The Information in the Term Structures of Bond Yields*

Andrew Meldrum† Marek Raczko‡ Peter Spencer§

Federal Reserve Board Bank of England University of York

30 January 2018

Abstract

While standard no-arbitrage term structure models are estimated using nominal yields from a single country, a growing literature estimates joint models of yields in multiple countries or nominal and real yields from a single country. This paper argues that the additional complexity involved in estimating these joint models may not be justified. For example, joint models of U.S. and German nominal yields do not offer economically significant advantages in fitting the cross-section of yields, or predicting future yields. We obtain similar results for joint models of U.S. nominal and real yields. Thus, we lose little if we simply estimate separate models of a single class of yields.

Keywords: Affine term structure model, International interest rate co-movement, real interest rates.

JEL: F30, G12, G15.

---

*We are grateful for comments from Iryna Kaminska and attendees at a seminar at the Bank of England. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Board of Governors of the Federal Reserve System; research staff at the Board; or the Bank of England or its committees.

†Division of Monetary Affairs, Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue N.W., Washington, DC 20551, USA. E-mail: andrew.c.meldrum@frb.gov.

‡Macro Financial Analysis Division, Bank of England, Threadneedle Street, London EC2R 8AH, U.K.. E-mail: marek.raczko@bankofengland.co.uk.

§Department of Economics and Related Studies, University of York, York, YO10 5DD, U.K.. E-mail: peter.spencer@york.ac.uk.

1 Introduction

Gaussian no-arbitrage affine term structure models (ATSMs) are a popular framework for modeling the dynamics of bond yields. In the majority of studies of ATSMs, nominal bond yields in a single country are driven by a small number of pricing factors extracted from those same yields. However, a growing number of studies jointly model nominal yields in more than one country (for example, Anderson et al. (2010), Egorov et al. (2011), Kaminska et al. (2013), and Diez de los Rios (2017)),[1] or nominal and real yields from a single country (for example, Joyce et al. (2010), D'Amico et al. (2014), and Abrahams et al. (2016)). The defining feature of these joint models is that current and / or future yields depend on pricing factors extracted from yields on multiple classes of bonds. However, this paper argues that the gains from two of the most common applications of joint models may be small relative to separate models. Thus, we prefer the simplicity of separate models.

The question of whether to model the term structures of yields on different types of bonds jointly or separately is related to an emerging literature on the role of "hidden" or "unspanned" factors in the term structure. The majority of studies of ATSMs assume that three principal components of U.S. nominal yields contain all of the relevant information for modeling both the cross section and the time-series dynamics of those yields. If this assumption is valid, then any marginal information in other variables—including other classes of yields—is redundant, and joint modeling is unnecessary. However, some recent studies have found evidence that factors hidden from the first three principal components of U.S. nominal yields may affect the time-series dynamics of yields; examples include other factors extracted from yields (as in Cochrane and Piazzesi (2005) and Duffee (2011)) and macroeconomic variables (as in Joslin et al. (2014)).[2] These findings raise the question of whether there are similar hidden factors that can be extracted from yields on other classes of bonds. A

[1]In addition to these studies using no-arbitrage ATSMs, Diebold et al. (2008) jointly model yields in different countries in a model that does not impose no-arbitrage.

[2]The presence of unspanned factors in U.S. yields is subject to some debate—for example, Bauer and Rudebusch (2017) find that models with unspanned macroeconomic factors are rejected statistically.

joint model would allow for this possibility, whereas a separate model would be misspecified. However, previous studies do not consider whether joint models have different implications for the dynamics of yields compared with separate models, so it is unclear whether the additional complexity involved with joint modeling is warranted. Our paper addresses this gap in the literature on joint models.

We start by setting out the circumstances in which joint ATSMs are observationally equivalent (or not) to two separate ATSMs of yields on a single asset class. In principle, we show that two conditions must hold in order for joint modelling to be necessary. First, at least one factor must be "unspanned" or "hidden" from one class of yields. And second, at least one of the hidden factors must Granger cause the yields from which they are hidden. In practice, however, it is not straightforward to compare the joint models estimated in previous studies with the most standard separate models—in part because there is no clear consensus about the appropriate factor structure of joint models, and in part because previous joint models do not nest separate models as a special case. We therefore propose a two-step approach. In the first step, we compare separate models with standard factor structures with a joint model that has the same specification for the cross section of yields—and which nests the separate models as a special case. However, by itself this exercise is quite limited, because it restricts us to joint models that where all of the hidden factors are hidden from one or the other class of yields. In a second step, we therefore examine the impact on joint models of also allowing for common "global factors" (while holding the number of factors spanned by each set of yields constant).

We apply this approach to two of the most popular applications of joint models. First, we show that there is little convincing evidence that we need to model U.S. and German nominal yields jointly. While the most flexible joint model is preferred statistically to two separate models, it offers no improvement in terms of the cross-sectional fit to yields, and no robust, economically significant improvements when it comes to predicting future yields. Reducing the number of factors in the joint model below six has only a modest impact on the results provided there are no more than two global factors; with three global factors the fit of the joint model is deteriorates substantially. As a robustness check, we show that these conclusions also hold if we allow as many as five factors to be spanned by yields in each country.

We next consider an application to joint models of U.S. nominal and real yields. This application faces more practical difficulties than the international term structure model, on account of the relatively short sample period and limited range of maturities for the real yields on Treasury Inflation Protected Securities (TIPS). Nevertheless, we again find that joint models do not offer any economically significant advantages relative to separate models. One caveat to this conclusion is that we can use a joint model in which the factors spanned by real yields are also spanned by nominal yields to extend the relatively short sample of real yields to cover a longer period. However, even here it is worth bearing in mind that such a joint model predicts both nominal and real yields less accurately than separate models.

The remainder of this paper proceeds as follows. In Section 2, we set out the separate and joint ATSMs. In Section 3 we explain the circumstances under which joint modeling is necessary and explain our approach to addressing the question of whether joint models offer any advantages relative to separate models. In Section 4, we describe our application to U.S. and German nominal yields, and in Section 5 our application to U.S. nominal and real yields. In Section 6 we summarize our conclusions.

2 Separate and Joint Affine Term Structure Models

Suppose we want to model the yields on two classes of default-risk-free bonds that have payments fixed in different units of account, such as different currencies. One option is to model each class of yields separately using the standard Gaussian ATSM of Duffee (2002), which we summarize in Section 2.1. Another option is to model the two classes of yields simultaneously within a joint model, which we present in Section 2.2.

2.1 Separate Models

Suppose each type of default-risk-free bond pays one unit of the j^{th} numeraire asset at maturity (where $j = 1, 2$). The standard approach in the no-arbitrage term structure literature is to model each class of yields separately using an ATSM. This model has four main assumptions. First, the short-term (i.e. one-period) risk-free interest rate relevant for pricing the j^{th} class of bonds ($r_{j,t}$) is an affine function of an $n_j \times 1$ vector of unobserved pricing factors ($\mathbf{x}_{j,t}$):

$$r_{j,t} = \delta_{j,0,\mathcal{S}} + \boldsymbol{\delta}'_{j,1,\mathcal{S}} \mathbf{x}_{j,t}. \tag{1}$$

Second, the factors follow a first-order vector autoregression (VAR) under the time-series measure (which we denote $\mathbb{P}$):

$$\mathbf{x}_{j,t+1} = \boldsymbol{\mu}_{j,\mathcal{S}} + \boldsymbol{\Phi}_{j,\mathcal{S}} \mathbf{x}_{j,t} + \boldsymbol{\Sigma}_{j,\mathcal{S}} \boldsymbol{\varepsilon}_{j,t+1}, \tag{2}$$

where $\boldsymbol{\varepsilon}_{j,t+1} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$ is an $n_j \times 1$ vector of Normally distributed shocks. Third, there are no arbitrage opportunities from investing in different maturity bonds, which implies that the time-t price of an n-period bond is given by

$$P_{j,n,t} = \mathbb{E}_t \left[M_{j,t+1} P_{j,n-1,t+1} \right], \tag{3}$$

where $M_{j,t+1}$ is the stochastic discount factor that prices the j^{th} class of bonds. Fourth, the stochastic discount factor takes the form

$$M_{j,t+1} = \exp\left(-r_{j,t} - \frac{1}{2} \boldsymbol{\lambda}'_{j,\mathcal{S},t} \boldsymbol{\lambda}_{j,\mathcal{S},t} - \boldsymbol{\lambda}'_{j,\mathcal{S},t} \boldsymbol{\varepsilon}_{j,t+1} \right), \tag{4}$$

where the $n_j \times 1$ prices of risk $\boldsymbol{\lambda}_{j,\mathcal{S},t}$ are affine in the factors, that is, $\boldsymbol{\lambda}_{j,\mathcal{S},t} \equiv \boldsymbol{\lambda}_{j,0,\mathcal{S}} + \boldsymbol{\Lambda}_{j,1,\mathcal{S}} \mathbf{x}_{j,t}$.

The assumption of no-arbitrage implies that there exists a unique risk-neutral probability measure (which we denote as $\mathbb{Q}_j$) such that the j^{th} class of bond prices satisfy

$$P_{j,n,t} = \mathbb{E}_t^{\mathbb{Q}_j} \left[\exp\left(-r_{j,t}\right) P_{j,n-1,t+1} \right], \tag{5}$$

where $\mathbb{E}_t^{\mathbb{Q}_j}$ denotes expectations with respect to the $\mathbb{Q}_j$ measure. The above assumptions also imply that the pricing factors follow a first-order VAR under $\mathbb{Q}_j$:

$$\mathbf{x}_{j,t+1} = \boldsymbol{\mu}_{j,\mathcal{S}}^{\mathbb{Q}_j} + \boldsymbol{\Phi}_{j,\mathcal{S}}^{\mathbb{Q}_j} \mathbf{x}_{j,t} + \boldsymbol{\Sigma}_{j,\mathcal{S}} \boldsymbol{\varepsilon}_{j,t+1}^{\mathbb{Q}_j}, \tag{6}$$

where $\boldsymbol{\varepsilon}_{j,t+1}^{\mathbb{Q}_j} \sim \mathcal{NID}\left(\mathbf{0}, \mathbf{I}\right)$ is an $n_j \times 1$ vector of Normally distributed shocks, $\boldsymbol{\mu}_{j,\mathcal{S}}^{\mathbb{Q}} \equiv \boldsymbol{\mu}_{j,\mathcal{S}} - \boldsymbol{\Sigma}_{j,\mathcal{S}} \boldsymbol{\lambda}_{j,\mathcal{S},0}$, and $\boldsymbol{\Phi}_{j,\mathcal{S}}^{\mathbb{Q}} \equiv \boldsymbol{\Phi}_{j,\mathcal{S}} - \boldsymbol{\Sigma}_{j,\mathcal{S}} \boldsymbol{\Lambda}_{j,\mathcal{S},1}$. The yield on an n-period bond ($y_{j,n,t} \equiv -\frac{1}{n} \log P_{j,n,t}$) is an affine function of the pricing factors—that is $y_{j,n,t} = -\frac{1}{n}\left(a_{j,n,\mathcal{S}} + \mathbf{b}_{j,n,\mathcal{S}}' \mathbf{x}_{j,t}\right)$, where

$$a_{j,n,\mathcal{S}} = a_{j,n-1,\mathcal{S}} + \mathbf{b}_{j,n-1,\mathcal{S}}' \boldsymbol{\mu}_{j,\mathcal{S}}^{\mathbb{Q}_j} + \frac{1}{2} \mathbf{b}_{j,n-1,\mathcal{S}}' \boldsymbol{\Sigma}_{j,\mathcal{S}} \boldsymbol{\Sigma}_{j,\mathcal{S}}' \mathbf{b}_{j,n-1,\mathcal{S}} - \delta_{j,0,\mathcal{S}}, \tag{7}$$

$$\mathbf{b}_{j,n,\mathcal{S}}' = \mathbf{b}_{j,n-1,\mathcal{S}}' \boldsymbol{\Phi}_{j,\mathcal{S}}^{\mathbb{Q}_j} - \boldsymbol{\delta}_{j,1,\mathcal{S}}', \tag{8}$$

and $a_{j,0,\mathcal{S}} = 0$ and $\mathbf{b}_{j,n,\mathcal{S}} = \mathbf{0}$ (see, for example, Joslin et al. (2011) or Appendix A of this paper for further details).

