

JFEC Invited Paper

Gaussian Macro–Finance Term Structure Models with Lags

SCOTT JOSLIN

University of Southern California, Marshall School of Business

ANH LE

Kenan-Flagler Business School, University of North Carolina

KENNETH J. SINGLETON

Graduate School of Business, Stanford University, and NBER

ABSTRACT

This article develops a new family of Gaussian macro–dynamic term structure models (MTSMs) in which bond yields follow a low-dimensional factor structure and the historical distribution of bond yields and macroeconomic variables is characterized by a vector-autoregression with order $p > 1$. Most formulations of MTSMs with $p > 1$ are shown to imply a much higher dimensional factor structure for yields than what is called for by historical data. In contrast, our “asymmetric” arbitrage-free MTSM gives modelers the flexibility to match historical lag distributions with $p > 1$ while maintaining a parsimonious factor representation of yields. Using our canonical family of MTSMs we revisit: (i) the impact of no-arbitrage restrictions on the joint distribution of bond yields and macro risks, comparing models with and without the restriction that macro risks are spanned by yield-curve information; and (ii) the identification of the policy parameters in Taylor-style monetary policy rules within MTSMs with macro risk factors and lags. (JEL: G12,E43, C58, E58)

KEYWORDS: Macro-finance term structure model, Lags, Taylor Rule Identification

Dynamic term structure models in which a subset of the pricing factors are macroeconomic variables (MTSMs) often have bond yields depending on lags of these factors.¹ As typically parameterized, such MTSMs imply that the cross-section

Address correspondence to Kenneth J. Singleton, Stanford Graduate School of Business Knight Management Center 655 Knight Way Stanford, CA 94305, or e-mail: kenneths@stanford.edu

¹See Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007), and Jardet, Monfort, and Pegoraro (2012), among others. Monfort and Pegoraro (2007) allow for higher order Markov processes in standard setting

doi:10.1093/jfinc/nbt012 Advanced Access publication August 23, 2013

© The Author, 2013. Published by Oxford University Press. All rights reserved.

For Permissions, please email: journals.permissions@oup.com

of bond yields is described by a high-dimensional factor model, one that seems over-parameterized relative to the low-dimension suggested by principal component analyses (Litterman and Scheinkman, 1991). Consider, for example, an MTSM in which the \mathcal{N} risk factors Z_t follow a vector-autoregression of order q under and risk-neutral (pricing) distribution. Then, by logic similar to the spanning arguments in Joslin, Le, and Singleton (2011) (hereafter JLS), this MTSM is observationally equivalent to one with $\mathcal{N}q$ portfolios of *contemporaneous* yields as risk factors. That is, if lag variables incrementally effect the yield curve then necessarily the contemporaneous yield curve subsumes the information in the laggard variables. Most MTSMs assume that \mathcal{N} is three or four. With $\mathcal{N}=3$ and $q=12$ (annual lags in monthly data), modelers are effectively adopting a thirty-six factor representation of the cross-section of bond yields!

In fact, using economic growth and inflation as macro factors and U.S. Treasury bond yields, and assuming that \mathcal{N} is three or four, we provide model-free evidence that the factor structure of yields calls for a first-order vector-autoregressive (VAR) model for Z_t under the risk-neutral distribution (\mathbb{Q}). These observations are not incompatible with evidence that the historical (\mathbb{P}) conditional mean of Z_t follows a $\text{VAR}^{\mathbb{P}}(p)$ with $p > 1$. Rather, it amounts to indirect evidence that the market prices of Z_t induce a lower dimensional VAR model for these risk factors under the pricing measure.

Motivated by these theoretical and empirical observations, this article develops a new family of MTSMs in which Z_t follows a $\text{VAR}^{\mathbb{P}}(p)$ (for arbitrary $p < \infty$) under the historical distribution, and a $\text{VAR}^{\mathbb{Q}}(1)$ ($q=1$) under the pricing distribution. With \mathcal{M} macro variables M_t included in Z_t , an MTSM can potentially capture the rich cyclical comovements of bond yields and the macroeconomy, while preserving the low-dimensional factor structure of yields that is so prevalent in the extant literature. Building upon JLS, the canonical form for this family of MTSMs has Z_t normalized so that the first \mathcal{M} factors are M_t and the remaining risk factors are the first $\mathcal{L} = \mathcal{N} - \mathcal{M}$ principal components (PCs) of bond yields. In this setting bond yields are affine functions of Z_t , and forecasts of future yields (and hence risk premiums) are based on a $\text{VAR}^{\mathbb{P}}(p)$ model of Z_t . Using this framework we explore two empirical questions about the properties of MTSMs.

First, affine MTSMs are often connected to the large literature on dynamic Taylor rules followed by monetary authorities.² This connection is natural given that the affine dependence of the short-term interest rate on output growth and inflation as well as other latent shocks resembles standard formulations of Taylor rules.

without macro pricing factors. This is in contrast to the vast majority of econometric studies of classical dynamic term structure models in which the latent pricing factors follow first-order Markov processes. The first-order Markov model can be traced back at least to Vasicek (1977) and Cox, Ingersoll, and Ross (1985), and this structure is adopted in Dai and Singleton (2000)'s characterization of affine models. Much of the recent literature on arbitrage-free term structure models extends the latter characterization while maintaining the first-order Markov specification.

²See, for examples, Ang, Dong, and Piazzesi (2007), Bikbov and Chernov (2010), Smith and Taylor (2009), and Ang *et al.* (2011).

However, using a benchmark formulation of this short-rate dependence on M_t , and allowing for a flexible $\text{VAR}^{\mathbb{P}}(p)$ specification of Z_t , we argue that the parameters of a Taylor rule are not econometrically identified within reduced-form MTSMs. This is illustrated with several *observationally equivalent* representations of the short rate that have very different weights on output growth and inflation.

Second, as emphasized by Joslin, Priebsch, and Singleton (2011) (hereafter JPS), MTSMs that include M_t as risk factors imply that these macro variables are fully contemporaneously spanned by the low-order PCs of bond yields. In fact, there are large components of many macro factors that are unspanned by yields and that also have predictive power for excess returns (Ludvigson and Ng (2010), JPS). This implies that the impulse responses (IRs) of M_t to yield-curve shocks and vice versa are more accurately portrayed in MTSMs that accommodate unspanned macro risks. Accordingly, of central importance to the literature that enforces macro-spanning is whether the joint distribution of bond yields, macro factors, and risk premiums are distorted by this constraint. To address this issue we compare the model-implied IRs within canonical MTSMs with $p=1$ and $p>1$ and macro-spanning enforced to those implied by MTSMs in which the macro-spanning constraint is relaxed.

The remainder of this article is organized as follows. Section 1 provides model-free evidence, based on commonly studied macro risk factors and bond yields, suggesting that the factor structure of affine MTSMs is best described by a low-dimensional Z_t following a $\text{VAR}^{\mathbb{Q}}(1)$. Our canonical MTSM with lags under the \mathbb{P} distribution and a $\text{VAR}^{\mathbb{Q}}(1)$ for Z_t under \mathbb{Q} , is developed in Section 2. In this section we also compare our framework to several complementary modeling strategies in the literature. The issue of whether insights about the structure of Taylor rules can be reliably extracted from MTSMs is taken up in Section 3. The impact of no-arbitrage restrictions on the fitted impulse response functions from MTSMs with and without macro-spanning enforced is explored empirically in Section 4.

To fix notation, suppose that an MTSM is evaluated using a set of J yields $y_t = (y_t^{m_1}, \dots, y_t^{m_J})'$ with maturities (m_1, \dots, m_J) with $J > \mathcal{N}$. We introduce a fixed, full-rank matrix of portfolio weights $W \in \mathbb{R}^{J \times J}$ and define the “portfolios” of yields $\mathcal{P}_t = Wy_t$ 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. The notation $\mathcal{C}^{\mathcal{N}}$ is used for the first \mathcal{N} PCs of the bond yields, a special case of $\mathcal{P}^{\mathcal{N}}$. We use the notation y_t^o to differentiate between observed yields and their theoretical counterpart y_t .

1 THE LAG STRUCTURE OF AFFINE BOND PRICING MODELS

Before exploring the properties of specific MTSMs with higher order lags, we review the spanning condition for the macro factors M_t derived in JLS for first-order Markov Z_t under \mathbb{Q} , and the observational equivalence of this class of models

to one in which yield portfolios are used as risk factors. Next, we examine the potential impact on spanning of introducing higher order lags. Finally, we examine some model-free evidence on the order of the VAR^Q called for by the standard risk factors in MTSMs. Throughout most of this section we are agnostic about the order of the Markov process Z_t under \mathbb{P} as our focus is on the spanning properties of MTSMs implied by arbitrage-free pricing. Indeed, as we discuss in Section 2, within a canonical setting, the order of the VAR^P(p) representation of Z_t is unconstrained by a researcher’s chosen order of the VAR^Q(q) process for Z_t .

1.1 Theoretical Spanning and Observational Equivalence for First-Order Markov Models

Suppose that \mathcal{M} macroeconomic variables M_t enter an 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 $\mathcal{P}_t^{\mathcal{L}}$,

$$r_t = \rho_0 + \rho_{1M} \cdot M_t + \rho_{1P} \cdot \mathcal{P}_t^{\mathcal{L}} \equiv \rho_0 + \rho_1 \cdot Z_t, \quad Z_t = (M_t', \mathcal{P}_t^{\mathcal{L}'})'. \tag{1}$$

The risk factors Z_t are assumed to follow a Gaussian VAR^Q(1) process:

$$Z_t = K_{0Z}^{\mathbb{Q}} + K_{1Z}^{\mathbb{Q}} Z_{t-1} + \sqrt{\Sigma_Z} \epsilon_t^{\mathbb{Q}}, \quad \epsilon_t^{\mathbb{Q}} \sim N(0, I). \tag{2}$$

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

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

where the loadings (A_{TS}, B_{TS}) are known functions of the parameters ($K_{0Z}^{\mathbb{Q}}, K_{1Z}^{\mathbb{Q}}, \rho_0, \rho_1, \Sigma_Z$) governing the risk-neutral distribution of yields in the term structure (“TS”) model.

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 . This follows immediately from inversion of the first \mathcal{N} rows of (3) to express Z_t as an affine function $\mathcal{P}_t^{\mathcal{N}}$:

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

Starting from the canonical form for yield-based \mathcal{N} -factor models of Joslin, Singleton, and Zhu (2011) (JSZ), (5) leads to a canonical form for this MTSM 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) (see JLS).

1.2 Theoretical Spanning and Observational Equivalence for Higher Order Markov Models under \mathbb{Q}

To extend this discussion to higher order Markov models, suppose that the \mathcal{N} risk factors Z_t follow the $\text{VAR}^{\mathbb{Q}}(q)$ process:

$$Z_t = K_{0Z}^{\mathbb{Q}} + K_{1Z}^{\mathbb{Q}}(L; q)Z_{t-1} + \sqrt{\Sigma_Z} \epsilon_t^{\mathbb{Q}}, \tag{6}$$

where $K_{1Z}^{\mathbb{Q}}(L; q)$ is a lag polynomial of order $\infty > q \geq 1$, and r_t is an affine function of the q -history $\mathcal{Z}_t^q \equiv (Z_t, Z_{t-1}, \dots, Z_{t-q+1})$. Here Σ_Z has full rank so \mathcal{N} is minimal in the sense of being the minimal number of shocks underlying variation of bond yields under the \mathbb{Q} measure.

