BANK OF ENGLAND

Staff Working Paper No. 772

The information in the joint term structures of bond yields

Andrew Meldrum, Marek Raczko and Peter Spencer

December 2018

BANK OF ENGLAND

Staff Working Paper No. 772

The information in the joint term structures of bond yields

Andrew Meldrum,[(1)] Marek Raczko[(2)] and Peter Spencer[(3)]

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. However, this paper argues that, in two of the most common applications joint modelling does not bring any material benefits in capturing the dynamics of bond yields. Joint models of US 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 US nominal and real yields.

Key words: Affine term structure model, international interest rate co-movement, real interest rates.

JEL classification: F30, G12, G15.

(1) Division of Monetary Affairs, Board of Governors of the Federal Reserve System. Email: andrew.c.meldrum@frb.gov
(2) Barclays Bank PLC. Email: mraczko@gmail.com
(3) Department of Economics and Related Studies, University of York. Email: peter.spencer@york.ac.uk

We are grateful for comments from Laura Coroneo, Antonio Diez de los Rios, Caitlin Hesser, Iryna Kaminska, Glenn Rudebusch, and Julieta Yung, as well as attendees at the International Association for Applied Econometrics 2018 Annual Conference in Montreal, the University of York 2018 Asset Pricing Workshop, and a seminar at the Bank of England. Marek Raczko's contributions to this paper were made when he was employed 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, Barclays (Mr Raczko's current employer), the Bank of England, or the committees of the Bank of England.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 3461 4030 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

Gaussian no-arbitrage affine term structure models (ATSMs) are a popular framework for modeling the dynamics of bond yields. The majority of previous studies of ATSMs model nominal government bond yields from a single country separately; in these models, yields are driven by a small number of pricing factors extracted from the same set of 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), Bauer and Diez de los Rios (2012), Kaminska et al. (2013), Pegoraro et al. (2014), 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. (2018), and Abrahams et al. (2016)). The defining feature of these joint models is that the pricing factors are extracted from the yields on more than one class of bond. However, despite growing interest in joint models, it is not yet clear whether they have materially different implications for the dynamics of yields compared with standard, separate models. This paper address this gap in the literature for two of the most common applications of joint models.

In general, there are three potential reasons why a joint model might provide a characterization of the dynamics of yields that is superior to separate models. First, if there are local factors that are hidden from (or "unspanned" by) one class of yields but affect the time-series dynamics of those yields, separate models may be misspecified. Any study of joint models is therefore related to the emerging literature on the role of hidden factors in the term structure. The role of hidden factors in separate models of a single class of yields has been examined by a number of studies (for example, Cochrane and Piazessi (2005), Duffee (2011), Joslin et al. (2014), and Bauer and Rudebusch (2017)). However, the potential presence of unspanned factors extracted from yields on other classes of bonds has not been explicitly considered by the literature on joint models. Second, if there are common factors, which are spanned by both classes of yields, then a joint model may deliver more precise estimates of those factors and potentially better predictions of future yields. And third, a joint model may be used to motivate over-identifying restrictions that

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

may improve the identification of the time-series dynamics of yields.

Whether these theoretical benefits of joint modeling arise in practice is likely to depend on the particular application considered. The aim of this paper is not to make general statements about the benefits of all possible joint models. Rather, we aim to highlight the importance of considering whether a particular joint model is necessary using two of the most popular applications: a model of U.S. and German nominal yields and a model of U.S. nominal and real yields. In each case, we first test for the presence of unspanned factors by exploiting the fact that two separate models are equivalent to a restricted joint model. For example, two standard, separate three-factor models of U.S. and German nominal yields are equivalent to a restricted six-factor joint model. In this restricted joint model, yields in each country are spanned by three local factors that are unspanned by yields in the other country, and the time-series dynamics of the two sets of local factors are independent. We can test whether the local factors contain unspanned information that is relevant for predicting yields in the other country by comparing two separate models with a joint model that has the same factor structure but has no restrictions on the time-series dynamics of the factors. We find that the more flexible joint model provides little in the way of robust, economically significant benefits relative to two separate models. Thus, we conclude that in this application there is no information spanned by foreign yields but unspanned by domestic yields that is relevant for modeling the time-series dynamics of domestic yields.

However, a potential limitation of considering joint models that have only local factors is that they may be unnecessarily over-parameterized and therefore prone to in-sample over-fitting. Indeed, most previous studies of joint models allow for at least some factors to be common to different classes of yields. We therefore also examine the implications of imposing that one or more of the three factors spanned by yields in each country are common factors. We find that the models with common factors are strongly rejected by standard approaches to model selection (consistent with the previous results of Golinski and Spencer (2018)) and offer no material advantages in fitting the conditional expectations of yields either in or out of sample.

A general problem when estimating ATSMs is that the time-series dynamics of the factors are weakly identified in small samples, as discussed by Kim and Orphanides (2012) and Bauer et al. (2012). One potential way of improving the identification of the time-series dynamics is to impose over-identifying restrictions on the model, such that there is a tighter link between the risk-neutral and real-world dynamics of yields, as proposed for separate ATSMs by Cochrane and Piazessi (2008) and Bauer (2018), among others. We therefore consider whether the over-identifying restrictions on the time-series dynamics of local factors in joint models proposed by Egorov et al. (2011) and Diez de los Rios (2017) can improve the out-of-sample performance of the models. We find that while these restrictions do not result in a substantial deterioration in the in-sample fit of joint models, they also do not bring any material out-of-sample improvement.

Our application to U.S. nominal and real yields is more limited in scope, partly due to the relatively limited availability of real yields implied by Treasury Inflation Protected Securities (TIPS). We focus on comparing the in-sample fit of two separate models of nominal and real yields with three and two factors, respectively, with a five-factor joint model that has three factors spanned by nominal yields and two factors spanned by real yields. We find that the separate models are preferred according to standard model selection criteria and there are essentially no economically significant differences between the fit of the two models. Thus, we conclude there is no clear evidence that there is any information in real yields that is both unspanned by nominal yields and relevant for the time-series dynamics of nominal yields, and vice versa.

It is worth stressing the precise question that we address in this paper: whether joint models imply materially different properties of bond yields. Another potential motivation for joint models is to study the link between the discount factors that price the yields on two different classes of bond— that is, the exchange rate in a joint international model or inflation in a joint nominal-real model. Indeed, a small number of previous studies of joint international models use exchange rate data alongside bond yields to estimate the models (for example, Bauer and Diez de los Rios (2012), Kaminska et al. (2013), and Yung (2017)). We do not analyze the predictions of joint international models for the exchange rate or joint nominal-real models for inflation in this paper, for three reasons. First, we would like to be as consistent as possible with the majority of studies on joint international and nominal-real models, which are focused on the properties of bond yields. Second, the question of whether there is relevant unspanned information in the exchange rate or inflation is separate from the question of whether there is unspanned information in another class of yields. We do not need a joint model of multiple classes of to incorporate relevant unspanned information in the exchange rate or inflation; rather, we can simply augment the factor vector to include an unspanned exchange rate or inflation factor in a *separate* model of a single class of yields. And, third, Yung (2017) has previously shown that a relatively parsimonious joint model that does not allow for any interactions between the yields in two countries, as in separate models, can achieve a reasonable fit to exchange rates in a model that does not incorporate a separate exchange rate factor.

It is also worth stressing that this paper is not about identifying the structural drivers of the joint dynamics of yields on multiple classes of bonds (such as the trade and financial linkages between countries). For our purposes, the relevant question is not about the source of the shocks but whether there is relevant marginal information in another set of yields. To illustrate the point, suppose a structural economic shock hits the U.S. economy, and investors believe this shock will eventually spillover to the German economy. If German yields adjust immediately in anticipation of these spillovers, all of the relevant information will be reflected in *current* German yields. Thus, we do not necessarily need to incorporate information from U.S. yields into a correctly specified model of German yields, even though the shock emanated from the United States.

The remainder of this paper proceeds as follows. In Section 2, we set out the separate and joint ATSMs and explain why joint models may have different implications for the dynamics of bond yields. In Section 3, we describe our application to U.S. and German nominal yields, and, in Section 4, our application to U.S. nominal and real yields. In Section 5 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. The "standard" approach would be to model each class of yields using two separate Gaussian ATSMs, which we present in Section 2.1. Alternatively, we can model the two classes of yields jointly using a single ATSM, which we present in Section 2.2. In Section 2.3, we explain why joint models might, in principle, offer advantages relative to separate models when it comes to modeling the dynamics of bond yields.

