we require

$$y_t^{24} = \frac{1}{24} E_t^{\mathbb{Q}} \left[\sum_{\tau=0}^{23} r_{t+\tau} \right] + \text{constant}. \tag{9}$$

The $\mathbb{Q}$-expectations in (9) can be computed according to the dynamics in (2) which give

$$E_t^{\mathbb{Q}}[r_{t+\tau}] = (0,1,0)E_t^{\mathbb{Q}}[Z_{t+\tau}] = (0,1,0)(I + K_1^{\mathbb{Q}})^{\tau} Z_t + \text{constant}.$$

Thus pricing consistency– the requirement that the loadings on Z_t be $(0, 0, 1)$– imposes non-linear restrictions on the $\mathbb{Q}$ parameters $K_1^{\mathbb{Q}}$ and ρ_1. Analysis of the constant term leads to additional nonlinear restrictions on the parameters $(K_0^{\mathbb{Q}}, \Sigma, \rho_0)$.

We specify the $\mathbb{Q}$ distribution in terms of the primitive parameters $\Theta_{TS}^{\mathbb{Q}}$. As such, the associated mapping from $\Theta_{TS}^{\mathbb{Q}}$ to the loadings on Z_t in the observation equation (7) automatically embeds these nonlinear constraints, thereby ensuring that pricing consistency always holds exactly.

Invariance of the theoretical model

Changing from one choice of the weight matrix W to another W^* has no impact on the distribution of the theoretical yields or macro-variables in a *MTSM* when the parameters are transformed appropriately. That is, consider the TS model and fix a portfolio matrix W and parameter vector $\Theta_{TS}(W) = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma, K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$. ($\Sigma_e$ has no role in this discussion.) For any other admissible weighting matrix W^*, the TS model with parameter vector $\Theta_{TS}^*(W^*) = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0^*, \gamma_1^*, \Sigma^*, K_0^{\mathbb{P}*}, K_1^{\mathbb{P}*})$, where for example $\gamma_1^* = \gamma_1 (W^{*\mathcal{N}} W^{-1} B_{TS})^{-1}$, implies exactly the same joint distribution for (M_t, y_t). Thus, the choice of weighting matrix can be based solely on what is convenient for the modeler. This analysis hold equally well for the FV model, provided one assumes the weighting matrix maintains non-singularity among the state yield portfolios.

Our framework and its invariance property extend immediately to the case where the risk factors are linear combinations of both the yields and macro variables. That is, we can recast our entire analysis in terms of the first $\mathcal{N}$ elements of the vector $\widetilde{W}(M_t', y_t')'$, where $\widetilde{W}$ is a full-rank $(\mathcal{M} + J) \times (\mathcal{M} + J)$ matrix. Our chosen normalization is the special case in which $\widetilde{W}$ is block diagonal with the first diagonal block being the $\mathcal{M} \times \mathcal{M}$ identity matrix and the second diagonal block being W. Exactly as above, any other canonical form based on a different choice $\widetilde{W}^*$ can be re-expressed in terms of our canonical form. Thus, once again, the joint distribution of (M_t, y_t) is not affected by the modeler's choice of $\widetilde{W}$.

Invariance with measurement errors

These invariance results also apply equally to models in which all of the bond yields are priced with errors. So long as the measurement error variance Σ_e for a TS model based on yield weights W is transformed to $\Sigma_e^* = A\Sigma_e A'$ when this model is reparametrized in terms of the weights $W^* = AW$, Kalman filtering will produce *identical* fitted distributions for (y_t^o, M_t^o). Thus, canonical models based on different choices of W give rise to observationally equivalent representations of bond yields. The same is true for model FV. Thus, comparisons between models TS and FV are fully invariant to the modeler's choice of W.

This invariance with respect to the choice of W carries over also to the case where restrictions are placed on a canonical model, provided that the restrictions are properly adjusted when rotating to risk factors based on a different W^*. For example, a common assumption in the literature is that the measurement errors are independent and of equal variance: $\Sigma_e = \sigma_e^2 I$. This form would be preserved by any orthogonal re-weighting matrix A. For example, if in one model $W = I_J$, so that the portfolios are individual yields, and in the second model W^* is given by the loadings of the yield PCs (an orthogonal matrix), then identical Kalman filter estimates will be obtained for the distribution of (y_t^o, M_t^o). For a general re-weighting A, identical estimates are obtained so long as Σ_e is replaced by $\sigma_e^2 AA'$.

In contrast, comparisons between model TS (with full rank Σ_e) and the associated model TS^n (with $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$) will depend on the modeler's choice of W. This is simply a consequence of the fact that the composition of $\mathcal{P}_t^{\mathcal{L}}$ depends on W. Assuming that individual bond yields are measured perfectly, as for example in Ang, Piazzesi, and Wei (2006), may lead to a very different impulse response of say the ten-year bond yield to an inflation shock than the corresponding impulse response obtained from a model in which PCs are assumed to be measured perfectly. We illustrate the practical implications of this point in Section 4.

The same logic of observational equivalence applies to the standard assumption that the macro-variables are observed without errors. The estimates of the joint distribution of (M_t^o, y_t^o) under this assumption will in general differ from those obtained when M_t is presumed to be measured with error. On the other hand, when both (M_t, y_t) are observed with measurement errors, observationally equivalent models will be obtained for arbitrary choices of the W used to construct $\mathcal{P}_t^{\mathcal{L}}$, so long as the joint distribution of the measurement errors for the yields and macro variables is properly matched to one's choice of W.

Verifying econometric identification and pricing consistency in practice

Verification that one has a well-specified *MTSM* is greatly facilitated by specifying a canonical form and then, within this form, imposing sufficient normalizations and restrictions to ensure econometric identification and internal (pricing) consistency. Instead, many studies of *MTSM*s proceed by imposing a mix of zero restrictions on the $\mathbb{P}$, $\mathbb{Q}$, and market price of risk parameters without explicitly mapping their models into a canonical form and verifying sufficient conditions for identification.[17]

Our canonical form reveals that a necessary "order" condition for identification is that the dimension of our Θ_{TS} (excluding Σ_e)– $1 + 2\mathcal{N} + \mathcal{N}^2 + \mathcal{M}(\mathcal{N}+1) + \mathcal{N}(\mathcal{N}+1)/2$– must be at least as large as the number of free parameters in any *MTSM* with $\mathcal{N}$ risk factors $\mathcal{M}$ of which are macro variables. It also leads to an easily imposed set of normalizations that ensure identification and pricing consistency. To our knowledge, ours is the only formally developed canonical form for the complete family of *MTSM*s.[18]

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 models TS and FV to be (nearly) identical to those implied by the FV^n model. Estimation of model FV^n conveniently reduces to two sets of OLS regressions: a VAR for the observed risk factors Z_t^o gives the parameters in (6), and an linear projection of $\mathcal{P}_t^o$ onto Z_t^o recovers the parameters characterizing (8).[19]

Importantly, as long as there exists one W^* such that the conditions that we derive 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 the fact that bilateral comparisons of the models (TS, FV^n) or the models (FV, FV^n) are rotation–dependent. Equally importantly, for such a W^*, everything that one can learn about the $\mathbb{P}$ distribution of Z_t from a canonical *MTSM* in which all bonds are measured with errors can be equally learned from analysis of the corresponding economics-free factor-VAR model FV^n in which $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$.

The filtering problem 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 matching is not sufficient for the Kalman filter estimates of the drift or the volatility of Z_t to match their OLS counterparts from model FV^n. The remainder of this section derives sufficient conditions for the efficient estimates of models TS and FV^n to (nearly) coincide.

To fix the notation, let $X_t^f = E[X_t|\mathcal{F}_t]$ and $X_t^s = E[X_t|\mathcal{F}_T]$ denote the filtered and smoothed version of any random variable X_t, where $\mathcal{F}_t$ is the observable information known at time t: $(y_1^o, M_1^o, \ldots, y_t^o, M_t^o)$.

---

[17] Recent examples include the *MTSM*s examined by Bikbov and Chernov (2010) and the constant parameter case in Ang, Boivin, Dong, and Loo-Kung (2010). The following necessary condition for identification suggests that the first of these models is in fact under-identified, while the second may be over-identified. Neither study verifies identification within a canonical form.

[18] Pericoli and Taboga (2008) attempt an adaptation of the canonical form for yield-only models in Dai and Singleton (2000) to *MTSM*s, but their forms are not identified models (Hamilton and Wu (2010)).

[19] This follows immediately from concentrating the likelihood function; see (14) below.

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.[20] The sizes of these errors depend on: (i) the magnitudes of the measurement errors on the yields; and (ii) the accuracy with which the yield portfolios can be forecasted based on current and lagged observables, excluding the current yields themselves.

The difference between the observed and filtered states, the filtered observation error

$$FOE_t \equiv E[e_t|\mathcal{F}_t] = E[\mathcal{P}_t - \mathcal{P}_t^o|\mathcal{F}_t] = \mathcal{P}_t^f - \mathcal{P}_t^o,$$

is an $\mathcal{F}_t$-measurable random variable. Consider the subvector $FOE_t^{\mathcal{L}}$ of the first $\mathcal{L}$ elements of FOE_t, and define the information set

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

where $\mathcal{P}_t^{-\mathcal{L}o}$ denotes the last $J - \mathcal{L}$ of the observed yield portfolios $\mathcal{P}_t^o$. Conditional on I_t, $e_t^{\mathcal{L}}$ and $\mathcal{P}_t^{\mathcal{L}o}$ are jointly normal and, therefore,[21]

$$FOE_t^{\mathcal{L}} = E[e_t^{\mathcal{L}}|I_t, \mathcal{P}_t^{\mathcal{L}o}] = \Sigma_{e\mathcal{L}} S_t^{-1}(\mathcal{P}_t^{\mathcal{L}o} - E[\mathcal{P}_t^{\mathcal{L}o}|I_t]), \tag{10}$$

where $\Sigma_{e\mathcal{L}}$ is the covariance matrix of $e_t^{\mathcal{L}}$ and $S_t = Var(\mathcal{P}_t^{\mathcal{L}o}|I_t)$ is the forecast-error variance of $\mathcal{P}_t^{\mathcal{L}o}$ based on I_t.