In this extended setting y_t is again affine in the pricing factors, though now with q lags:³

$$y_t = A_Z + B_Z(L; q)Z_t. \tag{7}$$

As long as $J \geq q\mathcal{N}$ —the number of bonds in the cross-section is at least as large as the number of current and lagged Z in (7)—and given A_Z and $B_Z(L; q)$, the expression for Wy_t implied by (7) can be solved for \mathcal{Z}_t^q using any full-rank $q\mathcal{N} \times J$ portfolio weight matrix W . Moreover, the model-implied Z_t inverted from Wy_t or from any Wy_{t+s} , for $0 \leq s \leq q-1$, are identical. It follows that \mathcal{Z}_t^q is fully spanned by the contemporaneous cross-section of yields y_t . This is the generalization of the spanning relation (5) that we derived for the case of $q=1$. In particular, any macro factors in Z_t , as well as their q lags that may appear in (7), remain fully spanned by contemporaneous bond yields. This is true regardless of whether the \mathcal{L} non-macro factors in Z_t are latent or observed portfolios of yields.

Though this formulation retains theoretical spanning, with $q > 1$ it is no longer possible to apply simple rotations to models with latent factors to obtain a counterpart to JLS’s canonical form for the case of $q=1$. For suppose that we start with an MTSM in which the risk factors X_t are a mix of latent and macro variables and satisfy (6). Then, in general, it is not possible to find an $\mathcal{N} \times J$ portfolio matrix W such that $\mathcal{P}_t^{\mathcal{N}} = Wy_t$ can be substituted for X_t in (6). Instead, premultiplying (7) by W and inverting $B_Z(L; q)$, we can express X_t as an infinite-order distributed lag of $\mathcal{P}_t^{\mathcal{N}}$, and when this expression is substituted into (6) the resulting time-series model for $\mathcal{P}_t^{\mathcal{N}}$ inherits this infinite order.⁴

An important practical implication of this observation is that two MTSMs with identical $(\mathcal{M}, \mathcal{L})$ and macro factors M , one with \mathcal{L} latent pricing factors ℓ_t and the other with \mathcal{L} yield portfolios $\mathcal{P}_t^{\mathcal{L}}$ as pricing factors, are not equivalent when

³Generically, $B(L; q) = B_1 + B_2L + \dots + B_qL^{q-1}$ generates weighted, summed q -histories.

⁴A simple illustration of this point is as follows: consider a one-factor model with r_t being the factor, and suppose that $Z_t = [r_t, r_{t-1}]$ is first-order Markov. Then in general the two-period bond yield will take the form $y_t^2 = a + b_1r_t + b_2r_{t-1}$ and this implies that $E_t^{\mathbb{Q}}[y_{t+1}]$ will be affine in $[r_t, r_{t-1}]$. This in turn implies that, except in degenerate cases, $E_t^{\mathbb{Q}}[y_{t+1}]$ is not affine in $[y_t, y_{t-1}]$ and so $[y_t, y_{t-1}]$ does not follow a first-order Markov process.

$q > 1$. In particular, the MTSMs studied by Jardet, Monfort, and Pegoraro (2012) and Ang, Dong, and Piazzesi (2007), when matched with identical $(\mathcal{M}, \mathcal{L})$ and M_t , are not equivalent theoretical models.

One special case where theoretical equivalence carries over to the presence of lags is when these lags appear only on the macro factors M_t . Specifically, consider the MTSM in which $Z'_t = (M'_t, \ell'_t)$ and r_t is an affine function of ℓ_t and the q -history \mathcal{M}_t^q of the macro factors:

$$r_t = \rho_0 + \rho_{\mathcal{M}} \cdot \mathcal{M}_t^q + \rho_{\mathcal{L}} \cdot \ell_t; \tag{8}$$

and $Z_t = (M'_t, \ell'_t)'$ follows the Gaussian process:

$$\begin{pmatrix} M_t \\ \ell_t \end{pmatrix} = \begin{pmatrix} K_{0M}^{\mathbb{Q}} \\ K_{0\ell}^{\mathbb{Q}} \end{pmatrix} + \begin{pmatrix} K_{1MM}^{\mathbb{Q}} & K_{1M\ell}^{\mathbb{Q}} \\ K_{1\ell M}^{\mathbb{Q}} & K_{1\ell\ell}^{\mathbb{Q}} \end{pmatrix} \begin{pmatrix} \mathcal{M}_{t-1}^q \\ \ell_{t-1} \end{pmatrix} + \sqrt{\Sigma_X} \epsilon_t^{\mathbb{Q}}, \tag{9}$$

where $\epsilon_t^{\mathbb{Q}}$ is distributed as $N(0, I)$. This formulation nests the Macro Lag MTSM of Ang and Piazzesi (2003) (it is the special case with $q=12$).

Given this structure, bonds yields satisfy

$$y_t = A_X(\Theta^{\mathbb{Q}}) + B_M(\Theta^{\mathbb{Q}}) \mathcal{M}_t^q + B_{\ell}(\Theta^{\mathbb{Q}}) \ell_t, \tag{10}$$

where $\Theta^{\mathbb{Q}}$ denotes the set of all risk-neutral parameters that govern (8) and (9). Theoretical spanning implies ℓ_t must be spanned by *any* \mathcal{L} portfolios of yields, $\mathcal{P}_t^{\mathcal{L}}$, and \mathcal{M}_t^q . Applying the weights $W^{\mathcal{L}}$ of $\mathcal{P}_t^{\mathcal{L}}$ to (10) and inverting for ℓ_t , we can write:

$$\ell_t = (W^{\mathcal{L}} B_{\ell})^{-1} (\mathcal{P}_t^{\mathcal{L}} - P^{\mathcal{L}} A_X - P^{\mathcal{L}} B_M \mathcal{M}_t^q). \tag{11}$$

Substituting (11) in (8) and (9), we recover an equivalent model where $\mathcal{P}_t^{\mathcal{L}}$ replaces ℓ_t in the vector of risk factors $Z'_t = (M'_t, \mathcal{P}_t^{\mathcal{L}'})$:

$$r_t = \rho_0 + \rho_{\mathcal{M}} \cdot \mathcal{M}_t^q + \rho_{\mathcal{L}} \cdot \mathcal{P}_t^{\mathcal{L}}, \text{ and} \tag{12}$$

$$\begin{pmatrix} M_t \\ \mathcal{P}_t^{\mathcal{L}} \end{pmatrix} = \begin{pmatrix} K_{0M}^{\mathbb{Q}} \\ K_{0\mathcal{P}}^{\mathbb{Q}} \end{pmatrix} + \begin{pmatrix} K_{1MM}^{\mathbb{Q}} & K_{1M\mathcal{P}}^{\mathbb{Q}} \\ K_{1\mathcal{P}M}^{\mathbb{Q}} & K_{1\mathcal{P}\mathcal{P}}^{\mathbb{Q}} \end{pmatrix} \begin{pmatrix} \mathcal{M}_{t-1}^q \\ \mathcal{P}_{t-1}^{\mathcal{L}} \end{pmatrix} + \sqrt{\Sigma_Z} \epsilon_t^{\mathbb{Q}} \text{ under } \mathbb{Q}. \tag{13}$$

Thus, in this setting, without loss of generality we can rotate the factors in an MTSM so that any latent factors are replaced by portfolios of bond yields of the modeler's choosing. This representation is canonical in the sense of being maximally flexible within this family of MTSMs with q lags of M_t and one lag of either ℓ_t or $\mathcal{P}_t^{\mathcal{L}}$ under \mathbb{Q} .

We emphasize that, in obtaining theoretical equivalence, the key transition from (10) to (11) relies critically on the assumption that Z_t depends only on a single lag of ℓ_t under \mathbb{Q} . If instead there is dependence on a history of ℓ_t , this transition will typically break down since the right-hand side of (11) would depend on an infinite number of lags of $\mathcal{P}_t^{\mathcal{L}}$.

1.3 Empirical Evidence on the Factor Structure of Bond Yields

If Z_t follows a $\text{VAR}^{\mathbb{Q}}(q)$ 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 $\mathcal{N}q$ portfolios of yields (see Section 1.2). It follows that setting $q > 1$ under the pricing distribution effectively increases the number of pricing factors from \mathcal{N} to $\mathcal{N}q$. In many MTSMs the choice of $q > 1$ under \mathbb{Q} arises as a consequence of the assumptions that Z_t follows a $\text{VAR}^{\mathbb{P}}(p)$ under \mathbb{P} with $p > 1$ in the presence of flexible market prices of risk. The question of whether the data call for $q > 1$ is often not addressed directly.

Within the family of MTSMs, 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^o . The null that Z_t follows the first-order $\text{VAR}^{\mathbb{Q}}(1)$ in (2) implies that projections of yields onto current and lagged values of Z_t^o 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} .

Now, strictly speaking, these projections only address variants of models in which the observed yield portfolios $\mathcal{P}_t^{\mathcal{L}o}$ and their theoretical model counterparts $\mathcal{P}_t^{\mathcal{L}}$ are identical. However, for our empirical analysis we set $\mathcal{L} = 2$ and choose the weight matrix W so that \mathcal{P}_t^2 is comprised of the first two PCs of $y_t^o, C_t^{2'o} = (\text{PC}1_t, \text{PC}2_t)$. For the reasons given in JLS, the filtered $(\text{PC}1_t^f, \text{PC}2_t^f)$ are nearly identical to $(\text{PC}1_t^o, \text{PC}2_t^o)$ in the variants of our models that accommodate measurement errors ($C_t^{2'o} \neq C_t^{2'o}$). Intuitively, the reason for this is that, even if individual bonds are measured with substantial errors, the diversification effect from the formation of portfolios implies accurate pricing of the low-order PCs. Therefore, we expect the following results to be robust to pricing errors on individual bond yields.

To address the order under \mathbb{Q} empirically we consider several variants of MTSMs with M_t comprised of a measure of real economic growth (g_t) and inflation (π_t). We follow Ang and Piazzesi (2003) and use the first PC of the help wanted index, unemployment, the growth rate of employment, and the growth rate of industrial production (REALPC) as our measure of g , 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 . The monthly zero yields are the unsmoothed Fama–Bliss series for maturities 3- and 6-months, and 1–10 years ($J = 12$) over the sample period 1972–2003.

Model $\text{GM}_3(g)$ has $(\mathcal{N} = 3, \mathcal{M} = 1)$; model $\text{GM}_3(g, \pi)$ has $(\mathcal{N} = 3, \mathcal{M} = 2)$; and model $\text{GM}_4(g, \pi)$ has $(\mathcal{N} = 4, \mathcal{M} = 2)$. Their associated sets of macro risk factors M_t of Z_t are given in Table 1. This table also presents the root-mean-squared projection errors in basis points, for lag lengths $q = 1, 6, 12$ months (i.e., $q = 12$ is 1 year). For two of the three cases, the improvements in fit from setting $q > 1$ are tiny, at most one or two basis points. The exception is $\text{GM}_3(g, \pi)$ with state vector (REALPC, INFPC, PC1). In this case the fit is very bad, with root-mean-squared errors (RMSEs) as large

Table 1 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^0 . For given q , the conditioning information is $(Z_t, Z_{t-1}, \dots, Z_{t-q+1})$.

Model	M_t	q	$y_t^{0.5 \text{ yr}}$	$y_t^{1 \text{ yr}}$	$y_t^{2 \text{ yr}}$	$y_t^{5 \text{ yr}}$	$y_t^{7 \text{ yr}}$	$y_t^{10 \text{ yr}}$
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, π)	REALPC, INFPC	1	60	45	22	25	36	47
GM ₃ (g, π)	REALPC, INFPC	6	57	42	20	24	34	45
GM ₃ (g, π)	REALPC, INFPC	12	53	37	18	22	31	40
GM ₄ (g, π)	REALPC, INFPC	1	9	15	17	10	10	18
GM ₄ (g, π)	REALPC, INFPC	6	8	14	16	9	10	17
GM ₄ (g, π)	REALPC, INFPC	12	8	13	15	9	9	16

as sixty basis points. Consequently, adding lags under \mathbb{Q} improves the RMSEs by up to eight basis points. In all cases, AIC and BIC model selection criteria select the lag length $q=1$.

