BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2017-33

Optimal Estimation of Multi-Country Gaussian Dynamic Term Structure Models Using Linear Regressions

by Antonio Diez de los Rios

Bank of Canada Staff Working Paper 2017-33

August 2017

Optimal Estimation of Multi-Country Gaussian Dynamic Term Structure Models Using Linear Regressions

by

Antonio Diez de los Rios

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
diez@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

I would like to thank Greg Bauer, Bruno Feunou, Jean-Sebastien Fontaine and Jonathan Witmer for their useful comments and suggestions on previous versions of this paper. The views expressed in this paper are those of the author and do not necessarily reflect those of the Bank of Canada.

Abstract

This paper proposes a novel asymptotic least-squares estimator of multi-country Gaussian dynamic term structure models that is easy to compute and asymptotically efficient, even when the number of countries is relatively large—a situation in which other recently proposed approaches lose their tractability. We illustrate our estimator within the context of a seven-country, 10-factor term structure model.

Bank topics: Asset pricing; Econometric and statistical methods; Exchange rates; Interest rates
JEL codes: E43, F31, G12, G15

Résumé

Nous proposons un nouvel estimateur des modèles dynamiques gaussiens de la structure par terme des taux d'intérêt à plusieurs pays. Cet estimateur fondé sur la méthode des moindres carrés ordinaires est facile à calculer et asymptotiquement efficient, même avec un assez grand nombre de pays, un cas pour lequel d'autres méthodes proposées récemment perdent leur simplicité. Nous illustrons l'emploi de l'estimateur dans un modèle de structure par terme à sept pays et dix facteurs.

Sujets : Évaluation des actifs ; Méthodes économétriques et statistiques ; Taux de change ; Taux d'intérêt
Codes JEL : E43, F31, G12, G15

Non-Technical Summary

In the wake of the financial crisis of 2007-08 and its transmission around the world, both academics and market practitioners have found a renewed interest in understanding the links among the yield curves denominated in different currencies. At the heart of this literature is the Gaussian dynamic term structure model (GDTSM), thanks to its tractability and relationship with the Gaussian vector autoregressive (VAR) model, a widely used empirical tool in macro-finance studies.

However, the estimation of these models in a multi-country setup tends to be problematic and researchers often face a myriad of numerical challenges when working with these models because of:

(i) the large number of parameters involved in these models,
(ii) the highly non-linear nature of the likelihood function, and/or
(iii) the existence of multiple local optima.

In fact, these issues are magnified in the case of multi-country models due to the increased number of parameters and factors needed to properly describe the joint dynamics of yield curves across different currencies. For example, the number of parameters one needs to jointly estimate in the case of a seven-country and 10-factor model, as in our empirical illustration, is 213, which renders traditional methods to estimate these models un-implementable.

In this paper, we overcome these issues by extending the linear estimator of Diez de los Rios (2015a) to the case of multi-country term structure models with unspanned exchange rate risk. This method completely avoids numerical optimization methods whenever yields on adjacent maturities are directly observed (i.e., whenever the researcher observes yields on both 16-quarter and 17-quarter bonds).

For illustrative purposes, we estimate a seven-country and 10-factor model and decompose 10-year zero-coupon bond yields into expectations and term premium components. Using this decomposition to analyze the covariation of the term premia across yield curves denominated in different currencies within a unified framework, we find that only 2 factors might be needed to explain most of the (economically interesting) variation in term premia: a result in line with studies in the United States.

1 Introduction

In the wake of the financial crisis of 2007-08 and its transmission around the world, both academics and market practitioners have found a renewed interest in understanding the links among the yield curves denominated in different currencies (see, e.g., Diebold, Li and Yue, 2009; Sarno, Schneider and Wagner, 2012; Dahlquist and Hasseltoft, 2013; Jotikasthira, Le and Lundblad, 2015; Meldrum, Raczko and Spencer, 2016). At the heart of this literature is the Gaussian dynamic term structure model (GDTSM), thanks to its tractability and relationship with the Gaussian vector autoregressive (VAR) model, a widely used empirical tool in macro-finance studies (see Ang and Piazzesi, 2003, for an extended discussion on this relationship).

The maximum likelihood (ML) approach has been traditionally considered as the most natural way to estimate GDTSMs, since such models provide a complete characterization of the joint distribution of yields. However, even in one-country studies, researchers often face a myriad of numerical challenges when using ML methods to estimate these models because of (i) the large number of parameters involved in these models, (ii) the highly non-linear nature of the likelihood function, and/or (iii) the existence of multiple local optima (e.g., the discussions in Duffee and Stanton, 2012; Hamilton and Wu, 2012). In fact, these issues are magnified in the case of multi-country models because of the increased number of parameters and factors needed to properly describe the joint dynamics of yield curves across different currencies. Consequently, the literature has been restricted to mainly two-country models (e.g., Backus, Foresi and Telmer, 2001), needed very computationally intensive methods for estimation (e.g., Sarno, Schneider and Wagner, 2012), used only domestic factors to fit the term structure of interest rates (e.g., Dahlquist and Hasseltoft, 2013), or even excluded exchange rate data from the analysis of these models (e.g., Jotikasthira, Le and Lundblad, 2015).

In this paper, we overcome these issues by extending the linear estimator of Diez de los Rios (2015a), which completely avoids numerical optimization methods whenever yields on adjacent maturities are directly observed (i.e., whenever the researcher observes yields on both 16-quarter and 17-quarter bonds), to the case of multi-country term structure models with unspanned exchange rate risk.[1] Importantly, we show how to overcome Golinski and

[1] A variable is unspanned if its value is not linearly related to the contemporaneous cross-section of bond yields.

Spencer's (2017) recent finding that this estimator tends to diverge when the number of bond pricing factors is larger than three, thus paving the way for its application to international term structure models with a large number of countries, exchange rates and bond pricing factors.

Specifically, our proposed estimator is an asymptotic least squares (ALS) estimator that exploits three features that characterize GDTSMs. First, these models have a reduced-form representation whose parameters can be easily estimated using ordinary least squares (OLS) regressions. Second, the no-arbitrage assumption upon which GDTSMs are built can be characterized as a set of implicit constraints between these reduced-form parameters and the parameters of interest. Third, this set of restrictions is linear in the parameters of interest. Consequently, we propose a two-step estimator, in which we first estimate the reduced-form parameters by OLS. In the second step, the parameters of the GDTSMs are inferred by forcing the no-arbitrage constraints, evaluated at the first-stage estimates of the reduced-form parameters, to be as close as possible to zero in the metric defined by a given weighting matrix. Note that, since the constraints are linear in the parameters of interest, the solution to the estimation problem in this second step is known in closed form. More importantly, our proposed estimator is asymptotically equivalent to maximum likelihood (ML) estimation under a suitably chosen weighting matrix.

While some recent approaches to the estimation of one-country GDTSMs have substantially lessened some of the numerical challenges faced by researchers, we argue that such approaches cannot really handle models where the number of countries is large. In particular, we derive a multi-country version of the canonical representation of Joslin, Singleton and Zhu (2011) (JSZ) and note that the ML estimator based on such representation still implies a numerical search over a very large dimensional space when either the number of countries or the number of factors is moderately large (e.g., 213 parameters in the case of a seven-country and 10-factor model as in our empirical illustration). This renders the MLE un-implementable in such cases, leaving the ALS methods proposed in this paper as the only reliable alternative for the estimation of international term structure models with either a large number of countries or factors.