As discussed by, for example, Dai and Singleton (2000), Joslin et al. (2011), and Hamilton and Wu (2012), a minimum set of normalization restrictions is required in order to identify a maximally flexible model. We impose that $\boldsymbol{\mu}_{j,\mathcal{S}}^{\mathbb{Q}_j} = \mathbf{0}$, $\boldsymbol{\Sigma}_{j,\mathcal{S}} = \mathbf{I}$, and $\boldsymbol{\Phi}_{j,\mathcal{S}}^{\mathbb{Q}_j}$ is a lower triangular matrix with diagonal elements ordered such that $\phi_{j,\mathcal{S},11}^{\mathbb{Q}_j} \geq \phi_{j,\mathcal{S},22}^{\mathbb{Q}_j} \geq \ldots \geq \phi_{j,\mathcal{S},n_x n_x}^{\mathbb{Q}_j}$.

2.2 Joint Models

Another option is to model the two classes of yields jointly. The starting point for a joint model is the observation that, in addition to equation (3), the prices of the second class of bond must also satisfy

$$P_{2,n,t} S_t = \mathbb{E}_t \left[M_{1,t+1} P_{2,n-1,t+1} S_{t+1} \right] \tag{9}$$

where S_t is the exchange rate—that is, the price of one unit of the second numeraire asset in terms of the first numeraire asset. For example, if we were considering bonds with payoffs in different currencies, S_t would be the domestic-currency price of one unit of foreign currency.

In a joint model, we collect all of the factors that affect either or both classes of yields into a single $n_x \times 1$ vector $\mathbf{x}_t$. The short rate relevant for pricing the first asset class is again affine in these pricing factors:

$$r_{1,t} = \delta_{1,0} + \boldsymbol{\delta}'_{1,1}\mathbf{x}_t. \tag{10}$$

Following Diez de los Rios (2008) and Abrahams et al. (2016), among others, the depreciation rate ($\Delta s_t = \log S_t - \log S_{t-1}$) is also affine in the factors:

$$\Delta s_t = s_0 + \mathbf{s}'_1\mathbf{x}_t. \tag{11}$$

The factors again follow a first-order Gaussian VAR under the $\mathbb{P}$ measure:

$$\mathbf{x}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}, \tag{12}$$

where $\boldsymbol{\varepsilon}_{t+1} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$. And the stochastic discount factor $M_{1,t+1}$ again takes the form

$$M_{1,t+1} = \exp\left(-r_{1,t} - \frac{1}{2}\boldsymbol{\lambda}'_{1,t}\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}'_{1,t}\boldsymbol{\varepsilon}_{t+1}\right), \tag{13}$$

where the $n_j \times 1$ prices of risk are again affine in the factors, that is, $\boldsymbol{\lambda}_{1,t} \equiv \boldsymbol{\lambda}_{1,0} + \boldsymbol{\Lambda}_{1,1}\mathbf{x}_t$.

Under these assumptions, the prices of the first and second class of bonds must also satisfy

$$P_{1,n,t} = \mathbb{E}_t^{\mathbb{Q}_1}\left[\exp\left(-r_{1,t}\right) P_{1,n-1,t+1}\right] \text{ and} \tag{14}$$

$$P_{2,n,t}S_t = \mathbb{E}_t^{\mathbb{Q}_1}\left[\exp\left(-r_{1,t}\right) P_{2,n-1,t+1}S_{t+1}\right], \tag{15}$$

respectively, and the factors again follow a first-order VAR under $\mathbb{Q}_1$:

$$\mathbf{x}_{t+1} = \boldsymbol{\mu}^{\mathbb{Q}_1} + \boldsymbol{\Phi}^{\mathbb{Q}_1}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}^{\mathbb{Q}_1}_{t+1}, \tag{16}$$

where $\boldsymbol{\varepsilon}^{\mathbb{Q}_1}_{t+1} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\mu}^{\mathbb{Q}_1} \equiv \boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,0}$, and $\boldsymbol{\Phi}^{\mathbb{Q}_1} \equiv \boldsymbol{\Phi} - \boldsymbol{\Sigma}\boldsymbol{\Lambda}_{1,1}$.

The pricing of the first class of bonds is directly analogous to Section 2.1, with yields given by

$$y_{1,n,t} = -\frac{1}{n}\left(a_{1,n} + \mathbf{b}'_{1,n}\mathbf{x}_t\right), \tag{17}$$

where

$$a_{1,n} = a_{1,n-1} + \mathbf{b}'_{1,n-1}\boldsymbol{\mu}^{\mathbb{Q}_1} + \frac{1}{2}\mathbf{b}'_{1,n-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{b}_{1,n-1} - \delta_0, \tag{18}$$

$$\mathbf{b}'_{1,n} = \mathbf{b}'_{1,n-1}\boldsymbol{\Phi}^{\mathbb{Q}_1} - \boldsymbol{\delta}'_1, \tag{19}$$

and $a_{2,0} = 0$ and $\mathbf{b}_{2,0} = \mathbf{0}$. Yields on the second class of bond are given by:

$$y_{2,n,t} = -\frac{1}{n}\left(a_{2,n} + \mathbf{b}'_{2,n}\mathbf{x}_t\right), \tag{20}$$

where

$$a_{2,n} = a_{2,n-1} - \delta_0 + s_0 + \left(\mathbf{s}_1 + \mathbf{b}_{2,n-1}\right)'\boldsymbol{\mu}^{\mathbb{Q}_1} + \frac{1}{2}\left(\mathbf{s}_1 + \mathbf{b}_{2,n-1}\right)'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\left(\mathbf{s}_1 + \mathbf{b}_{2,n-1}\right), \tag{21}$$

$$\mathbf{b}'_{2,n} = \left(\mathbf{s}_1 + \mathbf{b}_{2,n-1}\right)'\boldsymbol{\Phi}^{\mathbb{Q}_1} - \boldsymbol{\delta}'_1, \tag{22}$$

and $a_{2,0} = 0$ and $\mathbf{b}_{2,n} = \mathbf{0}$ (see Abrahams et al. (2016) or Appendix B of this paper for further details).

The identification restrictions on a maximally-flexible joint model are directly analogous to those for the the same as in the separate models—that is, $\boldsymbol{\mu}^{\mathbb{Q}_1} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}$, and $\boldsymbol{\Phi}^{\mathbb{Q}_1}$ is a lower triangular matrix with diagonal elements ordered such that $\phi^{\mathbb{Q}_1}_{11} \geq \phi^{\mathbb{Q}_1}_{22} \geq ... \geq \phi^{\mathbb{Q}_1}_{n_x n_x}$.

3 Hidden Factors in Joint Models

We now turn to the question of when it is necessary to model yields on multiple asset classes jointly rather than separately. In Section 3.1 we show that if one class of yields spans all pricing factors, and if yields are measured without error, information from other classes of yields is redundant. However, in Section 3.2 we show that the presence of "hidden" factors provides a rationale for joint modeling—this analysis is closely related to that of Duffee (2011), who studies the issue of hidden factors within separate ATSMs of U.S. nominal yields. In Section 3.3 we explain our approach for assessing whether joint models offer advantages relative to separate models in practice.

3.1 When is Joint Modeling Unnecessary?

If yields are measured without error then—except for a special case that we discuss below—other classes of yields are redundant and joint modeling is unnecessary. To see why, let $\mathbf{y}_t = \mathbf{A} + \mathbf{B}\mathbf{x}_t$ denote an $n_{yd} \times 1$ vector of yields on the first class of bonds, where the definitions of $\mathbf{A}$ and $\mathbf{B}$ follow from equation (17). In addition, let $\mathcal{P}_t$ denote an $n_x \times 1$ vector of independent linear combinations of the first class of yields—that is, $\mathcal{P}_t \equiv \mathbf{W}\mathbf{y}_t$, where $\mathbf{W}$ is a full-rank $n_x \times n_{yd}$ weighting matrix. Providing $\mathbf{B}$ has full column rank, we can follow Joslin et al. (2011) and apply invariant rotations to the factors such that the $\mathbb{Q}_1$ and $\mathbb{P}$ dynamics of the model are given by

$$\mathcal{P}_{t+1} = \boldsymbol{\mu}_{\mathcal{P}}^{\mathbb{Q}_1} + \boldsymbol{\Phi}_{\mathcal{P}}^{\mathbb{Q}_1}\mathcal{P}_t + \boldsymbol{\Sigma}_{\mathcal{P}}\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}_1} \text{ and} \tag{23}$$

$$\mathcal{P}_{t+1} = \boldsymbol{\mu}_{\mathcal{P}} + \boldsymbol{\Phi}_{\mathcal{P}}\mathcal{P}_t + \boldsymbol{\Sigma}_{\mathcal{P}}\boldsymbol{\varepsilon}_{t+1}, \tag{24}$$

respectively, where $\boldsymbol{\Phi}_{\mathcal{P}}^{\mathbb{Q}_1} \equiv \mathbf{W}\mathbf{B}\boldsymbol{\Phi}^{\mathbb{Q}_1}(\mathbf{W}\mathbf{B})^{-1}$, $\boldsymbol{\mu}_{\mathcal{P}}^{\mathbb{Q}_1} \equiv \left(\mathbf{I} - \boldsymbol{\Phi}_{\mathcal{P}}^{\mathbb{Q}_1}\right)\mathbf{W}\mathbf{A} + \mathbf{W}\mathbf{B}\boldsymbol{\mu}^{\mathbb{Q}_1}$, $\boldsymbol{\Phi}_{\mathcal{P}} \equiv \mathbf{W}\mathbf{B}\boldsymbol{\Phi}(\mathbf{W}\mathbf{B})^{-1}$, $\boldsymbol{\mu}_{\mathcal{P}} \equiv (\mathbf{I} - \boldsymbol{\Phi}_{\mathcal{P}})\mathbf{W}\mathbf{A} + \mathbf{W}\mathbf{B}\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}_{\mathcal{P}} \equiv \mathbf{W}\mathbf{B}\boldsymbol{\Sigma}$. The short rate for the first asset class in the rotated model is given by

$$r_t = \delta_{1,0,\mathcal{P}} + \boldsymbol{\delta}'_{1,1,\mathcal{P}} \mathcal{P}_t, \tag{25}$$

where $\boldsymbol{\delta}'_{1,1,\mathcal{P}} \equiv \boldsymbol{\delta}'_{1,1} (\mathbf{WB})^{-1}$, and $\delta_{1,0,\mathcal{P}} \equiv \delta_{1,0} - \boldsymbol{\delta}'_{1,1,\mathcal{P}} \mathbf{WA}$. Thus, the model is observationally equivalent to a model that has $\mathcal{P}_t$ as its pricing factors—that is, a model estimated using only factors extracted from the first asset class. Information from the second class is redundant.

3.2 When is Joint Modeling Necessary?

3.2.1 Completely Hidden Factors

The assumption that $\mathbf{B}$ has full column rank—such that the factors that drive the first class of yields can be inverted using only those yields—is crucial for the above result that the information contained in the second class of yields is redundant. Suppose instead that the k^{th} element of $\boldsymbol{\delta}_1$ in a joint model is equal to zero (that is, $\delta_{1,1,k} = 0$), and that $\phi_{ik}^{\mathbb{Q}_1} = 0$ for all i for which $\delta_{1,1,i} \neq 0$. In this case, the k^{th} column of $\mathbf{B}$ is equal to zero—that is, the first class of yields does not load on the k^{th} factor.[3] Thus, the k^{th} factor cannot be inverted from the first class of yields alone. Borrowing terminology from Duffee (2011), the k^{th} factor is "completely hidden" from the first class of yields.

However, even if there are some factors that are spanned by the second class of yields but completely hidden from the first class of yields, joint modeling may still be unnecessary. To see why, consider the case in which, in addition, $\phi_{ik} = 0$ for all i for which $\delta_{1,1,i} \neq 0$: in this case, the k^{th} factor will also not affect the $\mathbb{P}$ dynamics of the factors that are spanned by the first class of yields and it can be omitted from a model of those yields.