Equation (10) shows how the Kalman filter computes the filtered observation error as a projection on current and lagged information. To assess the magnitude of this filtered measurement error, define the filtering root mean squared error, $RMSFE_t$, by

$$RMSFE_t^2 \equiv E\left[(FOE_t)^2\right], \tag{11}$$

where we consider $RMSFE_t$ a J-dimensional vector and the square and square-root are defined element-by-element. To the extent that $RMSFE_t$ is small, the filtered yield portofolios and the observed yield portfolios will closely agree on average. Substituting (10) into (11), we see that $RMSFE_t^{\mathcal{L}} = diag\left[\Sigma_{e\mathcal{L}} S_t^{-1} \Sigma_{e\mathcal{L}}\right]$.

The "diversification" effect for portfolios of measurement errors on bond yields

$\Sigma_{e\mathcal{L}}$ is determined by the pricing errors on individual yields, the correlations among these errors, and the choice of W. The diversification effect from constructing $\mathcal{P}_t = W y_t$ will typically lead to diagonal elements of $\Sigma_{e\mathcal{L}}$ that are smaller than the corresponding RMSEs for individual yields. For example, if the individual yield errors are cross-sectionally independent and if the first row of W weights yields equally (corresponding to a level factor), then the RMSE will be reduced by a factor of $1/\sqrt{J}$.[22] Owing to this averaging effect, even if individual bonds are priced with sizable errors, the elements of $\Sigma_{e\mathcal{L}}$ can still be relatively small.

[20] For simplicity, we focus here on filtering. The same arguments with appropriate modifications apply to the smoothed yield portfolios as well.

[21] 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. Here $X = e_t^{\mathcal{L}}$ and $Y = \mathcal{P}_t^{\mathcal{L}o}$, and Σ_{XY} is simply the variance of the errors by independence.

[22] Typically PCs are normalized so that the sum of the squares of the weights is one. This condition also ensures the observational equivalence of Section 2.3 if one supposes that the individual yield measurement

The relative size of the forecast error variance

S_t reflects the uncertainty about $\mathcal{P}_t^{\mathcal{L}o}$ given past realizations of the yield curve and macro variables and the current information $(M_t, \mathcal{P}_t^{-\mathcal{L}o})$. Consider the case that the measurement errors are uncorrelated. Observe that $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}} + e_t^{\mathcal{L}}$, the pair $(\mathcal{P}_t^{\mathcal{L}}, e_t^{\mathcal{L}})$ are independent, and $e_t^{\mathcal{L}}$ is independent of I_t. It follows that S_t is always at least as large as Σ_e (that is, $S_t - \Sigma_e$ is positive semi-definite). That is, even if the theoretical state yield portfolios were perfectly forecastable based on I_t, it would still be the case that we would have a forecast variance of $\Sigma_{e\mathcal{L}}$ when forecasting $\mathcal{P}_t^{\mathcal{L}o}$ because the measurement errors cannot be forecasted.

Approximation errors in practice

These observations lead to (rough) average magnitudes of the differences between the filtered and observed states. For example, if all yields are observed with *i.i.d.* measurement errors of equal variance σ_y^2 and there is a single state portfolio ($\mathcal{L} = 1$) which is a level factor with equal weights $(1/J)$, then

$$RMSFE_t = \frac{\sigma_y}{J} \times \frac{\sigma_y}{\sqrt{S_t}}. \quad (12)$$

If, for example, $\sigma_y = 10$ basis points, the forecast errors are on the order of 20 basis points, and there are $J = 10$ yields use in the estimation, then $RMSFE_t$ would be about half a basis point. Quadrupling σ_y to 40 basis points, and increasing $\sqrt{S_t}$ to 50 basis points, holding J at 10, increases $RMSFE_t$ to only about two and one-half basis points.

We see then that when the measurement errors for the portfolios are small, the filtered and observed states will track each other closely. In particular, 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. Furthermore, S_t will be much larger than $\Sigma_{e\mathcal{L}}$ when there is substantial uncertainty about $\mathcal{P}_t^{\mathcal{L}o}$ based on the information in I_t. This uncertainty is likely to rise as the sampling frequency decreases.

Thus $FOE_t^{\mathcal{L}}$ will tend to decline when W is chosen so (i) that there is cancelation of measurement errors across maturities, (ii) more cross-sectional information is used in estimation, and (iii) the variance of the error in forecasting $\mathcal{P}_t^{\mathcal{L}o}$ based on I_t is large. This dependence of $FOE_t^{\mathcal{L}}$ on W means that, for a given model, some choices of W may imply that $\mathcal{P}_t^{\mathcal{L}o} \approx \mathcal{P}_t^{\mathcal{L}f}$, while for other choices the differences may be large. Choices of W that select individual yields are inherently handicapped in this regard, because they forego the diversification benefits of nontrivial portfolios.

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 $YTSM$s studied by Duffee (2011b) and JSZ may be attributable to the smaller forecast-error variances of the higher-order PCs. Under the typical assumption of *i.i.d.* measurement errors and normalized loadings,

errors are independent with equal variances. For ease of interpretation, it is convenient to rescale the PCs so that the sum of the weights is equal to one for the first PC. This rescaling gives an observationally equivalent model with the adjusted Σ_e.

the measurement error variances are the same for all PCs. However, the sample standard deviations of the fourth and fifth PCs, about 19 and 13 basis points respectively for our data, are much smaller than those for the first three PCs. Since the forecast-error variances of the fourth and fifth PCs must be smaller than their respective unconditional variances, it is likely that the elements of $\Sigma_{\mathcal{L}e} S_t^{-1}$ corresponding to these PCs are relatively large. Whence, the Kalman filter will tend to emphasize measurement error reduction, by smoothing 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)

With sufficient conditions for $\mathcal{P}_t^{\mathcal{L}o} \approx \mathcal{P}_t^{\mathcal{L}f}$ in hand, we turn next to establishing sufficient conditions for the Kalman filter estimators of models TS and FV to (nearly) coincide. For either of the models TS or FV, 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 integrating the joint density $f_m^{\mathbb{P}}(\vec{Z} = z, \vec{\mathcal{P}}^o, \vec{M}^o; \Theta_m)$ over the missing data $\vec{Z}$:

$$f_m^{\mathbb{P}}(\vec{\mathcal{P}}^o, \vec{M}^o; \Theta_m) = \int_z f_m^{\mathbb{P}}(\vec{Z} = z, \vec{\mathcal{P}}^o, \vec{M}^o; \Theta_m) dz, \tag{13}$$

for $m = TS$ or FV, with $\vec{X}$ denoting the full 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.

The density $\log f^{\mathbb{P}}(\vec{Z}, \vec{\mathcal{P}}^o, \vec{M}^o)$ in (13) is equal to

$$\sum_{t=1}^{T} \log f^{\mathbb{P}} \left(\mathcal{P}_t^o | Z_t; \Theta^{\mathbb{Q}}, \Sigma_e \right) + \sum_{t=1}^{T} \log f^{\mathbb{P}} \left(Z_t | Z_{t-1}; K_1^{\mathbb{P}}, K_0^{\mathbb{P}}, \Sigma \right). \tag{14}$$

This construction reveals that the conditional distribution of the risk factors Z_t 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 $f^{\mathbb{P}}(\mathcal{P}_t^o|Z_t)$. This shared property of the null model TS and the alternative model FV is immediately apparent in our canonical form, while being largely obscured in the standard identification schemes of *MTSM*s such as the one 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.

Taking the derivative of (13) 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[23]

$$0 = E\left[\partial_\Theta \log f^{\mathbb{P}}(\vec{Z}, \vec{\mathcal{P}}^o, \vec{M}^o; \hat{\Theta}) \Big| \mathcal{F}_T \right], \tag{15}$$

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

where T is the sample size and $\mathcal{F}_T$ is all of the observable information.[24] 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

$$[\hat{K}_0^{\mathbb{P}}, \hat{K}_1^{\mathbb{P}}]' = \left(\left(\tilde{Z}'\tilde{Z}\right)^{\mathrm{s}}\right)^{-1}\left(\tilde{Z}'\Delta Z\right)^{\mathrm{s}}, \tag{16}$$

where the "hats" indicate ML estimators, $\tilde{Z}_t = [1, Z_t']'$, 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 (16) 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 FV^n 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. Equation (16) 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}. Underlying the difference between (16) and the latter estimator is the fact that

$$(Z_tZ_t')^{\mathrm{s}} = \mathrm{Var}(Z_t|\mathcal{F}_T) + Z_t^{\mathrm{s}}Z_t^{\mathrm{s}\prime}. \tag{17}$$

This equation and the analogous extensions to $(Z_tZ_{t+1}')^{\mathrm{s}}$ 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 FV^n will be similar to those obtained by Kalman filtering within a *MTSM*. In Section 3.1 we have seen conditions where the filtered and observed state yield portfolios agree. In Appendix D, we show that these same conditions (with a few mild additional assumptions) imply that $\mathrm{Var}(Z_t|\mathcal{F}_T)$ is small as well. As with the approximation $\mathcal{P}_t^{\mathcal{L}o} \approx \mathcal{P}_t^{\mathcal{L}f}$, the near equality of the ML estimates of $[K_0^{\mathbb{P}}, K_1^{\mathbb{P}}]$ may arise even in the presence of large pricing errors on the individual bond yields.

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

$$E\left[vec\left((\hat{\Sigma}_{FV})^{-1} - (\hat{\Sigma}_{FV})^{-1}\hat{\Sigma}_{FV}^u(\hat{\Sigma}_{FV})^{-1}\right)\middle|\, \mathcal{F}_T\right] = 0, \tag{18}$$

where the sample covariance matrix $\hat{\Sigma}_{FV}^u$ is based on the residuals $\hat{i}_{FV,t}^u = \Delta Z_t - (\hat{K}_{0FV}^{\mathbb{P}} + \hat{K}_{1FV}^{\mathbb{P}}Z_{t-1})$ that are partially observed owing to their dependence on $\vec{Z}$. From (18), we obtain $\hat{\Sigma}_{FV} = (\hat{\Sigma}_{FV}^u)^{\mathrm{s}}$. Using the logic of our discussion of the conditional mean, as long as the estimated model FV accurately prices the risk factors, then $(\Sigma_{FV}^u)^{\mathrm{s}}$ will be nearly identical to the OLS estimator of Σ from the VAR model FV^n.