2.1 Separate Models

The Gaussian ATSM of a single class of yields is entirely standard in the literature on DTSMs (see, for example, Duffee (2002)). It makes four 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, there are no arbitrage opportunities from investing in different maturity bonds, which implies that there exists a unique risk-neutral probability measure ($\mathbb{Q}_j$) such that the prices of the j^{th} class of bond satisfy

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

where $\mathbb{E}_t^{\mathbb{Q}_j}$ denotes expectations with respect to the $\mathbb{Q}_j$ measure. Third, the pricing factors follow a first-order VAR under $\mathbb{Q}_j$:

$$\mathbf{x}_{j,t+1} = \boldsymbol{\mu}_{j,\mathcal{S}}^{\mathbb{Q}_j} + \boldsymbol{\Phi}_{j,\mathcal{S}}^{\mathbb{Q}_j}\mathbf{x}_{j,t} + \boldsymbol{\Sigma}_{j,\mathcal{S}}\boldsymbol{\varepsilon}_{j,t+1}^{\mathbb{Q}_j}, \tag{3}$$

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. Under these assumptions, 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

$$\begin{aligned} a_{j,n,\mathcal{S}} &= a_{j,n-1,\mathcal{S}} + \mathbf{b}'_{j,n-1,\mathcal{S}}\boldsymbol{\mu}_{j,\mathcal{S}}^{\mathbb{Q}_j} + \frac{1}{2}\mathbf{b}'_{j,n-1,\mathcal{S}}\boldsymbol{\Sigma}_{j,\mathcal{S}}\boldsymbol{\Sigma}'_{j,\mathcal{S}}\mathbf{b}_{j,n-1,\mathcal{S}} - \delta_{j,0,\mathcal{S}}, && (4)\\ \mathbf{b}'_{j,n,\mathcal{S}} &= \mathbf{b}'_{j,n-1,\mathcal{S}}\boldsymbol{\Phi}_{j,\mathcal{S}}^{\mathbb{Q}_j} - \boldsymbol{\delta}'_{j,1,\mathcal{S}}, && (5)\end{aligned}$$

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).

Finally, the factors follow a first-order vector autoregression (VAR) under the physical probability measure ($\mathbb{P}$):

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

where $\boldsymbol{\varepsilon}_{j,t+1} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$ is an $n_j \times 1$ vector of Normally distributed shocks.

2.2 Joint Model

We next turn to a joint model of two classes of yields. In Section 2.2.1 we set out the assumptions of the joint model and derive expressions for long-term bond yields. In Section 2.2.2 we discuss how to restrict the parameters to ensure that some factors are local to one particular class of yields, and we further show that two separate models are equivalent to a joint model that has only local factors and has over-identifying restrictions on the time-series dynamics of yields. In Section 2.2.3 we explain how we identify the models and estimate them by maximum likelihood.

2.2.1 Bond Pricing and $\mathbb{P}$ Dynamics

The starting point for a joint model is the observation that under the assumption of no arbitrage, the prices of the first and second classes of bond must satisfy

$$\begin{aligned} P_{1,n,t} &= \mathbb{E}_t^{\mathbb{Q}_1}\left[\exp\left(-r_{1,t}\right) P_{1,n-1,t+1}\right] \text{ and} && (7) \\ P_{2,n,t}S_t &= \mathbb{E}_t^{\mathbb{Q}_1}\left[\exp\left(-r_{1,t}\right) P_{2,n-1,t+1}S_{t+1}\right], && (8) \end{aligned}$$

respectively, where S_t is the relevant "exchange rate" between the two numeraire assets; for example, if we were considering bonds with payoffs in a domestic currency for $j = 1$ and a foreign currency for $j = 2$, 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 class 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:

$$r_{1,t} = \delta_{1,0} + \boldsymbol{\delta}_{1,1}'\mathbf{x}_t. \tag{9}$$

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

$$\Delta s_t = s_0 + \mathbf{s}_1' \mathbf{x}_t. \tag{10}$$

The factors again follow a first-order VAR under $\mathbb{Q}_1$, that is,

$$\mathbf{x}_{t+1} = \boldsymbol{\mu}^{\mathbb{Q}_1} + \boldsymbol{\Phi}^{\mathbb{Q}_1} \mathbf{x}_t + \boldsymbol{\Sigma} \boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}_1}, \tag{11}$$

where $\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}_1} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$. Under these assumptions, the pricing of the first class of bonds is directly analogous to the separate model of Section 2.1, with yields given by

$$y_{1,n,t} = -\frac{1}{n}\left(a_{1,n} + \mathbf{b}_{1,n}' \mathbf{x}_t\right), \tag{12}$$

where

$$\begin{aligned} a_{1,n} &= a_{1,n-1} + \mathbf{b}_{1,n-1}' \boldsymbol{\mu}^{\mathbb{Q}_1} + \frac{1}{2} \mathbf{b}_{1,n-1}' \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \mathbf{b}_{1,n-1} - \delta_0, && (13) \\ \mathbf{b}_{1,n}' &= \mathbf{b}_{1,n-1}' \boldsymbol{\Phi}^{\mathbb{Q}_1} - \boldsymbol{\delta}_1', && (14) \end{aligned}$$

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

$$y_{2,n,t} = -\frac{1}{n}\left(a_{2,n} + \mathbf{b}_{2,n}' \mathbf{x}_t\right), \tag{15}$$

where

$$\begin{aligned} a_{2,n} &= a_{2,n-1} + s_0 + \left(\mathbf{s}_1 + \mathbf{b}_{2,n-1}\right)' \boldsymbol{\mu}^{\mathbb{Q}_1} + \frac{1}{2}\left(\mathbf{s}_1 + \mathbf{b}_{2,n-1}\right)' \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \left(\mathbf{s}_1 + \mathbf{b}_{2,n-1}\right) - \delta_0, && (16) \\ \mathbf{b}_{2,n}' &= \left(\mathbf{s}_1 + \mathbf{b}_{2,n-1}\right)' \boldsymbol{\Phi}^{\mathbb{Q}_1} - \boldsymbol{\delta}_1', && (17) \end{aligned}$$

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).

Equations (15)-(17) imply that the short rate for the second class of bonds takes the form

$$r_{2,t} = \delta_{2,0} + \boldsymbol{\delta}_{2,1}' \mathbf{x}_t \tag{18}$$

where

$$\begin{aligned} \delta_{2,0} &= \delta_{1,0} - s_0 - \mathbf{s}_1' \boldsymbol{\mu}^{\mathbb{Q}_1} - \frac{1}{2} \mathbf{s}_1' \boldsymbol{\Sigma} \boldsymbol{\Sigma}' \mathbf{s}_1 \text{ and} && (19) \\ \boldsymbol{\delta}_{2,1} &= \boldsymbol{\delta}_{1,1} - \left(\boldsymbol{\Phi}^{\mathbb{Q}_1}\right)' \mathbf{s}_1 && (20) \end{aligned}$$

(see Appendix C for details). We equivalently parameterize the maximally flexible joint model in terms of $\delta_{2,0}$ and $\boldsymbol{\delta}_{2,1}$, rather than s_0 and $\mathbf{s}_1$.

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

$$\mathbf{x}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi} \mathbf{x}_t + \boldsymbol{\Sigma} \boldsymbol{\varepsilon}_{t+1}, \tag{21}$$

where $\boldsymbol{\varepsilon}_{t+1} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$.

2.2.2 Common and Local Factors

In a maximally-flexible joint model, all of the factors may be common to both classes of yields; that is, both classes of yields may load on all of the pricing factors. However, several studies of joint models impose restrictions such that some of the factors are local factors that have zero loadings for all but one class of yields (for example, Egorov et al. (2011) and Kaminska et al. (2013)). These local factors are "hidden" from ("unspanned" by) all but one class of yields. In this section we explain the restrictions required to ensure some factors are local.