Thus, in the absence of measurement error, two conditions must hold in order for joint modeling to be necessary. First, at least one factor must be completely hidden from the first

[3]For example, consider a two-factor joint model where $\delta_{1,1,1} = 0$ and $\delta_{1,1,2} \neq 0$—that is, where the short rate is given by, $r_{1,t} = \delta_{1,1,2} x_{2,t}$. If $\phi_{21}^{\mathbb{Q}_1} = 0$ then it follows from equation (19) that the first class of yields will have a zero loading on the first factor ($x_{t,1}$).

class of yields (that is, there must be some k such that $\delta_{1,1,k} = 0$ and $\phi_{ik}^{\mathbb{Q}_1} = 0$ for all i for which $\delta_{1,1,i} \neq 0$). And second, at least one of those hidden factors must Granger cause the factors that are not hidden from the first class of yields (that is, $\phi_{ik} \neq 0$ for some i for which $\delta_{1,1,i} \neq 0$).

3.2.2 Partly Hidden Factors

Again borrowing terminology from Duffee (2011), the restrictions required for a factor to be completely hidden are "knife-edge", in the sense that even a miniscule factor loading would render that factor invertible in the absence of measurement error. However, in practice yields are likely to be measured with error, which raises the possibility that some factors may be "partly hidden" from one of the classes of yields.

For example, in our applications below we allow for measurement error by assuming that observed yields are given by

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_t^* \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}^* \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{B}^* \end{bmatrix} \mathbf{x}_t + \begin{bmatrix} \mathbf{w}_t \\ \mathbf{w}_t^* \end{bmatrix}, \tag{26}$$

where $\mathbf{w}_t \sim \mathcal{NID}\left(\mathbf{0}, \sigma_w^2 \times \mathbf{I}\right)$ is an $n_{yd} \times 1$ vector of measurement errors on the first class of yields and $\mathbf{w}_t^* \sim \mathcal{NID}\left(\mathbf{0}, \sigma_{w^*}^2 \times \mathbf{I}\right)$ is an $n_{yo} \times 1$ vector of measurement errors on the second class of yields. Thus, equations (12) and (26) form a linear-Gaussian state-space system, and we can estimate the parameters of the model by maximum likelihood, using the Kalman filter to estimate the latent pricing factors for a given set of parameters.[4]

Even if the knife-edge restrictions required for a factor to be completely hidden do not hold exactly, the loadings of one class of yields on some factors may be sufficiently small that the Kalman filter estimate of the pricing factors is not precise—that is, in a short sample, the effects of those factors may be "partly hidden" by measurement errors. Thus, including a

[4] D'Amico et al. (2014) (among others) provide further details on the estimation of joint models by maximum likelihood using the Kalman filter. We use the same approach to estimate the separate models.

second class of yields as additional observed variables may ensure a more consistent estimate of the pricing factors.

3.3 How Should We Assess the Benefits of Joint Models in Practice?

One practical difficulty when assessing whether joint models offer any advantages relative to separate models is that it is unclear how we should determine the appropriate number of factors in a joint model. While most studies of nominal yields in a single country assume three pricing factors, there appears to be less consensus about the most appropriate number of factors in joint models. For example, Anderson et al. (2010) estimate five-factor joint models of yields in two countries; Egorov et al. (2011) a four-factor joint model of yields in two countries; Kaminska et al. (2013) a five-factor joint model of yields in three countries; and Diez de los Rios (2017) a ten-factor joint model of yields in seven countries. Moreover, some of these studies also impose restrictions to ensure that some factors are completely hidden from one class of yields and / or the $\mathbb{P}$ dynamics of yields. Thus, it is not straightforward to determine from the previous literature which factor structure we should focus on.

A further difficulty is that previous joint models do not actually nest the most standard separate models as special cases, which further complicates the comparison.[5] For example, while the five-factor joint model of Anderson et al. (2010) allows each set of yields to be spanned by more than the standard three factors, it offers less overall flexibility to fit the cross-sectional dimension of two classes of yields than two separate three-factor models. In addition, two separate three-factor models effectively assume independence between the factors driving yields in two countries, whereas the five-factor joint model does not. As a result, it is not straightforward to disentangle the effects of the various differences between non-nested joint and separate models.

[5] Yung (2017) estimates six-factor joint models of yields in two countries that nest two separate three-factor models. However, the two sets of factors in the joint model are restricted to be independent, so the joint model is effectively equivalent to two separate models.

In this section, we offer a practical way forward. Our approach breaks down the question of whether there are any advantages of joint modeling into two steps. In the first step, compare separate models with a joint model that has the same specification for the cross-section of yields as the separate models. For example, in Section 4 we start by comparing separate three-factor models of U.S. and German yields with a six-factor joint model that has three factors spanned by yields in each country. To do that, we must restrict the joint models to ensure that some factors are hidden, which we explain how to do in the following paragraphs

More generally, suppose that we want to restrict a joint model to ensure that we have n_g global factors spanned by yields on both classes of bonds, n_d local factors completely hidden from the second class of bonds, and n_o local factors completely hidden from the first class of bonds (with $n_g + n_d + n_o = n_x$). Note from equations (20)-(22) that the short rate for the second class of bonds takes the form

$$r_{2,t} = \delta_{2,0} + \boldsymbol{\delta}'_{2,1}\mathbf{x}_t \tag{27}$$

where $\delta_{2,0} = \delta_{1,0} - s_0 - \mathbf{s}'_1\boldsymbol{\mu}^{\mathbb{Q}_1} - \frac{1}{2}\mathbf{s}'_1\boldsymbol{\Sigma\Sigma}'\mathbf{s}_1$ and $\boldsymbol{\delta}_{2,1} = \boldsymbol{\delta}_{1,1} - \left(\boldsymbol{\Phi}^{\mathbb{Q}_1}\right)'\mathbf{s}_1$ (see Appendix C for details). Thus we can equivalently parameterize the joint model in terms of $\delta_{2,0}$ and $\boldsymbol{\delta}_{2,1}$, rather than s_0 and $\mathbf{s}_1$. We can ensure the desired factor structure by imposing that the short rate loadings take the forms

$$\begin{aligned} \boldsymbol{\delta}_{1,1} &= \left[\boldsymbol{\delta}'_{1,1,g}, \boldsymbol{\delta}'_{1,1,d}, \mathbf{0}'_{n_o\times 1}\right]' \text{ and} && (28) \\ \boldsymbol{\delta}_{2,1} &= \left[\boldsymbol{\delta}'_{2,1,g}, \mathbf{0}'_{n_d\times 1}, \boldsymbol{\delta}'_{2,1,o}\right]', && (29) \end{aligned}$$

where $\boldsymbol{\delta}_{1,1,g}$ and $\boldsymbol{\delta}_{2,1,g}$ are $n_g \times 1$, $\boldsymbol{\delta}_{1,1,d}$ is $n_d \times 1$, and $\boldsymbol{\delta}_{2,1,o}$ is $n_o \times 1$; and that $\boldsymbol{\Phi}^{\mathbb{Q}_1}$ takes the form

$$\mathbf{\Phi}^{\mathbb{Q}_1} = \begin{bmatrix} \mathbf{\Phi}_{gg}^{\mathbb{Q}_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Phi}_{dg}^{\mathbb{Q}_1} & \mathbf{\Phi}_{dd}^{\mathbb{Q}_1} & \mathbf{0} \\ \mathbf{\Phi}_{og}^{\mathbb{Q}_1} & \mathbf{0} & \mathbf{\Phi}_{oo}^{\mathbb{Q}_1} \end{bmatrix}, \tag{30}$$

where $\mathbf{\Phi}_{gg}^{\mathbb{Q}_1}$ is $n_g \times n_g$, $\mathbf{\Phi}_{dg}^{\mathbb{Q}_1}$ is $n_d \times n_g$, $\mathbf{\Phi}_{dd}^{\mathbb{Q}_1}$ is $n_d \times n_d$, $\mathbf{\Phi}_{og}^{\mathbb{Q}_1}$ is $n_o \times n_g$, and $\mathbf{\Phi}_{oo}^{\mathbb{Q}_1}$ is $n_o \times n_o$. These restructions ensure that the yield loadings take the forms $\mathbf{B}= \begin{bmatrix} \mathbf{B}_g & \mathbf{B}_d & \mathbf{0} \end{bmatrix}$ and $\mathbf{B}^*= \begin{bmatrix} \mathbf{B}_g^* & \mathbf{0} & \mathbf{B}_o^* \end{bmatrix}$, where $\mathbf{B}_g$ is $n_{yd} \times n_g$, $\mathbf{B}_g^*$ is $n_{yo} \times n_g$, $\mathbf{B}_d$ is $n_{yd} \times n_d$, and $\mathbf{B}_{yo}^*$ is $n_y \times n_o$. Thus we can partition the pricing factors conformably as $\mathbf{x}_t = \left[\mathbf{x}_{g,t}', \mathbf{x}_{d,t}', \mathbf{x}_{o,t}'\right]'$, where $\mathbf{x}_{g,t}$ are global factors, and $\mathbf{x}_{d,t}$ and $\mathbf{x}_{o,t}$ are factors hidden from one or the other set of yields.[6]

Next, we note that two separate models with n_1 and n_2 factors, respectively, have the same specification for the cross section of yields as a joint model with $n_g = 0$, $n_d = n_1$, and $n_o = n_2$ factors (see Appendix D for details). However, the most flexible joint model with this factor structure has unrestricted $\mathbb{P}$ dynamics—that is,

$$\begin{bmatrix} \mathbf{x}_{d,t+1} \\ \mathbf{x}_{o,t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_d \\ \boldsymbol{\mu}_o \end{bmatrix} + \begin{bmatrix} \mathbf{\Phi}_{dd} & \mathbf{\Phi}_{do} \\ \mathbf{\Phi}_{od} & \mathbf{\Phi}_{oo} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{d,t} \\ \mathbf{x}_{o,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{d,t+1} \\ \boldsymbol{\varepsilon}_{o,t+1} \end{bmatrix}, \tag{31}$$

whereas the separate models impose that $\mathbf{\Phi}_{do} = \mathbf{0}$ and $\mathbf{\Phi}_{od} = \mathbf{0}$. We can therefore view the question of whether a joint model offers an advantages relative to two separate models simply as a test of these zero restrictions—that is, as a test of whether the two sets of factors Granger cause the other.

However, simply imposing that there are *only* completely hidden (that is, local) factors foregoes the potential benefit that a joint models may obtain more consistent estimates of *partly* hidden pricing factors. For example, yields in different countries are often highly correlated, which suggests the presence of at least one global factor driving yields. In the second step of our approach, we therefore consider whether allowing for global factors—but

[6] With some factors restricted to be local we partly identify the ordering of the factors through the placing of the zero restrictions in equations (28) and (29). Thus, for identification we only need to impose ordering restrictions on the diagonal elements of $\mathbf{\Phi}_{gg}^{\mathbb{Q}}$, $\mathbf{\Phi}_{dd}^{\mathbb{Q}}$, and $\mathbf{\Phi}_{oo}^{\mathbb{Q}}$ separately, rather than on $\mathbf{\Phi}^{\mathbb{Q}}$ as a whole.

maintaining the same number of factors spanned by each set of yields as in separate models—affects the properties of the joint models. In this case, joint models with global factors are nested by a joint model that only has completely hidden factors.

4 An Application to Joint Models of U.S. and German Nominal Yields

We now turn to our application to joint models of U.S. and German nominal yields, which we compare to separate models of yields in each country. In Section 4.1, we describe our data set. In Section 4.2, we discuss the choices of factor structure in the separate and joint models. In Section 4.3, we present results concerning the in-sample fit of the various models to the cross-section of yields, and the predictive accuracy of the models for future yields. In Section 4.4 we present the results of an out-of-sample forecasting exercise. In Section 4.5 we discuss the implications of the separate and joint models for the term premiums reflected in bond yields. Finally, as a robustness check, in Section 4.6 we show that our main conclusions also hold in models with more pricing factors.