The ML estimator of Σ in model TS will in general be more efficient than in model FV^n and this is true even when there is no measurement error in the state yield portfolios. The first-order conditions for Σ in model TS have an additional term since the density $f_{TS}^{\mathbb{P}}(\mathcal{P}_t^o|Z_t;\Theta)$ also depends on Σ. Combining this term, derived in Appendix E as (A53), with (18) gives

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

[24]In model FV^n with our choice of W, $Z_t = Z_t^o$ and (15) holds without the conditional expectation.

where $\hat{\Sigma}^u_{TS}$ is the sample covariance of the residuals $\hat{i}^u_{TS,t} = \Delta Z_t - (\hat{K}^{\mathbb{P}}_{0TS} + \hat{K}^{\mathbb{P}}_{1TS} Z_t)$, $\hat{\beta}_Z$ is a the 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

$$\hat{\Sigma}_{FV} = \left(\hat{\Sigma}^u_{FV}\right)^{\mathrm{s}} \approx \hat{\Sigma}_{FV^n}. \tag{19}$$

To guarantee that the right hand side of (19) 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 $\hat{\Sigma}_{TS} \approx (\hat{\Sigma}^u_{FV})^{\mathrm{s}}$ and, hence, that the estimators from all three models TS, FV, and FV^n approximately agree with each other.

3.3 Discussion

Summarizing, we have just shown that the same conditions derived in Section 3.1 for $\mathcal{P}^{\mathcal{L}o}_t \approx \mathcal{P}^{\mathcal{L}f}_t$ also ensure that the ML estimators of the conditional mean parameters of the state process Z_t approximately coincide for all three models TS, FV, and FV^n. When, in addition, the sample average of the fitted pricing errors for $\mathcal{P}^o_t$, $T^{-1}\sum \hat{e}^o_t$, is small relative to the estimated covariance matrix $\hat{\Sigma}_e$ of these errors, the ML estimates of the conditional variance Σ of Z_t will also approximately coincide in these models.

These observations regarding the conditional distribution of 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. By the same reasoning as above, if $\mathcal{P}^{\mathcal{L}}_t$ is reasonably accurately priced, the OLS loadings are likely to be close to those implied by model FV.[25] Nevertheless, large errors 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.

Further intuition for our results comes from exploring two restrictive special cases: the state yield portfolios are observed without measurement error in the *MTSM* ($\mathcal{P}^{\mathcal{L}o}_t = \mathcal{P}^{\mathcal{L}}_t$) 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.

[25] 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:

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

which should be close to the loadings from projecting $\mathcal{P}^o_t$ on Z^o_t if $\mathcal{P}^{\mathcal{L}o}_t$ is accurately priced. Within the context of *YTSM*s in which $\mathcal{P}^{\mathcal{N}}_t$ is priced perfectly and measurement errors on yields are relatively small, Duffee (2011a) documents this point using Monte Carlo methods.

A stark version of our results is obtained under the assumption $\mathcal{P}_t^{\mathcal{L}o} = \mathcal{P}_t^{\mathcal{L}}$, in which case the relevant comparison is between models TS^n and FV^n. With exact pricing of $\mathcal{P}_t^{\mathcal{L}o}$, the *ML* estimates of $(K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$ from model TS^n exactly coincide with the *OLS* estimates from model FV^n, regardless of the magnitude of the mean-to-variance ratios of pricing errors.[26] 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 ratio $(\hat{\Sigma}_e)^{-1}(T^{-1}\sum \hat{e}^o_{TS,t})$ 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 $\hat{e}^o_{TS,t}$ is exactly equal to zero comes from Duffee (2011a)'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, and the mean of $\hat{e}^o_{TS,t}$ 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.

The same condition $(J = \mathcal{N}+1)$ in an $\mathcal{N}$-factor *MTSM* with $\mathcal{M}$ macro factors guarantees that $\hat{\Sigma}$ 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 $\hat{e}^o_{TS,t}$ 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.

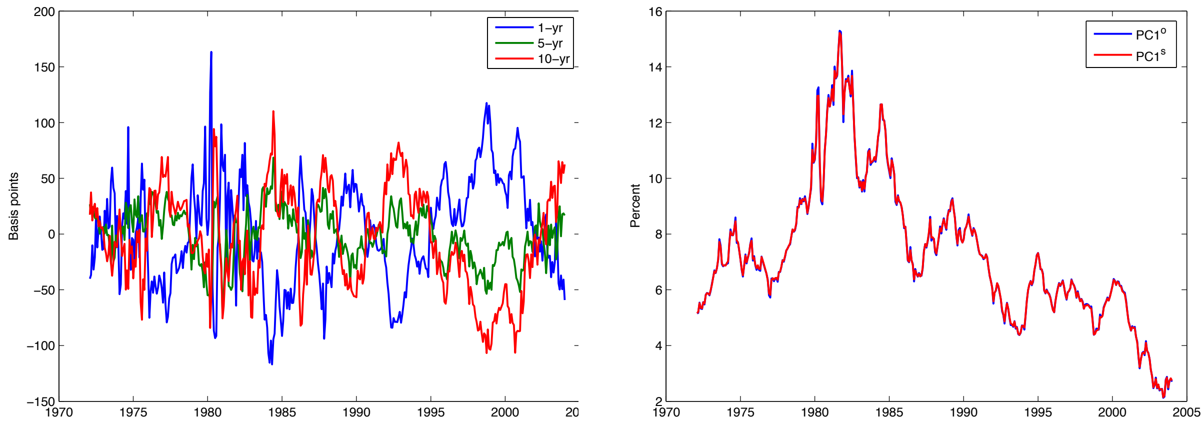
The first-order conditions of the *ML* estimators in our general setup (an over-identified *MTSM* with $J > \mathcal{N}+1$ imperfectly priced bond portfolios) do not set the sample mean of the pricing error $\hat{e}^u_{TS,t}$ to zero. However, it is easily verified that the first-order conditions with 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. 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.

4 Empirical (Near) Equivalence of *MTSM*s and VARs

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

We first focus on a *MTSM*– model $GM_3(g,\pi)$– with $\mathcal{N}=3$, $\mathcal{M}=2$, and $M_t = (g_t, \pi_t)'$, 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

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



(a) Comparison of individual yields

(b) Comparison of $PC1^o$ and $PC1^s$

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 $(y_t^m)^{\mathrm{s}}$. Panel (b) plots the observed $PC1^o$ and its smoothed version $PC1^{\mathrm{s}}$.

g, and the first PC of measures of inflation based on the CPI, the PPI of finished goods, and the spot market commodity prices ($INFPC$) for π.[27] The monthly zero yields are the unsmoothed Fama-Bliss series for maturities three- and six-months, and one through ten years over the sample period 1972 through 2003. The weighting matrix W is chosen to be the principal component loadings so that the state yield portfolio is the level of interest rates (PC1).[28] All bond yields are assumed be measured with *i.i.d.* Normal$(0, \sigma_y^2 I_{12})$ errors.

4.1 On the Need For Filtering PCs Within Canonical $\boldsymbol{MTSM}$s

A key part of our derivation of conditions under which the filtered versions of the state yield portfolios agree with their observed counterparts was the "diversification" effect of averaging the errors across maturities. Even when individual yields are very noisy with large measurement errors, the yield portfolios can be measured much more precisely.

Panel (a) of Figure 2 plots the time series of the differences between observed (annualized) yields, y_t^{mo}, and their smoothed counterparts $(y_t^m)^{\mathrm{s}}$, for m=12, 60, and 120 months. These pricing errors are large, occasionally exceeding 100 basis points, so this model clearly has difficulty matching *individual yields*. The Kalman filter estimate $\hat{\sigma}_y$ is 43.1 basis points. The reason for this poor fit is that the macro variables (g_t, π_t) replicate only a small portion of the variation in the slope and curvature of the yield curve.

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

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

Although the individual yields are poorly fit by the model, the model provides an excellent fit to $PC1$. Panel (b) of Figure 2 plots $PC1_t^o$ against the smoothed $PC1_t^s$. The sample standard deviation of the difference $\{PC1_t^o - PC1_t^s\}$ is only 1.7 basis points; the standard deviation of $\{PC1_t^o - PC1_t^f\}$ is only 4.3 basis points.

These numbers are fully anticipated by our theory in Section 3.1. Consider again the filtered observation error (10) and the expression (12) for the associated filtering root mean squared error $RMSFE$. For model $GM_3(g,\pi)$ the standard deviation $\sqrt{S_t} = \sqrt{Var(PC1_t^o|I_t)}$ is 40.7 basis points.[29] The estimated standard deviation of the measurement error on $PC1^o$ is 12.5 basis points, which approximately equals $\hat{\sigma}_y$ (43.1 basis points) divided by the square root of the number of yields used in estimation ($J = 12$). According to (12), then, the estimated standard deviation of $\{PC1_t^o - PC1_t^f\}$ is 3.8 basis points, close to the sample value of 4.3 basis points.

4.2 ML estimation of the conditional distribution

Section 3.2 arrived at several conditions for the Kalman filter estimator of model TS and the ML estimator of the factor-VAR FV^n to produce (nearly) identical fitted distributions of $(M_t, \mathcal{P}_t^{\mathcal{L}})$. They were that (i) $\mathcal{P}_t^{\mathcal{L}s}$ tracks $\mathcal{P}_t^{\mathcal{L}o}$ closely; (ii) there is a low amount of uncertainty about the (unobserved) theoretical $\mathcal{P}_t^{\mathcal{L}}$; and (iii) the time series average of the measurement errors, relative to their variances, should be small for the higher order portfolios $\mathcal{P}^{-\mathcal{L}o}$.

We have just seen that the first two of these conditions are satisfied at the Kalman filter estimates of model $GM_3(g,\pi)$. Intuitively, the second condition follows from the first and, indeed, the estimates indicate that the square root of $\text{Var}(\mathcal{P}_t^1|\mathcal{F}_t)$ is only 11.4 basis points. The final condition for equivalence is that the time series average of the measurement errors (relative to their variances) are small. 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 averages are small. In fact, for $GM_3(g,\pi)$ they are only 0.6, -1.4, and -4.6 basis points for the one-, five-, and ten-year yields, respectively!

Given that all three conditions are (approximately) satisfied, the ML estimates of $(K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma)$ should agree for all three models TS, FV, and FV^n. Table 1 displays the ratios of the estimated parameters from $GM_3(g,\pi)$ and its associated factor-VAR, with and without filtering. Consistent with our theory, they are all virtually identical.

4.3 Statistics of the distribution of (M_t, y_t)

It follows that the distributions of the risk factors are virtually the same across these different factor models. This, in turn, implies that all statistics of the distribution, such as the IRs, will be nearly identical as well. These results underlie Figure 1, where the IRs of $PC1$ to a shock to inflation in model $GM_3(g,\pi)$ and the associated model FV^n (nearly) coincide.