Specifically, suppose that we want to restrict a joint model with n_x factors such that there are n_c common factors, n_{l_1} factors local to the first class of yields, and n_{l_2} factors local to the second class of yields (with $n_x = n_c + n_{l_1} + n_{l_2}$). In such a model, $n_c + n_{l_1}$ factors are spanned by the first class of yields and $n_c + n_{l_2}$ factors are spanned by the second class of yields. Such a specification requires that the short rate loadings take the forms

$$\boldsymbol{\delta}_{1,1} = \left[\boldsymbol{\delta}'_{1,1,c}, \boldsymbol{\delta}'_{1,1,l_1}, \mathbf{0}'_{n_{l_2}\times 1}\right]' \text{ and} \tag{22}$$

$$\boldsymbol{\delta}_{2,1} = \left[\boldsymbol{\delta}'_{2,1,c}, \mathbf{0}'_{l_1\times 1}, \boldsymbol{\delta}'_{2,1,l_2}\right]', \tag{23}$$

where $\boldsymbol{\delta}_{1,1,c}$ and $\boldsymbol{\delta}_{2,1,c}$ are $n_c \times 1$, $\boldsymbol{\delta}_{1,1,l_1}$ is $n_{l_1} \times 1$, and $\boldsymbol{\delta}_{2,1,l_2}$ is $n_{l_2} \times 1$; and that $\boldsymbol{\Phi}^{\mathbb{Q}_1}$ takes the form

$$\boldsymbol{\Phi}^{\mathbb{Q}_1} = \begin{bmatrix} \boldsymbol{\Phi}^{\mathbb{Q}_1}_{cc} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Phi}^{\mathbb{Q}_1}_{l_1 c} & \boldsymbol{\Phi}^{\mathbb{Q}_1}_{l_1 l_1} & \mathbf{0} \\ \boldsymbol{\Phi}^{\mathbb{Q}_1}_{l_2 c} & \mathbf{0} & \boldsymbol{\Phi}^{\mathbb{Q}_1}_{l_2 l_2} \end{bmatrix}, \tag{24}$$

where $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{cc}$ is $n_c \times n_c$, $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{l_1 c}$ is $n_{l_1} \times n_c$, $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{l_1 l_1}$ is $n_{l_1} \times n_{l_1}$, $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{l_2 c}$ is $n_{l_2} \times n_c$, and $\boldsymbol{\Phi}^{\mathbb{Q}_1}_{l_2 l_2}$ is $n_{l_2} \times n_{l_2}$. Under the zero restrictions in equations (22)-(24) (which we refer to as the "$\mathbb{Q}_1$ restrictions"), we can partition the pricing factors conformably as $\mathbf{x}_t = \left[\mathbf{x}'_{c,t}, \mathbf{x}'_{l_1,t}, \mathbf{x}'_{l_2,t}\right]'$, where $\mathbf{x}_{c,t}$ are common factors, and $\mathbf{x}_{l_1,t}$ and $\mathbf{x}_{l_2,t}$ are factors local to the first and second class of bonds, respectively.

In general, a model with local factors need not impose any restrictions on the $\mathbb{P}$ dynamics of yields. Thus we can write $\boldsymbol{\Phi}$ as

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_{cc} & \boldsymbol{\Phi}_{cl_1} & \boldsymbol{\Phi}_{cl_2} \\ \boldsymbol{\Phi}_{l_1 c} & \boldsymbol{\Phi}_{l_1 l_1} & \boldsymbol{\Phi}_{l_1 l_2} \\ \boldsymbol{\Phi}_{l_2 c} & \boldsymbol{\Phi}_{l_2 l_1} & \boldsymbol{\Phi}_{l_2 l_2} \end{bmatrix}, \tag{25}$$

where all of the parameter blocks are unrestricted. In Section 2.3.2, we consider cases where certain blocks of $\boldsymbol{\Phi}$ are subject to over-identifying restrictions.

2.2.3 Identification and Estimation

As discussed by, for example, Dai and Singleton (2000), Joslin et al. (2011), and Hamilton and Wu (2012), before we take an ATSM to the data we need to impose a minimum set of identifying restrictions. A maximally flexible model would have only common factors (i.e. would not impose the $\mathbb{Q}_1$ or $\mathbb{P}$ restrictions mentioned in the previous section). To identify a model with only common factors, we can impose that $\boldsymbol{\mu}^{\mathbb{Q}_1} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}$, and $\boldsymbol{\Phi}_{cc}^{\mathbb{Q}_1}$ is a lower triangular matrix with ordered diagonal elements ($\phi_{cc,11}^{\mathbb{Q}_1} \geq \phi_{cc,22}^{\mathbb{Q}_1} \geq \ldots \geq \phi_{cc,n_c n_c}^{\mathbb{Q}_1}$). In models that have local factors we also require the identifying assumptions that $\boldsymbol{\Phi}_{l_1 l_1}^{\mathbb{Q}_1}$ and $\boldsymbol{\Phi}_{l_2 l_2}^{\mathbb{Q}_1}$ are lower triangular with ordered diagonal elements.[2]

When estimating the joint models, we assume that all yields are measured with error. Specifically, we allow for measurement error by assuming that observed yields are given by

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

Here, $\mathbf{y}_t$ is an $n_{y_1} \times 1$ vector of observed yields on the first class of bonds; $\mathbf{y}_t^*$ is an $n_{y_2} \times 1$ vector of observed yields on the second class of bonds; $\mathbf{w}_t \sim \mathcal{NID}\left(\mathbf{0}, \sigma_w^2 \times \mathbf{I}\right)$ and $\mathbf{w}_t^* \sim \mathcal{NID}\left(\mathbf{0}, \sigma_{w^*}^2 \times \mathbf{I}\right)$ are $n_{y_1} \times 1$ and $n_{y_2} \times 1$ vectors of measurement errors; and the definitions of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}^*$, and $\mathbf{B}^*$ follow from equations (13), (14), (16), and (17). Equations (21) and (26) form a linear-Gaussian state-space system, and we can therefore estimate the free parameters of the model by maximum likelihood, using the Kalman filter to estimate the latent pricing factors for a given set of parameters.[3]

2.3 Why Might Joint Models Better Capture the Dynamics of Bond Yields?

We now turn to the question of *why* we might want to model yields jointly with those in another country. In Section 2.3.1, we explain why the rationale has nothing to do with the ability of joint models to improve the cross-sectional fit to current yields. However, in Section 2.3.2, we explain how joint models may in principle offer improvements relative to nested separate models when it comes to predicting future yields.

[2] We impose that $\boldsymbol{\Phi}_{cc}^{\mathbb{Q}_1}$, $\boldsymbol{\Phi}_{l_1 l_1}^{\mathbb{Q}_1}$, and $\boldsymbol{\Phi}_{l_2 l_2}^{\mathbb{Q}_1}$ have only real eigenvalues. Strictly speaking, a maximally flexible model allows for complex eigenvalues (see Joslin et al. (2011)).

[3] D'Amico et al. (2018) (among others) provide further details on the estimation of joint models by maximum likelihood using the Kalman filter. Other approaches to estimating joint models exist that assume that some linear combinations of yields are measured without error (for example, Abrahams et al. (2016) and Diez de los Rios (2017)).

2.3.1 Cross-Sectional Accuracy

Most previous studies of separate ATSMs assume that a small number of factors extracted from a single class of yields are sufficient to explain all of the predictable cross-sectional variation in that same set of yields. Of course, if the number of yields used to estimate the model is greater than the number of pricing factors, then the cross-sectional fit to observed yields will not be perfect (that is, observed yields will be measured with errors relative to the values predicted by the model). However, if we want to improve the cross-sectional fit further, we can always simply increase the number of factors in a separate model.

To be sure, if there is some common variation in the two classes of yields, then two separate models with n_1 and n_2 factors may not fit the cross sections of current yields quite as well as a joint model with $n_x = n_c = n_1 + n_2$ factors. Indeed, preliminary (unreported) results showed that a maximally flexible joint model with six common factors fits the cross-section of U.S. and German yields marginally better than two separate three-factor models. But of course two separate six-factor models of U.S. and German yields achieve an even better cross-sectional fit than a maximally flexible six-factor joint model.[4]

It is worth emphasizing an important point here: The aim of this paper is not to choose an optimal factor structure for a joint model of multiple classes of yields. Rather, it is to consider whether joint models offer any improvements in our ability to capture the dynamics of yields relative to standard, separate models. When we compare a joint model with two separate models, it is important to ensure a fair comparison by being consistent in our assumptions about how many factors are spanned by yields in each country. For example, if we consider a maximally flexible six-factor joint model of U.S. and German yields, then we are taking a view that there may be as many as six factors spanned by the yields in each country.[5] If we compared this joint model with two separate three-factor models, we would be unfairly handicapping the separate models by omitting potentially relevant information. And if the maximally flexible six-factor joint model out-

[4] This finding is essentially the result reported in Table 2 in Egorov et al. (2011). They show that a given number of principal components of (dollar) Libor and Euribor rates explains less of the pooled data sets than the same number of principal components explains in each of the separate data sets.

[5] See Pegoraro et al. (2014) for a discussion of the optimal factor structure in joint international term structure models that allow each yield curve to be spanned by more than the standard number of factors.

performed the separate three-factor models, then we could not be sure whether that was because it incorporates the information in overseas yields, or because it uses more of the information in *domestic* yields (such as the fourth or fifth principal component of domestic yields, as proposed by Duffee (2011)); we might be able to achieve a similar improvement simply by increasing the number of factors in a separate model. Thus, when we are attempting to answer the question of whether there are benefits from joint models, the relevant comparison is not the total number of factors in the models, but the number of factors spanned by each class of yields. We return to this point in Section 3.3.