For illustrative purposes, we estimate a seven-country and 10-factor model and decompose 10-year zero-coupon bond yields into expectations and term premium components. Furthermore, using this decomposition to analyze the covariation of the term premia across yield curves denominated in different currencies within a unified framework, we find that only two factors might be needed to explain most of the (economically interesting) variation in term premia: a result in line with those in Duffee (2010) and Joslin Priebsch and Singleton (2014) for the U.S. case.

The structure of article is as follows. In section 2, we describe the class of multi-country GDTSMS with unspanned foreign exchange risk, and discuss its estimation using the ALS framework in section 3. In section 4, we discuss the relationship of our proposed approach with ML estimation. Our empirical illustration is presented in section 5. Section 6 concludes.

2 International Gaussian Term Structure Models

2.1 Basic Framework

We start by considering a world with $J+1$ countries and currencies where, without loss of generality, we consider the $J+1$st currency to be the numeraire (U.S. dollar in our case). Let $s_{j,t}$ be the (log) U.S. dollar price of a unit of foreign currency j and $\Delta s_{j,t} \equiv s_{j,t} - s_{j,t-1}$ be the rate of depreciation of currency j against the U.S. dollar, which we collect in the $(J \times 1)$ vector $\Delta \mathbf{s}_t = (\Delta s_{1,t}, \ldots, \Delta s_{J,t})'$.

For each country j, there is a set of n-period default-free discount bonds with prices in local currency given by $P_{j,t}^{(n)}$ for $n = 1, ..., N$, and (log) yields given $y_{j,t}^{(n)} = -\frac{1}{n} \log P_{j,t}^{(n)}$. Let $\mathbf{y}_{j,t} = (y_{j,t}^{(1)}, ..., y_{j,t}^{(N)})'$ be a $(N \times 1)$ vector that collects all yields in country j, and let $\mathbf{y}_t = \left(\mathbf{y}'_{\$,t}, \mathbf{y}'_{1,t}, ..., \mathbf{y}'_{J,t}\right)'$ be a $(\mathcal{N} \times 1)$ vector, with $\mathcal{N} = N \times (J+1)$, that collects all yields in the (global) economy.[2]

The state of the global economy is summarized by the following two vectors of state variables: (i) a $(F \times 1)$ vector $\mathbf{x}_{b,t}$, with $F \leq \mathcal{N}$, of bond pricing factors that completely describe the correlation structure of bond yields, and (ii) the $(J \times 1)$ vector $\Delta \mathbf{s}_t$ collecting the rates of depreciation of the J currencies against the U.S. dollar. Further, the joint dynamic evolution of these state variables under the physical measure, $\mathbb{P}$, is governed by a VAR(1) process with Gaussian innovations:

$$\begin{pmatrix} \mathbf{x}_{b,t+1} \\ \Delta \mathbf{s}_{t+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_b \\ \boldsymbol{\mu}_s \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Phi}_{bb} & \boldsymbol{\Phi}_{bs} \\ \boldsymbol{\Phi}_{sb} & \boldsymbol{\Phi}_{ss} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{b,t} \\ \Delta \mathbf{s}_t \end{pmatrix} + \begin{pmatrix} \mathbf{v}_{b,t+1} \\ \mathbf{v}_{s,t+1} \end{pmatrix}, \quad (1)$$

[2]Note that, for simplicity and without loss of generality, we have assumed that the number of bonds in each country is the same.

which can be represented in compact form as $\mathbf{x}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1}$, where $\mathbf{x}_t = (\mathbf{x}'_{b,t}, \Delta\mathbf{s}'_t)'$ is a $(M \times 1)$ vector with $M = F + J$, and $\mathbf{v}_t \sim iid\ N(0, \boldsymbol{\Sigma})$.

Let $r_{j,t}$ be the continuously compounded one-period interest rate in country j (i.e., the short rate), which is related to the set of bond pricing factors through the following affine relation:

$$r_{j,t} = \delta_j^{(0)} + \boldsymbol{\delta}_j^{(1)\prime}\mathbf{x}_{b,t}, \qquad j = \$, 1, \ldots, J. \tag{2}$$

Collecting the short rates into the $[(J+1) \times 1]$ vector $\mathbf{r}_t = (r_{\$,t}, r_{1,t}, \ldots r_{J,t})'$, we can represent equation (2) in compact form as $\mathbf{r}_t = \boldsymbol{\Delta}^{(0)} + \boldsymbol{\Delta}^{(b)}\mathbf{x}_{b,t}$, where $\boldsymbol{\Delta}^{(0)} = (\delta_\$^{(0)}, \delta_1^{(0)}, \ldots, \delta_J^{(0)})'$ and $\boldsymbol{\Delta}^{(b)} = (\boldsymbol{\delta}_\$^{(b)}, \boldsymbol{\delta}_1^{(b)}, \ldots, \boldsymbol{\delta}_J^{(b)})'$.[3]

Lastly, the model is completed by specifying the dynamics of the state variables under the risk-neutral probability measure, $\mathbb{Q}$, for the numeraire currency (i.e., the U.S). Specifically, we assume that the joint evolution of the bond and exchange rate factors under $\mathbb{Q}$ is characterized by the following VAR(1) process with Gaussian innovations:

$$\begin{pmatrix} \mathbf{x}_{b,t+1} \\ \Delta\mathbf{s}_{t+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_b^{\mathbb{Q}} \\ \boldsymbol{\mu}_s^{\mathbb{Q}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Phi}_{bb}^{\mathbb{Q}} & \mathbf{0} \\ \boldsymbol{\Phi}_{sb}^{\mathbb{Q}} & \boldsymbol{\Phi}_{ss}^{\mathbb{Q}} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{b,t} \\ \Delta\mathbf{s}_t \end{pmatrix} + \begin{pmatrix} \mathbf{v}_{b,t+1}^{\mathbb{Q}} \\ \mathbf{v}_{s,t+1}^{\mathbb{Q}} \end{pmatrix}, \tag{3}$$

which can be represented in compact form as $\mathbf{x}_{t+1} = \boldsymbol{\mu}^{\mathbb{Q}} + \boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{x}_t + \mathbf{v}_{t+1}^{\mathbb{Q}}$ with $\mathbf{v}_t^{\mathbb{Q}} \sim iid$ $N(0, \boldsymbol{\Sigma})$ and where $\mathbf{0}$ is a conformable matrix of zeros. Under the assumption of absence of arbitrage opportunities, this risk-neutral measure can be used to price any traded asset denominated in U.S. dollars using the following relation:

$$P_t = E_t^{\mathbb{Q}}\left[\exp(-r_{\$,t})X_{t+1}\right], \tag{4}$$

where P_t is the value of a claim to a stochastic cash flow of X_{t+1} U.S. dollars one period later.