[29] Note that the sample standard deviation of the first difference $\Delta PC1_t^o$ is 42.5 basis points. Comparing 40.7 to 42.5 it follows that very little of $\Delta PC1_t^o$ is predictable based on the information structure of $GM_3(g,\pi)$. This is consistent with the near-random walk behavior of the level of interest rates.

	$K_0^{\mathbb{P}}$	$I + K_1^{\mathbb{P}}$			Σ		
	1	1	1	0.999	1.01	-	-
$\frac{TS}{FV}$	1	1	1	1	0.987	1	-
	1	0.999	1	1	0.998	1.01	1
	1.12	0.998	1.04	1.07	1.07	-	-
$\frac{TS}{TS^n}$	0.999	0.999	1	1	0.90	1	-
	0.988	0.93	1	1	0.885	1.11	1.01
	1.12	0.998	1.04	1.02	1.07	-	-
$\frac{FV}{FV^n}$	0.999	0.999	1	1	0.898	1	-
	0.989	0.929	1	1	0.885	1.11	1.01

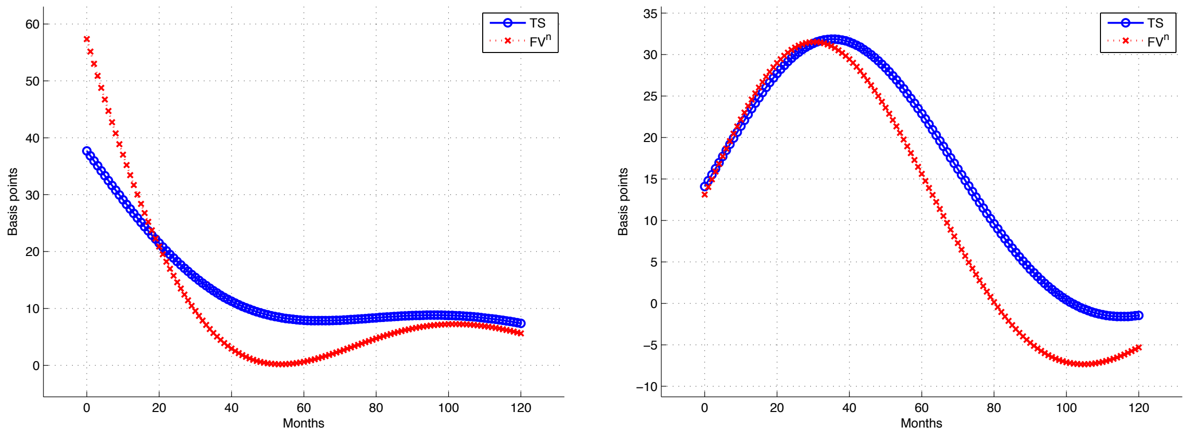
Table 1: Ratios of estimated $K_0^{\mathbb{P}}$, $I + K_1^{\mathbb{P}}$, and Σ for model $GM_3(g, \pi)$. The first block compares the estimates for models TS and FV, the second block compares models TS and TS^n, and the third compares models FV and FV^n.

Neither the no-arbitrage restrictions nor filtering in the presence of sizable measurement errors for the individual bond yields impact estimates of these responses.

4.4 Invariance of the distribution of (M_t, y_t)

For models TS and FV these empirical irrelevancy results extend to any full rank portfolio matrix W (Section 2.3). In particular, had we chosen to normalize the model so that $\mathcal{P}_t^1$ was any of the individual twelve yields instead of $PC1$, all of the results in Figure 2 would be exactly the same. The results in Table 1 would have been identical after rotation. The parameters governing the conditional distribution of Z_t would change, of course, since any such reweighting leads to different risk factors. Such renormalizations do not, however, affect the implied relationships among any given set of yields and macro variables.

As was discussed in Section 2.3, this invariance does not extend to comparisons across models constructed with different W and in which $\mathcal{P}_t^{\mathcal{L}}$ is assumed to be measured without error (models TS^n or FV^n). To illustrate this rotation sensitivity consider first the case where W is chosen so that y_t^3, the yield on three-month Treasury bills, is the state yield factor $\mathcal{P}_t^1$. This yield is one of the state yield factors in the models of Ang, Piazzesi, and Wei (2006) and Jardet, Monfort, and Pegoraro (2010), and in both studies y_t^3 is presumed to be measured without error. We compare results from $GM_3(g, \pi)$ (i.e., model TS) which has all bonds priced imperfectly and $\mathcal{P}_t^1 = y_t^3$, to those from its factor-VAR counterpart FV^n in which y_t^3 is presumed to be measured without error. Figure 3(a) displays the IRs of y_t^3 to its own innovation (in basis points) for these two models. Because of rotation invariance, the response for model TS is identical to what we would have obtained from estimation of this $MTSM$ normalized so that $\mathcal{P}_t^1 = PC1_t$. However the IR from model FV^n is very different: it is nearly fifty percent larger over very short horizons, decays much faster, and troughs at a



(a) Impulse Response of y_t^3 to y_t^3

(b) Impulse Response of y_t^3 to g_t

Figure 3: Impulse responses of y_t^3 to its own innovation (Panel (a)) and an innovation in g_t (Panel (b)) within models TS and FV^n for the family $GM_3(g, \pi)$.

lower value than the IR from model TS. The reason for these differences is that model FV^n captures the dynamic responses of the observed data, while the $MTSM$ (model TS) presumes that a portion of these responses are attributable to measurement error.

The IRs of y_t^3 to a shock in output growth $REALPC$ implied by models TS and FV^n follow similar patterns (Figure 3(b)), and the gap in responses is not as large as with the own responses. Yet the $MTSM$ implies a more persistent response that peaks later and dies out more slowly than what emerges from the factor-VAR.

The differences in attribution of dynamic responses to economic forces across a $MTSM$ and it factor-VAR counterpart can be extreme. Consider, for example, the version of $GM_3(g, \pi)$ in which $\mathcal{P}_t^1$ is normalized to be the third PC of bond yields ($PC3$).[30] Again, we stress that under the assumption that all bonds are measured with error, the Kalman-filter/ML estimates of the joint distribution of (M_t, y_t) under the rotations with $\mathcal{P}_t^1 = PC1$ or $\mathcal{P}_t^1 = PC3$ are *identical*. However, as can be seen from Figure 4, the model-implied IRs of $PC3$ to its own innovation are very different across models TS and FV^n. The $MTSM$ that enforces no arbitrage implies that there is essentially no response at all, whereas the factor-VAR characterization of history shows a large (though short-lived) response. This difference arises because, within $GM_3(g, \pi)$, the sufficient conditions for $PC3_t^o \approx PC3_t^f$ derived in Section 3.1 are not satisfied even though the differences $\{PC1_t^o - PC1_t^f\}$ are small (Figure 2). Essentially, $GM_3(g, \pi)$ does a poor job of replicating the historical time-series properties of $PC3^o$ owing to the presence of (g_t, π_t) as two of the three risk factors.

[30] For computing the impulse IRs for $PC3$ displayed in Figure 4 we scale its loadings so that $PC3$ has the same sample standard deviation as curvature measured as $y_t^{120} + y_t^3 - 2y_t^{24}$. Similarly, for $PC2$ in Figure 6 we scale the loadings so the it has the same sample volatility as slope measured as $y_t^{120} - y_t^3$.

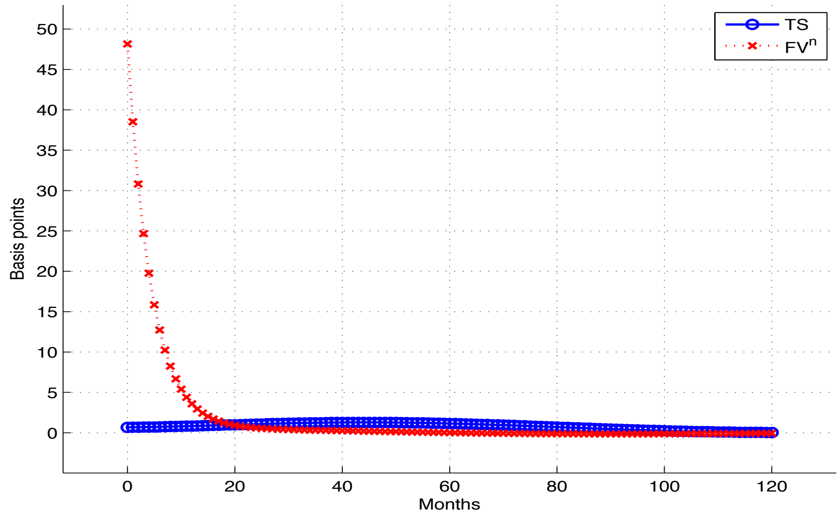


Figure 4: Impulse responses of $PC3$ to its own innovation within models TS and FV^n for the family $GM_3(g, \pi)$.

5 Extensions

This section explores an additional application and two extensions of our invariance propositions. As an additional illustration we examine the impact of no-arbitrage restrictions on recent resolutions of the failure of the expectations theory of the term structure (ETTS). This application is of interest both because it has received considerable attention in the literature and tests of the ETTS involve both the conditional means and variances of the distribution of Z_t. In exploring this question we focus on four-factor models to provide insight into the properties of *MTSM*s when one increases the number of yield-based risk factors from one to two. Separately, 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

According to the ETTS, changes in long-term bond yields should move one-to-one with changes in the slope of the yield curve. 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

$$Proj\left[y_{t+1}^{n-1} - y_t^n | y_t^n - r_t\right] = \alpha_n + \phi_n \left(\frac{y_t^n - r_t}{n-1}\right), \tag{20}$$

where $Proj[\cdot|\cdot]$ denotes linear least-squares projection. The ETTS implies that $\phi_n = 1$, for all maturities n. Following Dai and Singleton (2002), it is instructive to compare this relationship to the general premium-adjusted expression

$$E^{\mathbb{P}}\left[y_{t+1}^{n-1} - y_t^n - (c_{t+1}^{n-1} - c_t^{n-1}) + \frac{p_t^{n-1}}{n-1}\big|y_t\right] = \left(\frac{y_t^n - r_t}{n-1}\right), \tag{21}$$

where

$$c_t^n \equiv y_t^n - \frac{1}{n}\sum_{i=0}^{n-1} E^{\mathbb{P}}[r_{t+1}|y_t] \text{ and } p_t^n \equiv f_t^n - E^{\mathbb{P}}[r_{t+n}|y_t] \tag{22}$$

are the yield and forward term premiums, respectively, and f_t^n denotes the forward rate for one-period loans commencing at date $t+n$. A model is considered successful at explaining the failure of the ETTS if the term premiums it generates through time-varying market prices of risk reproduce (21) and, thereby, lead to a pattern in the model-implied ϕ_n^{TS} that matches the ϕ_n in the sample.