4.1 Data

Our data set consists of end-month U.S. and German zero-coupon nominal government bond yields, which are estimated from the prices of coupon-bearing bonds using the parametric method of Svensson (1994).[7] Our sample starts in January 1990 and ends in December 2007. Starting the sample in 1990 is broadly consistent with previous studies of ATSMs of U.S. nominal yields and avoids a potential structural break with German reunification. Ending the sample in December 2007 avoids complications caused by the proximity of nominal bond yields to the zero lower bound. At each point in time, we consider a cross section of yields

[7] The U.S. yields are from Gürkaynak et al. (2007), updates of which are published by the Board of Governors of the Federal Reserve System. The German yields are published by the Deutsche Bundesbank.

with maturities of six months and one, two, three, five, seven, and ten years.

4.2 Factor Structure

As is standard in the literature on ATSMs, we consider separate models with three pricing factors. As discussed in Section 3.3, to ensure consistency between the joint and separate models, we also maintain this standard assumption that three factors span yields in each country in the joint models. We therefore consider four joint ATSMs with between three and six factors in total, as summarized in Table 1 (which uses the notation JM(x) to refer to the joint model with $n_x = x$). At one extreme, JM(6) has three local factors for each country and no global factors. As discussed in Section 3.3, the separate models can be viewed as a special case of JM(6) with restrictions on the $\mathbb{P}$ dynamics. The other joint models (JM(3), JM(4), and JM(5)) have between one and three global factors. As discussed in Section 3.3, these joint models with global factors can also be viewed as restricted versions of JM(6)—for example, JM(3) is observationally equivalent to a version of JM(6) in which $\mathbf{x}_{o,t}$ is a linear rotation of $\mathbf{x}_{d,t}$.

Table 1: Factor Structures in Joint Models of U.S. and German Nominal Yields
This table shows the factor structures we consider for joint models of U.S. and German yields. In each case, we assume that three factors are spanned by the yields in each country. The total number of factors in the model is n_x, of which n_g factors are global and n_d and n_o are local to U.S. and German yields, respectively.

	n_x	n_g	n_d	n_o
JM(3)	3	3	0	0
JM(4)	4	2	1	1
JM(5)	5	1	2	2
JM(6)	6	0	3	3

4.3 In-Sample Fit

Since the two separate models together are equivalent to a restricted version of JM(6), we start by testing those restrictions by comparing the sum of the log likelihoods achieved by the two separate models with the log likelihood achieved by JM(6). Table 2 reports the log likelihood values achieved by the various models. JM(6) achieves a likelihood of $4,086$, while the sum of the likelihoods in the separate models (denoted SM(3,3) in Table 2) is only $4,063$; and this difference is significant at the 5 percent confidence level according to a standard likelihood ratio test. Thus, JM(6) is preferred to two separate models when evaluated purely in statistical terms, providing some evidence to support the case for joint modeling.

Table 2: Log Likelihoods for Models of U.S. and German Yields

This table reports the log likelihoods for and number of free parameters in each of the considered models of U.S. and German yields. It reports results for joint models with different total numbers of factors (JM(x), where x denotes the number of factors) and two separate models with three factors (where we have summed the log likelihoods from and the numbers of parameters in the separate models).

Model	Log likelihood	Number of free parameters
JM(3)	634	28
JM(4)	$1,559$	39
JM(5)	$2,923$	51
JM(6)	$4,086$	64
SM(3,3)	$4,063$	46

However, JM(6) appears to offer only limited economically significant advantages relative to two separate models. Panel $\mathcal{A}$ of Table 3 reports the root mean squared errors (RMSE) between observed and model-implied yields at selected maturities from the various models. As we would expect, JM(6) and the separate models achieve the same fit to the cross section of yields. Panels $\mathcal{B}$ and $\mathcal{C}$ of Table 3 report RMSEs between predictions of yields and the subsequent realizations of yields, at horizons of one and 12 months respectively. At the one-month horizon, the differences between JM(6) and the two separate models are barely perceptible. At the 12-month horizon, JM(6) does achieve a superior predictive accuracy for short-maturity German yields, where the RMSE is 97 basis points in the separate model and 81 basis points in JM(6). However, the differences in the predictive accuracy of JM(6) and the separate models are smaller for U.S. yields and long-maturity German yields.

As explained in Section 3.3, we next turn to the question of whether the cross-sectional and time-series fit of a joint model is affected by restricting some of the factors to be global. Table 2 shows that joint models with fewer than six factors achieve much lower likelihoods

Table 3: Models of U.S. and German Yields: Cross-Sectional and Time-Series Accuracy

This table reports root mean squared errors (in annualized percentage points) between model-implied and actual yields. Panel $\mathcal{A}$ reports the cross-sectional accuracy, that is, the difference between current-period model-implied and actual yields. Panels $\mathcal{B}$ and $\mathcal{C}$ report the time-series accuracy, that is, the difference between model-implied expected yields one and twelve months ahead and subsequent realized yields. In each panel we report results for joint models of U.S. and German yields with different total numbers of factors (JM(x), where x denotes the number of factors) and separate three-factor models (SM(3,3)).

Model	United States (maturity in years)			Germany (maturity in years)		
	0.5	5	10	0.5	5	10
$\mathcal{A}$: Cross-section						
JM(3)	0.10	0.06	0.08	0.23	0.23	0.42
JM(4)	0.12	0.05	0.09	0.14	0.06	0.12
JM(5)	0.12	0.04	0.09	0.02	0.02	0.02
JM(6)	0.02	0.02	0.03	0.02	0.02	0.02
SM(3,3)	0.02	0.02	0.03	0.02	0.02	0.02
$\mathcal{B}$: 1-step ahead						
JM(3)	0.20	0.29	0.27	0.30	0.28	0.44
JM(4)	0.22	0.29	0.27	0.21	0.22	0.24
JM(5)	0.22	0.29	0.27	0.18	0.22	0.20
JM(6)	0.19	0.28	0.25	0.18	0.22	0.20
SM(3,3)	0.19	0.28	0.26	0.19	0.22	0.20
$\mathcal{C}$: 12-step ahead						
JM(3)	1.25	1.01	0.79	0.95	0.82	0.73
JM(4)	1.25	0.98	0.78	0.81	0.81	0.77
JM(5)	1.22	0.91	0.71	0.82	0.79	0.67
JM(6)	1.18	0.90	0.71	0.81	0.78	0.68
SM(3,3)	1.26	0.99	0.78	0.97	0.91	0.73

than JM(6); and unreported likelihood ratio tests show that these differences are statistically significant at the 5 percent significance level, which provides some evidence against imposing that some factors are global. That said, the differences in the fit of JM(5) and JM(6) are fairly small in economic terms. Panel $\mathcal{A}$ of Table 3 shows that JM(5) achieves about the sample cross-sectional fit to German yields, although the RMSEs for U.S. yields rise modestly—to around 10 basis points at the shortest and longest maturities. And Panels $\mathcal{B}$ and $\mathcal{C}$ of Table 3 show there is barely any difference between the time-series accuracy of JM(5) and JM(6). Increasing the number of global factors to two, in JM(4), results in only a modest further deterioration in the fit of the joint model: in-sample RMSEs are between about 5 and 15 basis points across maturities in both countries, while the predictive accuracy remains broadly similar to JM(5), JM(6), and the two-single country models. However, a model with only three global factors cannot fit yields in both countries well simultaneously: while the accuracy of JM(3) for current and future U.S. yields is comparable to the other models, the fit to current and one-step-ahead German yields is substantially worse. That said, the 12-step-ahead predictive accuracy of JM(3) is roughly comparable to that of the other models, suggesting that at this horizon the errors in pricing the cross-section of yields are dwarfed by the errors in forecasting the pricing factors.

In summary, a joint model with the same cross-sectional dimension as two separate models offers only modest gains in terms of predictive power. If we maintain the assumption that three factors are spanned by yields in each country, imposing that there are either one or two global factors in a joint model does not appear to offer material advantages or disadvantages. However, imposing that all three factors are global does lead to a more substantial deterioration in fit.

4.4 Out-of-Sample Forecasts

While JM(6) does offer some modest gains in predicting future yields, particularly for Germany, it has 18 more free parameters compared with two separate models, as shown in Table

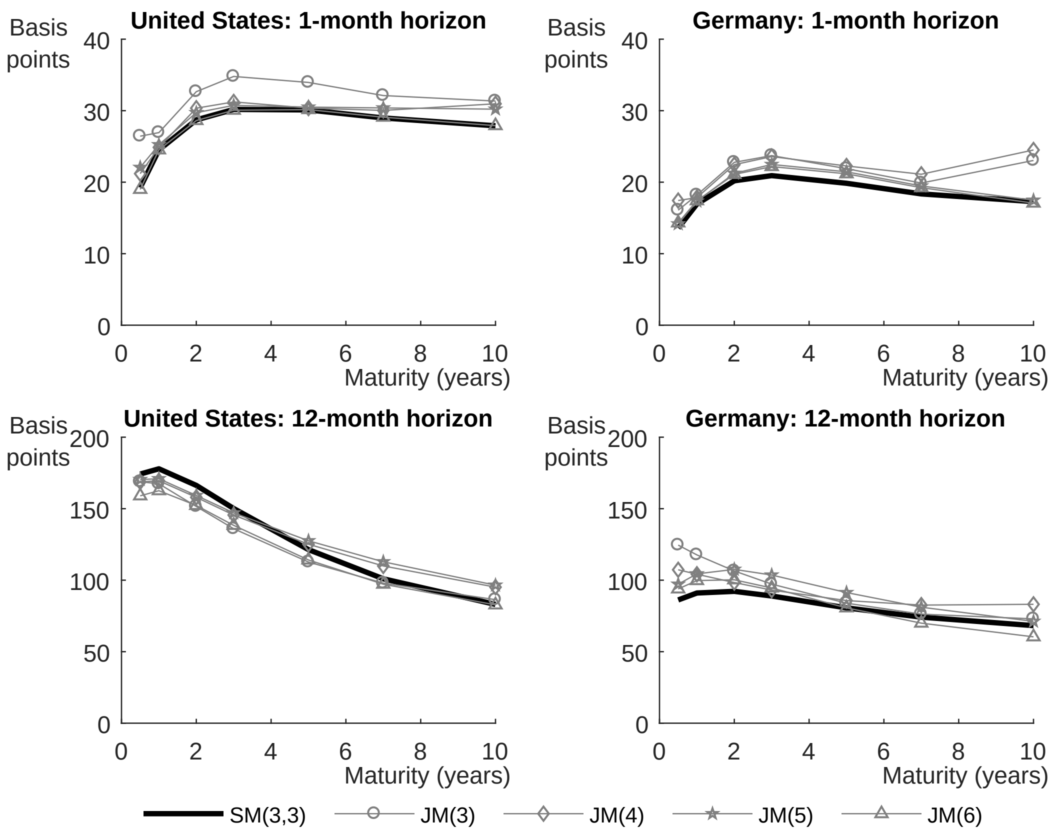
2. This observation raises the question of whether its greater predictive accuracy is a result of over-fitting the data within our sample. On the other hand, imposing that some factors in the joint model are global *reduces* the number of free parameters—and if we have at least two global factors the number of free parameters is actually smaller than the total number of free parameters in two separate models. This second observation raises the additional question of whether joint models with global factors might offer some out-of-sample forecasting advantages relative to larger joint models or even separate models.

We explore the answers to these questions using the following recursive out-of-sample forecasting exercise. We first estimate the models using the first ten years' of data (that is, from January 1990 to December 1999). We then produce forecasts of yields at horizons of up to 12 months and compute the prediction errors relative to subsequent realizations of yields. We then recursively add one more month at a time to the estimation sample, repeating the forecasting exercise at each step. Our final estimation sample runs from January 1990 to December 2006, leaving the final 12 months' of data for evaluating the final set of forecasts.

Figure 1 shows the root mean squared prediction errors (RMSPEs) across our forecasting period for different maturity yields at one- and 12-month forecast horizons. The heavy black lines show the RMSPEs from the separate models, while the remaining lines show the RMSPEs from the various joint models. In general, the differences between the various models are modest. For the United States, none of the joint models can beat a separate model at the one-month forecast horizon. At the 12-month horizon, the joint models perform slightly better than the separate model for short-maturity yields, with JM(6) achieving the lowest RMSPE. [However, even here the improvements are not statistically significant according to unreported Giacomini and White (2006) tests.] For Germany, the separate model out-forecasts all of the joint models at both horizons, with the exception that JM(6) achieves a marginally lower RMSPE for long-maturity yields at the 12-month horizon.