Our specification of the $\mathbb{Q}$-measure has three ingredients. First, given that we focus on models where the exchange rate risks are unspanned, we assume that the bond pricing factors, $\mathbf{f}_t$, follow an autonomous Gaussian VAR(1) process under the risk-neutral measure (i.e., $\boldsymbol{\Phi}_{bs}^{\mathbb{Q}} = \mathbf{0}$). In the absence of this restriction, no-arbitrage pricing would imply that bond yields would be affine functions of all $\mathbf{x}_{b,t}$, and $\Delta\mathbf{s}_t$ (cf equations 9 and 13 below), which is contrary to our assumption that only the bond pricing factors, $\mathbf{f}_t$, are needed to

[3] We assume that there are no redundant factors. That is, for every factor $x_{b,k,t}$, there is at least one country j for which its loading with respect to this factor is different from zero, $\delta_{jk}^{(1)} \neq 0$. Otherwise, we would be contradicting our assumption of an F-factor structure for bond yields.

adequately represent the correlation structure of bond yields.[4]

Second, we note that the nominal expected return to currency speculation, conditional on the available information, must be equal to zero under the risk-neutral measure. This is a consequence of the pricing of a foreign one-period bond by a U.S. investor. In particular, using equation (4), we have that $P_{j,t}^{(1)} \times S_t = E_t^{\mathbb{Q}} \left(e^{-r_{\$,t}} \times S_{t+1} \times 1 \right)$, which in its log form implies that the uncovered interest parity must be satisfied under the $\mathbb{Q}$-measure:

$$E_t^{\mathbb{Q}} \Delta s_{j,t+1} = -\frac{1}{2} Var_t^{\mathbb{Q}} \left(\Delta s_{j,t+1} \right) + (r_{\$,t} - r_{j,t}), \qquad j = 1, \ldots, J,$$

where $-\frac{1}{2} Var_t \left(\Delta s_{j,t+1} \right)$ is a Jensen's inequality term which, in turn, pins down the coefficients in $\boldsymbol{\mu}_s^{\mathbb{Q}}$, $\boldsymbol{\Phi}_{sb}^{\mathbb{Q}}$, and $\boldsymbol{\Phi}_{ss}^{\mathbb{Q}}$:

$$\mathbf{e}_j' \boldsymbol{\mu}_s^{\mathbb{Q}} = -\frac{1}{2} \mathbf{e}_j' \boldsymbol{\Sigma}_{ss} \mathbf{e}_j + \left[\delta_{\$}^{(0)} - \delta_j^{(0)} \right], \tag{5}$$

$$\mathbf{e}_j' \boldsymbol{\Phi}_{sb}^{\mathbb{Q}} = \left[\boldsymbol{\delta}_{\$}^{(b)} - \boldsymbol{\delta}_j^{(b)} \right]', \tag{6}$$

$$\mathbf{e}_j' \boldsymbol{\Phi}_{ss}^{\mathbb{Q}} = \mathbf{0}', \tag{7}$$

for $j = 1, \ldots, J$, where $\mathbf{e}_j$ is a conformable vector of zeros with a one in the j-th position.

Third, consistent with the literature on risk-neutral valuation, we have assumed that the conditional variance-covariance matrices of the innovations to the pricing factors, $\mathbf{x}_t$, are the same under both the physical and risk-neutral distribution (see Monfort and Pegoraro, 2012, for a relaxation of this hypothesis): $Var_t(\mathbf{v}_{t+1}) = Var_t(\mathbf{v}_{t+1}^{\mathbb{Q}}) = \boldsymbol{\Sigma}$.

Bond pricing in the numeraire country We can now use risk-neutral valuation to price zero-coupon bonds by specializing equation (4) to the case of zero-coupon bonds in the numeraire country. Specifically:

$$P_{\$,t}^{(n)} = E_t^{\mathbb{Q}} \left[\exp(-r_{\$,t}) P_{\$,t+1}^{(n-1)} \right], \tag{8}$$

where $P_{\$,t}^{(n)}$ is the price of a U.S. zero-coupon bond of maturity n periods at time t. Note that, by recursive substitution of equation (8), we find that:

$$P_{\$,t}^{(n)} = E_t^{\mathbb{Q}} \left[\exp \left(-\sum_{i=0}^{n-1} r_{\$,t+i} \right) \right],$$

[4] We note that, while the evidence on macro risk (un)spanning is mixed (see, e.g., Bauer and Hamilton, 2015, and Bauer and Rudebusch, 2017), there is clear evidence that foreign exchange risk is not spanned by interest rates. For example, Brandt and Santa-Clara (2002) introduce an exchange rate factor that is orthogonal to both interest rates and the SDFs in order to match the high degree of exchange rate volatility.

That is, one can price a zero-coupon bond as if agents were risk neutral by using the (local) expectations hypothesis once the law of motion of the state variables has been modified to account for the fact that agents are not risk neutral.

Solving (8), we show in Appendix A.1 that the continuously compounded yield on an n-period zero-coupon bond denominated in U.S. dollars at time t, $y_{\$,t}^{(n)} = -\frac{1}{n}\log P_{\$,t}^{(n)}$, is given by

$$y_{\$,t}^{(n)} = a_{\$}^{(n)} + \mathbf{b}_{\$}^{(n)\prime}\mathbf{x}_{b,t}, \tag{9}$$

where $a_{\$}^{(n)} = -A_{\$}^{(n)}/n$ and $\mathbf{b}_{\$}^{(n)} = -\mathbf{B}_{\$}^{(n)}/n$, and $A_{\$}^{(n)}$ and $\mathbf{B}_{\$}^{(n)}$ satisfy the following set of recursive relations:

$$\mathbf{B}_{\$}^{(n)\prime} = \mathbf{B}_{\$}^{(n-1)\prime}\mathbf{\Phi}_{bb}^{\mathbb{Q}} + \mathbf{B}_{\$}^{(1)\prime}, \tag{10}$$

$$A_{\$}^{(n)} = A_{\$}^{(n-1)} + \mathbf{B}_{\$}^{(n-1)\prime}\boldsymbol{\mu}_{b}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}_{\$}^{(n-1)\prime}\mathbf{\Sigma}_{bb}\mathbf{B}_{\$}^{(n-1)} + A_{\$}^{(1)}, \tag{11}$$

for $n = 2, ..., N$. The recursion is started by exploiting the fact that the affine pricing relationship is trivially satisfied for one-period bonds (i.e., $y_{t}^{(1)} = r_t$), which implies that $A_{\$}^{(1)} = -\delta_{\$}^{(0)}$ and $\mathbf{B}_{\$}^{(1)} = -\boldsymbol{\delta}_{\$}^{(b)}$.

Bond pricing in the foreign country In a similar fashion, we can use again the risk-neutral approach to price the zero-coupon bonds in the rest of the countries:

$$P_{j,t}^{(n)} = E_t^{\mathbb{Q}}\left[\exp(-r_{\$,t})\frac{S_{t+1}}{S_t}P_{j,t+1}^{(n-1)}\right], \tag{12}$$

where $P_{j,t}^{(n)} \times S_t$ is the price in U.S. dollars of the zero-coupon bond of maturity n periods at time t in country j, and $P_{j,t+1}^{(n-1)} \times S_{t+1}$ is the payoff in U.S. dollars that a U.S. investor will obtain by selling the n-period zero-coupon bond one period later.