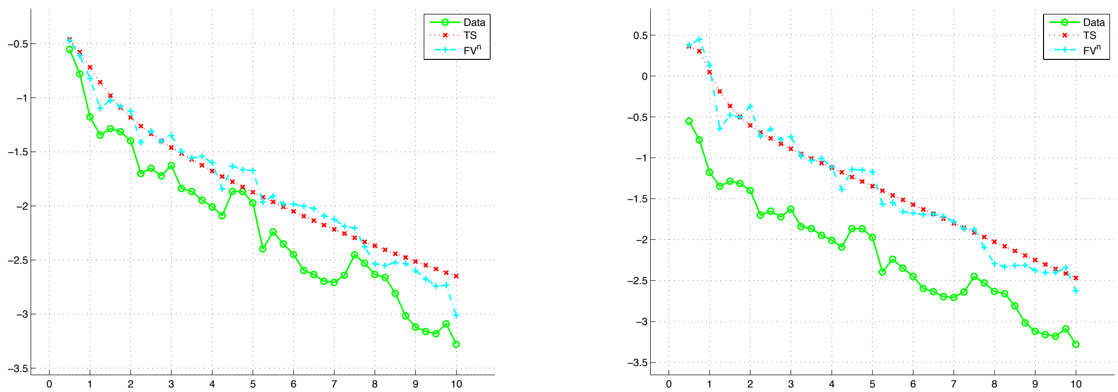
To investigate the impact of no-arbitrage restrictions and Kalman filtering on tests of the ETTS we estimate two classes of four-factor models. Model GY_4 is a standard four-factor $YTSM$ normalized as in JSZ so that $Z_t = \mathcal{P}_t^4$, the first four PCs of bond yields. Model $GM_4(g,\pi)$ has the same M_t as $GM_3(g,\pi)$ ($REALPC$ and $INFPC$), and is normalized so that other two factors are $\mathcal{P}_t^2 = (PC1_t, PC2_t)$. Inclusion of $PC2$ as a state yield factor is important for matching the observed violations of the ETTS (Dai and Singleton (2002)). Additionally, the fit of $GM_3(g,\pi)$ to the cross-section of yields was modest at best so by increasing N to four in model $GM_4(g,\pi)$ we potentially improve its fit and, thereby, provide additional perspective on the irrelevancy issue in $MTSM$s.

We estimate the models assuming that the yields are priced with $i.i.d.$ $N(0,\sigma_y^2 I_{12})$ errors. Using the covariances of the steady-state distribution of $\mathcal{P}_t$ implied by model TS evaluated at the ML estimates, we compute the projection coefficients ϕ_n^{TS}. For comparison we compute the coefficients $\phi_n^{FV^n}$ for model FV^n.[31] The data are again the unsmoothed Fama-Bliss zero yields on US Treasury bonds for the period January, 1972 through December, 2003.

The results for the case where the short-term positions are rolled every three months are displayed in Figure 5.[32] Consistent with the extant evidence, these low-dimensional term

[31] A practical problem that arises in computing the regression coefficients for model FV^n is that, from the twelve yields used in estimation of model TS 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 FV^n-implied coefficients.

[32] 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 longer holding periods are qualitatively similar.



(a) Model GY_4 (b) Model $GM_4(g, \pi)$

Figure 5: Projection coefficients ϕ_n implied by models GY_4 and $GM_4(g, \pi)$ and their corresponding unconstrained factor-VAR models FV^n. The horizontal axis is maturity in years.

structure models do resolve the expectations puzzle: the implied ϕ_n^{TS} track the estimated ϕ_n from the data quite closely. The *YTSM* GY_4 matches the historical data even more closely than does the *MTSM*, but the differences are not large.

More to the point of our analysis, the implied ϕ_n from models TS and FV^n are virtually on top of each other for both models, regardless of their distances from the sample projection coefficients. It follows that the reason these term structure models are successful at resolving the ETTS puzzle is because the unconstrained factor models FV^n resolve this puzzle. In other words, inherent in the reduced-form factor structure (8) is a pattern of projection coefficients $\phi_n^{FV^n}$ that approximately matches those from the regression equations (20). The term structure models also match the regression slopes simply because models TS and FV^n produce almost identical conditional distributions of bond yields. The *MTSM* estimated by Kalman filtering adds nothing to the insights gleaned from the factor-VAR.

Note that implicit in this finding is a very close similarity between the cross-sectional patterns of factor loadings produced by *OLS* projections of y_t onto Z_t and the loadings produced by the arbitrage-free term structure models. Duffee (2011a), using Monte Carlo methods and focusing on *YTSM*s, obtains a similarly close correspondence between these two estimates of the risk-factor loadings. He argues that this finding arises as a consequence of the small measurement errors on yields (his $\hat{\sigma}_y$ is only a few basis points). While $\hat{\sigma}_y \approx 0$ is clearly sufficient for this finding, it is not necessary. For each of the families $GM_3(g, \pi)$ and $GM_4(g, \pi)$, the estimated loadings from models TS and FV^n (not shown) are virtually indistinguishable, and yet the measurement errors on yields in these models can be large. What drives this result is our finding that $\mathcal{P}_t^{\mathcal{L}o} \approx \mathcal{P}_t^{\mathcal{L}f}$, and we have seen that this arises even when $y_t^o - y_t^f$ is large. Thus, the (near) observational equivalence of models TS and FV extends to all yields through the factor loadings A and B.

5.2 Models with Unspanned Risk Factors

The *MTSM*s considered so far have the macro variables entering directly as risk factors determining interest rates, as is the case with the large majority of the extant literature. Joslin, Priebsch, and Singleton (2010) have developed a different class of models that allow for unspanned macro risks– risks that cannot be replicated by linear combinations of bond yields.[33] Their canonical model with unspanned risks shares two important properties with *MTSM*s with spanned risks: (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, an implication of property (1) is that when the risk factors are observed without error, forecasts agree identically with the corresponding factor-VAR. Likewise, property (2) implies that the deviation of the ML estimate of Σ in model TS from its counterpart in model FV^n 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 of the models TS or FV^n will be nearly identical.[34]

5.3 Higher-Order VAR Models of Risk Factors

Up to this point we have focused on the class of *MTSM*s in which Z_t follows first-order Markov processes 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 historical distributions of macro-factors and bond yields as those implied by their associated factor-VARs.

Central to our construction of a revealing canonical form for *MTSM*s with first-order Markov risk factors Z_t was their property that Z_t is linearly spanned by contemporaneous bond yields. Before exploring the properties of specific *MTSM*s with higher order lags, it is instructive to inquire whether such models give rise to a spanning condition for the macro factors M_t analogous to (4). If Z_t follows a VAR$^{\mathbb{Q}}$(q) under the pricing measure, then it is no longer the case that any $\mathcal{N}$ portfolios of yields formed with a full rank weight matrix $W^{\mathcal{N}}$ span M_t or any latent factors in Z_t. However, outside of knife-edge cases, it will be the case that the macro and latent factors are spanned by $\mathcal{N} \times q$ portfolios of yields. It follows that setting $q > 1$ under the pricing distribution effectively increases the number of pricing factors from $\mathcal{N}$ to $\mathcal{N} \times q$.[35]

The choice of $q > 1$ under $\mathbb{Q}$ is typically supported by indirect evidence under the historical distribution. Evidence that Z_t follows a VAR$^{\mathbb{P}}$(p) under $\mathbb{P}$ with $p > 1$, and the presumption

[33]For additional applications of their framework, see Wright (2009) and Barillas (2010). Duffee (2011b) discusses a complementary model of unspanned risks in yield-only models.

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

[35]In this case M_t is spanned by $\mathcal{N} \times q$ linear combinations of bond yields. However, when $q > 1$, the number of free parameters is less than in a canonical $\mathcal{N} \times q$ factor model, because the increase in the number of factors comes from inclusion of lagged values of Z_t.

Model	M_t	q	$y_t^{0.5yr}$	y_t^{1yr}	y_t^{2yr}	y_t^{5yr}	y_t^{7yr}	y_t^{10yr}
$GM_3(g)$	$REALPC$	1	10	15	17	10	10	18
$GM_3(g)$	$REALPC$	6	9	14	16	10	10	18
$GM_3(g)$	$REALPC$	12	9	14	16	9	9	17
$GM_3(g,\pi)$	$REALPC, INFPC$	1	60	45	22	25	36	47
$GM_3(g,\pi)$	$REALPC, INFPC$	6	57	42	20	24	34	45
$GM_3(g,\pi)$	$REALPC, INFPC$	12	53	37	18	22	31	40
$GM_4(g,\pi)$	$REALPC, INFPC$	1	9	15	17	10	10	18
$GM_4(g,\pi)$	$REALPC, INFPC$	6	8	14	16	9	10	17
$GM_4(g,\pi)$	$REALPC, INFPC$	12	8	13	15	9	9	16

Table 2: 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 $(Z_t, Z_{t-1}, \ldots, Z_{t-q+1})$.

of flexible market prices of risk, together imply that Z_t follows a VAR$^{\mathbb{Q}}$(p) as well. The question of whether the data call for $q > 1$ is often not addressed directly.

Within the family of *MTSM*s, guidance on the lag structure of the $\mathbb{Q}$ distribution of (r_t, Z_t) is provided by the projections of yields onto current and lagged values of Z. The null that Z_t follows the first-order VAR$^{\mathbb{Q}}$(1) in (2) implies that projections of yields onto current and lagged values of Z_t do not improve the explained variation in yields relative to the contemporaneous projections of y_t^o onto Z_t^o. On the other hand, evidence of improved fits would suggest that the bond data call for setting $q > 1$ under $\mathbb{Q}$.