This evidence is period and bond-market specific. For a different sample period spanning 1964–1995, Monfort and Pegoraro (2007) found that $q > 1$ within a parametric regime-switching VAR model was useful for fitting the dramatic changes in the distributions of bond yields that arose from changes in the Federal Reserve’s policy rules. Our empirical analysis focuses on what is effectively one of the post-war regimes (see Dai, Singleton, and Yang, 2007). For this regime, $q=1$ appears adequate.

2 MTSMs WITH LAGS UNDER \mathbb{P}

Supported by the evidence in Table 1, we proceed to explore asymmetric formulations of MTSMs in which Z_t follows a VAR $^{\mathbb{P}}$ (p) under \mathbb{P} , and a VAR $^{\mathbb{Q}}$ (1) under \mathbb{Q} , the family MTSM $^{\mathcal{N}}_{\mathcal{M}}(p)$.⁵ Since Z_t follows the first-order Markov process (2) under \mathbb{Q} we adopt the JLS normalization under which $Z_t' = (M_t', c_t^{\mathcal{L}'})$, so $\mathcal{P}_t^{\mathcal{L}}$ is rotated to the first \mathcal{L} PCs of y_t^0 , M_t continues to satisfy the spanning condition (4), and r_t is given by (1). Furthermore, building upon JSZ, the parameters governing (1) and (2) are expressed as explicit functions of $\Theta_{TS}^{\mathbb{Q}} = (r_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \gamma_0, \gamma_1, \Sigma)$, where $r_{\infty}^{\mathbb{Q}}$ is the long-run \mathbb{Q} -mean of r_t and (γ_0, γ_1) are the weights in the macro-spanning condition (4).

⁵While, as we have seen, we could accommodate lags in M under \mathbb{Q} , we restrict our attention to the case of $q=1$ based on the evidence in Section 1.

Where our MTSM with lags differs from the standard models discussed in Dai and Singleton (2000), Collin-Dufresne, Goldstein, and Jones (2008), and JSZ is in the specification of the \mathbb{P} -dynamics of Z_t . Instead of their first-order Markov models, we allow the conditional \mathbb{P} distribution of Z_t to be governed by the $\text{VAR}^{\mathbb{P}}(p)$ process

$$Z_t = K_{0Z}^{\mathbb{P}} + K_{1Z}^{\mathbb{P}} Z_{t-1}^p + \sqrt{\Sigma_Z} \epsilon_t^{\mathbb{P}}. \tag{14}$$

This formulation nests the \mathbb{P} distributions studied by Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007), and Jardet, Monfort, and Pegoraro (2012), among others.

Note that having Z_t follow a $\text{VAR}^{\mathbb{Q}}(1)$ process under the pricing measure and a $\text{VAR}^{\mathbb{P}}(p)$ process under the data-generating process are fully compatible assumptions. In fact, given $q \geq 1$, we are free to set p in the $\text{VAR}^{\mathbb{P}}(p)$ model for Z_t to any non-negative integer value through our choice of the market prices of Z_t risks. To illustrate this point, consider the case where Z_t includes the first \mathcal{N} PCs of bond yields and y_t is an affine function of Z_t ($q=1$). Since $Z_t = Wy_t = C_t^{\mathcal{N}}$, for suitably chosen portfolio matrix W , it follows that the loading matrices in (3) satisfy $WA_{TS} = 0$ and $WB_{TS} = I_{\mathcal{N}}$. However, the $\text{VAR}^{\mathbb{P}}$ structure of Z remains unconstrained, since the market prices of risk may depend on multiple lagged values of Z_t , in which case so will the \mathbb{P} distribution of Z_t . Similar reasoning applies to MTSMs in which some of the risk factors are macro variables.

2.1 A Canonical MTSM with Lags under \mathbb{P}

Specializing (12) and (13) to $q=1$ gives

$$r_t = \rho_0 + \rho_{\mathcal{M}} \cdot M_t + \rho_{\mathcal{L}} \cdot \mathcal{P}_t^{\mathcal{L}}, \tag{15}$$

$$\begin{pmatrix} M_t \\ \mathcal{P}_t^{\mathcal{L}} \end{pmatrix} = \begin{pmatrix} K_{0M}^{\mathbb{Q}} \\ K_{0\mathcal{P}}^{\mathbb{Q}} \end{pmatrix} + \begin{pmatrix} K_{1MM}^{\mathbb{Q}} & K_{1M\mathcal{P}}^{\mathbb{Q}} \\ K_{1\mathcal{P}M}^{\mathbb{Q}} & K_{1\mathcal{P}\mathcal{P}}^{\mathbb{Q}} \end{pmatrix} \begin{pmatrix} M_{t-1} \\ \mathcal{P}_{t-1}^{\mathcal{L}} \end{pmatrix} + \sqrt{\Sigma_Z} \epsilon_t^{\mathbb{Q}} \text{ under } \mathbb{Q}. \tag{16}$$

We complete our specification by assuming that Z_t follows the $\text{VAR}^{\mathbb{P}}(p)$ process under \mathbb{P} .⁶

$$\begin{pmatrix} M_t \\ \mathcal{P}_t^{\mathcal{L}} \end{pmatrix} = \begin{pmatrix} K_{0M}^{\mathbb{P}} \\ K_{0\mathcal{P}}^{\mathbb{P}} \end{pmatrix} + \begin{pmatrix} K_{1MM}^{\mathbb{P}} & K_{1M\mathcal{P}}^{\mathbb{P}} \\ K_{1\mathcal{P}M}^{\mathbb{P}} & K_{1\mathcal{P}\mathcal{P}}^{\mathbb{P}} \end{pmatrix} \begin{pmatrix} M_{t-1}^p \\ \mathcal{P}_{t-1}^{\mathcal{L}p} \end{pmatrix} + \sqrt{\Sigma_Z} \epsilon_t^{\mathbb{P}}. \tag{17}$$

To ensure that we end up with a canonical and econometrically identified MTSM, normalizations are required. Here we follow the literature and impose

⁶With some slight abuse of our notation, $\mathcal{P}_t^{\mathcal{L}p}$ denotes the p -history of $\mathcal{P}_t^{\mathcal{L}}$. Extension to the case where the number of lags of \mathcal{P}_t and M_t are different in (17) is straightforward.

normalizations under \mathbb{Q} and, since $q=1$, our pricing distribution is identical to the one examined in JLS. That is, (15) and (16) constitute a standard \mathbb{Q} -affine pricing model in which (effectively, as with essentially all \mathcal{N} -factor MTSMs without lags under \mathbb{Q}) the priced risks are the first \mathcal{N} yield portfolios $\mathcal{P}^{\mathcal{N}}$.

Given this pricing equivalence, we proceed by adopting the computationally convenient JLS normalization scheme. Specifically, consider the latent \mathcal{N} -factor model with

$$r_t = \rho_{0X} + \rho_{1X} \cdot X_t, \tag{18}$$

$$X_t = K_{0X}^{\mathbb{Q}} + K_{1X}^{\mathbb{Q}} X_{t-1} + \sqrt{\Sigma_X} \epsilon_t^{\mathbb{Q}}, \epsilon_t^{\mathbb{Q}} \sim N(0, I). \tag{19}$$

We assume that the eigenvalues of $K_{1X}^{\mathbb{Q}}$ are non-zero, real and distinct,⁷ whence we impose the normalizations

$$K_{0X}^{\mathbb{Q}} = 0, K_{1X}^{\mathbb{Q}} = \text{diag}(\lambda^{\mathbb{Q}}), \rho_{1X} = (1, \dots, 1)', \rho_{0X} = r_{\infty}^{\mathbb{Q}}, \tag{20}$$

where $\lambda^{\mathbb{Q}}$ is the \mathcal{N} -vector of eigenvalues of $K_{1X}^{\mathbb{Q}}$ and $r_{\infty}^{\mathbb{Q}}$ is the long-run \mathbb{Q} -mean of r_t . Under these normalizations we can rotate X_t so that the \mathcal{N} factors are $C_t^{\mathcal{N}}$, the first \mathcal{N} PCs of bond yields (see JSZ). The loadings A_{TS} and B_{TS} in (3) are fully determined by the parameters $(r_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_{CC})$ and the weights $W^{\mathcal{N}}$, where Σ_{CC} is the conditional covariance matrix of $C_t^{\mathcal{N}}$. The pricing piece of our MTSM is then completed by enforcing the macro-spanning condition (4), the incremental economic content of an MTSM. The full set of parameters governing the \mathbb{Q} distribution of $Z'_t = (M'_t, C_t^{\mathcal{L}'})$ is $\Theta^{\mathbb{Q}} = (r_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_Z, \gamma_0, \gamma_1)$.

Implicit in this construction is the assumption 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} . In terms of the parameters of our canonical form, we require that none of the eigenvectors of the risk-neutral feedback matrix $K_{1X}^{\mathbb{Q}}$ is orthogonal to the loadings ρ_{1X} for the short rate (see Joslin, 2011). Finally, to maintain valid transformations between alternative choices of risk factors, we require that the matrices $W^{\mathcal{N}} B_{TS}$ and γ_1 be full rank.

We summarize the above in the following theorem:

Theorem 1: Fix a full-rank portfolio matrix $W \in \mathbb{R}^{J \times J}$, and let $\mathcal{P}_t = Wy_t$. Any canonical form for the family of models $\text{MTSM}_{\mathcal{M}}^{\mathcal{N}}(p)$ is observationally equivalent to a unique member of $\text{MTSM}_{\mathcal{M}}^{\mathcal{N}}(p)$ in which the first \mathcal{M} components of the pricing factors are the macro variables M_t , and the remaining \mathcal{L} components are $C_t^{\mathcal{L}}$; r_t is given by (15); M_t is spanned by $C_t^{\mathcal{N}}$ as long as $J \geq \mathcal{M} + \mathcal{L}$; the risk factors follow the Gaussian process (16) under \mathbb{Q} and (17) under \mathbb{P} , where $K_{0Z}^{\mathbb{Q}}, K_{1Z}^{\mathbb{Q}}, \Sigma_{CC}, \rho_0, \rho_{\mathcal{M}}$, and $\rho_{\mathcal{L}}$ are explicit functions of $\Theta^{\mathbb{Q}}$. For given W , our canonical form is parametrized by $\Theta^{\text{TS}} = (r_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_Z, \gamma_0, \gamma_1, K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}})$.

⁷These assumptions can be relaxed along the lines of the analysis in JSZ.

The proof follows the development in JLS for the case of $p=1$, suitably adjusted to accommodate the lags in Z_{t-1}^p .

Though we have structured our canonical form to have Z_t following a $\text{VAR}^{\mathbb{Q}}(1)$, from the discussion in Section 1.2 it is evident that our canonical form is easily extended to the case of (9) where Z_t depends only on the first lag of the yield portfolios $\mathcal{P}_t^{\mathcal{L}}$ (equivalently, on the first lag of a set of latent factors), but on $q > 1$ lags of the macro factors M_t . In this extended model we would normalize $K_{1\mathcal{P}\mathcal{P}}^{\mathbb{Q}}$ to be in Jordan form and the coefficient matrices on the lagged values of M_t under \mathbb{Q} become additional free parameters of the pricing distribution.

Analogously to the construction in JLS, our canonical form 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. Relative to the JSZ canonical form with pricing factors $\mathcal{P}_t^{\mathcal{N}}$, an MTSM adds the spanning property (4) with its $\mathcal{M}(\mathcal{N} + 1)$ free parameters (γ_0, γ_1) . 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. This is true *regardless of the modeler's choice of p* , the lag structure under \mathbb{P} , because we have set $q=1$ under the pricing distribution.

2.2 The Conditional Distribution of the Pricing Factors

In taking the model to the data, we accommodate the fact that the observed data $\{M_t^o, C_t^o\}$ are not perfectly matched by a theoretical no-arbitrage model by supposing that the observed yield portfolios C_t^o are equal to their theoretical values plus a mean-zero measurement error.