Specifically, we show in Appendix A.2 that, solving (12), the continuously compounded yield on a foreign n-period zero-coupon bond at time t, $y_{j,t}^{(n)}$, is also affine in the set of bond pricing factors, $\mathbf{x}_{b,t}$:

$$y_{j,t}^{(n)} = a_{j}^{(n)} + \mathbf{b}_{j}^{(n)\prime}\mathbf{x}_{b,t}, \tag{13}$$

where $a_{j}^{(n)} = -A_{j}^{(n)}/n$ and $\mathbf{b}_{j}^{(n)} = -\mathbf{B}_{j}^{(n)}/n$, and the scalar $A_{j}^{(n)}$ and vector $\mathbf{B}_{j}^{(n)\prime}$ satisfy a set of recursive relations similar to those for the numeraire country:

$$\mathbf{B}_{j}^{(n)\prime} = \mathbf{B}_{j}^{(n-1)\prime}\mathbf{\Phi}_{bb}^{\mathbb{Q}} + \mathbf{B}_{j}^{(1)\prime}, \tag{14}$$

$$A_{j}^{(n)} = A_{j}^{(n-1)} + \mathbf{B}_{j}^{(n-1)\prime}\left[\boldsymbol{\mu}_{b}^{\mathbb{Q}} + \mathbf{\Sigma}_{bs}\mathbf{e}_j\right] + \frac{1}{2}\mathbf{B}_{j}^{(n-1)\prime}\mathbf{\Sigma}_{bb}\mathbf{B}_{j}^{(n-1)} + A_{j}^{(1)}, \tag{15}$$

for $n = 2, ..., N$. Once again, the recursion is started by exploiting the fact that the affine pricing relationship is trivially satisfied for one-period bonds ($n = 1$), which implies that $A_j^{(1)} = -\delta_j^{(0)}$, and $\mathbf{B}_j^{(1)} = -\boldsymbol{\delta}_j^{(b)}$.

2.2 A reduced-form representation

As noted by Hamilton and Wu (2012), GDTSMs have a reduced-form representation that can be exploited to estimate the parameters of interest of the model. In particular, our model admits the following state-space representation of the observed bond yields:

$$\mathbf{y}_t^o = \mathbf{a} + \mathbf{b}\mathbf{x}_{b,t} + \boldsymbol{\eta}_t, \tag{16}$$

$$\begin{pmatrix} \mathbf{x}_{b,t+1} \\ \Delta\mathbf{s}_{t+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_b \\ \boldsymbol{\mu}_s \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Phi}_{bb} & \boldsymbol{\Phi}_{bs} \\ \boldsymbol{\Phi}_{sb} & \boldsymbol{\Phi}_{ss} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{b,t} \\ \Delta\mathbf{s}_t \end{pmatrix} + \begin{pmatrix} \mathbf{v}_{b,t+1} \\ \mathbf{v}_{s,t+1} \end{pmatrix}, \tag{17}$$

where $\mathbf{y}_t$ is the vector of model-implied yields that stack the affine mappings in equations (9) and (13), for all maturities and countries, $\mathbf{y}_t^o$ is the corresponding vector of observed yields and $\boldsymbol{\eta}_t$ is a zero-mean measurement error that is *i.i.d.* across time and that has a covariance matrix $\boldsymbol{\Omega}$. Note that $\mathbf{a} = \mathbf{a}(\boldsymbol{\mu}_b^{\mathbb{Q}}, \boldsymbol{\Phi}_{bb}^{\mathbb{Q}}, \boldsymbol{\Sigma}_{bb}, \boldsymbol{\Sigma}_{bs})$ and $\mathbf{b} = \mathbf{b}(\boldsymbol{\Phi}_{bb}^{\mathbb{Q}})$ are non-linear functions of $\boldsymbol{\mu}_b^{\mathbb{Q}}, \boldsymbol{\Phi}_{bb}^{\mathbb{Q}}, \boldsymbol{\Sigma}_{bb}, \boldsymbol{\Sigma}_{bs}$.

The parameters of this reduced-form representation can be trivially estimated when the bond pricing factors are observable. Specifically, we follow Joslin, Singleton and Zhu (2011) in working with bond state variables that are linear combinations (i.e., portfolios) of the observed yields, $\mathbf{x}_{b,t} = \mathbf{P}'\mathbf{y}_t^o$, where $\mathbf{P}$ is a $(\mathcal{N} \times F)$ full-rank matrix of weights, and by further assuming that $\mathbf{x}_{b,t}$ is observed perfectly. That is, $\mathbf{P}'(\mathbf{y}_t^o - \mathbf{y}_t) = \mathbf{P}'\boldsymbol{\eta}_t = 0$ $\forall t$. Since the errors of the model are conditionally homoskedastic, this assumption allows us to obtain maximum likelihood (ML) estimates of the reduced-form parameters via a set of OLS regressions (see Sentana, 2002, Hamilton and Wu, 2012, and Diez de los Rios, 2015a): (i) the (cross-sectional) coefficients $\mathbf{a}$ and $\mathbf{b}$ could be estimated from the OLS regression of $\mathbf{y}_t^o$ on a constant and $\mathbf{x}_{b,t}$; (ii) the (time-series) coefficients $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$ could be estimated from the OLS regression of $\mathbf{f}_t$ on a constant and its lag.[5]

Then, similar to the case of one-country GDSTMs in Diez de los Rios (2015a), one can use Gourieroux, Monfort and Trognon's (1982, 1985) (GMT hereafter) ALS estimation framework to obtain estimates of the model parameters by trying to force the pricing

[5] We further assume that $\boldsymbol{\Omega} = \boldsymbol{\sigma}_\eta^2 \times (\mathbf{P}_\perp \mathbf{P}_\perp')$ where $\mathbf{P}_\perp'$ is a basis for the orthogonal component of the row span of $\mathbf{P}'$. This guarantees that $\mathbf{P}'\boldsymbol{\Omega}\mathbf{P} = \mathbf{0}$ and allows concentrating $\boldsymbol{\sigma}_\eta^2$ from the likelihood function through $\widehat{\boldsymbol{\sigma}}_\eta^2 = \sum_{t=1}^T \sum_{j=\$}^J \sum_{n=1}^N (y_{t,n}^o - y_{t,n})^2 / (T \times J \times (N - M))$.

recursions in (10), (11), (14), (15), evaluated at the estimates of the reduced-form parameters, to be as close as possible to zero. We discuss such an ALS estimator of multi-country GDTSMs in the next section.

3 Asymptotic least squares estimation of international GDTSMs

3.1 The asymptotic least squares estimation framework

As noted by GMT, many empirical models can be formalized as a set of G implicit equations $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ between a set of parameters of interest $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^K$ and a set of auxiliary parameters $\boldsymbol{\pi} \in \boldsymbol{\Pi} \subset \mathbb{R}^H$.[6] In the case of the estimation of GDTSMs, we advance that $\boldsymbol{\theta}$ is related to the parameters of the no-arbitrage model in equations (1), (2), and (3); $\boldsymbol{\pi}$ is related to the set of parameters from the reduced-form model in equations (16) and (17); the set equations $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ is related to the pricing recursions in equations (10), (11), (14) and (15); and $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$.