To address the order under $\mathbb{Q}$ empirically we consider several variants of *MTSM*s using the real and nominal macro factors $REALPC$ and $INFPC$.[36] Model $GM_3(g)$ has $(\mathcal{N} = 3, \mathcal{M} = 1)$; model $GM_3(g,\pi)$ has $(\mathcal{N} = 3, \mathcal{M} = 2)$; and model $GM_4(g,\pi)$ has $(\mathcal{N} = 4, \mathcal{M} = 2)$. Their associated sets of risk factors Z_t are given in Table 2. This table also presents the root mean-squared projection errors in basis points, for lag lengths $q = 1, 6, 12$ (one year in our monthly data).[37] For three of the four cases, the improvements in fit from setting $q > 1$ are tiny, at most one or two basis points. The only exception is model $GM_3(g,\pi)$ with state vector $(REALPC, INFPC, PC1)$. In this case the fit is so poor, with $RMSE$s as large as sixty basis points, that adding lags under $\mathbb{Q}$ improves the $RMSE$s more, up to eight basis points. In all cases, the (AIC, BIC) model selection criteria select the lag length $q = 1$.

Supported by the evidence in Table 2, we proceed to explore asymmetric formulations of *MTSM*s in which Z_t follows a VAR$^{\mathbb{P}}$(p) under $\mathbb{P}$, and a VAR$^{\mathbb{Q}}$(1) under $\mathbb{Q}$. Since Z_t follows the first-order Markov process (2) under $\mathbb{Q}$ we can once again normalize this model

[36]Very similar results are obtained using the help-wanted index as the real series and CPI inflation as in Bikbov and Chernov (2010).

[37]Strictly speaking, these projections speak directly to variants of models that assume $\mathcal{P}_t^{2o} = \mathcal{P}_t^2$. However, the filtered $(PC1_t^f, PC2_t^f)$ are nearly identical to $(PC1_t^o, PC2_t^o)$ in the variants with measurement errors $(\mathcal{P}_t^{2o} \neq \mathcal{P}_t^2)$, so the following observations are relevant for both cases.

so that $Z'_t = (M'_t, \mathcal{P}^{\mathcal{L}\prime}_t)$, M_t continues to satisfy the spanning condition (4), and r_t is given by (1). Furthermore, the parameters governing (1) and (2) are explicit functions of $\Theta^{\mathbb{Q}}_{TS} = (r^{\mathbb{Q}}_\infty, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma)$. These functions are identical to those in Section 2 and are provided in Appendix A. Our *MTSM* with lags is completed by specifying the $\mathbb{P}$-dynamics of Z_t as the unrestricted VAR

$$Z_t = K^{\mathbb{P}}_0 + K^{\mathbb{P}}_1 \overrightarrow{Z}_{t-1,p} + \sqrt{\Sigma}\epsilon^{\mathbb{P}}_t, \tag{23}$$

which nests the $\mathbb{P}$ distributions in the *MTSM*s with lags studied by Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007), and Jardet, Monfort, and Pegoraro (2010), among others.[38]

By the same reasoning as before,[39] this formulation is canonical and, for given W, it is parametrized by $\Theta_{TS} = (\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_\infty, K^{\mathbb{P}}_0, K^{\mathbb{P}}_1, \Sigma, \gamma_0, \gamma_1)$ with the dimension of $K^{\mathbb{P}}_1$ suitably adjusted to accommodate the lags in $\overrightarrow{Z}_{t-1,p}$. Moreover, our main analysis comparing the properties of no-arbitrage models with their associated factor-VARs applies directly to this generalized *MTSM* with lags. Under the same conditions set forth in Section 3, the estimated $\mathbb{P}$-distributions of Z_t from models TS and FV^n will be (nearly) identical.

To examine the impact on the joint distribution of Z^o_t of adopting a VAR$^{\mathbb{P}}$(p) specification in the family $GM_4(g, \pi)$, VAR models are fit over a wider range of p and the optimal number of lags is chosen using BIC and AIC criteria. The optimal p according to the BIC (AIC) criterion is 2 (3). Given our objective of examining the sensitivity of the conditional distribution of bond yields to high-order lag structures we choose $p = 3$ and denote this family by $GM_{4,3}(g, \pi)$. With $p = 3$ the no-arbitrage and factor-VAR 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.[40] Figure 6 displays the IRs implied by models TS and FV^n within family $GM_{4,3}(g, \pi)$.[41] As before, the imposition of no arbitrage and the use of filtering is virtually inconsequential for how shocks to macro factors impact the yield curve.

6 Concluding Remarks

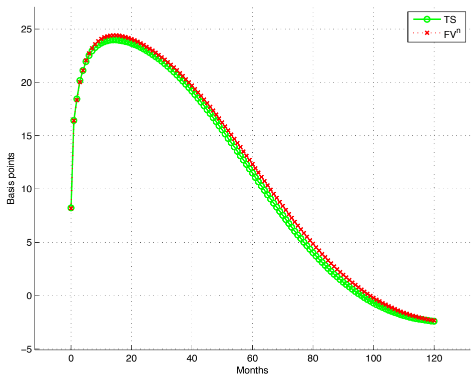
We have shown theoretically and documented empirically that the no-arbitrage restrictions of canonical *MTSM*s, and the accommodation of measurement errors on all bond yields through Kalman filtering, have essentially no impact on the ML estimates of the joint conditional distribution of the macro and yield-based risk factors, including models that nest some of the

[38]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 in Table 2, we omit these additional lags from our analysis.

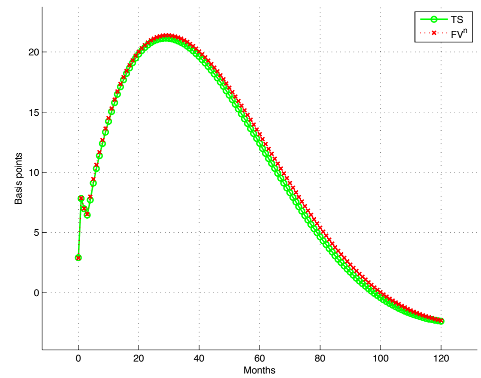
[39]The applicability of our prior reasoning to the case of a VAR$^{\mathbb{P}}$(p) process is critically dependent on our assumption that Z_t follows a VAR$^{\mathbb{Q}}$(1) process. With a more flexible lag structure under $\mathbb{Q}$, the normalization strategies in JSZ or in our Theorem A1 would no longer apply. Even with $q > 1$ lags under $\mathbb{Q}$, however, it would still be the case that Z_t is fully spanned by $\mathcal{N} \times q$ portfolios of yields.

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

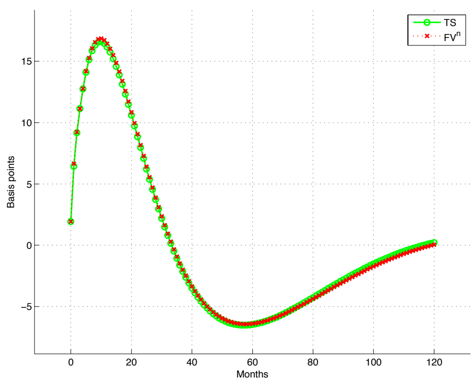
[41]The choppy behavior over short horizons for some of the responses in Figure 6 is also evident in the IRs reported in Ang and Piazzesi (2003) for their *MTSM* with lags.



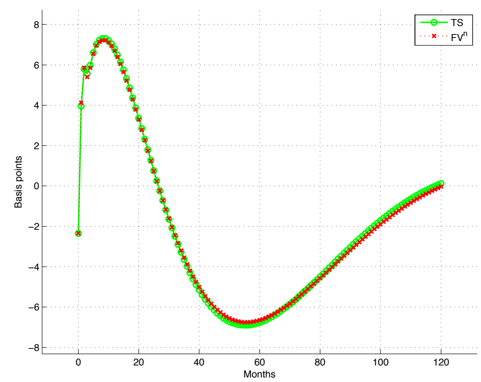
(a) Response of $PC1$ to $REALPC$



(b) Response of $PC1$ to $INFPC$



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



(d) Response of $PC2$ to $INFPC$

Figure 6: Impulse responses of $PC1$ and $PC2$ to innovations in $REALPC$ and $INFPC$ based on ML estimates of model $GM_{4,3}(g, \pi)$. The risk factors are ordered as $(INFPC, REALPC, PC1, PC2)$.

most widely studied *MTSM*s in the literature. 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 estimates of the $\mathbb{P}$ distribution of yields within a *MTSM* are more efficient than those from the factor-VAR comprised of (6) and (8).

Whether such efficiency gains are large 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 imposed 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.

On the other hand, the *MTSM*s in Joslin, Priebsch, and Singleton (2010) and Jardet, Monfort, and Pegoraro (2010) that enforce near cointegration under $\mathbb{P}$ of the risk factors have very different dynamic properties than their unconstrained factor-VAR counterparts. Similarly, unit-root or cointegration-type restrictions imposed directly on the $\mathbb{P}$ distribution of the risk factors in *YTSM*s were shown by JSZ and Duffee (2011a) to lead to improved out-of-sample forecasts of bond yields. 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 on the efficiency of ML/Kalman filter estimates.

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.

Appendices

A A Canonical Form for *MTSM*s

Our objective is to show that each *MTSM* where

$$r_t = \rho_0^{\mathcal{L}} + \rho_1^{\mathcal{L}} \cdot Z_t^{\mathcal{L}} \tag{A1}$$

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

$$\begin{aligned} \Delta Z_t^{\mathcal{L}} &= \kappa_0^{\mathbb{Q}} + \kappa_1^{\mathbb{Q}} Z_{t-1}^{\mathcal{L}} + \sqrt{\Omega}\epsilon_t^{\mathbb{Q}} \text{ under } \mathbb{Q} \text{ and} && \text{(A2)} \\ \Delta Z_t^{\mathcal{L}} &= \kappa_0^{\mathbb{P}} + \kappa_1^{\mathbb{P}} Z_{t-1}^{\mathcal{L}} + \sqrt{\Omega}\epsilon_t^{\mathbb{P}} \text{ under } \mathbb{P}, && \text{(A3)} \end{aligned}$$

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

$$\begin{aligned} r_t &= \rho_0 + \rho_1 \cdot Z_t, && \text{(A4)} \\ \Delta Z_t &= K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}} Z_{t-1} + \sqrt{\Sigma}\epsilon_t^{\mathbb{Q}} \text{ under } \mathbb{Q} \text{ and} && \text{(A5)} \\ \Delta Z_t &= K_0^{\mathbb{P}} + K_1^{\mathbb{P}} Z_{t-1} + \sqrt{\Sigma}\epsilon_t^{\mathbb{P}} \text{ under } \mathbb{P} && \text{(A6)} \end{aligned}$$

where $(\rho_0, \rho_1, K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}})$ are explicit functions of some underlying parameter set $\Theta_{TS}^{\mathbb{Q}} = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma)$ to be described. We will make precise the sense in which $\Theta_Z = (\Theta_{TS}^{\mathbb{Q}}, K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$ uniquely characterizes the latter *MTSM*.