2.3.2 Time-Series Accuracy

Although joint models cannot offer advantages relative to separate models in terms of their fit to the cross section of yields, joint models may in principle out-perform separate models when it comes to predicting future yields, for three reasons: First, joint models allow for the possibility that there are hidden (unspanned) factors in other classes of yields. Second, we may be able to obtain more efficient estimates of any pricing factors common to both sets of yields. And, third, the factor structure of some joint models provides an economic rationale for certain over-identifying restrictions that may help estimate the time series of yields more precisely. In the remainder of this section we discuss these three motivations for joint modeling in turn.

Hidden Factors In joint models that have some local factors, it is possible that factors unspanned by (say) the first class of yields may nevertheless enter the time-series dynamics of the factors that are spanned by the first class of yields and therefore affect expectations of future yields. Because these unspanned factors are hidden from the first class of yields, a separate model would necessarily be misspecified because it omits important information.[6] As mentioned in the introduction, studies of joint models are therefore related to the emerging literature on unspanned factors in the term structure, which has found that factors unspanned by the first three principal components of U.S. nominal yields nevertheless appear to explain the time-series dynamics of

[6] Following the terminology of Duffee (2011), these unspanned factors may be "completely hidden," in that the $\mathbb{Q}_1$ restrictions set out above are exactly satisfied, or they may be "partly hidden," in that measurement error obscures the effect of some factors that have sufficiently small loadings.

those yields.

We test for unspanned factors in other classes of yields by comparing two separate models with a joint model that has only local factors but allows for time-series interactions between two sets of local factors. Two separate models with n_1 and n_2 factors are equivalent to a joint model that imposes the $\mathbb{Q}_1$ restrictions such that there are no common factors ($n_c = 0$), and n_1 and n_2 local factors ($n_{l_1} = n_1$ and $n_{l_2} = n_2$); and that imposes restrictions on the $\mathbb{P}$ dynamics such that the two sets of local factors are independent, that is, with $\mathbf{\Phi}_{l_1 l_2} = \mathbf{0}$ and $\mathbf{\Phi}_{l_2 l_1} = \mathbf{0}$ in in equation (25). (Appendix D provides further details.) If we can accept these restrictions, then we would have evidence that there is no relevant marginal information in each class of yields for modeling the time-series dynamics of the other class. Alternatively, if we cannot accept these restrictions, then we would have evidence that the time-series dynamics of yields are misspecified in separate models because they do not allow for interactions between the factors spanned by different classes of yields.

Common Factors While two separate models are equivalent to a restricted joint model that has only local factors, if some of the factors are common to both classes of yields, then such a joint model would be unnecessarily heavily parameterized. If we hold the number of factors spanned by each class of yields fixed, then introducing common factors means that we must reduce the *total* number of factors and the model may be less prone to in-sample over-fitting. In particular, assuming that some factors are common may allow us to increase the precision of the estimates of those factors, and therefore the parameters of the model, by imposing that some of the factors are common to both classes of yields. However, it is also possible that by imposing that some factors are common we discard important information about the time-series dynamics of yields that would result in worse out-of-sample performance. We explore the advantages and disadvantages of allowing for common factors in our application to U.S. and German nominal yields.

Over-Identifying Restrictions As discussed in the introduction, a well-known problem with estimating the time-series dynamics of bond yields is that the available samples of yields are typically short in relation to the high persistence of yields. Unfortunately, the no-arbitrage restrictions do not materially help improve the identification of the time-series dynamics of yields, because they do not imply any restrictions on $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$. One way of mitigating this identification problem that has been considered by some previous studies is to impose constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$ directly or on the parameters of the prices of risk that relate $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$ to $\boldsymbol{\mu}^{\mathbb{Q}_1}$ and $\boldsymbol{\Phi}^{\mathbb{Q}_1}$, respectively.

In a model of a single yield curve, it is not clear how we can motivate such restrictions on theoretical grounds (although Joslin et al. (2014) and Bauer (2018) propose statistical approaches for choosing which restrictions we should put the most weight on). In contrast, previous studies of joint models have suggested two alternative approaches for setting *a priori* over-identifying restrictions in joint models using the distinction between common and local factors as motivation. First, Egorov et al. (2011) propose that local factors spanned only by one class of yields should also not affect the factors spanned by the other class of yields under $\mathbb{P}$. To obtain a model with such a feature, we must impose the additional over-identifying restrictions (which we refer to as the "$\mathbb{P}$ restrictions") that $\boldsymbol{\Phi}$ must take the form

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_{cc} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Phi}_{l_1 c} & \boldsymbol{\Phi}_{l_1 l_1} & \mathbf{0} \\ \boldsymbol{\Phi}_{l_2 c} & \mathbf{0} & \boldsymbol{\Phi}_{l_2 l_2} \end{bmatrix}, \tag{27}$$

where the various blocks of parameters are partioned conformably with the partioning of $\mathbf{x}_t = \left[\mathbf{x}'_{c,t}, \mathbf{x}'_{l_1,t}, \mathbf{x}'_{l_2,t}\right]'$. As discussed above, in the case where there are no common factors, a joint model that imposes these $\mathbb{P}$ restrictions is equivalent to two separate models with n_1 and n_2 factors. In our application to U.S. and German yields below, we also consider the impact of imposing the $\mathbb{P}$ restrictions in joint models that have a mixture of common and local factors.

Second, Bauer and Diez de los Rios (2012) alternatively consider restrictions on the prices of risk. Under our normalization restrictions, the $n_x \times 1$ prices of risk can be written as

$$\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}_1 \mathbf{x}_t,$$

where $\boldsymbol{\lambda}_0 = \boldsymbol{\mu} - \boldsymbol{\mu}^{\mathbb{Q}_1}$ is an $n_x \times 1$ vector and $\Lambda_1 = \boldsymbol{\Phi} - \boldsymbol{\Phi}^{\mathbb{Q}_1}$ is an $n_x \times n_x$ matrix. Bauer and Diez de los Rios (2012) assume that: (1) any local factors are not only unspanned by yields in other countries, but they are also idiosyncratic, such that they do not carry a price of risk; and (2) that the local factors do not affect the price of risk of common factors. The price of risk parameters therefore take the form

$$\boldsymbol{\lambda}_0 = \begin{bmatrix} \boldsymbol{\lambda}_{0,c} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \text{ and } \boldsymbol{\Lambda}_1 = \begin{bmatrix} \boldsymbol{\Lambda}_{1,c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{28}$$

where $\boldsymbol{\lambda}_{0,c}$ is $n_c \times 1$ and $\boldsymbol{\Lambda}_{1,c}$ is $n_c \times n_c$. These restrictions imply that $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$ take the form

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_c \\ \boldsymbol{\mu}_{l_1}^{\mathbb{Q}_1} \\ \boldsymbol{\mu}_{l_2}^{\mathbb{Q}_1} \end{bmatrix} \text{ and } \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_{cc} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Phi}_{l_1c}^{\mathbb{Q}_1} & \boldsymbol{\Phi}_{l_1l_1}^{\mathbb{Q}_1} & \mathbf{0} \\ \boldsymbol{\Phi}_{l_2c}^{\mathbb{Q}_1} & \mathbf{0} & \boldsymbol{\Phi}_{l_2l_2}^{\mathbb{Q}_1} \end{bmatrix}. \tag{29}$$

Thus, when we impose these "price of risk restrictions," the identification of the $\mathbb{P}$ dynamics of the local factors is improved by making use of information in the cross-sectional dimension of yields. In our application to U.S. and German yields below, we consider the impact of imposing the price of risk restrictions in joint models with local factors.

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

We now turn to our application to U.S. and German nominal yields. In Section 3.1, we describe our data set. In Section 3.2, we examine whether there is evidence of relevant unspanned information in the yields of the other country. In Sections 3.3 and 3.4, we explore whether the performance of the joint models can be improved by imposing that some factors are common or by imposing over-identifying restrictions on the $\mathbb{P}$ dynamics, respectively.

3.1 Data

Our data set consists of month-end 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 with maturities of six months and one, two, three, five, seven, and ten years.

Figure 1 shows the time series of the one- and ten-year yields. The cross-country correlation between movements in long-term yields in particular appears to be quite high, which provides some preliminary support for the idea that there may be common factors driving U.S. and German yields.

3.2 Hidden Factors

We start by comparing two separate three-factor models (which, taken together, we refer to as "model SM") with a joint model that has three local factors spanned by yields in each country but unrestricted $\mathbb{P}$ dynamics ("model JM6"). In both cases, the three factors spanned by each yield curve should contain essentially all of the information in domestic yields; the only difference between SM and JM6 is that the latter allows for interactions between the two sets of local factors under $\mathbb{P}$.