Further, we assume the existence of a strongly consistent and asymptotically normal estimator of the auxiliary parameters $\widehat{\boldsymbol{\pi}}$, such that as $T \to \infty$, $\widehat{\boldsymbol{\pi}} \to \boldsymbol{\pi}^0$, P_{θ^0} almost surely; and $\sqrt{T}(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0) \xrightarrow{d} N\left[\mathbf{0}, \mathbf{V}_\pi(\boldsymbol{\theta}^0)\right]$, where T denotes the number of observations in the sample and $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$ denote the true value of the parameters of interest and auxiliary parameters respectively, i.e., $\mathbf{g}(\boldsymbol{\pi}^0, \boldsymbol{\theta}^0) = \mathbf{0}$.

The ALS estimation principle consists of minimizing a quadratic form in the distance function evaluated at the estimates of the auxiliary parameters, $\widehat{\boldsymbol{\pi}}$:

$$\widehat{\boldsymbol{\theta}}_{ALS} = \arg\min_{\boldsymbol{\theta}} T\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})'\mathbf{W}_T\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}), \tag{18}$$

where $\mathbf{W}_T$ is a positive semi-definite weighting matrix that possibly depends on the observations. In other words, GMT propose forcing the G implicit equations evaluated at $\widehat{\boldsymbol{\pi}}$ to be as close as possible to zero in the metric defined by $\mathbf{W}_T$. Further, notice that, when the distance function is linear in the set of parameters of interest (as in the case of the estimation of GDTSMs), the solution to the optimization problem in (18) is known in closed form.

[6] To be more specific, we assume that the set of G implicit equations $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ has a unique solution for $\boldsymbol{\theta}$ given $\boldsymbol{\pi}$ so that the parameters of interest can be determined without ambiguity from the auxiliary parameters.

Further, assuming that (i) $\mathbf{g}(\boldsymbol{\pi},\boldsymbol{\theta})$ is twice continuously differentiable, (ii) $\mathbf{W}_T$ converges P_{θ^0} almost surely to $\mathbf{W}$, a non-stochastic semi-definite weighting matrix of size G, and rank greater or equal than K, (iii) the true values of the parameters of interest and auxiliary parameters, $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$, both belong to the interior of $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$, respectively, (iv) and $\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\theta}}\mathbf{W}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}$ evaluated at $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$ is non-singular (which implies that the rank of $\partial \mathbf{g}/\partial \boldsymbol{\theta}' = K$ and that $K \leq G$), then (see GMT for the proof) $\widehat{\boldsymbol{\theta}}_{ALS}$ is strongly consistent for every choice of $\mathbf{W}_T$, and its asymptotic distribution is given by

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{ALS}-\boldsymbol{\theta}^0) \xrightarrow{d} N\left[\mathbf{0},\left(\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\theta}}\mathbf{W}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}\right)^{-1}\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\theta}}\mathbf{W}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\pi}'}\mathbf{V}_{\pi}\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\pi}}\mathbf{W}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}\left(\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\theta}}\mathbf{W}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}\right)^{-1}\right], \tag{19}$$

where the various matrices in this equation are evaluated at $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$.

3.2 The case of GDTSMs

In the specific example of GDTSMs, we have that the vector of auxiliary parameters is given by $\boldsymbol{\pi} = (\boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \boldsymbol{\pi}_3')'$ (i.e., the reduced-form parameters), where $\boldsymbol{\pi}_1 = vec\left[(\mathbf{a}\ \mathbf{b})'\right]$, $\boldsymbol{\pi}_2 = vec\left[(\boldsymbol{\mu}\ \boldsymbol{\Phi})'\right]$, and $\boldsymbol{\pi}_3 = vech\left(\boldsymbol{\Sigma}^{1/2}\right)'$. In order to guarantee the positivity of the covariance matrix $\boldsymbol{\Sigma}$, we focus on its Cholesky decomposition, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2\prime}$ rather than on $\boldsymbol{\Sigma}$ itself. Thus, we have a total of $H = \mathcal{N} \times (M+1) + (M+J) \times (M+J+1) + (M+J) \times (M+J+1)/2$ auxiliary parameters.

As previously noted in section 2.2, the maximum likelihood estimation of the reduced-form parameters coincides with OLS estimation equation-by-equation, and therefore there is a consistent and asymptotically normal estimate $\widehat{\boldsymbol{\pi}}$ available. Specifically, we have that

$$\sqrt{T}\left[\begin{pmatrix}\widehat{\boldsymbol{\pi}}_1\\ \widehat{\boldsymbol{\pi}}_2\\ \widehat{\boldsymbol{\pi}}_3\end{pmatrix}-\begin{pmatrix}\boldsymbol{\pi}_1^0\\ \boldsymbol{\pi}_2^0\\ \boldsymbol{\pi}_3^0\end{pmatrix}\right] \xrightarrow{d} N\left[\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{pmatrix},\begin{pmatrix}\mathbf{V}_{\pi_1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{V}_{\pi_2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{V}_{\pi_3}\end{pmatrix}\right], \tag{20}$$
$$\sqrt{T}\left(\widehat{\boldsymbol{\pi}}-\boldsymbol{\pi}^0\right) \xrightarrow{d} N\left(\mathbf{0},\mathbf{V}_{\pi}\right),$$

where $\mathbf{V}_{\pi_1} = \boldsymbol{\Omega} \otimes E(\widetilde{\mathbf{x}}_{b,t}\widetilde{\mathbf{x}}_{b,t}')^{-1}$, $\mathbf{V}_{\pi_2} = \boldsymbol{\Sigma} \otimes E(\widetilde{\mathbf{x}}_t\widetilde{\mathbf{x}}_t')^{-1}$, $\mathbf{V}_{\pi_3} = 2\mathbf{E}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{E}'$; with $\widetilde{\mathbf{x}}_{b,t} = (1\ \mathbf{x}_{b,t}')'$, $\widetilde{\mathbf{x}}_t = (1\ \mathbf{x}_t')'$, $\mathbf{E} = \left[\mathbf{L}_M(\mathbf{I}+\mathbf{K}_{MM})(\boldsymbol{\Sigma}^{1/2}\otimes\mathbf{I})\mathbf{L}_M'\right]^{-1}\mathbf{D}_M^{+}$, where $\mathbf{L}_M$ is an "elimination matrix" such that $vech\,(\boldsymbol{\Sigma}) = \mathbf{L}_M vec\,(\boldsymbol{\Sigma})$, $\mathbf{K}_{MM}$ is a "commutation matrix" such that $\mathbf{K}_{MM}vec(\mathbf{F}) = vec(\mathbf{F}')$ for any $(M \times M)$ matrix $\mathbf{F}$, and $\mathbf{D}_M^{+} = (\mathbf{D}_M'\mathbf{D}_M)^{-1}\mathbf{D}_M'$ where $\mathbf{D}_M$ is a "duplication matrix" satisfying $\mathbf{D}_M vech\,(\boldsymbol{\Sigma}) = vec\,(\boldsymbol{\Sigma})$ (see Lütkepohl, 1989).