Observational Equivalence

Assuming, for ease of exposition, that $\kappa_1^{\mathbb{Q}}$ has nonzero, real and distinct eigenvalues with the standard eigendecomposition$\kappa_1^{\mathbb{Q}} = A^{\mathbb{Q}} diag(\lambda^{\mathbb{Q}}) A^{\mathbb{Q}^{-1}}$, we follow Joslin (2006) and JSZ by adopting the rotation:[42]

$$X_t = \mathcal{V}^{-1}\left(Z_t^{\mathcal{L}} + (\kappa_1^{\mathbb{Q}})^{-1}\kappa_0^{\mathbb{Q}}\right) \text{ where } \mathcal{V} = A^{\mathbb{Q}} diag((\rho_1^{\mathcal{L}})' A^{\mathbb{Q}})^{-1} \tag{A7}$$

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

$$r_t = r_\infty^{\mathbb{Q}} + \iota \cdot X_t, \text{ and } \Delta X_t = diag(\lambda^{\mathbb{Q}}) X_{t-1} + \sqrt{\Sigma_X}\epsilon_t^{\mathbb{Q}} \tag{A8}$$

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

$$r_\infty^{\mathbb{Q}} = \rho_0^{\mathcal{L}} + (\rho_1^{\mathcal{L}})'(\kappa_1^{\mathbb{Q}})^{-1}\kappa_0^{\mathbb{Q}} \text{ and } \Omega = \mathcal{V}\Sigma_X\mathcal{V}'.$$

From (A8), the $J \times 1$ vector of yields y_t is affine in X_t:

$$y_t = A_X(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) + B_X(\lambda^{\mathbb{Q}}) X_t \tag{A9}$$

[42] 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 $\mathcal{P}_t = Wy_t$. Focusing on the first $\mathcal{N}$ portfolios $\mathcal{P}_t^{\mathcal{N}}$, we have:

$$\mathcal{P}_t^{\mathcal{N}} = W^{\mathcal{N}} A_X + W^{\mathcal{N}} B_X X_t. \tag{A10}$$

Based on (A7) and (A10), there is a linear mapping between M_t and $\mathcal{P}_t^{\mathcal{N}}$:

$$M_t = \gamma_0 + \gamma_1 \mathcal{P}_t^{\mathcal{N}} \tag{A11}$$

where

$$\gamma_1 = \mathcal{V}_{\mathcal{M}}(W^{\mathcal{N}} B_X)^{-1} \text{ and } \gamma_0 = -\gamma_1 W^{\mathcal{N}} A_X - A_{\mathcal{M}}^{\mathbb{Q}}(\lambda^{\mathbb{Q}})^{-1} A^{\mathbb{Q}^{-1}} \kappa_0^{\mathbb{Q}}, \tag{A12}$$

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:

$$Z_t = \Gamma_0 + \Gamma_1 \mathcal{P}_t^{\mathcal{N}} = \Gamma_0 + \Gamma_1(W^{\mathcal{N}} A_X + W^{\mathcal{N}} B_X X_t) = \mathcal{U}_0 + \mathcal{U}_1^{-1} X_t \tag{A13}$$

where

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

Combining (A8) and (A13), the $\mathbb{Q}$-specification of Z_t is:

$$r_t = \rho_0 + \rho_1 \cdot Z_t \text{ and } \Delta Z_t = K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}} Z_{t-1} + \sqrt{\Sigma}\epsilon_t^{\mathbb{Q}}, \tag{A14}$$

where

$$\rho_1 = (\mathcal{U}_1)'\iota \text{ and } \rho_0 = r_\infty^{\mathbb{Q}} - \rho_1 \cdot \mathcal{U}_0,$$
$$K_1^{\mathbb{Q}} = \mathcal{U}_1^{-1}\lambda^{\mathbb{Q}}\mathcal{U}_1, \;\; K_0^{\mathbb{Q}} = -K_1^{\mathbb{Q}}\mathcal{U}_0 \;\; (\text{and } \Sigma_X = \mathcal{U}_1 \Sigma \mathcal{U}_1').$$

Based on (A7) and (A13), 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 (A6).

To summarize, the *MTSM* with mixed macro-latent risk factors $Z_t^{\mathcal{L}}$, described by (A1), (A2), and (A3), is observationally equivalent to one with observable mixed macro-yield-portfolio risk factors Z_t, characterized by (A4), (A5), and (A6). The *primitive* parameter set is $\Theta_Z = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma, K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$. The mappings between $(\rho_0, \rho_1, K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}})$ and $\Theta_{TS}^{\mathbb{Q}} = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma)$ are:

$$\rho_1 = (\mathcal{U}_1)'\iota, \;\; \rho_0 = r_\infty^{\mathbb{Q}} - \rho_{1Z} \cdot \mathcal{U}_0, \;\; K_1^{\mathbb{Q}} = \mathcal{U}_1^{-1}\lambda^{\mathbb{Q}}\mathcal{U}_1, \;\; K_0^{\mathbb{Q}} = -K_1^{\mathbb{Q}}\mathcal{U}_0, \tag{A15}$$

where

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

Uniqueness

Consider two parameter sets, Θ_Z and $\tilde{\Theta}_Z$, that give rise to two observationally equivalent *MTSM*'s with risk factors Z_t. Since Z_t is observable, the parameters, $\Sigma, K_0^{\mathbb{P}}, K_1^{\mathbb{P}}$, describing the $\mathbb{P}$-dynamics of Z_t must be identical. Additionally, based on (A11), the following identity must hold state by state:

$$M_t \equiv \gamma_0 + \gamma_1 \mathcal{P}_t^{\mathcal{N}} \equiv \tilde{\gamma}_0 + \tilde{\gamma}_1 \mathcal{P}_t^{\mathcal{N}}. \tag{A16}$$

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

$$\gamma_0 = \tilde{\gamma}_0 \text{ and } \gamma_1 = \tilde{\gamma}_1. \tag{A17}$$

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

$$y_t = A_X + B_X(W^{\mathcal{N}} B_X)^{-1}(P_t^{\mathcal{N}} - W^{\mathcal{N}} A_X), \tag{A18}$$

it follows that

$$B_X(W^{\mathcal{N}} B_X)^{-1} = \tilde{B}_X(W^{\mathcal{N}} \tilde{B}_X)^{-1}, and \tag{A19}$$
$$(I_J - B_X(W^{\mathcal{N}} B_X)^{-1} W^{\mathcal{N}})A_X = (I_J - \tilde{B}_X(W^{\mathcal{N}} \tilde{B}_X)^{-1} W^{\mathcal{N}})\tilde{A}_X. \tag{A20}$$

Now (A19) is equivalent to:

$$diag(\frac{1-\lambda_i^n}{1-\lambda_i})(W^{\mathcal{N}} B_X)^{-1} = diag(\frac{1-\tilde{\lambda}_i^n}{1-\tilde{\lambda}_i})(W^{\mathcal{N}} \tilde{B}_X)^{-1} \tag{A21}$$

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 (A20), we note that

$$A_X = \iota\, r_\infty^{\mathbb{Q}} + \beta_X vec(\Sigma_X) \tag{A22}$$

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 $(\gamma_1, \lambda^{\mathbb{Q}}, \Sigma)$, 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 assumptions can be easily relaxed - see JSZ for detailed treatments. Second, we assume that the *MTSM*'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 $(\rho_1^{\mathcal{L}})' A^{\mathbb{Q}}$ are non-zero. In terms of the parameters of our canonical form, we require that none of the eigenvectors of the risk-neutral feedback matrix $K_1^{\mathbb{Q}}$ is orthogonal to the loadings vector ρ_1 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, respectively.

The following theorem summarizes the above derivations:

Theorem A1. *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* MTSM *is observationally equivalent to a unique* MTSM *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* $\mathcal{P}_t^{\mathcal{L}}$*;* r_t *is given by (A4);* M_t *is related to* $\mathcal{P}_t$ *through*

$$M_t = \gamma_0 + \gamma_1 \mathcal{P}_t^{\mathcal{N}}, \tag{A23}$$

for $\mathcal{M} \times 1$ *vector* γ_0 *and* $\mathcal{M} \times \mathcal{N}$ *matrix* γ_1*; and* Z_t *follows the Gaussian* $\mathbb{Q}$ *and* $\mathbb{P}$ *processes (A5), and (A6), where* $K_0^{\mathbb{Q}}$*,* $K_1^{\mathbb{Q}}$*,* ρ_0*, and* ρ_1 *are explicit functions of* $\Theta_{TS}^{\mathbb{Q}} = (r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma)$*, given by (A15). For given* W*, our canonical form is parametrized by* $\Theta_{TS} = (\Theta_{TS}^{\mathbb{Q}}, K_0^{\mathbb{P}}, K_1^{\mathbb{P}})$*.*

B Bond Pricing in *MTSM*s

Under (A4–A6), the price of an m-year zero-coupon bond is given by

$$D_{t,m} = E_t^{\mathbb{Q}}[e^{-\sum_{i=0}^{m-1} r_{t+i}}] = e^{\mathcal{A}_m + \mathcal{B}_m \cdot Z_t}, \tag{A24}$$

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

$$\mathcal{A}_{m+1} - \mathcal{A}_m = K_0^{\mathbb{Q}'} \mathcal{B}_m + \frac{1}{2} \mathcal{B}_m' H_0 \mathcal{B}_m - \rho_0 \tag{A25}$$

$$\mathcal{B}_{m+1} - \mathcal{B}_m = K_1^{\mathbb{Q}'} \mathcal{B}_m - \rho_1 \tag{A26}$$

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

C Invariant Transformations of *MTSM*s

As in Dai and Singleton (2000), given a *MTSM* with parameters as in (A4–A6) and state Z_t, application of the invariant transformation $\hat{Z}_t = C + DZ_t$ gives an observationally equivalent term structure model with state $\hat{Z}_t$ and parameters

$$K_{0\hat{Z}}^{\mathbb{Q}} = DK_0^{\mathbb{Q}} - DK_1^{\mathbb{Q}} D^{-1} C, \tag{A27}$$

$$K_{1\hat{Z}}^{\mathbb{Q}} = DK_1^{\mathbb{Q}} D^{-1}, \tag{A28}$$

$$\rho_{0\hat{Z}} = \rho_0 - \rho_1' D^{-1} C \tag{A29}$$

$$\rho_{1\hat{Z}} = (D^{-1})' \rho_1, \tag{A30}$$

$$K_{0\hat{Z}}^{\mathbb{P}} = DK_0^{\mathbb{P}} - DK_1^{\mathbb{P}} D^{-1} C, \tag{A31}$$

$$K_{1\hat{Z}}^{\mathbb{P}} = DK_1^{\mathbb{P}} D^{-1}, \tag{A32}$$

$$\Sigma_{\hat{Z}} = D \Sigma D'. \tag{A33}$$

D Filtering Invariance of the Mean Parameters

This appendix shows that when $\Sigma_{e\mathcal{L}} S_t^{-1}$ is small the filtered version of equation (16),

$$[\hat{K}_0^{\mathbb{P}}, I + \hat{K}_1^{\mathbb{P}}]' = \left(\frac{1}{T} \sum_t Z_{t+1}^f, \frac{1}{T} \sum_t (Z_{t+1} Z_t')^f \right) \left(\begin{array}{cc} 1 & \frac{1}{T} \sum_t Z_t^{f\prime} \\ \frac{1}{T} \sum_t Z_t^f & \frac{1}{T} \sum_t (Z_t Z_t')^f \end{array} \right)^{-1}, \tag{A34}$$

gives (under mild assumptions) estimates that are close to the OLS estimates. 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 (16) will also give approximately the OLS estimates of $K_0^{\mathbb{P}}$ and $K_1^{\mathbb{P}}$.