In summary, our out-of-sample forecasting results suggest that none of the joint models offer any robust, economically significant advantages in predicting yields relative to two separate models.

Figure 1: Models of U.S. and German Yields: Out-of-Sample Forecasting Results
This figure reports root mean squared prediction errors between out-of-sample forecasts and realized bond yields. To construct out-of-sample forecasts we first estimate the models using the first ten years' of data (that is, from January 1990 to December 1999) and construct forecasts of yields at horizons up to 12 months ahead. We then recursively add one more month of data, re-estimate the model parameters and construct new forecasts. The final forecasts are produced for a sample ending in December 2006, leaving the final 12 months' of data for forecast evaluation. The left column shows results for forecasts of U.S. yields and the right column shows forecasts of German yields. The heavy black line shows estimates from three-factor models estimated using yields from a single country (SM(3,3). The remaining lines show estimates from joint models with different total numbers of factors (JM(x), where x denotes the number of factors).



4.5 Term Premiums

Another way of comparing the implications of the models for the time-series dynamics of short-term interest rates is to examine model-implied term premiums from the various models. Following Dai and Singleton (2002), we define the term premium in the n-period yield ($TP_{j,n,t}$) as the difference between the model-implied n-period yield and the average expected short-term interest rate over the next n periods, that is, $TP_{j,n,t} = y_{j,n,t} - \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}_t\left[r_{j,t+i}\right]$.

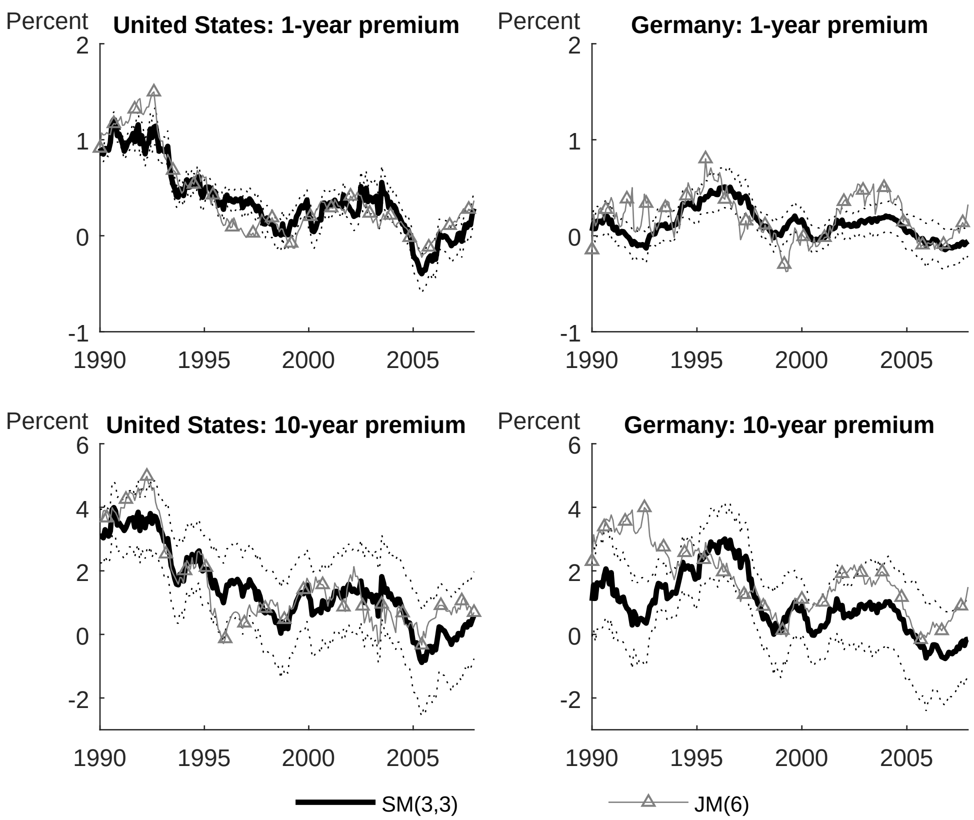
Figure 2 shows the term premiums implied by the separate models (shown by the heavy black lines) and JM(6) (shown by the gray lines). The dashed lines show the 95 percent confidence interval for the estimates from the separate models. Broadly speaking, there is little to distinguish between the term premiums from JM(6) and the separate models. For the United States, the estimates from JM(6) track those from the separate models fairly closely for most of the sample, although the joint model estimates do fall a little outside the 95 percent confidence interval for the separate model at times. The differences for Germany are more substantial at times—particularly during the early 1990s, when the term premiums implied by all of the joint models are significantly higher than those from the separate model. Finally, we note that unreported results (which we omit in the interests of keeping Figure 2 legible) show that the results highlighted in this section are broadly similar for all of the joint models.

4.6 Robustness to Increasing the Number of Spanned Factors

In this section, we verify the robustness of our results to increasing the number of factors spanned by yields in each country. While specifications with three spanned factors are representative of the large majority of the recent literature on separate ATSMs of nominal yields in a single country, some recent studies find that factors that are not spanned by the first three principal components of nominal yields can nevertheless help predict future bond yields or returns. For example, Cochrane and Piazessi (2005, 2008) find that a "return-forecasting factor" extracted from five U.S. nominal forward rates has predictive power for future U.S. bond returns; and Dahlquist and Hasseltoft (2013) find that the average of these return-forecasting factors for Germany, Switzerland, the United Kingdom, and the United States has greater predictive power for bond returns than factors extracted from domestic yields. However, Dahlquist and Hasseltoft (2013) also find that this "global CP factor"

Figure 2: Models of U.S. and German Yields: Term Premiums

This figure reports model-implied estimates of term premiums for United States (left column) and Germany (right column). The top and bottom rows show estimates of one- and ten-year term premiums respectively. The heavy black lines show estimates from three-factor models estimated using yields from a single country (SM(3,3)), with the dashed lines showing a 95 percent confidence interval for these estimates. The grey lines show estimates from a joint model with three local factors for each set of yields (JM(6)).



explains a fairly small proportion of the variance of yield forecast errors—for example, it explains 11 and 1 percent of the variance of ten-year-ahead forecast errors for U.S. one-month and five-year yields, respectively; while for Germany the equivalent proportions are even smaller, at 3 and 2 percent, respectively. Nevertheless, their results raise the question of whether joint models with only three factors spanned by yields in each country omit information that helps to predict future yields in other countries, and therefore understate the potential benefits of joint models. As a robustness check, we therefore examine how our main results change if we allow five factors to be spanned by the yields in each country, as in the models of U.S. nominal yields studied by Duffee (2011) and Adrian et al. (2013).

We start by comparing separate three- and five-factor models of yields for a single country. The results in Panel $\mathcal{A}$ of Tables 3 and 4 show that the scope for improving the cross-sectional fit by adding additional factors in separate models is limited, since the three-factor models already achieve RMSEs of 2 to 3 basis points, and the improvements in the cross-sectional fit from adding the additional factors are therefore modest. Separate three- and five-factor models also achieve similar predictive accuracy, as shown by the results in Panels $\mathcal{B}$ and $\mathcal{C}$ of Tables 3 and 4. At a one-month horizon, there are essentially no reductions in the RMSE from adding two additional factors, while at the 12-month horizon the reductions in the RMSEs are less than 10 basis points. Thus, our results broadly confirm the findings of Duffee (2011), who also shows that increasing the number of factors from three to five in ATSMs of U.S. nominal yields results in only small improvements in cross-sectional and time-series accuracy.

We next consider whether there are any benefits from joint modeling once we have allowed five factors to be spanned by the yields in each country. Here, we limit our attention to a single joint model that has five local factors for each country (that is, ten factors in total), which we compare with the two separate five-factor models. With 154 parameters, such a joint model is almost certainly unnecessarily heavily parameterized. However, it does at least ensure that we can be confident that the joint model will capture any marginal information

Table 4: Models of U.S. and German Yields with Five Spanned Factors: Cross-Sectional and Time-Series Accuracy

This table reports root mean squared errors (in annualized percentage points) between model-implied and actual yields. Panel $\mathcal{A}$ reports the cross-sectional accuracy, that is, the difference between current-period model-implied and actual yields. Panels $\mathcal{B}$ and $\mathcal{C}$ report the time-series accuracy, that is, the difference between model-implied expected yields one and twelve months ahead and subsequent realized yields. In each panel we report results for a joint model of U.S. and German yields with ten factors (JM(10)) and separate models with five factors (SM(5,5)).

Model	United States (maturity in years)			Germany (maturity in years)		
	0.5	5	10	0.5	5	10
$\mathcal{A}$: Cross-section						
JM(10)	0.00	0.00	0.00	0.00	0.00	0.00
SM(5,5)	0.00	0.00	0.00	0.00	0.00	0.00
$\mathcal{B}$: 1-step ahead						
JM(10)	0.19	0.28	0.25	0.18	0.21	0.20
SM(5,5)	0.19	0.28	0.25	0.18	0.22	0.20
$\mathcal{C}$: 12-step ahead						
JM(10)	1.06	0.83	0.68	0.80	0.76	0.66
SM(5,5)	1.16	0.94	0.76	0.97	0.93	0.74

in U.S. yields that is relevant for forecasting German yields (and *vice versa*). Similar to the three-factor case, a likelihood ratio test rejects the (50) zero restrictions implied by the separate five-factor models at standard confidence levels (the log likelihood of the joint model is 5,676 and the sum of the log likelihoods for the two separate models is 5,627). However, Panels $\mathcal{A}$ and $\mathcal{B}$ of Table 4 show that the cross-sectional and predictive accuracies at a one-month horizon are almost identical for the joint and separate models. At the 12-month horizon, the largest reductions in RMSEs from using a joint model rather than separate models are for short-maturity German yields. However, these (in-sample) gains for German yields are almost identical whether there are three or five spanned factors by the yields of each country. Thus, we conclude that our main conclusion is robust to increasing the size of joint models.

5 An Application to U.S. Nominal and Real Yields

We now turn to our application to joint models of U.S. nominal and real yields. In Section 5.1, we briefly describe our data set and the factor structures we consider. In Section 5.2, we present results on the in-sample fit and predictive accuracy of the models. In Section 5.3, we discuss the potential use of joint models as a means of extending the short available time-series of real yields.

5.1 Data and Factor Structure

We use the same sample of U.S. nominal yields described in Section 4.1—that is, end-month yields with maturities of six months and one, two, three, five, seven, and ten years over the period January 1990 to December 2007. Our sample of U.S. real yields is derived from the yields on Treasury Inflation Protected Securities using the method of Gürkaynak et al. (2010).[8] One practical difficulty with this application is the limited availability of TIPS-implied real yields; we use a sample of end-month real yields that runs from January 1999 to December 2007, with maturities of five, seven, and ten years. Given this much smaller range of maturities, we assume that only two factors are required to fit the term structure of real yields. Specifically, we compare separate models of nominal and real yields with three and two factors, respectively, with joint models that also have three factors spanned by nominal yields and two factors spanned by real yields. Thus, we consider three different joint models, with: (1) three factors local to nominal yields and two local to real yields (that is, five in total, denoted JM(5)); (2) one global factor, two factors local to nominal yields, and one local to real yields (JM(4)); and (3) two global factors and one local to nominal yields (JM(3)).

[8] Updates of these yields are published by the Board of Governors of the Federal Reserve System.

5.2 In-Sample Fit

Table 5 reports results on the fit of our joint and separate models of nominal and real yields. In summary, the main conclusions are similar to those we drew for our application to joint models of U.S. and German nominal yields: Panel $\mathcal{A}$ shows that the separate models and JM(5) achieve the same in-sample fit to the cross-section of yields (as expected), while Panels $\mathcal{B}$ and $\mathcal{C}$ show that the predictive accuracy of the two models is also similar. Imposing that one or two factors are common (that is, JM(4) and JM(3)) results in a modest deterioration in the cross-sectional fit of the model and a more substantial deterioration in the 12-month-ahead predictive accuracy. Thus, we conclude that there is also little of economic significance to be gained from joint modeling of U.S. nominal and real yields.