Next, we consider the pricing recursions in equations (10), (11), (14) and (15). Let $\boldsymbol{\Theta}^{\mathbb{Q}}$ be a matrix that collects the parameters driving the dynamics under the risk-neutral measure in the following way:

$$\boldsymbol{\Theta}^{\mathbb{Q}\prime} = \begin{pmatrix} \boldsymbol{\Delta}^{(0)} & \boldsymbol{\Delta}^{(b)} \\ \boldsymbol{\mu}_b^{\mathbb{Q}} & \boldsymbol{\Phi}_{bb}^{\mathbb{Q}} \end{pmatrix},$$

and let $\boldsymbol{\theta}$ be the vector of parameters of interest such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \boldsymbol{\theta}_3')'$ with $\boldsymbol{\theta}_1 = vec(\boldsymbol{\Theta}^{\mathbb{Q}})$, $\boldsymbol{\theta}_2 = vec\left[(\boldsymbol{\mu}\ \boldsymbol{\Phi})'\right]$, and $\boldsymbol{\theta}_3 = vech\left(\boldsymbol{\Sigma}^{1/2}\right)$. Thus, we have a total of $K = (J+M)\times(M+1)+(M+J)\times(M+J+1)+(M+J)\times(M+J+1)/2$ parameters of interest.

Then, by stacking equations for all bond yields and countries, we can express the restrictions implied by the no-arbitrage model in compact form as

$$\mathbf{G}(\boldsymbol{\pi}, \boldsymbol{\theta})' = \mathbf{Y}(\boldsymbol{\pi}) - \mathbf{X}(\boldsymbol{\pi})\boldsymbol{\Theta}^{\mathbb{Q}\prime} = \mathbf{0}, \tag{21}$$

where $\mathbf{Y}(\boldsymbol{\pi}) = \left[\mathbf{Y}_{\$}(\boldsymbol{\pi})', \mathbf{Y}_1(\boldsymbol{\pi})', ..., \mathbf{Y}_J(\boldsymbol{\pi})'\right]'$ and $\mathbf{X}(\boldsymbol{\pi}) = \left[\mathbf{X}_{\$}(\boldsymbol{\pi})', \mathbf{X}_1(\boldsymbol{\pi})', ..., \mathbf{X}_J(\boldsymbol{\pi})'\right]'$ with

$$\mathbf{Y}_{\$}(\boldsymbol{\pi}) = \begin{pmatrix} A_{\$}^{(1)} & \mathbf{B}_{\$}^{(1)\prime} \\ A_{\$}^{(2)} - A_{\$}^{(1)} - \frac{1}{2}\mathbf{B}_{\$}^{(1)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_{\$}^{(1)} - A_{\$}^{(1)} & \mathbf{B}_{\$}^{(2)\prime} - \mathbf{B}_{\$}^{(1)\prime} \\ \vdots & \vdots \\ A_{\$}^{(n)} - A_{\$}^{(n-1)} - \frac{1}{2}\mathbf{B}_{\$}^{(n-1)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_{\$}^{(n-1)} - A_{\$}^{(1)} & \mathbf{B}_{\$}^{(n)\prime} - \mathbf{B}_{\$}^{(1)\prime} \\ \vdots & \vdots \\ A_{\$}^{(N)} - A_{\$}^{(N-1)} - \frac{1}{2}\mathbf{B}_{\$}^{(N-1)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_{\$}^{(N-1)} - A_{\$}^{(1)} & \mathbf{B}_{\$}^{(n)\prime} - \mathbf{B}_{\$}^{(1)\prime} \end{pmatrix},$$

$$\mathbf{X}_{\$}(\boldsymbol{\pi}) = \begin{pmatrix} -1 & \mathbf{0} & 0 \\ 0 & \mathbf{0} & \mathbf{B}_{\$}^{(1)\prime} \\ \vdots & \vdots & \vdots \\ 0 & \mathbf{0} & \mathbf{B}_{\$}^{(n-1)\prime} \\ \vdots & \vdots & \vdots \\ 0 & \mathbf{0} & \mathbf{B}_{\$}^{(N-1)\prime} \end{pmatrix},$$

for the numeraire country and

$$\mathbf{Y}_j(\boldsymbol{\pi}) = \begin{pmatrix} A_j^{(1)} & \mathbf{B}_j^{(1)\prime} \\ A_j^{(2)} - A_j^{(1)} - \mathbf{B}_j^{(1)\prime}\boldsymbol{\Sigma}_{bs}\mathbf{e}_j - \frac{1}{2}\mathbf{B}_j^{(1)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_j^{(1)} - A_j^{(1)} & \mathbf{B}_\$^{(2)\prime} - \mathbf{B}_\$^{(1)\prime} \\ \vdots & \vdots \\ A_j^{(n)} - A_j^{(n-1)} - \mathbf{B}_j^{(n-1)\prime}\boldsymbol{\Sigma}_{bs}\mathbf{e}_j - \frac{1}{2}\mathbf{B}_j^{(n-1)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_j^{(n-1)} - A_j^{(1)} & \mathbf{B}_\$^{(n)\prime} - \mathbf{B}_\$^{(1)\prime} \\ \vdots & \vdots \\ A_j^{(N)} - A_j^{(N-1)} - \mathbf{B}_j^{(N-1)\prime}\boldsymbol{\Sigma}_{bs}\mathbf{e}_j - \frac{1}{2}\mathbf{B}_j^{(N-1)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_j^{(N-1)} - A_j^{(1)} & \mathbf{B}_\$^{(N)\prime} - \mathbf{B}_\$^{(1)\prime} \end{pmatrix},$$

$$\mathbf{X}_j(\boldsymbol{\pi}) = \begin{pmatrix} 0 & -\mathbf{e}_j' & 0 \\ 0 & \mathbf{0} & \mathbf{B}_j^{(1)\prime} \\ \vdots & \vdots & \vdots \\ 0 & \mathbf{0} & \mathbf{B}_j^{(n-1)\prime} \\ \vdots & \vdots & \vdots \\ 0 & \mathbf{0} & \mathbf{B}_j^{(N-1)\prime} \end{pmatrix},$$

for the rest of the countries, i.e., $j = 1, ..., J$.

Then, vectorizing equation (21) and adding the set of identities $\boldsymbol{\theta}_2 = \boldsymbol{\pi}_2$ and $\boldsymbol{\theta}_3 = \boldsymbol{\pi}_3$, we arrive at the following expression for $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})$:

$$\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \boldsymbol{\gamma}(\boldsymbol{\pi}) - \boldsymbol{\Gamma}(\boldsymbol{\pi})\boldsymbol{\theta}, \tag{22}$$

where $\boldsymbol{\gamma}(\boldsymbol{\pi}) = \left\{vec\left[\mathbf{Y}(\boldsymbol{\pi})'\right]', \boldsymbol{\pi}_2', \boldsymbol{\pi}_3'\right\}$ and

$$\boldsymbol{\Gamma}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{X}(\boldsymbol{\pi}) \otimes \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

Thus, we have that, in total, there are $G = \mathcal{N} \times (M+1) + (M+J) \times (M+J+1) + (M+J) \times (M+J+1)/2$ distance functions. Further, we have that the number of distance functions is equal to the number of auxiliary parameters, that is, $G = H$.