The \mathbb{P} distribution of Z implied by Theorem 1 is

$$f^{\mathbb{P}}(Z_t | Z_{t-1}^p; K_{1Z}^{\mathbb{P}}, K_{0Z}^{\mathbb{P}}, \Sigma_Z) = (2\pi)^{-\mathcal{N}/2} |\Sigma_Z|^{-1/2} \exp\left(-\frac{1}{2} \|\Sigma_Z^{-1/2} (Z_t - E_{t-1}^{\mathbb{P}}[Z_t])\|_2^2\right), \tag{21}$$

where $E_{t-1}^{\mathbb{P}}[Z_t] = K_{0Z}^{\mathbb{P}} + K_{1Z}^{\mathbb{P}} Z_{t-1}^p$. Importantly, the only parameters from the set $\Theta^{\mathbb{Q}}$ that appear in (21) are those in Σ_Z . This distinctive feature of our canonical form, inherited from the JSZ normalization, will play a prominent role in our discussion of the implications of no arbitrage for the conditional distributions of bond yields.

Initially, suppose that the first \mathcal{L} entries of C_t are priced perfectly by the model ($C_t^{o\mathcal{L}} = C_t^{\mathcal{L}}$), and the last $J - \mathcal{L}$ entries, say C_t^e , are priced up to the *i.i.d.* measurement errors $e_t = C_t^{eo} - C_t^e$, $e_t \sim N(0, \Sigma_e)$.⁸ In this case, the joint (conditional) density of the data can be written as

$$f^{\mathbb{P}}(Z_t^o, C_t^{eo} | Z_{t-1}^p; \Theta^{TS}, \Sigma_e) = f^{\mathbb{P}}(C_t^{eo} | Z_t^o; \Theta^{\mathbb{Q}}, \Sigma_e) \times f^{\mathbb{P}}(Z_t^o | Z_{t-1}^p; K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}}, \Sigma_Z). \tag{22}$$

⁸In the literature on DTSMs this assumption was introduced by Chen and Scott (1993) and it has been adopted in many empirical studies of latent-factor term structure models. It is also maintained by Ang, Piazzesi, and Wei (2006) and Jardet, Monfort, and Pegoraro (2012), among others, in studies of MTSMs.

Since $Z_t^o = Z_t$, the conditional density $f^{\mathbb{P}}(Z_t | \tilde{Z}_{t-1}^p; \Theta^{\text{TS}})$ is given by (21). It depends on $(K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}}, \Sigma_Z)$, but not on $(\lambda^{\mathbb{Q}}, r_{\infty}^{\mathbb{Q}}, \gamma_0, \gamma_1)$. On the other hand, the density of C_t^{eo} is

$$f^{\mathbb{P}}(C_t^{eo} | Z_t^o; \Theta^{\mathbb{Q}}, \Sigma_e) = (2\pi)^{-(J-L)/2} |\Sigma_e|^{-1/2} \exp\left(-\frac{1}{2} e_t(\Theta^{\mathbb{Q}})' \Sigma_e^{-1} e_t(\Theta^{\mathbb{Q}})\right), \quad (23)$$

and it depends only on Σ_e and the risk-neutral parameters $\Theta^{\mathbb{Q}}$.

Optimization of (22) proceeds from initial starting values of Θ^{TS} as follows. Given $\Theta^{\mathbb{Q}}$ we compute $K_{0Z}^{\mathbb{Q}}, K_{1Z}^{\mathbb{Q}}, \rho_{0Z}$, and ρ_{1Z} using the known mappings from JSZ and the invariant transformation (5). These parameters are then used to compute bond yields as functions of the *observed* state Z_t^o using standard affine pricing formulas. Finally, the model-implied yields along with the portfolio matrix W are used to compute the conditional density $f^{\mathbb{P}}(\mathcal{P}_t^{eo} | Z_t^o; \Theta^{\mathbb{Q}}, \Sigma_e)$. As will be discussed as part of our subsequent empirical analyses, reliable starting values are easily obtained for most of the parameters in Θ^{TS} . We have found in practice that our search algorithm converges to the global optimum extremely quickly, usually in just a few seconds.

It is becoming standard practice to relax the exact pricing assumption $\mathcal{P}_t^o = P_t$ by introducing *i.i.d.* measurement errors for the entire vector of yields y_t^o . With the addition of measurement errors, ML estimation involves the use of the Kalman filter. To set up the Kalman filtering problem we start with a given set of portfolio weights $W \in \mathbb{R}^{J \times J}$ and our canonical normalization with theoretical pricing factors $Z_t' = (M_t, C_t^{\mathcal{L}'})$. From W and $\Theta^{\mathbb{Q}}$, we construct $(K_{0Z}^{\mathbb{Q}}, K_{1Z}^{\mathbb{Q}}, \rho_{0Z}, \rho_{1Z})$. Based on the no-arbitrage pricing of bonds we then construct $A_{\text{TS}}(\Theta_{\text{TS}}^{\mathbb{Q}}) \in \mathbb{R}^J$ and $B_{\text{TS}}(\Theta_{\text{TS}}^{\mathbb{Q}}) \in \mathbb{R}^{J \times \mathcal{N}}$ with $\mathcal{P}_t = A_{\text{TS}} + B_{\text{TS}} Z_t$. The observation equation is then (3) adjusted for measurement errors:

$$C_t^o = A_{\text{TS}}(\Theta^{\mathbb{Q}}) + B_{\text{TS}}(\Theta^{\mathbb{Q}}) Z_t + e_t, \quad e_t \sim N(0, \Sigma_e), \quad (24)$$

and the state equation is (17). Together (17) and (24) comprise the state-space representation of the MTSM. Note that the dimension of e_t is now J , and not $J-L$, because all J yields y_t^o are presumed to be measured with error.

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 can be relaxed as in JLS. We do not pursue this relaxation here, because JLS found that introducing measurement errors on the macro factors effectively led to M_t being driven out of the model. That is, the filtered M_t^f bore only very weak resemblance to their historical counterparts M_t^o . Enforcing $M_t = M_t^o$ allows us to preserve a central role for the macro factors in pricing and, equally importantly, it aligns our empirical analysis with the extant literature on MTSMs.

Though estimation with filtering can be challenging in terms of finding a global optimum to the likelihood function, the factorization of the likelihood function under our normalization leads directly to quite accurate starting values. First, the parameters governing the conditional \mathbb{Q} -mean of Z_t enter only through the density

of the pricing errors. Therefore, the conditional \mathbb{P} -mean of Z_t is invariant to the imposition of restrictions on the conditional \mathbb{Q} -mean of the risk factors (see also JSZ). In other words, ML estimates of the \mathbb{P} -mean parameters ($K_{0Z}^{\mathbb{P}}, K_{1Z}^{\mathbb{P}}$) can be obtained from a regression of Z_t^o on z_{t-1}^p , regardless of Σ_Z or other constraints on the \mathbb{Q} dynamics. Second, with or without lags, Σ_Z only affects the level of yields and not their factor loadings.⁹ Consequently, an estimate of Σ_Z based on the residuals from OLS estimation of a $\text{VAR}^{\mathbb{P}}(p)$ model for Z_t usually gives starting values that are very close to the ML estimates at the global optimum.

A complementary approach to parameterizing arbitrage-free factor models for excess returns on bonds has been proposed by Adrian, Crump, and Moench (2012) (ACM). The computational tractability of the (JSZ, JPS) approach is inherited by the ACM baseline parameterizations, because they normalize their pricing factors to be portfolios of yields. However, the ACM implementation of this normalization does not yield a logically consistent pricing model. There is an explicit link between the eigenvalues $\lambda^{\mathbb{Q}}$ of the risk-neutral feedback matrix in the $\text{VAR}^{\mathbb{Q}}(1)$ representation of $C_t^{\mathcal{N}}$ and the loadings B_{TS} on Z_t that we enforce. On the other hand, ACM leave the parameters of their B_{TS} free and, as such, the model-implied $C_t^{\mathcal{N}}$ will not in general match the historical PCs even though their model and estimation strategy presume such a match. Put differently, they have more free parameters than in the maximally flexible, canonical Gaussian model (Dai and Singleton, 2000), and this induces an inconsistency between actual and model-implied prices of certain portfolios of yields.

There is another important sense in which ACM follow a “limited information” estimation strategy relative to the known structure of their affine pricing models. By focusing on projections of excess returns onto the risk factors $\mathcal{P}_t^{\mathcal{N}}$, ACM do not enforce the known link between the level of bond yields and the parameters of their model. That is, they do not enforce the model-implied constraints on the conditional mean of their pricing kernel. Consequently, their estimators will in general be inefficient relative to the full-information maximum-likelihood estimators implemented in JSZ and in our analysis.

Our framework and the ACM approach offer equal flexibility in modeling the pricing errors on individual and portfolios of yields. In particular, the introduction of serially correlated measurement errors on bond yields, which alters the implied degree of serial correlation in the errors in measured holding period returns, introduces no new conceptual or practical issues for econometric analysis of our canonical model. We illustrate this point in Section 2.4.

2.3 State-Space Formulations under Alternative Hypotheses

Throughout our subsequent analysis we compare the MTSMs characterized in Theorem 1 to their “unconstrained alternatives.” Since an MTSM involves multiple

⁹Only $K_{1Z}^{\mathbb{Q}}$ and $\rho_{\mathcal{M}}$ determine the loadings of risk factors on yields.

over-identifying restrictions, the relevant alternative model depends on which of these restrictions one is interested in relaxing. We follow JLS and distinguish between three alternative formulations. The arbitrage-free “null model” for the case where all yields are measured with errors is denoted by TS^f , with the superscript f meaning Kalman filter.

The FV^f (“factor-VAR”) alternative maintains the state Equation (17), but generalizes the observation equation to

$$\mathcal{P}_t^o = A_{FV} + B_{FV}Z_t + e_t, \quad (25)$$

for conformable matrices A_{FV} and B_{FV} , with e_t normally distributed from the same family as the MTSM. For identification we normalize the first \mathcal{L} entries of A_{FV} to zero and the first \mathcal{L} rows of B_{FV} to the corresponding standard basis vectors. Except for this, A_{FV} and B_{FV} are free from any restrictions. The full parameter set $\Theta^{FV} = (A_{FV}, B_{FV}, K_{0Z}^p, K_{1Z}^p, \Sigma_Z, \Sigma_e)$ is estimated using the Kalman filter.

Special cases of models TS and FV that are also of interest are when 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). The Kalman filtering problem then simplifies to conventional ML estimation. In particular, for the FV^n model, estimation conveniently reduces to two sets of OLS regressions: a VAR for the observed risk factors Z_t^o gives the parameters in (17),¹⁰ and an OLS regression of C_t^o on Z_t^o recovers the parameters characterizing (25).

Relative to model TS^f , model FV^f 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^f 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^f to model FV^n one is relaxing both the no-arbitrage restrictions and the presumption that the state yield portfolios are measured without errors ($C_t^{\mathcal{L}} = C_t^{\mathcal{L}o}$ in model FV^n), while again maintaining the low-dimensional factor structure.

2.4 No-Arbitrage Restrictions in MTSMs with Lags

Prior to delving into various substantive economic and modeling issues, we address the basic question of whether the no-arbitrage restrictions of an MTSM impact the joint distributions of the macro factors and yield portfolios $\mathcal{P}_t^{\mathcal{L}}$. JLS found, in the

¹⁰The ML estimators of K_{0Z}^p and K_{1Z}^p are the standard OLS estimators, and the ML estimator of Σ_Z is the usual sample covariance matrix based on the OLS residuals.

Table 2 Estimated conditional standard deviations along the diagonals and correlations along the off-diagonals. The risk factors PC1, PC2, INFPC, and REALPC follow a VAR^P(3) process and estimation is by Kalman filtering.