As shown in Section D.1, when $\Sigma_{e\mathcal{L}} S_t^{-1}$ is small, convergence to the steady-state distribution will be fast. As such we can treat $\Omega_t = Var(Z_t^o | \mathcal{P}_t^{-\mathcal{L}o}, \mathcal{F}_{t-1})$ as a constant matrix, with $\mathcal{P}_t^{-\mathcal{L}o}$ being the $J - \mathcal{L}$ higher order PCs. Post-multiplying both terms on the right-hand side of (A34) by $\left(\begin{array}{cc} 1 & 0 \\ 0 & \Omega_t^{-1} \end{array} \right)$ leads to:

$$\left(\frac{1}{T} \sum_t Z_{t+1}^f, \frac{1}{T} \sum_t (Z_{t+1} Z_t')^f \Omega_t^{-1} \right) \left(\begin{array}{cc} 1 & \frac{1}{T} \sum_t Z_t^{f\prime} \Omega_t^{-1} \\ \frac{1}{T} \sum_t Z_t^f & \frac{1}{T} \sum_t (Z_t Z_t')^f \Omega_t^{-1} \end{array} \right)^{-1}. \tag{A35}$$

Now,

$$\begin{aligned} (Z_t Z_t')^f \Omega_t^{-1} =& Var(Z_t | \mathcal{F}_t) \Omega_t^{-1} + Z_t^f (Z_t^f)' \Omega_t^{-1} \\ =& Var(Z_t | \mathcal{F}_t) \Omega_t^{-1} + Z_t^o (Z_t^o)' \Omega_t^{-1}, \end{aligned} \tag{A36}$$

where the second line follows from results in Section 3.1. Using block inversion, the non-zero block of the first term is:

$$\Sigma_{e\mathcal{L}} S_t^{-1} - \Sigma_{e\mathcal{L}} S_t^{-1} \Sigma_{e\mathcal{L}} S_t^{-1}$$

which under our assumption must be close to zero. Therefore we can replace the term $\frac{1}{T} \sum_t (Z_t Z_t')^f \Omega_t^{-1}$ in (A35) by $\frac{1}{T} \sum_t Z_t^o Z_t^{o\prime} \Omega_t^{-1}$. Using a similar argument, $\frac{1}{T} \sum_t (Z_{t+1} Z_t')^f \Omega_t^{-1}$ can also be replaced by $\frac{1}{T} \sum_t Z_{t+1}^o Z_t^{o\prime} \Omega_t^{-1}$. Furthermore, results in Section 3.1 allow us to replace Z_t^f in (A35) by its observed counter-part:

$$\left(\frac{1}{T} \sum_t Z_{t+1}^o, \frac{1}{T} \sum_t Z_{t+1}^o Z_t^{o\prime} \Omega_t^{-1} \right) \left(\begin{array}{cc} 1 & \frac{1}{T} \sum_t Z_t^{o\prime} \Omega_t^{-1} \\ \frac{1}{T} \sum_t Z_t^o & \frac{1}{T} \sum_t Z_t^o Z_t^{o\prime} \Omega_t^{-1} \end{array} \right)^{-1}. \tag{A37}$$

Finally, if $Var_T(Z_t^o) Var(Z_t^o | \mathcal{P}_t^{-\mathcal{L}o}, \mathcal{F}_{t-1})^{-1}$ is non-degenerate relative to $\Sigma_{e\mathcal{L}} S_t^{-1}$, then all Ω_t's cancel out and (A37) reduces to the the familiar OLS estimates.

D.1 Speed of Convergence to Steady States

Consider the following *generic* state space system:

$$Z_{t+1} = K_0 + K_1 Z_t + \sqrt{\Sigma} \epsilon_{t+1}, \tag{A38}$$

$$Z_{t+1}^o = Z_{t+1} + e_{Z,t+1}, \tag{A39}$$

$$Y_{t+1}^o = A + B Z_{t+1} + e_{Y,t+1} \tag{A40}$$

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:

$$\Sigma_{t+1} = \Sigma + K_1(\Sigma_t - \Sigma_t \tilde{B}'(\tilde{B}\Sigma_t \tilde{B}' + \Sigma_e)^{-1}\tilde{B}\Sigma_t)K_1' \tag{A41}$$

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

Standard linear algebra allows us to express the term between K_1 and K_1' in (A41) as:

$$\Sigma_{Ze} - (\Sigma_{Ze}, 0)\left(\tilde{B}\Sigma_t\tilde{B}' + \Sigma_e\right)^{-1}\begin{pmatrix} \Sigma_{Ze} \\ 0 \end{pmatrix}. \tag{A42}$$

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

$$\partial\Sigma_{t+1} = \Phi\partial\Sigma_t\Phi' \text{ with } \Phi = K_1(\Sigma_{Ze}, 0)\left(\tilde{B}\Sigma_t\tilde{B}' + \Sigma_e\right)^{-1}\begin{pmatrix} I \\ B \end{pmatrix}. \tag{A43}$$

Replacing $\left(\tilde{B}\Sigma_t\tilde{B}' + \Sigma_e\right)$ by $Var\begin{pmatrix} Z^o_t \\ Y^o_t \end{pmatrix} |\mathcal{F}_{t-1}\Big)$ and applying block-wise inversion to this matrix, gives:

$$\Phi = K_1\Sigma_{Ze}\Omega_t^{-1}(I - \Sigma_t B'(B\Sigma_t B' + \Sigma_{Ye})^{-1}B). \tag{A44}$$

As a result, as $\Sigma_{Ze}\Omega_t^{-1}$ approaches zeros, so do the eigenvalues of Φ. Since the recursion (A41) can be written approximately as:

$$vec(\Sigma_{t+1} - \bar{\Sigma}) \approx (\Phi \otimes \Phi)vec(\Sigma_t - \bar{\Sigma}), \tag{A45}$$

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

For *MTSM*s we assume that M_t is perfectly observed, and the $\mathcal{M}$ rows and columns of Σ_e corresponding to M_t are zeros. Applying block inversion to Ω_t and collecting the $\mathcal{L} \times \mathcal{L}$ block corresponding to the yield portfolios $\mathcal{P}^{\mathcal{L}}_t$, it can be seen that we need $\Sigma_{e\mathcal{L}}S_t^{-1}$ to be small.

E Filtering Invariance of the Variance Parameters

The term structure corresponding to our canonical form with the observable risk factors Z_t can be obtained by substituting (A13) into (A18):

$$y_t = A_X + B_X(W^{\mathcal{N}}B_X)^{-1}(\Gamma_1^{-1}(Z_t - \Gamma_0) - W^{\mathcal{N}}A_X). \tag{A46}$$

From this we can write $\mathcal{P}_t = A_{TS} + B_{TS} Z_t$, where

$$A_{TS} = \mathcal{G}\gamma_r + \beta_Z vec(\Sigma), \tag{A47}$$

$$B_{TS} = W B_X \mathcal{U}_1, \tag{A48}$$

$$\mathcal{G} = W\left((I_J - B_X(W^{\mathcal{N}} B_X)^{-1} W^{\mathcal{N}})\iota, B_X \mathcal{U}_{1,\mathcal{M}}\right), \tag{A49}$$

$$\beta_Z = W(I_J - B_X(W^{\mathcal{N}} B_X)^{-1} W^{\mathcal{N}})\beta_X(\mathcal{U}_1 \otimes \mathcal{U}_1), \tag{A50}$$

$\gamma_r' = (r_\infty^{\mathbb{Q}}, \gamma_0')$, 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. Therefore, from (7), the errors in pricing $\mathcal{P}_t$ are given by

$$e_t = \mathcal{P}_t^o - \mathcal{G}\gamma_r - \beta_Z vec(\Sigma) - B_{TS} Z_t. \tag{A51}$$

Since

$$f(\mathcal{P}_t^o | Z_t; \Theta^{\mathbb{Q}}, \Sigma_e) = (2\pi)^{-J/2} |\Sigma_e|^{-1/2} \exp\left(-\frac{1}{2} e_t' \,\Sigma_e^{-1}\, e_t\right), \tag{A52}$$

it follows that

$$\sum_t \partial \log f(\mathcal{P}_t^o | Z_t; \Theta^{\mathbb{Q}}, \Sigma_e)/\partial vec(\Sigma) = \hat{\beta}_Z' (\hat{\Sigma}_e)^{-1} \sum_t \hat{e}_t^u, \tag{A53}$$

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

References

Ang, A., J. Boivin, S. Dong, and R. Loo-Kung, 2010, "Monetary Policy Shifts and the Term Structure," Discussion paper, forthcoming, *Review of Economic Studies*.

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.

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

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

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

Duffee, G., 2011a, “Forecasting with the Term Structure: the Role of No-Arbitrage,” Discussion paper, Johns Hopkins University.

Duffee, G., 2011b, “Information In (and Not In) The Term Structure,” Discussion paper, forthcoming, *Review of Financial Studies*.

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.

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

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*.