Specializing equation (18) to the case of the distance functions given by equation (22) and an identity weighting matrix, $\mathbf{W}_T = \mathbf{I}$, we obtain the following OLS estimator:

$$\widehat{\boldsymbol{\theta}}_{OLS} = \left(\widehat{\boldsymbol{\Gamma}}'\widehat{\boldsymbol{\Gamma}}\right)^{-1}\left(\widehat{\boldsymbol{\Gamma}}'\widehat{\boldsymbol{\gamma}}\right), \tag{23}$$

where $\widehat{\boldsymbol{\gamma}} \equiv \boldsymbol{\gamma}(\widehat{\boldsymbol{\pi}})$ and $\widehat{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}(\widehat{\boldsymbol{\pi}})$. Asymptotic standard errors for $\widehat{\boldsymbol{\theta}}_{OLS}$ can be obtained by specializing equation (19) to the case of $\mathbf{W} = \mathbf{I}$ and $\partial\mathbf{g}/\partial\boldsymbol{\theta}' = -\boldsymbol{\Gamma}(\boldsymbol{\pi}^0)$.

Note, however, that $\widehat{\boldsymbol{\theta}}_{OLS}$ does not deliver a self-consistent model in the sense that the model-implied yields will not reproduce the bond pricing factors. In other words, one should guarantee that, when choosing state variables that are linear combinations (portfolios) of the yields, $\mathbf{f}_t = \mathbf{P}'\mathbf{y}_t^o$, the state variables that come out of the model need to be the same as the state variables that we started with (Cochrane and Piazzesi, 2005). Therefore, it is necessary to ensure that the pricing of portfolios of yields in equations (9) and (13) is consistent with $\mathbf{x}_{b,t} = \mathbf{P}'\mathbf{y}_t = \mathbf{P}'\mathbf{a}(\boldsymbol{\theta}) + \mathbf{P}'\mathbf{b}(\boldsymbol{\theta})\mathbf{x}_{b,t}$, which amounts to imposing the following set of constraints when estimating the model:

$$\mathbf{P}'\mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}, \qquad \mathbf{P}'\mathbf{b}(\boldsymbol{\theta}) = \mathbf{I}. \tag{24}$$

Let $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ denote the set of $S = M \times (M+1)$ self-consistency restrictions implicit in equation (24). We analyze the implications of these restrictions for the optimality of our estimator in the next section.

3.3 Optimal asymptotic least squares of GDTSMs

As in the case of generalized method of moments (GMM) estimation, an identity weighting matrix is not necessarily optimal and (asymptotic) efficiency gains can be achieved by selecting an appropriate weighting matrix. In particular, GMT show that when $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\pi}'}\mathbf{V}_{\boldsymbol{\pi}}\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\pi}}$ and $\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\theta}}\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\pi}'}\mathbf{V}_{\boldsymbol{\pi}}\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\pi}}\right)^{-1}\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}$ are non-singular when evaluated at $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$ (which implies that the rank of $\partial\mathbf{g}/\partial\boldsymbol{\pi}' = G$ and that $G \leq H$), then an optimal estimator exists. Such an estimator is optimal in the sense that the difference between the asymptotic variance of the resulting ALS estimator and another ALS estimator based on any other quadratic form in the same distance function is negative semidefinite. In particular, the optimal ALS estimator corresponds to the choice of a weighting matrix $\mathbf{W}_T$ that converges to $\mathbf{W} = \left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\pi}'}\mathbf{V}_{\boldsymbol{\pi}}\frac{\partial \mathbf{g}'}{\partial \boldsymbol{\pi}}\right)^{-1}$. Note that, by the delta method, we have that $\mathbf{V}_g(\boldsymbol{\theta}^0) = avar\left[\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)\right] = \left[\frac{\partial \mathbf{g}(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)}{\partial \boldsymbol{\pi}'}\mathbf{V}_{\boldsymbol{\pi}}(\boldsymbol{\pi}^0)\frac{\partial \mathbf{g}(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)}{\partial \boldsymbol{\pi}}\right]^{-1}$, so the optimal weighting matrix is simply the inverse of the asymptotic covariance of the distance function. Similarly, given that $\mathbf{r}(\boldsymbol{\theta}_0) = \mathbf{0}$, one would expect efficiency gains by imposing the self-consistency restrictions in (24) when estimating the parameters of interest. Therefore, optimal ALS estimation should, in principle, involve both choosing an optimal weighting matrix and simultaneously imposing the self-consistency constraints when estimating the model.

However, the self-consistency restrictions combined with the assumption that the bond state variables are observed perfectly imply that $\boldsymbol{\Omega}$, the covariance of the measurement errors in equation (16) is singular. In particular, note that $\boldsymbol{\Omega}$ appears in the expression of the asymptotic covariance matrix of the estimator of $\widehat{\boldsymbol{\pi}}_1$ in equation (20). Thus, the reduced rank structure in $\boldsymbol{\Omega}$ translates into a reduced-rank structure in $\mathbf{V}_\pi$, which can be seen by the fact that the OLS estimates of the reduced-form coefficients automatically satisfy the set of self-consistent restrictions:

$$\mathbf{P}'\widehat{\mathbf{a}} = \mathbf{0}, \qquad \mathbf{P}'\widehat{\mathbf{b}} = \mathbf{I}. \tag{25}$$

More important, given that $\partial\mathbf{g}/\partial\boldsymbol{\pi}'$ is a non-singular $H \times H$ matrix, the singularity in $\mathbf{V}_\pi$ also carries over to $\mathbf{V}_g$.

To overcome this problem, we follow Peñaranda and Sentana (2012), who study the problem of obtaining an optimal GMM estimator when the asymptotic variance of the moment conditions is singular in the population. Specifically, we (i) replace the ordinary inverse of $\mathbf{V}_g(\boldsymbol{\theta}^0)$ by any of its generalized inverses $\widehat{\mathbf{V}}_g^{+}(\boldsymbol{\theta}^0)$ and, (ii) simultaneously, impose the self-consistency restrictions in equation (24) when estimating the model.

In order to provide intuition on the optimality of this approach (see Diez de los Rios, 2015b, for a formal proof), let the spectral decomposition of $\mathbf{V}_g(\boldsymbol{\theta}^0)$ be written as

$$\mathbf{V}_g(\boldsymbol{\theta}^0) = \begin{pmatrix} \mathbf{T}_1 & \mathbf{T}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_2' \end{pmatrix} = \mathbf{T}_1\boldsymbol{\Lambda}\mathbf{T}_1',$$

where $\boldsymbol{\Lambda}$ is a $(G-S) \times (G-S)$ positive definite diagonal matrix. Therefore, we can split our set of distance functions into two groups: (i) the set of $K-S$ distance functions $\mathbf{T}_1'\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})$ whose asymptotic long-run variance is the non-singular matrix $\boldsymbol{\Lambda}$, and (ii) the set of degenerate S distance functions $\mathbf{T}_2'\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})$ that converge in mean square to zero due to the fact that the set of parameters of interest satisfy the self-consistent restrictions $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$.