The final two rows of Table 1 provide the estimated log likelihood, number of parameters, and the Schwarz Information Criterion (SIC) for models SM and JM6 (we refer back to the other rows of the table in later sections). The SIC clearly favors the two separate models, which have 18 fewer parameters.

We next consider whether the differences in fit are economically significant, starting with the

[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.

Figure 1: U.S. and German Nominal Yields

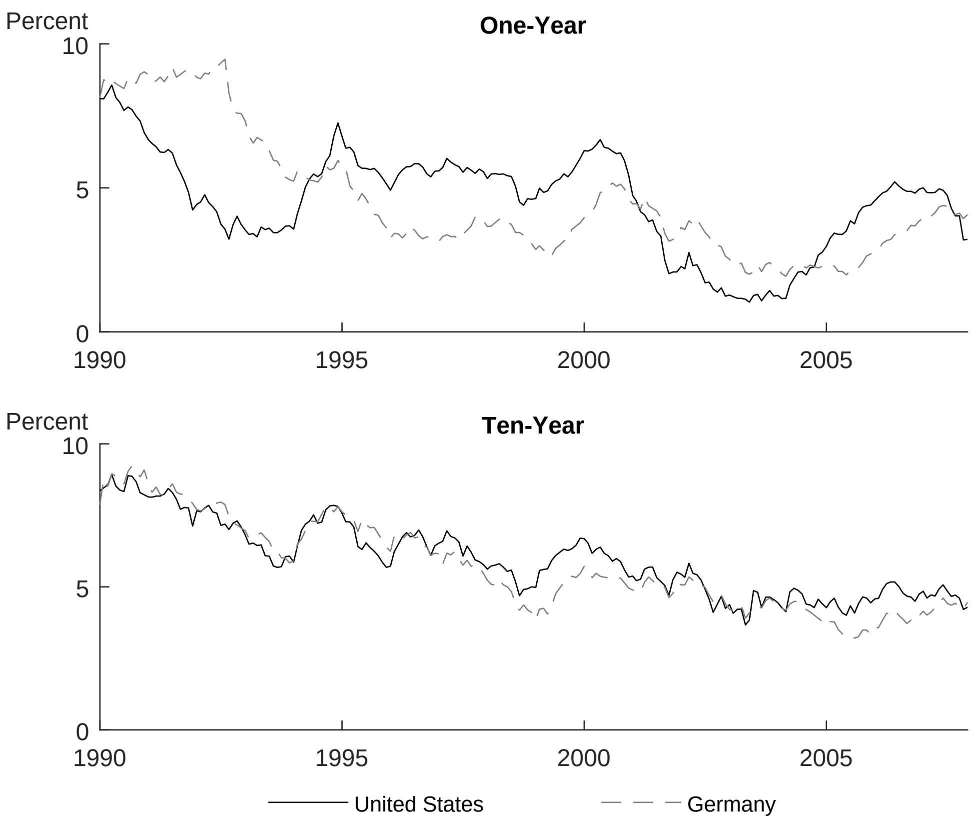


Table 1: Log Likelihoods for Models of U.S. and German Yields

This table reports the log likelihoods, number of free parameters, and Schwarz Information Criteria (SIC) for joint models of U.S. and German yields.

Model	Log Likelihood	Number of Free Parameters	SIC
JM3	634	28	-1043
JM4	1875	39	-3438
JM4-$\mathbb{P}$	1869	33	-3474
JM4-Λ	1856	25	-3511
JM5	2923	51	-5437
JM5-$\mathbb{P}$	2914	39	-5516
JM5-Λ	2599	23	-5014
JM6	4086	64	-7658
SM	4063	46	-7758

fit of the models to the cross section of yields. Panel $\mathcal{A}$ of Table 2 reports root-mean-squared errors (RMSEs) between observed and model-implied yields at selected maturities. As we would expect given that both SM and JM6 have three local factors to explain the cross section of yields in each country, there are essentially no differences in their cross-sectional fit, with RMSEs between 2 and 3 basis points at all considered maturities. Panel $\mathcal{B}$ of Table 2 reports the RMSEs between observed and one-step-ahead predictions of yields (when we estimate the models using the Kalman filter, we are effectively minimizing a weighted sum of one-step-ahead fitting errors). Again, two separate models obtain practically identical RMSEs to model JM6, which again suggests that the more flexible specification of the $\mathbb{P}$ dynamics in model JM6 does not bring any benefits in terms of the in-sample fit.

Table 2: Models of U.S. and German Yields with Three Local Factors

This table reports results for two separate models (SM) and a six-factor joint model with three local factors spanned by yields in each country (JM6). Panel $\mathcal{A}$ reports the cross-sectional accuracy, that is, the root mean squared error (RMSE) between current-period model-implied and actual yields. Panels $\mathcal{B}$ and $\mathcal{C}$ report the time-series accuracy, that is, the RMSEs between model-implied expected yields one and twelve months ahead and subsequent realized yields when the model parameters are estimated using the full sample. Panels $\mathcal{D}$ and $\mathcal{F}$ report the RMSEs between model-implied expected yields one and twelve months ahead and subsequent realized yields when the model parameters are estimated recursively.

This table reports root mean squared errors (in annualized percentage points) between the cross sections of model-implied and actual yields at selected maturities. Model JM6-R, highlighted in bold, is equivalent to two separate three-factor models.

	United States (maturity in years)			Germany (maturity in years)		
	1	5	10	1	5	10
$\mathcal{A}$: Cross section						
JM6	0.03	0.02	0.03	0.03	0.02	0.02
SM	0.03	0.02	0.03	0.03	0.02	0.02
$\mathcal{B}$: 1-step ahead						
JM6	0.24	0.28	0.25	0.19	0.22	0.20
SM	0.24	0.28	0.26	0.20	0.22	0.20
$\mathcal{C}$: 12-step ahead						
JM6	1.20	0.90	0.71	0.86	0.78	0.68
SM	1.30	0.99	0.78	1.04	0.91	0.73
$\mathcal{D}$: 1-step ahead (recursive forecasting)						
JM6	0.25	0.30	0.28	0.17	0.21	0.17
SM	0.25	0.30	0.28	0.17	0.20	0.17
$\mathcal{E}$: 12-step ahead (recursive forecasting)						
JM6	1.63	1.14	0.83	1.00	0.81	0.60
SM	1.78	1.21	0.83	0.91	0.81	0.68

Although JM6 does not bring any benefits in terms of in-sample fit, it is possible that the more flexible specification of the $\mathbb{P}$ dynamics would allow the model to better match features of the data not included in the estimation. However, it is possible that the greater number of parameters in JM6 will result in in-sample over-fitting and weaker out-of-sample performance than two separate models. We therefore consider two exercises comparing the model's ability to match out-of-sample features of the data.

First, we examine how well the models predict yields at 12-month horizon, while still using parameters estimated from the full sample. Panel $\mathcal{C}$ of Table 2 reports the RMSEs between observed yields and model-implied 12-step-ahead predictions. At this horizon, model JM6 achieves a moderate reduction in RMSEs relative to two separate models, with the largest gains at shorter maturities. For example, JM6 achieves an RMSE about 10 basis points lower than two separate models for the U.S. 1-year yield and almost 20 basis points lower for the German 1-year yield. This result provides some evidence that modeling yields in two countries separately may miss out on some important interactions between the factors. However, the differences between the models are insignificant at the 5 percent level, according to unreported Diebold and Mariano (1995) tests.

Second, we consider a recursive out-of-sample forecasting exercise. We start by estimating the models using the first ten years' of data (that is, from January 1990 to December 1999) and compute model-implied forecasts of yields at horizons of up to 12 months. 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. Panels $\mathcal{D}$ and $\mathcal{E}$ of Table 2 report root-mean-squared prediction errors (RMSPEs) at 1- and 12-month forecast horizons, respectively. At the 1-month horizon, there is essentially no difference between the forecasting performance of two separate models and JM6. There are some larger differences at the 12-month horizon, with the forecasts from JM6 being slightly better for U.S. yields and slightly worse for German yields. However, none of the differences are significant at the 5 percent level, according to Diebold and Mariano (1995) tests.

In summary, a comparison of two separate three-factor models with a joint model that has three local factors spanned by each yield curve and unrestricted $\mathbb{P}$ dynamics suggests there is no strong evidence that there is any relevant unspanned information in overseas yields.