Table 5: Models of U.S. Nominal and Real Yields: Cross-Sectional and Time-Series Accuracy

This table reports root mean squared errors (in annualized percentage points) between model-implied and actual yields. Panel $\mathcal{A}$ reports the cross-sectional accuracy, that is, the difference between current-period model-implied and actual yields. Panels $\mathcal{B}$ and $\mathcal{C}$ report the time-series accuracy, that is, the difference between model-implied expected yields one and twelve months ahead and subsequent realized yields. In each panel we report results for joint models of nominal and real yields with different total numbers of factors (JM(x), where x denotes the number of factors) and separate models (SM(3,2)) of nominal yields (with three factors) and real yields (with two factors).

	Nominal (maturity in years)			Real (maturity in years)	
	0.5	5	10	5	10
$\mathcal{A}$: Cross section					
JM(3)	0.11	0.04	0.09	0.06	0.06
JM(4)	0.11	0.04	0.09	0.01	0.01
JM(5)	0.02	0.02	0.03	0.01	0.01
SM(3,2)	0.02	0.02	0.03	0.01	0.01
$\mathcal{B}$: 1-step ahead					
JM(3)	0.22	0.29	0.26	0.23	0.20
JM(4)	0.22	0.29	0.27	0.24	0.18
JM(5)	0.19	0.28	0.25	0.22	0.17
SM(3,2)	0.19	0.28	0.26	0.23	0.18
$\mathcal{C}$: 12-step ahead					
JM(3)	1.57	1.04	0.75	0.83	0.56
JM(4)	1.34	0.97	0.72	1.00	0.56
JM(5)	1.25	0.93	0.71	0.59	0.43
SM(3,2)	1.26	0.99	0.78	0.61	0.42

5.3 Using Joint Models to Extend the Sample of TIPS-implied Real Yields

One caveat to this conclusion is that the joint model of nominal and real yields has another potential use: to extrapolate the sample period for real yields to cover a longer period, as proposed previously by D'Amico et al. (2014). For example, in JM(3), all of the factors spanned by real yields are also spanned by nominal yields. Thus, when using this joint model we can use nominal yields to filter the factors before the start of the TIPS market and hence compute model-implied real yields before 1999, (although it is of course worth remembering that JM(3) predict yields less accurately than separate models or a joint model with only local factors.

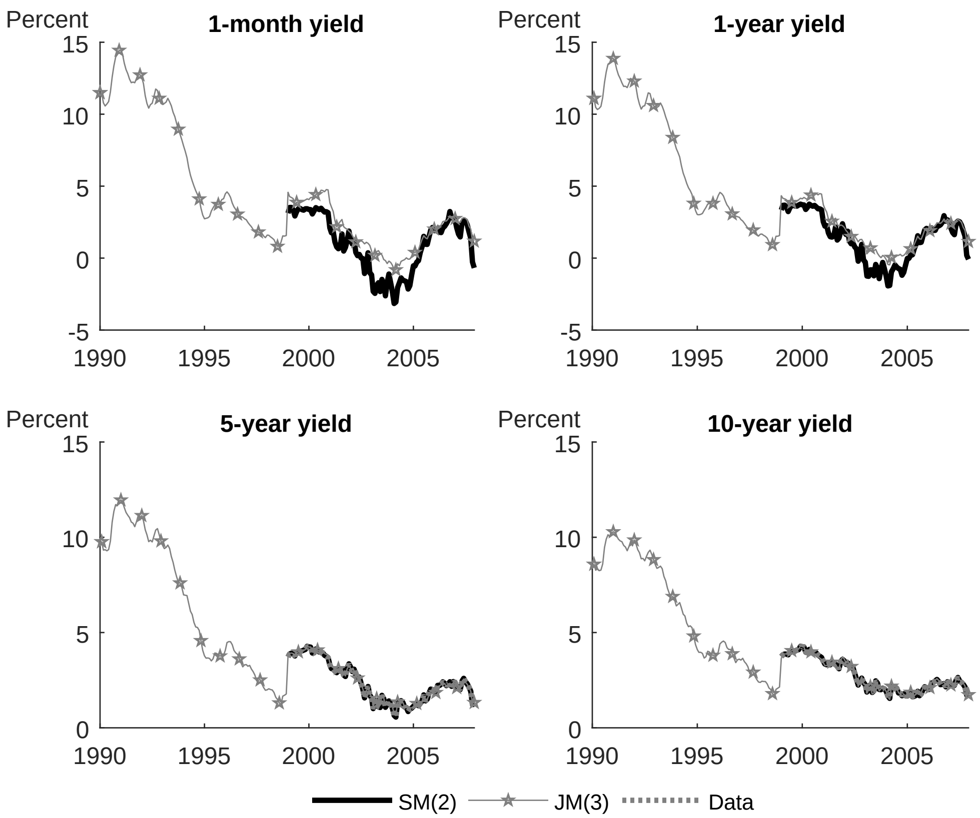
Figure 3 shows the fitted yields from the separate models (the heavy black lines) and JM(3) (the gray lines) at maturities of one month and one, five, and ten years. Over the period since 1999, implied five- and ten-year real yields from both the separate models and JM(3) have closely matched the observed yields (shown by the heavy broken line). At shorter maturities, where there are no observed real yields, we do observe some differences between the separate models and JM(3). Figure 3 also shows that JM(3) implies that hypothetical real yields have fallen substantially at all maturities since 1990.

6 Conclusion

While the large majority of studies of affine term structure models estimate models of nominal yields in a single country, a growing number of studies have estimated joint models of yields in multiple countries or of nominal and real yields within a single country. This paper argues that the case for such joint models is not particularly compelling. There is nothing to be gained from joint models in terms of the in-sample fit to yields, since we can fit any term structure of yields closely using a small number of factors extracted from those same yields. And joint models also have very similar predictive accuracy compared with separate models.

Figure 3: Models of U.S. Nominal and Real Yields: Fitted TIPS Yields

This figure reports model-implied fitted U.S. real yields at different maturities. The heavy black lines show results from a two-factor model estimated using only real yields (SM(2)) over a sample from January 1999 to December 2007. The gray lines show results from a joint nominal-real model with two global factors and one factor local to nominal yields (JM(3)). Observed data are not available for the one-month and one-year real yields.



References

Abrahams, M., Adrian, T., Crump, R. K. & Moench, E. (2016), 'Decomposing Real and Nominal Yield Curves', *Journal of Monetary Economics* **84**, 182–200.

Adrian, T., Crump, R. K. & Moench, E. (2013), 'Pricing the Term Structure with Linear Regressions', *Journal of Financial Economics* **110**(1), 110–138.

Anderson, B., Hammon, P. J. & Ramezani, C. A. (2010), 'Affine Models of the Joint Dynamics of Exchange Rates and Interest Rates', *Journal of Financial and Quantitative Analysis* **45**(5), 1341–1365.

Backus, D. K., Foresi, S. & Telmer, C. I. (2001), 'Affine Term Structure Models and the Forward Premium Anomaly', *Journal of Finance* **51**(1), 279–304.

Bauer, M. D. & Rudebusch, G. D. (2017), 'Resolving the Spanning Puzzle in Macro-Finance Term Structure Models', *Review of Finance* **21**(2), 511–553.

Cochrane, J. H. & Piazessi, M. (2005), 'Bond Risk Premia', *American Economic Review* **95**(1), 138–160.

Cochrane, J. H. & Piazessi, M. (2008), 'Decomposing the yield curve', *Working paper* .

Dahlquist, M. & Hasseltoft, H. (2013), 'International Bond Risk Premia', *Journal of International Economics* **90**(1), 17–32.

Dai, Q. & Singleton, K. J. (2000), 'Specification Analysis of Affine Term Structure Models', *Journal of Finance* **55**(5), 1943–1978.

Dai, Q. & Singleton, K. J. (2002), 'Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure', *Journal of Financial Economics* **63**(3), 415–441.

D'Amico, S., Kim, D. H. & Wei, M. (2014), 'Tips from TIPS: the informational content of Treasury Inflation-Protected Security prices', *Federal Reserve Board Finance and Economics Discussion Series* **2014-24**.

Diebold, F. X., Li, C. & Yue, V. Z. (2008), 'Global Yield Curve Dynamics and Iinteractions: A Dynamic Nelson-Siegel Approach', *Journal of Econometrics* **146**(2), 351–363.

Diez de los Rios, A. (2008), 'McCallum Rules, Exchange Rates, and the Term Structure of Interest Rates', *Bank of Canada Staff Working Paper* **2008-43**.

Diez de los Rios, A. (2017), 'Optimal Estimation of Multi-Country Gaussian Dynamic Term Structure Models Using Linear Regressions', *Bank of Canada Staff Working Paper* **2017-33**.

Duffee, G. R. (2002), 'Term Premia and Interest Rate Forecasts in Affine Models', *Journal of Finance* **57**(1), 405–443.

Duffee, G. R. (2011), 'Information in (and not in) the Term Structure', *Review of Financial Studies* **24**(9), 2895–2934.

Egorov, A. V., Li, H. & Ng, D. (2011), 'A Tale of Two Yield Curves: Modeling the Joint Term Structure of Dollar and Euro Interest Rates', *Journal of Econometrics* **162**(1), 55–70.

Giacomini, R. & White, H. (2006), 'Tests of Conditional Predictive Ability', *Econometrica* **74**(6), 1545–1578.

Gürkaynak, R. S., Sack, B. & Wright, J. H. (2007), 'The US Treasury Yield Curve: 1961 to the Present', *Journal of Monetary Economics* **54**(8), 2291–2304.

Gürkaynak, R. S., Sack, B. & Wright, J. H. (2010), 'The TIPS Yield Curve and Inflation Compensation', *American Economic Journal: Macroeconomics* **2**(1), 70–92.

Hamilton, J. D. & Wu, J. C. (2012), 'Identification and Estimation of Affine Term Structure Models', *Journal of Econometrics* **168**(2), 315–331.

Joslin, S., Priebsch, M. A. & Singleton, K. J. (2014), 'Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks', *Journal of Finance* **69**(3), 1197–1233.

Joslin, S., Singleton, K. J. & Zhu, H. (2011), 'A New Perspective on Gaussian Dynamic Term Structure Models', *Review of Financial Studies* **24**(3), 926–970.

Joyce, M. A. S., Lildholdt, P. & Sorensen, S. (2010), 'Extracting Inflation Expectations and Inflation Risk Premia from the Term Structure: A Joint Model of the UK Nominal and Real Yield Curves', *Journal of Banking and Finance* **34**(2), 281–294.

Kaminska, I., Meldrum, A. C. & Smith, J. (2013), 'A Global Model of International Yield Curves: No-Arbitrage Term Structure Approach', *International Journal of Finance and Economics* **18**(4), 352–374.

Svensson, L. E. O. (1994), 'Estimating and Interpreting Forward Rates: Sweden 1992-4', *National Bureau of Economic Research Working Paper* **4871**.

Yung, J. (2017), 'Can Interest Rate Factors Explain Exchange Rate Fluctuations?', *Working Paper* .