	Model GM _{3,4} (g, π)				Corresponding model FV			
No filtering	0.0138	—	—	—	0.0138	—	—	—
	0.169	0.0052	—	—	0.172	0.00519	—	—
	0.0728	-0.0516	0.201	—	0.072	-0.0509	0.201	—
	0.204	0.0424	-0.0136	0.233	0.204	0.043	-0.0137	0.233
Filtering	0.0136	—	—	—	0.0136	—	—	—
	0.185	0.00478	—	—	0.19	0.00472	—	—
	0.0737	-0.0556	0.201	—	0.073	-0.0523	0.201	—
	0.207	0.0466	-0.014	0.233	0.209	0.0375	-0.0142	0.233

context of canonical MTSMs with $p=1$, that the impulse response functions among these factors were virtually identical to those obtained from an unconstrained factor-VAR, model FVⁿ. We address this same question in the context of a $\mathcal{N}=4$ factor MTSM with $p=3$, model GM_{3,4}(g, π). The risk factors are (INFPC, REALPC, PC1, PC2), and impulse response (IR) functions are computed using this ordering. To arrive at our choice of p we fit VAR models over a wide range of p and evaluated the optimal lag length using standard model selection criteria. The optimal p chosen by the BIC (AIC) model selection criterion was 2 (3), so we chose $p=3$ to offer maximal flexibility consistent with these criteria.

The MTSMs and factor-VARs give very similar estimates for the parameters governing the conditional mean¹¹ (not shown) and conditional covariance (Table 2) of the risk factors, even when allowing for measurement errors on all yields. Estimates of the standard deviations and correlations implied by the covariance matrix Σ_Z were computed for model GM_{3,4}(g, π) estimated with and without filtering on P_t^2 and for corresponding VAR models FV^f and FVⁿ. The estimates across all four models are very similar, consistent with our priors that Kalman filter estimates of model GM_{3,4}(g, π) would be very close to those of model FVⁿ even though the latter model ignores the no-arbitrage restrictions and the measurement errors on the yields.

Though the estimates of each individual element of the parameter set ($K_{0Z}^P, K_{1Z}^P, \Sigma_Z$) are close across models with and without arbitrage restrictions and with and without filtering, it remains to document that any small differences do not compound to impact the IRs. Figure 1 displays the IRs implied by models TS and FVⁿ within family GM_{3,4}(g, π).¹² There it is seen that the imposition of no arbitrage and the use of filtering is virtually inconsequential for how shocks to macro factors

¹¹This confirms, for this family of canonical models, that VAR estimates of (K_{0Z}^P, K_{1Z}^P) are likely to be good starting values for ML estimation of MTSMs.

¹²The choppy behavior over short horizons for some of the responses in Figure 1 is also evident in the IRs reported in Ang and Piazzesi (2003) for their MTSM with lags.

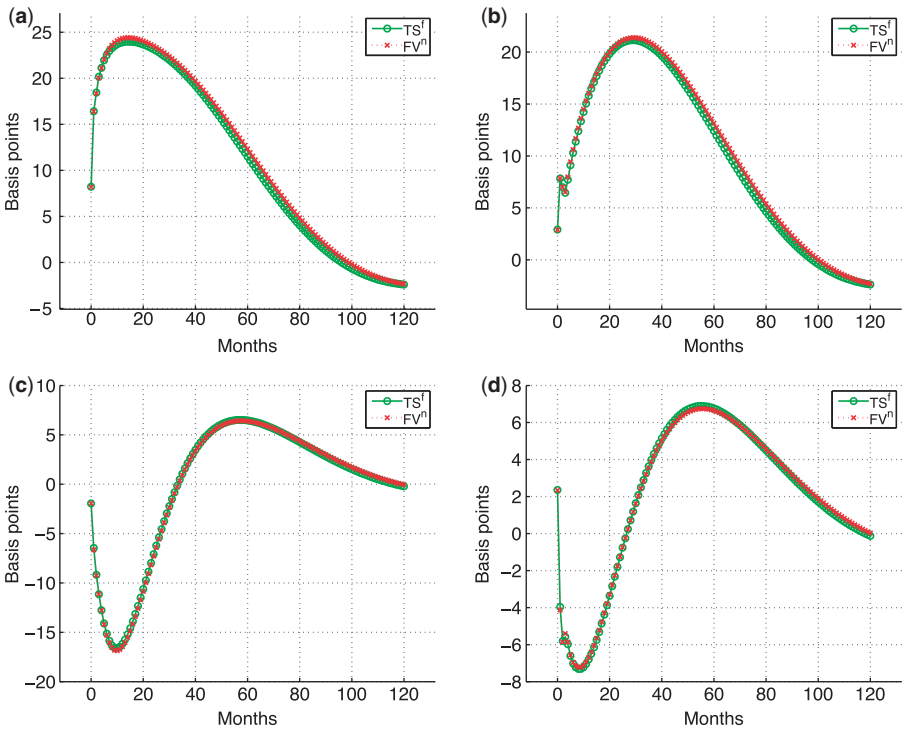


Figure 1 Impulse responses of PC1 and PC2 to innovations in REALPC and INFPC based on ML estimates of model $GM_{3,4}(g, \pi)$. (a) Response of PC1 to REALPC; (b) Response of PC1 to INFPC; (c) Response of PC2 to REALPC; and (d) Response of PC2 to INFPC.

impact the yield curve. This finding demonstrates that the results on the irrelevance of no-arbitrage restrictions of IRs for the case of $p=1$ carry over to models in which the \mathbb{P} distribution of the risk factors is governed by a higher order VAR.

Up to this point we have assumed the measurement errors on the higher order PCs appearing in (24) are serially uncorrelated. Though this assumption is commonplace in the literature, it implies that the measurement errors on holding period returns, which are quasi-differences of the yields, are persistent (ACM). Our prior, informed by the results of Dai and Singleton (2000) for yield-only models, is that the introduction of serial correlation will have little impact on the dynamic properties of the model-implied yields in fitted MTSMs.

We examine this issue in model $GM_{3,4}(g, \pi)$ by introducing first-order autoregressive measurement errors in (24). The fitted errors on the higher order PCs are in fact persistent—typical estimates of the autocorrelation coefficients are around 0.75. Figure 2 displays the responses of PC2 to shocks to REALPC and INFPC, the counterparts to panels (c) and (d) of Figure 1. It is immediately apparent that the IRs from the models with and without serially correlated measurement

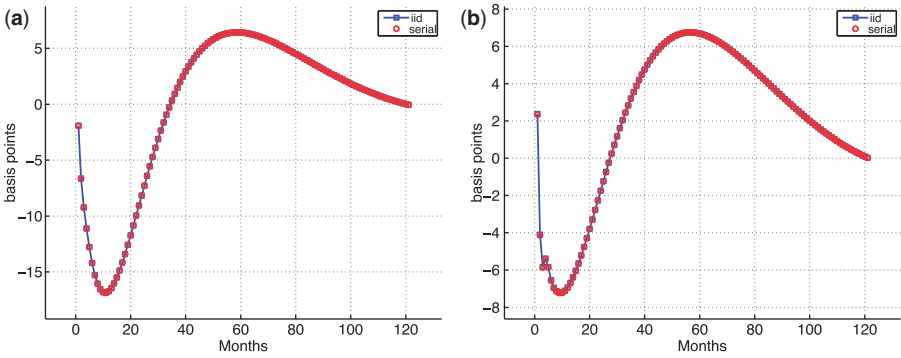


Figure 2 Impulse responses of PC2 to innovations in REALPC and INFPC based on ML estimates of model $GM_{3,4}(g, \pi)$, with measurement errors following AR(1) processes. (a) Response of PC2 to REALPC; and (b) Response of PC2 to INFPC.

errors are virtually identical. Though not shown, this is true of the other cases displayed in Figure 1, as well as the responses of the individual bond yields to innovations in the macro variables. We conclude that the MTSM-implied IRs are robust to alternative assumptions about the persistence of the measurement errors.

3 TAYLOR-STYLE RULES AND MTSMs

Several recent studies interpret their versions of the short-rate Equation (1) as a Taylor-style rule. This is premised, we suspect, on the virtually identical appearance of special cases of (1) with $M'_t = (g_t, \pi_t)$ and $\mathcal{L} = 1$,

$$r_t = \rho_0 + \rho_\pi \pi_t + \rho_g g_t + \rho_\ell \ell_t, \tag{26}$$

and policy rules that have central banks setting the short rate according to targets for inflation and the output gap. Furthermore, just as in standard structural formulations of a Taylor rule, there is also a latent shock ℓ_t to the short rate. However, without imposing additional *economic* structure on an MTSM, the parameters (ρ_π, ρ_g) are not meaningfully interpretable as the reaction coefficients of a central bank.

To see this in the simplest, benchmark case we focus on the family $GM_3(g, \pi)$ with the short rate given by (26). The latent shock ℓ_t is allowed to be correlated with M_t . Cochrane (2007) argues that it is an inherent feature of new-Keynesian models that the policy shock “jumps” in response to changes in inflation or output gaps.

Within this setting, an immediate implication of Theorem 1 is that each choice of full-rank portfolio matrix $W \in \mathbb{R}^{N \times J}$ leads to a canonical model for $GM_3(g, \pi)$ in which r_t is given by

$$r_t = \rho_0^W + \rho_\pi^W \pi_t + \rho_g^W g_t + \rho_{\mathcal{P}}^W \mathcal{P}_t^{1W}. \tag{27}$$

One choice of W places the first PC (the “level” factor) in place of ℓ_t in (26)– sets $\mathcal{P}_t^1 = \text{PC1}_t$. Another choice sets $\mathcal{P}_t^1 = \text{PC2}_t$ (the “slope” factor), etc. All choices give rise to *theoretically equivalent* representations of the short-term rate and, through the no-arbitrage term structure model, identical bond yields at all maturities. Moreover, when all yields are measured with errors, all choices of W give rise to *observationally equivalent* versions of (27). Absent the imposition of additional economic structure, there appears to be no basis for interpreting any one of these equivalent rotations (27) as *the* Taylor rule of a structural model.

Put differently, when modelers include a latent-factor ℓ_t in their short-rate specification, then the normalizations they choose select one among an uncountable infinity of equivalent normalizations. A particular normalization may lead to estimates of the loadings ρ_π and ρ_g that bear some resemblance to estimates of the policy parameters in the literature on fitting structural Taylor rules. However, this would be coincidental. Further, even in this case, ℓ_t is representable as a portfolio of bond yields and the plausibility of this portfolio being the central bank’s policy shock warrants evaluation.

For instance, Ang, Dong, and Piazzesi (2007) estimate a benchmark MTSM with a short-rate equation similar to (26). They treat their shock ℓ_t as latent, consistent with their interpretation of ℓ_t as a policy shock. At their model estimates they find that the fitted $\hat{\ell}_t$ has a correlation of 94% with the short (one-quarter) rate and 98% with the 5-year rate, indicating that $\hat{\ell}_t$ is highly correlated with the level factor PC1. Thus, through the lens of our canonical form it is seen that their normalization scheme and estimation algorithm were effectively selecting out a portfolio matrix W that set ℓ_t to the first PC of bond yields.

Motivated by their experience, we proceed to estimate the canonical version of $\text{GM}_{p,3}(g, \pi)$ under three equivalent models: \mathcal{P}_t^1 is rotated to one of PC1, PC2, or PC3. Proceeding in this manner is without loss of generality because, with all yields filtered, the fitted yields will be exactly the same regardless of which rotation of \mathcal{P}_t^1 to a theoretical PC is chosen. This is true even though the macro factors (g_t, π_t) may not (in general will not) be spanned to equal degrees by the individual low-order PCs of yields, because the spanning constraint links M_t to the entire vector \mathcal{C}_t^N . We estimate the model with \mathcal{P}_t^1 rotated to PC1; and then rotate the estimated model to obtain the cases in which \mathcal{P}_t^1 is rotated to PC2 or PC3.