3.3 Common Factors

As we discussed in Section 2.3.2, a joint model with three local factors spanned by yields in each country may be unnecessarily heavily parameterized, because it ignores the possibility that some of the factors may be common to both countries. In this section, we therefore allow for one or more factors spanned by the yields in each country to be common, while maintaining the standard assumption that three factors are spanned by yields in each country, as in the joint models of Egorov et al. (2011) and Kaminska et al. (2013), for the reasons explained in Section 2.3.1.[8]

This assumption narrows down the set of permissible factor structures for joint models to four, which we list in Table 3. The columns of the table refer to the total number of factors in the model (n_x), the number of common factors (n_c), and the number of local factors spanned by yields in each country (n_{l_1} and n_{l_2}). JM6, considered in the previous section, is the most flexible joint model, with three factors spanned by each yield curve. At the opposite extreme, in model JM3 the three spanned factors are common to the yields in both country, such that the model has a total of three factors. The two intermediate cases (model JM4 and model JM5) have two and one common factors, respectively.

Table 3: Permissible Factor Structures in Joint Models of U.S. and German Yields
This table lists the permissible factor structures in joint models of U.S. and German yields under the assumption that three factors are spanned by the yields in each country. It shows the total number of factors (n_x), the number of common factors (n_c), and the number of local factors for each yield curve (n_{l_1} and n_{l_2}).

Model	n_x	n_c	n_{l_1}	n_{l_2}
JM3	3	3	0	0
JM4	4	2	1	1
JM5	5	1	2	2
JM6	6	0	3	3

[8] Some studies of joint models allow for more than three factors to be spanned by yields in each country. For example, Anderson et al. (2010) estimate two-country models with five common factors. As discussed in Section 2.3.1, if three factors are sufficient to span yields in each country, then we know *a priori* that a specification with more than three spanned factors for each yield curve must be over-parameterized because it must be possible to rotate the joint model such that only three factors have a non-zero loading on each class of yields. In addition, if we allow more than three spanned factors for each yield curve, then we cannot be sure whether any improvements offered by joint models are because we are incorporating unspanned information in overseas yields or because we are incorporating more information from domestic yields.

Of course, the models with common factors are by construction less flexible than the six-factor models and will not fit the data as well in the sample. However, the fact that they are less heavily parameterized means that they may perform better out of sample. But if the models with common factors omit important information from yields in one or both countries, it is possible that they will perform worse than model JM6 out of sample.

Table 1 shows that models JM3, JM4, and JM5 are not preferred to two separate models or model JM6, according to the SIC, confirming the previous findings of Pegoraro et al. (2014). Panel $\mathcal{A}$ of Table 4 reports RMSEs for the cross-sectional fit. Comparing these results with those in Table 2 shows that imposing that one factor is common to yields in both countries does not result in a very substantial increase in cross-sectional RMSEs, but that joint models with two or more common factors struggle to fit yields in one or both countries. A similar result arises for 1- and 12-step-ahead prediction errors, which are reported in Panels $\mathcal{B}$ and $\mathcal{C}$. Finally, Panels $\mathcal{D}$ and $\mathcal{E}$ of the tables reveal that joint models with fewer than six factors offer no consistent advantages over six-factor models when it comes to out-of-sample forecasting. Thus, we conclude that joint models with common factors offer little benefit relative to joint models that only have local factors and can result in material reductions in the in-sample fit if there is more than one common factor.

3.4 Over-Identifying Restrictions

We now consider whether imposing the $\mathbb{P}$ restrictions and Λ restrictions discussed in Section 2.3.2 materially worsens the in-sample fit of the models and whether it materially improves or worsens the out-of-sample fit. Here we consider only restricted versions of models JM4 and JM5. Model JM3 has only common factors, so neither the $\mathbb{P}$ restrictions nor the Λ restrictions are relevant. Model JM6 with the $\mathbb{P}$ restrictions imposed is equivalent to two separate models, which we consider above. Finally, model JM6 with the Λ restrictions imposed implies zero risk premiums, which we rule out on the basis of overwhelming previous evidence against the expectations hypothesis of the term structure in previous studies.

Table 1 also reports the log likelihood, number of free parameters, and the SIC for the restricted four- and five-factor joint models. JM4-$\mathbb{P}$ denotes JM4 with the $\mathbb{P}$ restrictions imposed and JM4-

Table 4: Models of U.S. and German Yields with Common Factors

This table reports results for joint models with between one and three common factors. Panel $\mathcal{A}$ reports the cross-sectional accuracy, that is, the root mean squared error (RMSE) between current-period model-implied and actual yields. Panels $\mathcal{B}$ and $\mathcal{C}$ report the time-series accuracy, that is, the RMSEs between model-implied expected yields one and twelve months ahead and subsequent realized yields when the model parameters are estimated using the full sample. Panels $\mathcal{D}$ and $\mathcal{F}$ report the RMSEs between model-implied expected yields one and twelve months ahead and subsequent realized yields when the model parameters are estimated recursively. This table reports root mean squared errors (in annualized percentage points) between the cross sections of model-implied and actual yields at selected maturities. Model JM6-R, highlighted in bold, is equivalent to two separate three-factor models.

	United States (maturity in years)			Germany (maturity in years)		
	1	5	10	1	5	10
$\mathcal{A}$: Cross section						
JM3	0.05	0.06	0.08	0.16	0.23	0.42
JM4	0.12	0.24	0.33	0.03	0.02	0.02
JM5	0.04	0.04	0.09	0.03	0.02	0.02
$\mathcal{B}$: 1-step ahead						
JM3	0.24	0.29	0.27	0.28	0.28	0.44
JM4	0.26	0.37	0.41	0.19	0.22	0.20
JM5	0.24	0.29	0.27	0.19	0.22	0.20
$\mathcal{C}$: 12-step ahead						
JM3	1.28	1.01	0.79	0.97	0.82	0.73
JM4	1.39	0.98	0.75	0.89	0.79	0.66
JM5	1.22	0.91	0.71	0.87	0.79	0.67
$\mathcal{D}$: 1-step ahead (recursive forecasting)						
JM3	0.27	0.34	0.31	0.18	0.22	0.23
JM4	0.25	0.30	0.31	0.18	0.22	0.25
JM5	0.25	0.31	0.30	0.17	0.21	0.17
$\mathcal{E}$: 12-step ahead (recursive forecasting)						
JM3	1.67	1.13	0.86	1.18	0.84	0.73
JM4	1.70	1.25	0.95	1.04	0.86	0.83
JM5	1.71	1.28	0.96	1.04	0.91	0.71

Λ denotes JM4 with the Λ restrictions imposed, with corresponding notation for the five-factor models. According to the SICs, both JM4-$\mathbb{P}$ and JM4-Λ are preferred to JM4. JM5-$\mathbb{P}$ is similarly preferred to JM5, but JM5-Λ is not; this result suggests that with only one common factor, the Λ restrictions imply an overly restrictive specification for the prices of risk. None of the restricted joint models are preferred to two separate models.

Table 5 reports the fit of the models that impose the $\mathbb{P}$ and Λ restrictions to the cross-section of yields and 1- and 12-month ahead predictions of yields, both in- and out-of-sample. Comparing these results with Tables 2 and 4 shows that while the restricted joint models do not perform materially worse than their unrestricted counterparts when it comes to fitting yields in-sample, they also offer no material out-of-sample benefits relative to more flexible joint models.

4 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. We use the same sample of U.S. nominal yields described in Section 3.1, that is, month-end 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).[9] One practical difficulty with this application is the limited availability of TIPS-implied real yields; we use a sample of month-end 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 spanned by the real yield curve. Figure 2 plots the nominal and real yields at five- and ten-year maturities.

In the interests of conciseness, in this application we limit our attention to a single joint model: a five-factor joint model that has three local factors spanned by the nominal yield curve and two local factors spanned by the real yield curve (model JM5). We compare this model with two separate models of nominal and real yields, with three and two factors, respectively. In view of the short sample of available TIPS yields, we do not consider a recursive out-of-sample forecasting