Appendix A: Solution for Bond Yields in a Separate Model

In this appendix we show that the solution for domestic nominal bond yields takes the form in equations (7) and (8). We first guess that the solution for bond prices takes the exponential affine form

$$P_{j,n,t} = \exp\left(a_{j,n,\mathcal{S}} + \mathbf{b}'_{j,n,\mathcal{S}}\mathbf{x}_{j,t}\right).$$

Substituting this guess into equation (5) and taking logarithms gives

$$a_{j,n,\mathcal{S}} + \mathbf{b}'_{j,n,\mathcal{S}}\mathbf{x}_{j,t} = \log \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-r_{j,t}\right)\exp\left(a_{j,n-1,\mathcal{S}} + \mathbf{b}'_{j,n-1,\mathcal{S}}\mathbf{x}_{j,t+1}\right)\right],$$

and combining with equations (1) and (6) gives

$$\begin{aligned}
a_{j,n,\mathcal{S}} + \mathbf{b}'_{j,n,\mathcal{S}}\mathbf{x}_{j,t} &= \log \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\delta_{j,0,\mathcal{S}} - \boldsymbol{\delta}'_{j,1,\mathcal{S}}\mathbf{x}_{j,t}\right)\exp\left(\begin{array}{c} a_{j,n-1,\mathcal{S}} + \\ \mathbf{b}'_{j,n-1,\mathcal{S}}\left(\boldsymbol{\mu}^{\mathbb{Q}_j}_{j,\mathcal{S}} + \boldsymbol{\Phi}^{\mathbb{Q}_j}_{j,\mathcal{S}}\mathbf{x}_{j,t} + \boldsymbol{\Sigma}_{j,\mathcal{S}}\boldsymbol{\varepsilon}^{\mathbb{Q}}_{j,t+1}\right)\end{array}\right)\right] \\
&= -\delta_{j,0,\mathcal{S}} - \boldsymbol{\delta}'_{j,1,\mathcal{S}}\mathbf{x}_{j,t} + a_{j,n-1,\mathcal{S}} + \mathbf{b}'_{j,n-1,\mathcal{S}}\boldsymbol{\mu}^{\mathbb{Q}_j}_{j,\mathcal{S}} + \mathbf{b}'_{j,n-1,\mathcal{S}}\boldsymbol{\Phi}^{\mathbb{Q}_j}_{j,\mathcal{S}}\mathbf{x}_{j,t} \\
&\quad + \frac{1}{2}\mathbf{b}'_{j,n-,\mathcal{S}1}\boldsymbol{\Sigma}_{j,\mathcal{S}}\boldsymbol{\Sigma}'_{j,\mathcal{S}}\mathbf{b}_{j,n-1,\mathcal{S}}.
\end{aligned}$$

Matching coefficients gives equations (7) and (8). The boundary conditions that $a_{j,0,\mathcal{S}} = 0$ and $\mathbf{b}_{j,n,\mathcal{S}} = \mathbf{0}$ follow from the fact that the price of a zero-period bond paying one unit at maturity must be equal to one.

Appendix B: Solution for Yields on the Second Asset Class in a Joint Model

In this appendix we show that the solution for foreign bond yields take the form in equations (21)-(22). We first guess that the solution for foreign bond prices takes the exponential affine form

$$P_{2,n,t} = \exp\left(a_{2,n} + \mathbf{b}'_{2,n}\mathbf{x}_t\right).$$

Substituting this solution into equation (15) and taking logarithms gives

$$a_{2,n} + \mathbf{b}'_{2,n}\mathbf{x}_t = \log \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-r_{1,t} + \Delta s_{t+1}\right)\exp\left(a_{2,n-1} + \mathbf{b}'_{2,n-1}\mathbf{x}_{t+1}\right)\right],$$

and combining with equations (10), (11), and (16) gives

$$\begin{aligned} a_{2,n} + \mathbf{b}'_{2,n}\mathbf{x}_t &= \log \mathbb{E}_t^{\mathbb{Q}}\left[\begin{array}{c} \exp\left(-\delta_{1,0} - \boldsymbol{\delta}'_{1,1}\mathbf{x}_t + s_0 + \mathbf{s}'_1\left(\boldsymbol{\mu}^{\mathbb{Q}_1} + \boldsymbol{\Phi}^{\mathbb{Q}_1}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}^{\mathbb{Q}}_{t+1}\right)\right) \times \ldots \\ \exp\left(a_{2,n-1} + \mathbf{b}'_{2,n-1}\left(\boldsymbol{\mu}^{\mathbb{Q}_1} + \boldsymbol{\Phi}^{\mathbb{Q}_1}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}^{\mathbb{Q}}_{t+1}\right)\right) \end{array}\right] \\ &= -\delta_{1,0} - \boldsymbol{\delta}'_{1,1}\mathbf{x}_t + s_0 + \mathbf{s}'_1\boldsymbol{\mu}^{\mathbb{Q}_1} + \left(\mathbf{s}_1 + \mathbf{b}'_{2,n-1}\right)'\boldsymbol{\Phi}^{\mathbb{Q}_1}\mathbf{x}_t + a_{2,n-1} + \mathbf{b}'_{2,n-1}\boldsymbol{\mu}^{\mathbb{Q}_1}\ldots \\ &\quad + \frac{1}{2}\left(\mathbf{s}_1 + \mathbf{b}'_{2,n-1}\right)'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\left(\mathbf{s}_1 + \mathbf{b}'_{2,n-1}\right). \end{aligned}$$

Matching coefficients gives equations (21) and (22). The boundary conditions that $a_{2,0} = 0$ and $\mathbf{b}_{2,0} = \mathbf{0}$ follow from the fact that the price of a zero-period bond paying one unit at maturity must be equal to one.

Appendix C: Second Short Rate in the Joint Model

In this appendix, we show that the short rate for the second class of bonds takes the form in equation (27). Note that from equations (20)-(22) the short rate that prices the second class of bonds $r_{2,t} \equiv y_{2,1,t}$ is given by

$$\begin{aligned} r_{2,t} &= -a_{2,1} - \mathbf{b}_{2,1}\mathbf{x}_t \\ &= \delta_{1,0} - s_0 - \mathbf{s}'_1\boldsymbol{\mu}^{\mathbb{Q}_1} - \frac{1}{2}\mathbf{s}'_1\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 - \left(\mathbf{s}'_1\boldsymbol{\Phi}^{\mathbb{Q}_1} - \boldsymbol{\delta}'_{1,1}\right)\mathbf{x}_t. \end{aligned}$$

Thus, $r_{2,t}$ takes the form in equation (27) where

$$\delta_{2,0} = \delta_{1,0} - s_0 - \mathbf{s}_1' \boldsymbol{\mu}^{\mathbb{Q}_1} - \frac{1}{2} \mathbf{s}_1' \boldsymbol{\Sigma} \boldsymbol{\Sigma}' \mathbf{s}_1 \text{ and} \tag{32}$$

$$\boldsymbol{\delta}_{2,1} = \boldsymbol{\delta}_{1,1} - \left(\boldsymbol{\Phi}^{\mathbb{Q}_1} \right)' \mathbf{s}_1, \tag{33}$$

as stated in the main text.

Appendix D: SM(3,3) as a Special Case of JM(6)

In this appendix we show that under the assumption of complete markets two separate three-factor models with factors $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$, respectively, can be written as a six-factor joint model. We first show that the joint model is symmetric—that is, that we can equivalently price the second class of bonds using the second asset as the numeraire, as is the case in a separate model of the second class of bonds. We then show that we can write two separate three-factor models as a restricted case of JM(6).

Symmetry of the Joint Model

As shown by Backus et al. (2001), in the presence of complete markets the stochastic discount factor that prices the second class of bonds under the $\mathbb{P}$ measure—that is $M_{2,t+1}$ such that $P_{2,n,t} = \mathbb{E}_t \left[M_{2,t+1} P_{2,n-1,t+1} \right]$—must satisfy

$$\log M_{2,t+1} = \Delta s_{t+1} + \log M_{1,t+1}. \tag{34}$$

Following Diez de los Rios (2008), combining equations (11), (13), and (34) gives

$$\log M_{2,t+1} = s_0 + \mathbf{s}_1' \mathbf{x}_{t+1} - r_{1,t} - \frac{1}{2} \boldsymbol{\lambda}_{1,t}' \boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}' \boldsymbol{\varepsilon}_{t+1},$$

and substituting in equation (12) gives

$$\log M_{2,t+1} = s_0 + \mathbf{s}_1' \left(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}\right) - r_{1,t} - \frac{1}{2}\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}'\boldsymbol{\varepsilon}_{t+1},$$

Using the mapping between the $\mathbb{P}$ measure and the $\mathbb{Q}_1$ measure, that is, $\boldsymbol{\mu} = \boldsymbol{\mu}^{\mathbb{Q}_1} + \boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,0}$ and $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\mathbb{Q}_1} + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_{1,1}$, gives

$$\begin{aligned}
\log M_{2,t+1} &= s_0 + \mathbf{s}_1' \left(\boldsymbol{\mu}^{\mathbb{Q}_1} + \boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,0} + \left(\boldsymbol{\Phi}^{\mathbb{Q}_1} + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_{1,1}\right)\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}\right) \\
&\quad - r_{1,t} - \frac{1}{2}\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}'\boldsymbol{\varepsilon}_{t+1} \\
&= s_0 + \mathbf{s}_1' \left(\boldsymbol{\mu}^{\mathbb{Q}_1} + \boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,0} + \boldsymbol{\Phi}^{\mathbb{Q}_1}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_{1,1}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}\right) \\
&\quad - r_{1,t} - \frac{1}{2}\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}'\boldsymbol{\varepsilon}_{t+1} \\
&= s_0 + \mathbf{s}_1'\boldsymbol{\mu}^{\mathbb{Q}_1} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,0} + \mathbf{s}_1'\boldsymbol{\Phi}^{\mathbb{Q}_1}\mathbf{x}_t + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_{1,1}\mathbf{x}_t + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&\quad - r_{1,t} - \frac{1}{2}\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}'\boldsymbol{\varepsilon}_{t+1}
\end{aligned}$$

Substituting in the definition of the short rates in equations (10) and (27) gives

$$\begin{aligned}
\log M_{2,t+1} &= \delta_{1,0} - \delta_{2,0} - \mathbf{s}_1'\boldsymbol{\mu}^{\mathbb{Q}_1} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\mu}^{\mathbb{Q}_1} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,0} + \left(\boldsymbol{\delta}_{1,1} - \boldsymbol{\delta}_{2,1}\right)'\mathbf{x}_t \\
&\quad + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_{1,1}\mathbf{x}_t + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} - \delta_{1,0} - \boldsymbol{\delta}_{1,1}'\mathbf{x}_t - \frac{1}{2}\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}'\boldsymbol{\varepsilon}_{t+1} \\
&= -\delta_{2,0} - \boldsymbol{\delta}_{2,1}'\mathbf{x}_t - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,0} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_{1,1}\mathbf{x}_t + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&\quad - \frac{1}{2}\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}'\boldsymbol{\varepsilon}_{t+1} \\
&= -r_{2,t} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,0} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_{1,1}\mathbf{x}_t + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&\quad - \frac{1}{2}\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}'\boldsymbol{\varepsilon}_{t+1}
\end{aligned}$$

Substituting in the definition of the price of risk $\boldsymbol{\lambda}_{1,t} = \boldsymbol{\lambda}_{1,0} + \boldsymbol{\Lambda}_{1,1}\mathbf{x}_t$ gives

$$\log M_{2,t+1} = -r_{2,t} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{1,t} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} - \frac{1}{2}\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t} - \boldsymbol{\lambda}_{1,t}'\boldsymbol{\varepsilon}_{t+1}$$

If we define

$$\boldsymbol{\lambda}_{2,t} = \boldsymbol{\lambda}_{1,t} - \boldsymbol{\Sigma}'\mathbf{s}_1 \tag{35}$$

and substitute this into the previous equation we obtain

$$\begin{aligned}
\log M_{2,t+1} &= -r_{2,t} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\left(\boldsymbol{\lambda}_{2,t} + \boldsymbol{\Sigma}'\mathbf{s}_1\right) + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&\quad -\frac{1}{2}\left(\boldsymbol{\lambda}_{2,t} + \boldsymbol{\Sigma}'\mathbf{s}_1\right)'\left(\boldsymbol{\lambda}_t^* + \boldsymbol{\Sigma}'\mathbf{s}_1\right) - \left(\boldsymbol{\lambda}_{2,t} + \boldsymbol{\Sigma}'\mathbf{s}_1\right)'\boldsymbol{\varepsilon}_{t+1} \\
&= -r_{2,t} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{2,t} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&\quad -\frac{1}{2}\left(\boldsymbol{\lambda}_{2,t}' + \mathbf{s}_1'\boldsymbol{\Sigma}\right)\left(\boldsymbol{\lambda}_{2,t}' + \boldsymbol{\Sigma}'\mathbf{s}_1\right) - \left(\boldsymbol{\lambda}_{2,t}' + \mathbf{s}_1'\boldsymbol{\Sigma}\right)\boldsymbol{\varepsilon}_{t+1} \\
&= -r_{2,t} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{2,t} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&\quad -\frac{1}{2}\left(\boldsymbol{\lambda}_{2,t}'\left(\boldsymbol{\lambda}_{2,t} + \boldsymbol{\Sigma}'\mathbf{s}_1\right) + \mathbf{s}_1'\boldsymbol{\Sigma}\left(\boldsymbol{\lambda}_{2,t} + \boldsymbol{\Sigma}'\mathbf{s}_1\right)\right) - \boldsymbol{\lambda}_{2,t}'\boldsymbol{\varepsilon}_{t+1} - \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&= -r_{2,t} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{2,t} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&\quad -\frac{1}{2}\left(\boldsymbol{\lambda}_{2,t}'\boldsymbol{\lambda}_{2,t} + \boldsymbol{\lambda}_{2,t}'\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{2,t} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1\right) - \boldsymbol{\lambda}_{2,t}'\boldsymbol{\varepsilon}_{t+1} - \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&= -r_{2,t} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{2,t} + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 + \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&\quad -\frac{1}{2}\boldsymbol{\lambda}_{2,t}'\boldsymbol{\lambda}_{2,t} - \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\lambda}_{2,t} - \frac{1}{2}\mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 - \boldsymbol{\lambda}_{2,t}'\boldsymbol{\varepsilon}_{t+1} - \mathbf{s}_1'\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1} \\
&= -r_{2,t} - \frac{1}{2}\boldsymbol{\lambda}_{2,t}'\boldsymbol{\lambda}_{2,t} - \boldsymbol{\lambda}_{2,t}'\boldsymbol{\varepsilon}_{t+1}.
\end{aligned}$$