Equipped with the Kalman filter/ML estimates of these models, we report the loadings on g and π as well as the standard deviation of the “residual” $\rho_{\hat{P}}^W \hat{P}_t^1$ implied by these choices of W . The version of the “Taylor rule” we examine is what is often referred to as the *benchmark* model that expresses r_t as a contemporaneous affine function of the risk factors. We set r_t to the actual 3-month Treasury bill rate for the purpose of computing residual “policy” shocks $e_t \equiv \rho_{\mathcal{P}^1}^W \mathcal{P}^{1W}$, as this is the shortest maturity we used in constructing our PCs and in estimation of the models. Comparable results were obtained with r_t replaced by the fitted 1- or 3-month yields from the MTSMs. The AIC and BIC model selection criteria pointed to different lag lengths, and we chose the longer $p=4$ in light of the use of lags out to a year in previous studies.

Table 3 Coefficients and moments of the variables in the short-rate equation under different rotations of the latent shock L_t implied by the $GM_{4,3}(g, \pi)$ model. The residual “policy” shock $e_t \equiv \rho_{\mathcal{P}^1}^W \mathcal{P}^{1W}$.

Rotation	ρ_0^W	$\rho_{\mathcal{P}^1}^W$	ρ_{π}^W	ρ_g^W	std(e)	corr(e, π)	corr(e, g)
$\mathcal{P}^1 = PC1$	-0.0111	0.2930	0.1697	0.0233	0.0271	0.4275	-0.0483
$\mathcal{P}^1 = PC2$	-0.1920	-6.299	2.1836	0.4318	0.0631	-0.7737	-0.5184
$\mathcal{P}^1 = PC3$	0.7575	-100.880	1.5159	-1.2429	0.0982	-0.2671	0.9247

The model-implied loadings for the short rate are displayed in Table 3. To facilitate comparisons of the estimated loadings to results in the literature, we scale INFPC (REALPC) to have the same standard deviation as the annualized CPI inflation rate (annualized growth rate of industrial production) over our sample period. In all three rotations $\rho_{\pi}^W > \rho_g^W$, which is consistent with MTSM-based estimates in Ang, Dong, and Piazzesi (2007) and the instrumental variables estimates reported in Clarida, Gali, and Gertler (2000) for their post-Volker disinflation periods. Despite differences in sample periods and estimation methods, the loadings reported by Ang, Dong, and Piazzesi (2007) ($\rho_{\pi}^W = 0.322$ (s.e. = 0.143) and $\rho_g^W = 0.091$ (s.e. = 0.064)) are quite close to our estimates for the rotation with $\mathcal{P}^1 = PC1$, particularly so in light of their reported standard errors.

Clearly the loadings on (g, π) change substantially with the choice of weight matrix W . This in turn has a large effect on the associated standard deviations of the shock e_t . To put the values of $\text{std}(e)$ in perspective, note that the sample standard deviation of the 3-month Treasury bill rate is 2.96%. Thus, under the rotation $\mathcal{P}^1 = PC1$, e_t and r_t have comparable volatilities. This is supported by the negative correlation between e_t and g_t . When W is chosen to select higher order PCs for \mathcal{P}^1 , the volatility of e_t increases substantially. The correlations between e_t and (g_t, π_t) —the within-month responses of a monetary authority to an increase in inflation or output growth—also change in magnitude and sign.

We stress that none of these representations of r_t can be viewed as the most “economically plausible” since, by construction, all three are observationally equivalent within the family of MTSMs being studied. The results for the rotation with $\mathcal{P}^1 = PC1$ appear to correspond most closely (in qualitative terms) with the estimates of Taylor rules in the literature, and it has the within-month response of the U.S. monetary authority focusing mostly on inflation (r_t rises with inflation). Yet absent additional identifying restrictions that pin down a unique W (and hence \mathcal{P}^1 as the policy shock), there is no basis for selecting this rotation over the infinity of other rotations with similarly plausible features.

Up to this point we have focused on model $GM_{4,3}(g, \pi)$ with $p=4$ lags under \mathbb{P} . A shortcoming of this model, evident from Table 1, is its poor fit to individual yields: root-mean-squared pricing errors are as large as 60 basis points. To investigate whether adding an additional risk factor—equivalently, under a

Table 4 Coefficients and moments of the variables in the short-rate equation under different rotations of the latent shock L_t implied by the $GM_{3,4}(g, \pi)$ model. The residual “policy” shock $e_t \equiv \rho_{\mathcal{P}^2}^W \cdot \mathcal{P}^2 W$.

Rotation	ρ_0^W	$\rho_{\mathcal{P}^2}^W$	ρ_{π}^W	ρ_g^W	$std(e)$	$corr(e, \pi)$	$corr(e, g)$	
$\mathcal{P}^2 = (PC1, PC2)$	0.0036	0.3184	0.5229	-0.0055	-0.0117	0.0297	0.5648	0.0425
$\mathcal{P}^2 = (PC1, PC3)$	-1.207	0.8459	153.960	-2.8458	2.1920	0.1897	0.5133	-0.8367
$\mathcal{P}^2 = (PC2, PC3)$	0.7348	0.8385	-92.937	1.7075	-1.3401	0.1064	-0.2995	0.9190

Taylor rule interpretation, allowing for a two-dimensional policy shock—impacts estimates of the loadings (ρ_g, ρ_{π}) , we undertake a similar exercise using the four factor model $GM_{3,4}(g, \pi)$. This model is similar in structure to those examined in Bikbov and Chernov (2010) and Chernov and Mueller (2012) in which multiple latent factors enter a short-rate equation described as a Taylor rule and in which output growth and inflation are spanned by a small number of yield PCs.

In Table 4, we report the model-implied loadings for the short rate under three different rotations of the (now two) yield portfolios \mathcal{P}_t^2 : to (PC1, PC2), (PC1, PC3), or (PC2, PC3). Under none of these rotations do we obtain estimates of (ρ_g, ρ_{π}) consistent with standard priors about interest rate rules of the Federal Reserve. For instance, with the rotation $\mathcal{P}^2 = (PC1, PC2)$, which matches the standard deviation of e_t in the previous case of three factors and $\mathcal{P}^1 = PC1$, neither of the reaction coefficients are positive! Their magnitudes are also notably smaller than in the first row of Table 3. The reason for these, possibly counter-intuitive loadings is that the shock e_t is linearly dependent on PC2 and this yield portfolio is positively correlated with output growth (note that $corr(e_t, g_t)$ has changed signs relative to the first row of Table 3). Therefore, the roles of both g_t and π_t in the short-rate equation are much attenuated and the loadings flip signs relative to priors for a Taylor rule. This is to some extent mechanical: (PC1, PC2) explain most of the variation in bond yields, with relatively weak explanatory power attributed to M_t . Yet it underscores the weak foundations of interpretations of the loadings (ρ_g, ρ_{π}) as policy reaction coefficients. Importantly, the primary change in moving from model $GM_{4,3}(g, \pi)$ to model $GM_{3,4}(g, \pi)$ was the introduction of a more flexible, two-dimensional policy shock.

4 MACRO-YIELD-CURVE DYNAMICS IN MTSMs WITH LAGS

Perhaps the most notable empirical weakness of MTSMs that include macro factors M_t in the set of pricing factors Z_t , with or without lags, is that they typically imply that M_t is spanned by the information in the current yield curve. Enforcement of the spanning condition (4) is what underlies the poor pricing performance of the $\mathcal{N} = 3$ model $GM_{p,3}(g, \pi)$ documented in JLS and Section 2.4. Expanding to $\mathcal{N} = 4$

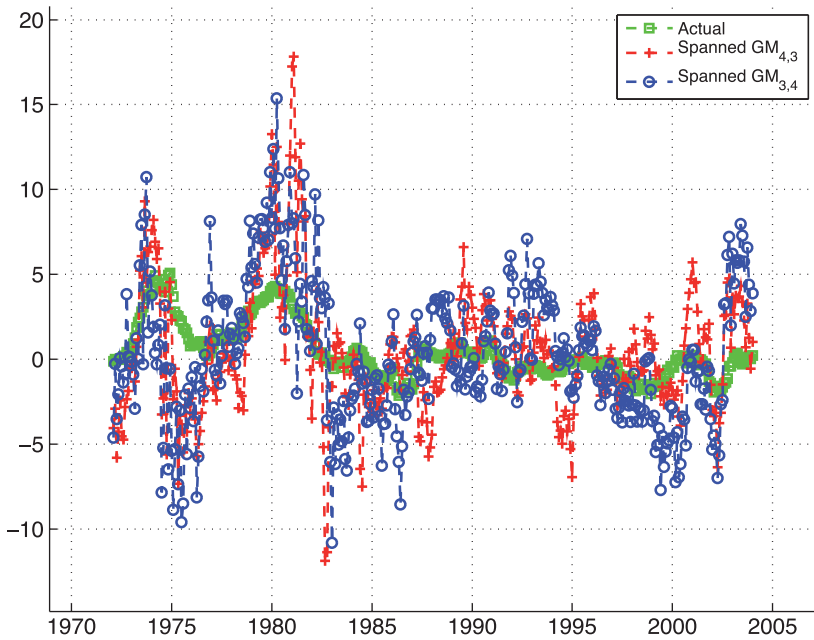


Figure 3 Actual and spanned INFPIC.

factors, holding M fixed, improves the pricing performance of MTSMs, but the counterfactual spanning restriction remains.

Which aspects of the joint distribution of macro risks and the yield curve does the spanning restriction distort? As a first look at this issue, Figures 3 and 4 display the spanned values of INFPIC and REALPC implied by models $GM_{4,3}^S(g, \pi)$ and $GM_{3,4}^S(g, \pi)$ examined in Section 3 (now with a superscript “S” to highlight its spanning property) against their historical values. Enforcement of the macro-spanning restriction leads to substantially distorted representations of the historical macro series. The spanned series, which are linear combinations of the PCs of zero-coupon bond yields, show much more volatility and exhibit much wider swings at turning points in the macro series.

Next, we examine the extent to which enforcement of the spanning restriction also distorts the IRs of the macro factors to yield-curve shocks and vice versa. Specifically, we compare the IRs implied by two three-factor MTSMs, one that enforces the spanning restrictions and one that does not. The model with spanned macro risks is the model $GM_{4,3}^S(g, \pi)$ with state vector (the factors that forecast excess returns) $Z_t^{S'} = (g_t, \pi_t, PC1_t)$. The model with unspanned macro risks, denoted $GM_{2,3}^U(g, \pi)$, maintains the assumption that there are three priced risks, but it presumes that risk premiums depend on the expanded state $Z_t^{U'} = (g_t, \pi_t, PC1_t, PC2_t, PC3_t)$. The order-selection criteria call for a higher order VAR

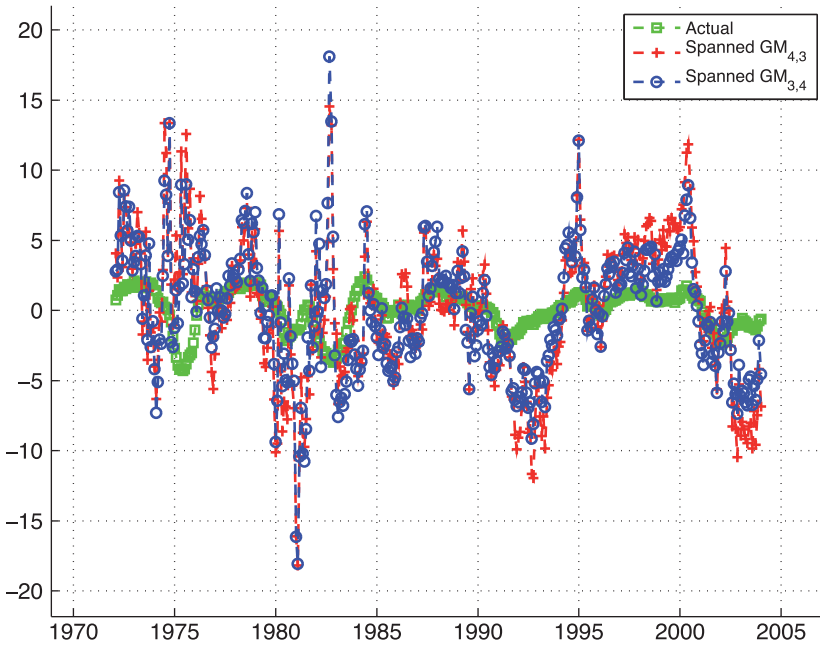


Figure 4 Actual and spanned REALPC.

representation for Z_t^S ($p=4$) than for Z_t^U ($p=2$), which is not surprising given the higher dimension of the latter state vector.