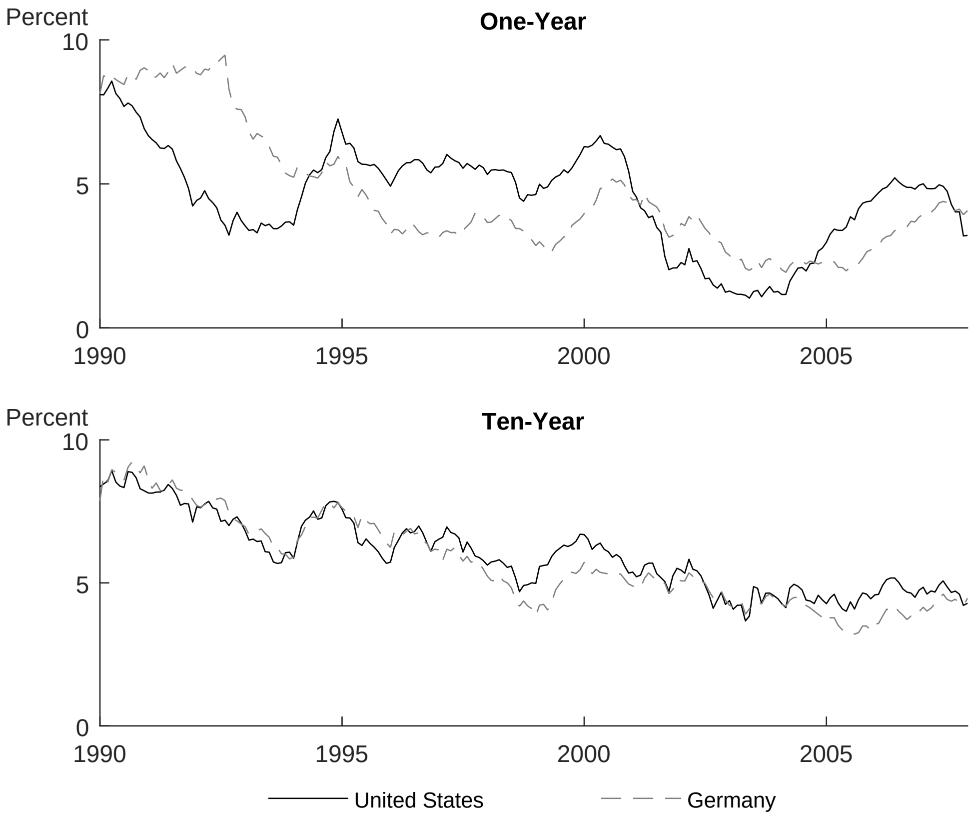
[9] Updates of these yields are published by the Board of Governors of the Federal Reserve System.

Table 5: Models of U.S. and German Yields: Results for Models with $\mathbb{P}$ and Λ Restrictions

This table reports results for four- and five-factor models with the $\mathbb{P}$ and Λ restrictions. Panel $\mathcal{A}$ reports the cross-sectional accuracy, that is, the root mean squared error (RMSE) between current-period model-implied and actual yields. Panels $\mathcal{B}$ and $\mathcal{C}$ report the time-series accuracy, that is, the RMSEs between model-implied expected yields one and twelve months ahead and subsequent realized yields when the model parameters are estimated using the full sample. Panels $\mathcal{D}$ and $\mathcal{F}$ report the RMSEs between model-implied expected yields one and twelve months ahead and subsequent realized yields when the model parameters are estimated recursively.

	United States (maturity in years)			Germany (maturity in years)		
	1	5	10	1	5	10
$\mathcal{A}$: Cross section						
JM4-$\mathbb{P}$	0.12	0.24	0.33	0.03	0.02	0.02
JM5-$\mathbb{P}$	0.04	0.04	0.09	0.03	0.02	0.02
JM4-Λ	0.13	0.23	0.33	0.03	0.02	0.02
JM5-Λ	0.03	0.02	0.03	0.04	0.06	0.12
$\mathcal{B}$: 1-step ahead						
JM4-$\mathbb{P}$	0.27	0.37	0.39	0.20	0.23	0.21
JM5-$\mathbb{P}$	0.25	0.29	0.27	0.20	0.23	0.21
JM4-Λ	0.30	0.37	0.39	0.20	0.23	0.20
JM5-Λ	0.26	0.29	0.26	0.22	0.23	0.24
$\mathcal{C}$: 12-step ahead						
JM4-$\mathbb{P}$	1.42	1.02	0.79	1.04	0.94	0.77
JM5-$\mathbb{P}$	1.37	0.97	0.78	1.03	0.95	0.80
JM4-Λ	1.62	1.12	0.83	1.07	0.95	0.77
JM5-Λ	1.55	1.10	0.87	1.21	1.03	0.90
$\mathcal{D}$: 1-step ahead (recursive forecasting)						
JM4-$\mathbb{P}$	0.25	0.30	0.31	0.19	0.21	0.23
JM5-$\mathbb{P}$	0.26	0.30	0.29	0.17	0.20	0.17
JM4-Λ	0.29	0.36	0.32	0.17	0.20	0.17
JM5-Λ	0.26	0.30	0.28	0.19	0.20	0.24
$\mathcal{E}$: 12-step ahead (recursive forecasting)						
JM4-$\mathbb{P}$	1.42	1.02	0.79	1.04	0.94	0.77
JM5-$\mathbb{P}$	1.37	0.97	0.78	1.03	0.95	0.80
JM4-Λ	1.77	1.18	0.84	0.96	0.84	0.71
JM5-Λ	1.68	1.16	0.88	0.99	0.86	0.87

Figure 2: U.S. Nominal and Real Yields



exercise.

Table 6 shows that the two separate models are preferred to the joint model, according to the SIC. Panels $\mathcal{A}$ and $\mathcal{B}$ of Table 7 show that the joint model and the two separate models obtain essentially the same cross-sectional fit to yields and one-step-ahead prediction errors. Panel $\mathcal{C}$ shows that the joint model achieves a small reduction in 12-month-ahead predictions for longer-maturity nominal yields, although the differences are not significant according to unreported Diebold-Mariano tests. Thus, we conclude that there is no relevant unspanned information in U.S. nominal and real yields that would justify joint modeling.

Table 6: Log Likelihoods for Models of U.S. Nominal and Real Yields

This table reports the log likelihoods, number of free parameters, and Schwarz Information Criterion (SIC) for joint models of U.S. nominal and real yields. Model JM5-R, highlighted in bold, is equivalent to separate models of nominal and real yields with three and two factors, respectively.

Model	Log likelihood	Number of free parameters	SIC
JM5	2413	48	−4441
SM	2390	36	−4491

Table 7: 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 root mean squared error (RMSE) between current-period model-implied and actual yields. Panels $\mathcal{B}$ and $\mathcal{C}$ report the time-series accuracy, that is, the RMSEs between model-implied expected yields one and twelve months ahead and subsequent realized yields when the model parameters are estimated using the full sample.

	Nominal (maturity in years)			Real (maturity in years)	
	1	5	10	5	10
$\mathcal{A}$: Cross section					
JM5	0.03	0.02	0.03	0.01	0.01
SM	0.03	0.02	0.03	0.01	0.01
$\mathcal{B}$: 1-step ahead					
JM5	0.24	0.28	0.25	0.22	0.17
SM	0.24	0.28	0.26	0.23	0.18
$\mathcal{C}$: 12-step ahead					
JM5	1.28	0.93	0.71	0.59	0.43
SM	1.30	0.99	0.78	0.61	0.42

5 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 in joint models of U.S. and German nominal and yields, and of U.S. nominal and real yields, the joint models offer no obvious advantages from the perspective of modeling the dynamics of bond yields relative to standard, separate models.

In both applications, we start by comparing separate models of a single class of yields with a joint model that has the same factor structure but allows for interactions between the factors spanned by each yield curve. Unsurprisingly, there is nothing to be gained from joint models in terms of the in-sample fit to yields. But joint models also have very similar predictive accuracy compared with separate models, with no statistically signicant improvements either in or out of sample.

In our application to U.S. and German yields, we also consider two approaches for potentially improving the out-of-sample properties of joint models by reducing the number of parameters: allowing for common factors to be spanned by both yield curves and imposing restrictions on the time-series dynamics of the factors. Neither approach brings any material improvement in the out-of-sample performance of the joint model, while imposing more than one common factor materially worsens the in-sample fit of a joint model.

While our analysis focuses on perhaps the two most popular applications of joint models in the previous literature, an obvious question is whether it generalizes to other applications, such as yields in other combinations of countries or yields on other types of assets. For example, it is possible that there may be greater advantages to joint modeling of yields in the United States and emerging markets. It is beyond the scope of this paper to consider every possible combination, so we leave such questions for further research. However, the results in this paper suggest that it is important to evaluate any proposed joint model according to whether it offers material advantages to separate models, particularly when a structure with common factors is being considered.

References

Abrahams, M., T. Adrian, R. K. Crump, and E. Moench (2016). Decomposing Real and Nominal Yield Curves. *Journal of Monetary Economics 84*, 182–200.

Anderson, B., P. J. Hammon, and C. A. Ramezani (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., S. Foresi, and C. I. Telmer (2001). Affine Term Structure Models and the Forward Premium Anomaly. *Journal of Finance 51*(1), 279–304.

Bauer, G. H. and A. Diez de los Rios (2012). An International Dynamic Term Structure Model with Economic Restrictions and Unspanned Risks. *Bank of Canada Working Paper 12-5.*

Bauer, M. D. (2018). Restrictions on Risk Prices in Dynamic Term Structure Models. *Journal of Business and Economic Statistics 36*(2), 196–211.