Thus, the stochastic discount factor that prices the second class of bonds ($M_{2,t+1}$) takes the same form as it would in a separate model if the factors follow the law of motion under the $\mathbb{Q}_2$ measure, that is,

$$\mathbf{x}_{t+1} = \boldsymbol{\mu}^{\mathbb{Q}_2} + \boldsymbol{\Phi}^{\mathbb{Q}_2}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}_2}, \tag{36}$$

where $\boldsymbol{\mu}^{\mathbb{Q}_2} = \boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{\lambda}_{2,0}$, $\boldsymbol{\Phi}^{\mathbb{Q}_2} = \boldsymbol{\Phi} - \boldsymbol{\Sigma}\boldsymbol{\Lambda}_{2,1}$, $\boldsymbol{\Lambda}_{2,1} = \boldsymbol{\Lambda}_{1,1}$, and $\boldsymbol{\lambda}_{2,0} = \boldsymbol{\lambda}_{1,0} - \boldsymbol{\Sigma}'\mathbf{s}_1$. Re-arranging these restrictions and combining them with the definition of the price of risk gives

$$\begin{aligned}
\boldsymbol{\mu}^{\mathbb{Q}_2} &= \boldsymbol{\mu}^{\mathbb{Q}_1} + \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 \text{ and} && (37) \\
\boldsymbol{\Phi}^{\mathbb{Q}_2} &= \boldsymbol{\Phi}^{\mathbb{Q}_1}. && (38)
\end{aligned}$$

Writing the Separate Models as a Joint Model

We next show that we can write two separate models as a joint model that satisfies the parameter restrictions in equations (32), (33), (37), and (38).

It will be convenient if we first apply a invariant level-shift to the factors in the separate model of the second class of yields $\widetilde{\mathbf{x}}_{2,t} = \mathbf{x}'_{2,t} + \boldsymbol{\theta}$. Given our normalization restrictions, the $\mathbb{Q}_2$ dynamics of the model written in terms of the level-shifted factors are

$$\widetilde{\mathbf{x}}_{2,t+1} = \widetilde{\boldsymbol{\mu}}^{\mathbb{Q}_2}_{2,\mathcal{S}} + \boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}} \widetilde{\mathbf{x}}_{2,t} + \boldsymbol{\varepsilon}_{2,t},$$

where $\widetilde{\boldsymbol{\mu}}^{\mathbb{Q}_2}_{2,\mathcal{S}} = \left(\mathbf{I} - \boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)\boldsymbol{\theta}$. The short rate equation is given by

$$r_{2,t} = \widetilde{\delta}_{2,0,\mathcal{S}} + \boldsymbol{\delta}'_{2,1,\mathcal{S}} \widetilde{\mathbf{x}}_{2,t},$$

where $\widetilde{\delta}_{2,0,\mathcal{S}} = \delta_{2,0,\mathcal{S}} - \boldsymbol{\delta}'_{2,1,\mathcal{S}}\boldsymbol{\theta}$.

The next step is to re-write each of the separate models using the augmented factor vector $\mathbf{x}_t = \left[\mathbf{x}'_{1,t}, \widetilde{\mathbf{x}}'_{2,t}\right]'$. Under our normalization, the $\mathbb{Q}_1$ and $\mathbb{Q}_2$ dynamics in the two separate models (that is, equation (6)) are given by

$$\begin{bmatrix} \mathbf{x}_{1,t+1} \\ \widetilde{\mathbf{x}}'_{2,t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \widetilde{\boldsymbol{\mu}}^{\mathbb{Q}_1}_{2,\mathcal{S}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}} & \mathbf{0} \\ \boldsymbol{\Phi}^{\mathbb{Q}_1}_{1,21,\mathcal{S}} & \boldsymbol{\Phi}^{\mathbb{Q}_1}_{1,22,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t} \\ \widetilde{\mathbf{x}}'_{2,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{\varepsilon}_{2,t} \end{bmatrix} \text{ and}$$
$$\begin{bmatrix} \mathbf{x}_{1,t+1} \\ \widetilde{\mathbf{x}}'_{2,t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{\mathbb{Q}_2}_{1,\mathcal{S}} \\ \widetilde{\boldsymbol{\mu}}^{\mathbb{Q}_2}_{2,\mathcal{S}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,11,\mathcal{S}} & \boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,12,\mathcal{S}} \\ \mathbf{0} & \boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t} \\ \widetilde{\mathbf{x}}'_{2,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{\varepsilon}_{2,t} \end{bmatrix},$$

respectively. And the short rates in the two separate models (that is, equation (1)) can be written as

$$r_{1,t} = \delta_{1,0,\mathcal{S}} + \begin{bmatrix} \boldsymbol{\delta}'_{1,1,\mathcal{S}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t} \\ \widetilde{\mathbf{x}}'_{2,t} \end{bmatrix} \text{ and}$$

$$r_{2,t} = \widetilde{\delta}_{2,0,\mathcal{S}} + \begin{bmatrix} \mathbf{0} & \boldsymbol{\delta}'_{2,1,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t} \\ \widetilde{\mathbf{x}}'_{2,t} \end{bmatrix},$$

respectively. Because $\widetilde{\mathbf{x}}'_{2,t}$ are unspanned factors in the first model and $\mathbf{x}_{1,t}$ are unspanned factors in the second model, it must be the case that the parameters $\widetilde{\boldsymbol{\mu}}^{\mathbb{Q}_1}_{2,\mathcal{S}}$, $\boldsymbol{\mu}^{\mathbb{Q}_2}_{1,\mathcal{S}}$, $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{1,21,\mathcal{S}}$, $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{1,22,\mathcal{S}}$, $\boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,11,\mathcal{S}}$, and $\boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,12,\mathcal{S}}$ are unidentified. In addition, each of the two separate models leaves the parameters s_0 and $\mathbf{s}_1$ unidentified. We are therefore free to set these parameters to any values without affecting the properties of the separate models.

First, with $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{1,21,\mathcal{S}} = \boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,12,\mathcal{S}} = \mathbf{0}$, $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{1,22,\mathcal{S}} = \boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}$, and $\boldsymbol{\Phi}^{\mathbb{Q}_2}_{2,11,\mathcal{S}} = \boldsymbol{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}$ equation (38) is satisfied. We can then set $\mathbf{s}_1$ in order satisfy equation (33), that is,

$$
\begin{aligned}
\mathbf{s}_1 &= \left(\mathbf{\Phi}^{\mathbb{Q}_1}\right)'^{-1}\left(\begin{bmatrix}\boldsymbol{\delta}_{1,1,\mathcal{S}}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\delta}_{2,1,\mathcal{S}}\end{bmatrix}\right)\\
&= \left(\begin{bmatrix}\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\end{bmatrix}\right)'^{-1}\left(\begin{bmatrix}\boldsymbol{\delta}_{1,1,\mathcal{S}}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\delta}_{2,1,\mathcal{S}}\end{bmatrix}\right)\\
&= \begin{bmatrix}\left(\left(\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}\right)'\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1}\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{\delta}_{1,1,\mathcal{S}}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\delta}_{2,1,\mathcal{S}}\end{bmatrix}\right)\\
&= \left(\begin{bmatrix}\left(\left(\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}\right)'\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}_{1,1,\mathcal{S}}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\left(\left(\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}\right)'\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\delta}_{2,1,\mathcal{S}}\end{bmatrix}\right)\\
&= \left(\begin{bmatrix}\left(\left(\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}\right)'\right)^{-1}\boldsymbol{\delta}_{1,1,\mathcal{S}}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1}\boldsymbol{\delta}_{2,1,\mathcal{S}}\end{bmatrix}\right)\\
&= \begin{bmatrix}\left(\left(\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}\right)'\right)^{-1}\boldsymbol{\delta}_{1,1,\mathcal{S}}\\ -\left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1}\boldsymbol{\delta}_{2,1,\mathcal{S}}\end{bmatrix}
\end{aligned}
$$

Thus we can set $\boldsymbol{\mu}^{\mathbb{Q}_2}_{1,\mathcal{S}} = -\left(\left(\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}\right)'\right)^{-1}\boldsymbol{\delta}_{1,1,\mathcal{S}}$ and $\widetilde{\boldsymbol{\mu}}^{\mathbb{Q}_1}_{2,\mathcal{S}} = \widetilde{\boldsymbol{\mu}}^{\mathbb{Q}_2}_{2,\mathcal{S}} + \left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1}\boldsymbol{\delta}_{2,1,\mathcal{S}}$ in order to satisfy equation (37). Further, because the second separate model is invariant to any level-shift $\boldsymbol{\theta}$, it must be invariant to the particular level-shift

$$\boldsymbol{\theta} = -\left(\mathbf{I} - \mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)^{-1}\left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1}\boldsymbol{\delta}_{2,1,\mathcal{S}}$$

which ensures that $\widetilde{\boldsymbol{\mu}}^{\mathbb{Q}_1}_{2,\mathcal{S}} = \mathbf{0}$, as required under our normalization of JM(6). Finally, we can set s_0 in order to satisfy equation (32).

In summary, we can write two separate models as a joint model in which

$$
\begin{aligned}
\mathbf{x}_t &= \left[\mathbf{x}'_{1,t}, \mathbf{x}_{2,t} - \left(\mathbf{I} - \mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)^{-1} \left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1} \boldsymbol{\delta}_{2,1,\mathcal{S}}\right]', \\
\delta_{1,0} &= \delta_{1,0,\mathcal{S}}, \\
\boldsymbol{\delta}_{1,1} &= \begin{bmatrix} \boldsymbol{\delta}'_{1,1,\mathcal{S}} & \mathbf{0}' \end{bmatrix}', \\
\boldsymbol{\mu}^{\mathbb{Q}_1} &= \mathbf{0}, \\
\mathbf{\Phi}^{\mathbb{Q}_1} &= \begin{bmatrix} \mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}} \end{bmatrix}, \\
s_0 &= \delta_{1,0} - \delta_{2,0} - \mathbf{s}'_1 \boldsymbol{\mu}^{\mathbb{Q}_1} - \frac{1}{2}\mathbf{s}'_1 \mathbf{\Sigma}\mathbf{\Sigma}'\mathbf{s}_1, \\
\mathbf{s}_1 &= \begin{bmatrix} \left(\left(\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}\right)'\right)^{-1} \boldsymbol{\delta}_{1,1,\mathcal{S}} \\ -\left(\left(\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}\right)'\right)^{-1} \boldsymbol{\delta}_{2,1,\mathcal{S}} \end{bmatrix}, \\
\mathbf{\Sigma} &= \mathbf{I},
\end{aligned}
$$

where $\mathbf{\Phi}^{\mathbb{Q}_1}_{1,\mathcal{S}}$ and $\mathbf{\Phi}^{\mathbb{Q}_2}_{2,\mathcal{S}}$ are lower triangular matrices with ordered diagonal elements. Thus the two separate models written as a joint model take exactly the same form as JM(6). However, when the separate models are written as a single joint model, the $\mathbb{P}$ dynamics are a restricted case of JM(6), as explained in the main text.