For an economy with unspanned macro risks, the focus on the subvector of states $(g_t, \pi_t, PC1_t)$ in model $GM_{4,3}^S(g, \pi)$ will in general lead to misspecified risk premiums since relevant conditioning information is omitted. Similarly, IRs among the included states will in general be inaccurate. Finally, the model with spanned macro risks is, by construction, completely silent about the impact of unspanned macro risks on the yield curve and vice versa. The remainder of this section illustrates the quantitative importance of these points.

To set up this analysis we briefly review the structure of MTSMs with unspanned M developed in JPS. Their canonical form replaces (15) and (16) with

$$r_t = \rho_{0P} + \rho_{1P} \cdot \mathcal{P}_t^N, \tag{28}$$

$$\mathcal{P}_t^N = K_{0P}^Q + K_{PP}^Q \mathcal{P}_{t-1}^N + \sqrt{\Sigma_{PP}^Q} \epsilon_{\mathcal{P}t}^Q. \tag{29}$$

Since, as we have seen, an N -factor MTSM that enforces spanning can always be rotated so that the N pricing factors are \mathcal{P}^N , this premise leads to identical priced risks as in the N -factor models encompassed by Theorem 1.

Where MTSMs with unspanned risks differ from those with spanned risks is in their specifications of the distribution of the state under \mathbb{P} . Instead of imposing the

spanning condition (4), the pricing factors $\mathcal{P}_t^{\mathcal{N}}$ and macro variables M_t are assumed to jointly follow an unconstrained VAR^{UP}(p):

$$\begin{bmatrix} \mathcal{P}_t^{\mathcal{N}} \\ M_t \end{bmatrix} = \begin{bmatrix} K_{0\mathcal{P}}^{\mathbb{P}} \\ K_{0M}^{\mathbb{P}} \end{bmatrix} + \begin{bmatrix} K_{\mathcal{P}\mathcal{P}}^{\mathbb{P}} & K_{\mathcal{P}M}^{\mathbb{P}} \\ K_{\mathcal{P}X}^{\mathbb{P}} & K_{MM}^{\mathbb{P}} \end{bmatrix} \begin{bmatrix} \mathcal{P}_{t-1}^{\mathcal{N}p} \\ \mathcal{M}_{t-1}^p \end{bmatrix} + \sqrt{\Sigma_Z} \epsilon_{Z_t}^{\mathbb{P}}, \tag{30}$$

where now $Z_t^U = (\mathcal{P}_t^{\mathcal{N}'}, M_t')$, $\epsilon_{Z_t}^{\mathbb{P}} \sim N(0, I_{\mathcal{N}+\mathcal{M}})$, the $(\mathcal{N} + \mathcal{M}) \times (\mathcal{N} + \mathcal{M})$ matrix Σ_Z is non-singular, and $\Sigma_{\mathcal{P}\mathcal{P}}$ is the upper $\mathcal{N} \times \mathcal{N}$ block of Σ_Z . Accordingly, M_t is not deterministically spanned by \mathcal{P}_t and forecasts of \mathcal{P} are conditioned on the full set of $\mathcal{N} + \mathcal{M}$ variables Z_t^U . JPS consider the special case of $p = 1$ but, using the logic of Theorem 1, their canonical form is easily extended to the case of lags under \mathbb{P} .

Analogously to (22), the conditional density of (Z_t^U, \mathcal{P}_t^e) is given by

$$f(Z_t^U, \mathcal{P}_t^e | Z_{t-1}^U; \Theta) = f(\mathcal{P}_t^e | \mathcal{P}_t^{\mathcal{N}'}; r_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \Sigma_e) \times f(Z_t^U | Z_{t-1}^U; K_{1Z}^{\mathbb{P}}, K_{0Z}^{\mathbb{P}}, \Sigma_Z). \tag{31}$$

Just as in the models with spanning, the associated likelihood function factors into a piece dependent on the risk-neutral parameters $(r_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$ and a second piece involving the parameters of the conditional \mathbb{P} -mean of Z_t^U , with the only overlap being the parameters governing Σ_Z . Therefore, the no-arbitrage structure of an MTSM with unspanned risk is also likely to be (nearly) irrelevant for studying the joint distribution Z_t^U under \mathbb{P} . This will be the case even when we allow all yields to be priced with errors, so long as the average pricing errors are small relative to their volatilities in the sense made precise in JLS.

Panels (a) and (b) of Figure 5 display the IRs of PC1—the level of the Treasury yield curve—to one standard deviation shocks in (g_t, π_t) (measured by REALPC and INFPC). Panels (c) and (d) display the reverse response of the macro factors to a shock to PC1. For comparison we have also included the IR for model $GM_{1,3}^S(g, \pi)$, since many MTSMs adopt first-order Markov representations of Z_t under both \mathbb{P} and \mathbb{Q} .

An immediately striking feature of the IRs from models $GM_{4,3}^S(g, \pi)$ and $GM_{1,3}^S(g, \pi)$ is how different they are as a consequence of changing the order of the VAR^P(p) from $p = 1$ to $p = 4$. Particularly in panel (c) the responses within the first 2 years are in opposite directions, and differences persist for nearly 8 years. One might be inclined to attribute these differences to over-fitting in the case of $p = 4$, but we reiterate that model selection criteria called for $p > 1$ and, moreover, the response patterns for model $GM_{4,3}^S(g, \pi)$ are qualitatively similar to those for model $GM_{2,3}^S(g, \pi)$ (not displayed). Inference about the joint distribution of macro risks and bond yields can be very sensitive to the order of the VAR^P(p) representation of the risk factors in MTSMs that enforce the macro-spanning restriction.

Comparing panels (a) and (b) to their counterparts in Figure 1 it is seen that models $GM_{4,3}^S(g, \pi)$, $GM_{3,4}^S(g, \pi)$, and $GM_{2,3}^U(g, \pi)$ all imply qualitatively similar responses of PC1 to the macro shocks. The timings and magnitudes of the peak responses differ somewhat across models, but the broad patterns are similar.

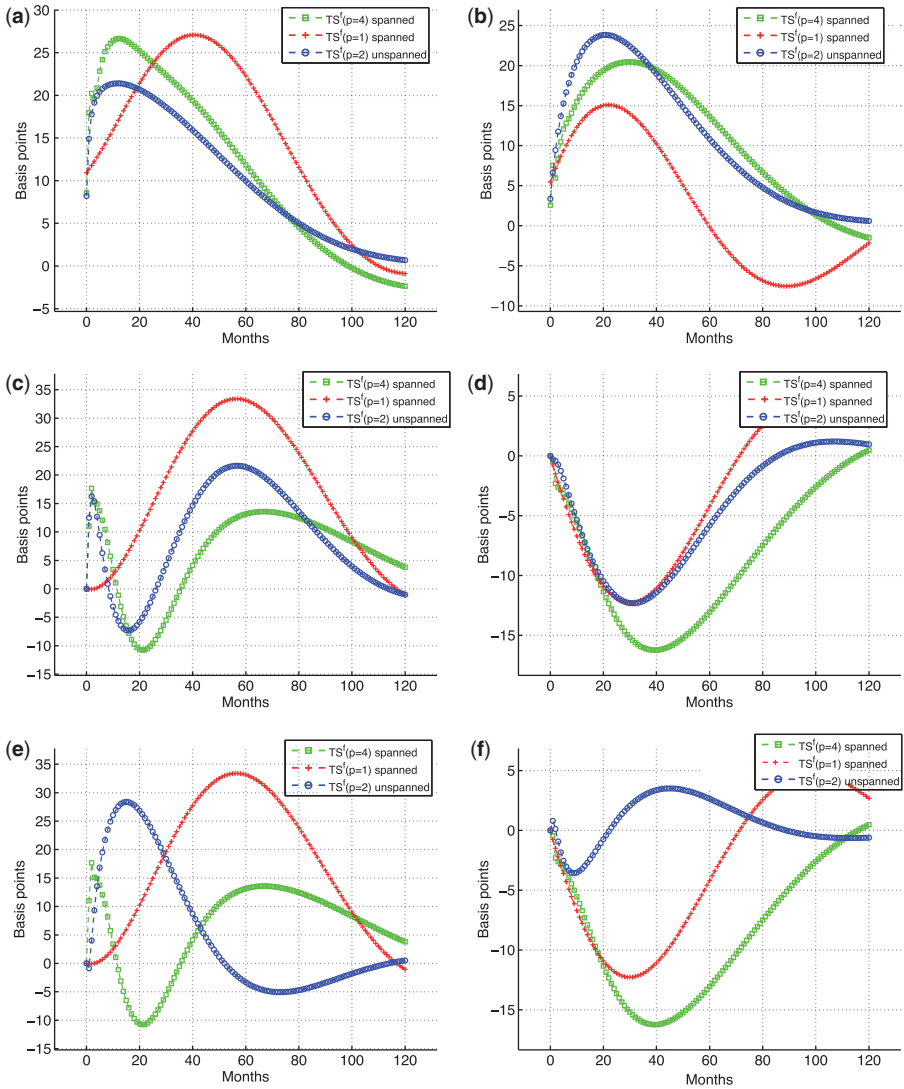


Figure 5 The states are ordered as (INFPC, REALPC, PC1, PC2, PC3) for the unspanned models and (INFPC, REALPC, PC1) for the spanned models. To draw IRs from and to PC2 implied by the spanned models, we first rotate the model to one in which the states are (INFPC, REALPC, PC2). The models with spanned risks are $GM_{4,3}^S(g, \pi)$ ($p=4$) and $GM_{1,3}^S(g, \pi)$ ($p=1$), and the model with unspanned risks is $GM_{2,3}^U(g, \pi)$. (a) Response of PC1 to REALPC; (b) Response of PC1 to INFPC; (c) Response of REALPC to PC1; (d) Response of INFPC to PC1; (e) Response of REALPC to PC2; and (f) Response of INFPC to PC2.

Evidently the joint dynamics of $(g_t, \pi_t, PC1_t)$ implied by VARs (with model-specific lag lengths) with and without the inclusion of $(PC2, PC3)$ are roughly comparable, owing to the inclusion in all three models of $(g_t, \pi_t, PC1_t)$ in the state vector Z . Similar observations apply to the responses of the macro factors to a shock in $PC1$ as displayed in Figure 5 panels (c) and (d).

At the same time, along other dimensions, the model-implied joint distributions of (M_t, y_t) are strikingly different. Consider, for instance, the joint distribution of $PC2$ and M_t . The slope of the yield-curve $PC2$ is not included as an independent source of risk in the models $GM_{p,N}^S(g, \pi)$ with spanned macro risks. Equivalently, the theoretical risk factors are the first N PCs C_t^N and (g_t, π_t) are linearly spanned by C_t^N according to (4). In contrast, $PC2$ enters both as a risk factor and as an element of the Z_t^U in the model with unspanned macro risks so it is a distinct source of risk relative to $(M_t, PC1_t)$.

Panels (e) and (f) of Figure 5 compare the responses of $(REALPC, INFPC)$ to a $PC2$ shock for our illustrative MTSMs with lags. Focusing first on output growth, after a shared initial positive response, the response patterns are almost mirror images of each other (highly negatively correlated). The unconstrained VAR with unspanned macro risks has $REALPC$ rising for about 18 months, and then gradually declining towards zero over a span of about 3 years. In contrast, the positive impact under the macro-spanning constraint dies out within a year, turns negative over the following year, and gradually cycles to a positive effect that persists for an additional 7 years or so.