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

Bauer, M. D., G. D. Rudebusch, and J. C. Wu (2012). Correcting estimation bias in dynamic term structure models. *Journal of Business & Economic Statistics 30*, 454–467.

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

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

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

D'Amico, S., D. H. Kim, and M. Wei (2018). Tips from TIPS: the Informational Content of Treasury Inflation-Protected Security Prices. *Journal of Financial and Quantitative Analysis 53*(1), 395–436.

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

Diebold, F. X. and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics 13*(3), 253–263.

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., H. Li, and D. Ng (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.

Golinski, A. and P. Spencer (2018). Estimating the Term Structure with Linear Regressions: Getting to the Roots of the Problem. Working Paper.

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

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

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

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

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

Joyce, M. A. S., P. Lildholdt, and S. Sorensen (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., A. C. Meldrum, and J. Smith (2013). A Global Model of International Yield Curves: No-Arbitrage Term Structure Approach. *International Journal of Finance and Economics 18*(4), 352–374.

Kim, D. H. and A. Orphanides (2012). Term structure estimation with survey data on interest rate forecasts. *Journal of Financial and Quantitative Analysis 47*, 241–272.

Pegoraro, F., A. F. Siegel, and L. Pezzoli (2014). Specification Analysis of International Treasury Yield Curve Factors. *Banque de France Working Paper 490.*

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 (4) and (5). 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 (2) 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 (3) gives

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

Matching coefficients gives equations (4) and (5). 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 (16)-(17). 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 (8) 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 (9), (10), and (11) gives

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

Matching coefficients gives equations (16) and (17). 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 (18). Note that from equations (15)-(17) 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 (18) where

$$\delta_{2,0} = \delta_{1,0} - s_0 - \mathbf{s}'_1\boldsymbol{\mu}^{\mathbb{Q}_1} - \frac{1}{2}\mathbf{s}'_1\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 \text{ and} \tag{30}$$

$$\boldsymbol{\delta}_{2,1} = \boldsymbol{\delta}_{1,1} - \left(\boldsymbol{\Phi}^{\mathbb{Q}_1}\right)'\mathbf{s}_1, \tag{31}$$

as stated in the main text.

Appendix D: Two Separate Models as a Special Case of JM6

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 JM6.

Symmetry of the Joint Model

We start by noting that we can equivalently price bonds under the $\mathbb{P}$ measure. The prices of the first and second classes of bonds must also satisfy

$$\begin{aligned} P_{1,n,t} &= \mathbb{E}_t\left[M_{1,t+1}P_{1,n-1,t+1}\right] \text{ and} \qquad (32)\\ P_{2,n,t} &= \mathbb{E}_t\left[M_{2,t+1}P_{2,n-1,t+1}\right], \qquad (33)\end{aligned}$$

where $M_{j,t+1}$ is the stochastic discount factor that prices the j^{th} class of bonds. The assumptions in the main text imply that $M_{j,t+1}$ takes the form

$$M_{j,t+1} = \exp\left(-r_{j,t} - \frac{1}{2}\boldsymbol{\lambda}'_{j,t}\boldsymbol{\lambda}_{j,t} - \boldsymbol{\lambda}'_{j,t}\boldsymbol{\varepsilon}_{j,t+1}\right), \qquad (34)$$

where the prices of risk satisfy $\boldsymbol{\lambda}_{j,t} \equiv \boldsymbol{\lambda}_{j,0} + \boldsymbol{\Lambda}_{j,1}\mathbf{x}_t$, $\boldsymbol{\mu}^{\mathbb{Q}_1} \equiv \boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{\lambda}_{j,0}$, and $\boldsymbol{\Phi}^{\mathbb{Q}_1} \equiv \boldsymbol{\Phi} - \boldsymbol{\Sigma}\boldsymbol{\Lambda}_{j,1}$.

The prices of the second class of bonds must also satisfy

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

As shown by Backus et al. (2001), in the presence of complete markets $M_{2,t+1}$ must satisfy

$$\log M_{2,t+1} = \Delta s_{t+1} + \log M_{1,t+1}. \tag{35}$$

Following Diez de los Rios (2008), combining equations (10), (34), and (35) 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 (21) 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 (9) and (18) gives

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

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

$$
\boldsymbol{\lambda}_{2,t} = \boldsymbol{\lambda}_{1,t} - \boldsymbol{\Sigma}'\mathbf{s}_1 \tag{36}
$$

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,

$$\mathbf{x}_{t+1}=\boldsymbol{\mu}^{\mathbb{Q}_2}+\boldsymbol{\Phi}^{\mathbb{Q}_2}\mathbf{x}_t+\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}_2}, \tag{37}$$

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

$$\begin{aligned}
\boldsymbol{\mu}^{\mathbb{Q}_2} &= \boldsymbol{\mu}^{\mathbb{Q}_1}+\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{s}_1 \text{ and} && (38)\\
\boldsymbol{\Phi}^{\mathbb{Q}_2} &= \boldsymbol{\Phi}^{\mathbb{Q}_1}. && (39)
\end{aligned}$$

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 (30), (31), (38), and (39).

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 (3)) 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

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

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 (39) is satisfied. We can then set $\mathbf{s}_1$ in order satisfy equation (31), that is,

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

Thus we can set $\boldsymbol{\mu}_{1,\mathcal{S}}^{\mathbb{Q}_2} = -\left(\left(\boldsymbol{\Phi}_{1,\mathcal{S}}^{\mathbb{Q}_1}\right)'\right)^{-1}\boldsymbol{\delta}_{1,1,\mathcal{S}}$ and $\widetilde{\boldsymbol{\mu}}_{2,\mathcal{S}}^{\mathbb{Q}_1} = \widetilde{\boldsymbol{\mu}}_{2,\mathcal{S}}^{\mathbb{Q}_2} + \left(\left(\boldsymbol{\Phi}_{2,\mathcal{S}}^{\mathbb{Q}_2}\right)'\right)^{-1}\boldsymbol{\delta}_{2,1,\mathcal{S}}$ in order to satisfy equation (38). Further, because the second separate model is invariant to any level-shift $\boldsymbol{\theta}$, it must be invariant to the particular level-shift

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

which ensures that $\widetilde{\boldsymbol{\mu}}_{2,\mathcal{S}}^{\mathbb{Q}_1} = \mathbf{0}$, as required under our normalization of JM(6). Finally, we can set s_0 in order to satisfy equation (30).

Finally, we can write the $\mathbb{P}$ dynamics of the joint model as

$$\begin{bmatrix} \mathbf{x}_{1,t+1} \\ \widetilde{\mathbf{x}}'_{2,t+1} - \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{1,\mathcal{S}} \\ \boldsymbol{\mu}_{2,\mathcal{S}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}_{1,\mathcal{S}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{2,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t} \\ \widetilde{\mathbf{x}}'_{2,t} - \boldsymbol{\theta} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Sigma}_{1,\mathcal{S}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{2,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t+1} \\ \boldsymbol{\varepsilon}_{2,t+1} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{x}_{1,t+1} \\ \widetilde{\mathbf{x}}'_{2,t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{1,\mathcal{S}} \\ \boldsymbol{\mu}_{2,\mathcal{S}} + \boldsymbol{\theta}\left(\mathbf{I} - \boldsymbol{\Phi}_{2,\mathcal{S}}\right) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}_{1,\mathcal{S}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{2,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t} \\ \widetilde{\mathbf{x}}'_{2,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Sigma}_{1,\mathcal{S}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{2,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t+1} \\ \boldsymbol{\varepsilon}_{2,t+1} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{x}_{1,t+1} \\ \widetilde{\mathbf{x}}'_{2,t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{1,\mathcal{S}} \\ \boldsymbol{\mu}_{2,\mathcal{S}} + \boldsymbol{\theta}\left(\mathbf{I} - \boldsymbol{\Phi}_{2,\mathcal{S}}\right) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}_{1,\mathcal{S}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{2,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t} \\ \widetilde{\mathbf{x}}'_{2,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Sigma}_{1,\mathcal{S}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{2,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t+1} \\ \boldsymbol{\varepsilon}_{2,t+1} \end{bmatrix}$$

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

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

where $\mathbf{\Phi}_{1,\mathcal{S}}^{\mathbb{Q}_1}$ and $\mathbf{\Phi}_{2,\mathcal{S}}^{\mathbb{Q}_2}$ are lower triangular matrices with ordered diagonal elements. Thus, relative to a maximally-flexible joint model, the two separate models therefore imply additional zero restrictions on the parameters of $\boldsymbol{\delta}_{1,1}$, $\mathbf{\Phi}^{\mathbb{Q}_1}$, and $\mathbf{\Phi}$.