The positive response of $REALPC$ in model $GM_{2,3}^U(g, \pi)$ is consistent with the earlier findings of Estrella and Hardouvelis (1991). Stock and Watson (2003) survey more recent evidence of a positive relationship, and also review the wide use of the yield-curve slope as a leading indicator of real economic activity. These positive responses do not match the response in the MTSM examined in Rudebusch and Wu (2008) and, indeed, their pattern more closely resembles the negative response in model $GM_{4,3}^S(g, \pi)$ (absent the initial positive response). Their model, which has a more structural foundation than our reduced-form MTSMs, implicitly enforces macro-spanning just as in model $GM_{4,3}^S(g, \pi)$.

Turning to the inflation responses in Figure 5, when accommodating unspanned inflation risk there is at most a small response of $INFPC$ to a shock to the slope of the yield curve. This is consistent with the descriptive time-series evidence discussed in Stock and Watson (2003). However a very different response pattern emerges from model $GM_{4,3}^S(g, \pi)$ with spanned macro risks. In this case inflation responds negatively with a trough occurring over 3 years after the shock.

Why do MTSMs with spanned and unspanned macro risks provide such different evidence on nature of the comovement among the slope of the yield curve, inflation, and output growth? For the reasons discussed previously (and elaborated on in JLS), the IRs implied by model $GM_{2,3}^U(g, \pi)$ are essentially those implied by the unconstrained VAR^P(2) model for Z_t^U —the no-arbitrage restrictions are (nearly) irrelevant for these calculations. That is, this MTSM is essentially just replicating

what is in the data for our choice of macro factors, bond yields, and sample period.

On the other hand, MTSMs that enforce macro-spanning fundamentally change the connection between the information in the yield curve and the macroeconomy. In our illustrative model $GM_{4,3}^S(g, \pi)$, PC2 is effectively omitted from the set of pricing factors ($Z_t^{S'} = (g_t, \pi_t, PC1_t)$). Therefore, this model has to synthesize PC2 as an affine function of Z_t^S through the macro-spanning constraint. The filtered values $PC2^f$ of the theoretical construct PC2 in this model typically bear little resemblance to the historical time series $PC2^o$. Consequently, assessments of the responses of macro factors to the slope of the yield curve under spanning are highly inaccurate in this model.

This same reasoning suggests that a model with macro-spanning and an expanded state vector $(g_t, \pi_t, PC1_t, PC2_t)$ would give similar responses among the $(PC2_t, M_t)$. To confirm this we display in Figure 6 the IRs of REALPC and INFPC to PC2 shocks from the models $GM_{3,4}^S(g, \pi)$, $GM_{1,4}^S(g, \pi)$, and $GM_{2,3}^U(g, \pi)$. The patterns of responses of REALPC are similar across these three models, with the IRs for models $GM_{3,4}^S(g, \pi)$ and $GM_{2,3}^U(g, \pi)$ being especially close. The responses of INFPC are also similar across the latter two models.

These findings suggest that in some circumstances the joint dynamics of included macro and yield-based factors can be approximated within a model that enforces macro-spanning by adding additional lags of Z_t under \mathbb{P} . However, this improved fit may, as in the illustrative cases here, require the addition of extraneous pricing factors (more than are needed to fit the cross-section of bond yields) and extra lags (relative to the more parsimonious model with unspanned macro risks). This last point is illustrated by the very large differences between the responses of INFPC in models $GM_{3,4}^S(g, \pi)$ and $GM_{1,4}^S(g, \pi)$. Put differently, our findings suggest that the large values of p often adopted in the literature on MTSMs may well have been necessary owing to the imposition of counterfactual macro-spanning restrictions. Furthermore, the spanning models $GM_{p,4}^S(g, \pi)$ omit PC3 as a risk factor and, therefore, the IRs of PC3 to macro shocks will likely be different from those in the model $GM_{2,3}^U(g, \pi)$.

5 CONCLUDING REMARKS

We have extended the family of MTSMs to accommodate higher order (beyond first order) lags in the parameterization of the historical distribution of the risk factors, while preserving the parsimonious first-order Markov structure of these factors called for by the cross-sectional distribution of yields. Our framework encompasses both the models with spanned macro risks studied in JLS and the models with unspanned risks examined in JPS and Duffee (2011). Since the $VAR^{\mathbb{P}}(p)$ representation of the state can have p of any finite order, our canonical form nests most of the extant MTSMs with lags.

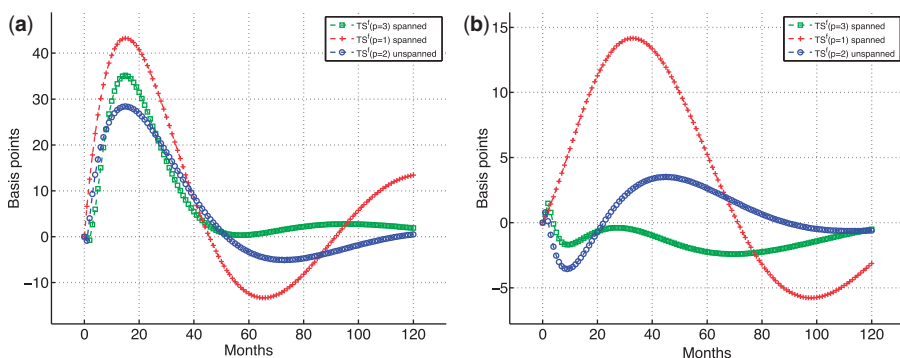


Figure 6 Impulse responses of REALPC and INFC given one standard deviation shock to PC2. The states are ordered as INFC, REALPC, PC1, PC2, PC3 for the unspanned models and INFC, REALPC, PC1, PC2 for the spanned models. The INFC, REALPC variables are normalized to have the same mean and standard deviation as annualized CPI, and IP growth, respectively. The models with spanned risks are $GM_{3,4}^S(g, \pi)$ ($p=3$) and $GM_{1,4}^S(g, \pi)$ ($p=1$), and the model with unspanned risks is $GM_{2,3}^U(g, \pi)$. (a) Response of REALPC to PC2; and (b) Response of INFC to PC2.

Our empirical analysis has addressed several distinct issues related to the specification of MTSMs with lags. First we argued, based on model-free evidence, that the macro and yield-curve data suggest a low-dimensional cross-sectional factor structure for bond yields. This structure is well described by a first-order Markov process under the pricing measure for the vector of risk factors. On the other hand, in several settings, statistical model selection criteria called for a higher order VAR representation of the state under the historical measure (for the data-generating process). Together, these observations motivated our asymmetric formulation of MTSMs based on a $VAR^{\mathbb{P}}(p)$ model with $p > 1$.

With the canonical version of this family of MTSMs in hand, we document through several examples that the no-arbitrage restrictions of an MTSM have very little impact on the properties of the joint distribution of the risk factors when the risk factors are rotated so that the non-macro factors are low-order PCs of bond yields. This shows that the prior results on the irrelevance of no-arbitrage restrictions in a canonical setting documented in JLS for the case of $p = 1$ extend to MTSMs with more lags under \mathbb{P} .

Next we examine the econometric identification of the reaction coefficients of a central bank policy rule—the Taylor rule—within MTSMs that enforce macro-spanning. We show that the policy parameters are in general not identified, absent the imposition of additional economic structure that is rarely imposed in the analysis of reduced-form MTSMs. The practical implication of this observation is that we are able to present several observationally equivalent (within an MTSM) representations of the short-term rate, many of which have loadings on output growth and inflation, and (what some have referred to as) policy shocks, that bear no resemblance to their counterparts in standard Taylor rules. Moreover, our

empirical results illustrate that expanding the number of risk factors from three to four—thereby effectively allowing for a richer distribution of the monetary policy shock—can fundamentally change the signs and magnitudes of loadings for the short-term rate, and the covariances of the shock with output and inflation.

Finally, drawing upon the literature on macro risks that are not spanned by the current information in the yield curve (unspanned macro risks), we addressed the question of how enforcement of the macro-spanning constraint distorts the impulse response functions among the macro and yield-curve factors. Particularly for yield-curve constructs that are not included in the set of risk factors (e.g., the slope of the yield curve in the family of models $GM_{p,3}(g, \pi)$) the distortions were large. Models with macro-spanning are also, by construction, completely silent about the impact of unspanned macro risks on risk premiums in bond markets.

Received April 17, 2012; revised February 6, 2013; accepted February 19, 2013.

REFERENCES

- Adrian, T., R. Crump, and E. Moench. 2012. Pricing the Term Structure with Linear Regressions. *Journal of Financial Economics*, forthcoming.
- Ang, A., J. Boivin, S. Dong, and R. Loo-Kung. 2011. Monetary Policy Shifts and the Term Structure. *Review of Economic Studies* 78: 429–457.
- 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.
- Bikbov, R., and M. Chernov. 2010. No-Arbitrage Macroeconomic Determinants of the Yield Curve. *Journal of Econometrics*, 159: 166–182
- 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. 2012. The Term Structure of Inflation Expectations. *Journal of Financial Economics* 106: 367–394.
- Clarida, R., J. Gali, and M. Gertler. 2000. Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics* 115: 147–180.
- Cochrane, J. 2007. “Identification with Taylor Rules: A Critical Examination.” Discussion paper, University of Chicago.
- Collin-Dufresne, P., R. Goldstein, and C. Jones. 2008. Identification of Maximal Affine Term Structure Models. *Journal of Finance* LXIII: 743–795.
- Cox, J., J. Ingersoll, and S. Ross. 1985. A Theory of the Term Structure of Interest Rates. *Econometrica* 53: 385–407.

- Dai, Q., and K. Singleton. 2000. Specification Analysis of Affine Term Structure Models. *Journal of Finance* 55: 1943–1978.
- Dai, Q., K. Singleton, and W. Yang. 2007. Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields. *Review of Financial Studies* 20: 1669–1706.
- Duffee, G. 2011. Information in (and not in) the Term Structure. *Review of Financial Studies* 24: 2895–2934.
- Duffie, D., and R. Kan. 1996. A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6: 379–406.
- Estrella, A., and G. Hardouvelis. 1991. The Term Structure as a Predictor of Real Economic Activity. *Journal of Finance* 46: 555–576.
- Jardet, C., A. Monfort, and F. Pegoraro. 2012. “No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth.” Discussion paper, Banque de France.
- Joslin, S. 2011. “Pricing and Hedging Volatility in Fixed Income Markets.” Discussion paper, Working Paper, USC.
- Joslin, S., A. Le, and K. Singleton. 2011. “Why Gaussian Macro-Finance Term Structure Models are (Nearly) Unconstrained Factor-VARs.” Discussion paper, University of Southern California.
- Joslin, S., M. Priebsch, and K. Singleton. 2011. “Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks.” Discussion paper, Stanford University.
- Joslin, S., K. Singleton, and H. Zhu. 2011. A New Perspective on Gaussian Dynamic Term Structure Models. *Review of Financial Studies* 24: 926–970.
- Litterman, R., and J. Scheinkman. 1991. Common Factors Affecting Bond Returns. *Journal of Fixed Income* 1: 54–61.
- Ludvigson, S., and S. Ng. 2010. Macro Factors in Bond Risk Premia. *Review of Financial Studies* 22: 5027–5067.
- Monfort, A., and F. Pegoraro. 2007. Switching VARMA Term Structure Models. *Journal of Financial Econometrics* 5: 105–153.
- Rudebusch, G., and T. Wu. 2008. A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy. *Economic Journal* 118: 906–926.
- Smith, J., and J. Taylor. 2009. The Term Structure of Policy Rules. *Journal of Monetary Economics* 56: 907–919.
- Stock, J., and M. Watson. 2003. Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature* XLI: 788–829.
- Vasicek, O.. 1977. An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics* 5: 177–188.