Focus now, for convenience and without loss of generality, on the Moore-Penrose generalized inverse of $\mathbf{V}_g(\boldsymbol{\theta}^0)$, such that

$$\mathbf{V}_g^{MP+}(\boldsymbol{\theta}^0) = \mathbf{T}_1\boldsymbol{\Lambda}^{-1}\mathbf{T}_1'.$$

Then, the optimal ALS estimator in this singular setup is equivalent to the constrained ALS estimator that works with the reduced set of $K-S$ distance functions $\mathbf{T}_1'\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})$ and the restrictions $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$. However, note that the ALS estimator that uses the generalized inverse of $\mathbf{V}_g(\boldsymbol{\theta}^0)$ alone without the self-consistency restrictions will not likely be optimal, since it drops the S asymptotically degenerate, i.e., most informative, linear combinations of $\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})$. In fact, it might even be the case that $\boldsymbol{\theta}$ is not identified from the set of reduced implicit relations $\mathbf{T}_1'\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})$. This will occur, for example, if $K > G - S$.

Consequently, we have that the optimal estimator of the parameters of interest is

$$\widehat{\boldsymbol{\theta}}_{CGLS} = \arg\min_{\boldsymbol{\theta}} T\left[\boldsymbol{\gamma}(\widehat{\boldsymbol{\pi}}) - \boldsymbol{\Gamma}(\widehat{\boldsymbol{\pi}})\boldsymbol{\theta}\right]' \widehat{\mathbf{V}}_g^{+} \left[\boldsymbol{\gamma}(\widehat{\boldsymbol{\pi}}) - \boldsymbol{\Gamma}(\widehat{\boldsymbol{\pi}})\boldsymbol{\theta}\right] \qquad \text{s.t. } \mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}, \tag{26}$$

where, by stacking and vectorizing (24), we have that $\mathbf{r}(\boldsymbol{\theta}) = vec\left(\mathbf{P}' \otimes \mathbf{I}\right)\mathbf{p}_1(\boldsymbol{\theta}) - \bar{\mathbf{r}}_1$, with $\mathbf{p}_1(\boldsymbol{\theta}) = vec\left\{\left[\mathbf{a}(\boldsymbol{\theta})\ \mathbf{b}(\boldsymbol{\theta})\right]'\right\}$, and $\bar{\mathbf{r}}_1 = vec\left[(\mathbf{0}\ \mathbf{I})'\right]$. We refer to this (optimal) estimator as the constrained generalized least squares (CGLS) estimator. The asymptotic distribution of this estimator is given by:

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{CGLS} - \boldsymbol{\theta}^0) \xrightarrow{d} N\left[\mathbf{0}, \mathbf{J}^{-1} - \mathbf{J}^{-1}\frac{\partial \mathbf{r}'}{\partial \boldsymbol{\theta}}\left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}'}\mathbf{J}^{-1}\frac{\partial \mathbf{r}'}{\partial \boldsymbol{\theta}}\right)^{-1}\frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}'}\mathbf{J}^{-1}\right], \tag{27}$$

where $\mathbf{J} = \boldsymbol{\Gamma}'\mathbf{V}_g^{+}\boldsymbol{\Gamma}$ and $\partial \mathbf{r}/\partial \boldsymbol{\theta}'$ are both evaluated at $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$ (see chapter 10 in Gourieroux and Monfort, 1995). Further, as in the case of GMM, the optimized value of the ALS criterion function has an asymptotic χ^2 distribution with degrees of freedom equal to the number of overidentifying restrictions $(G - K)$.

Unfortunately, the solution to the optimal ALS (i.e., the CGLS) estimator in equation (26), $\widehat{\boldsymbol{\theta}}_{CGLS}$, is not known in closed form because $\mathbf{r}(\boldsymbol{\theta})$ is not linear in the set of parameters of interest, $\boldsymbol{\theta}$. Still, as noted by Newey and McFadden (1994) and Gourieroux and Monfort (1995) among others, estimating the model subject to a linearized version of the constraint (around a consistent estimate of $\boldsymbol{\theta}$) delivers an estimator that is asymptotically equivalent to the one that uses the non-linear constraint.

For this reason, we focus instead on the (feasible) linearized constrained GLS estimator, $\widetilde{\boldsymbol{\theta}}_{LCGLS}$, defined as:

$$\widetilde{\boldsymbol{\theta}}_{LCGLS} = \arg\min_{\boldsymbol{\theta}} T\left[\boldsymbol{\gamma}(\widehat{\boldsymbol{\pi}}) - \boldsymbol{\Gamma}(\widehat{\boldsymbol{\pi}})\boldsymbol{\theta}\right]' \widehat{\mathbf{V}}_g^{+} \left[\boldsymbol{\gamma}(\widehat{\boldsymbol{\pi}}) - \boldsymbol{\Gamma}(\widehat{\boldsymbol{\pi}})\boldsymbol{\theta}\right], \tag{28}$$
$$\text{s.t. } \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS}) = \frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS})}{\partial \boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_{OLS} - \boldsymbol{\theta}),$$

where, as a difference with Diez de los Rios (2015a), the constraint $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ has been linearized around the unconstrained OLS estimate of $\boldsymbol{\theta}$ defined above in equation (23). The main advantage of such linearization is that, since the objective function is quadratic and the restrictions are now linear in the parameters of interest, the solution of the estimation problem is known in closed form:

$$\widetilde{\boldsymbol{\theta}}_{LCGLS} = \widehat{\boldsymbol{\theta}}_{GLS} - \widehat{\mathbf{J}}^{-1}\frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS})'}{\partial \boldsymbol{\theta}}\left(\frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS})}{\partial \boldsymbol{\theta}'}\widehat{\mathbf{J}}^{-1}\frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS})'}{\partial \boldsymbol{\theta}}\right)^{-1}\mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS}), \tag{29}$$

where $\widehat{\boldsymbol{\theta}}_{GLS} = \left(\widehat{\boldsymbol{\Gamma}}'\widehat{\mathbf{V}}_g^{+}\widehat{\boldsymbol{\Gamma}}\right)^{-1}\left(\widehat{\boldsymbol{\Gamma}}'\widehat{\mathbf{V}}_g^{+}\widehat{\boldsymbol{\gamma}}\right)$ is the (suboptimal) ALS estimator that uses a consistent estimate of the generalized inverse of $\mathbf{V}_g(\boldsymbol{\theta})$ as weighting matrix, but that does not impose the restrictions $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$, and $\widehat{\mathbf{J}} = \left(\widehat{\boldsymbol{\Gamma}}'\widehat{\mathbf{V}}_g^{+}\widehat{\boldsymbol{\Gamma}}\right)$.

However, $\widetilde{\boldsymbol{\theta}}_{LCGLS}$ still does not satisfy the constraint $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ exactly, even though $\widetilde{\boldsymbol{\theta}}_{LCGLS}$ is asymptotically equivalent to the estimator that uses the non-linear constraint. This is why we follow Bekaert and Hodrick (2001) in iterating equation (29) when constructing our constrained estimates. Specifically, we start by obtaining a first restricted estimate of $\boldsymbol{\theta}$ using equation (29) and linearizing the constraint $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ around $\widehat{\boldsymbol{\theta}}_{OLS}$. Denote this first restricted estimate $\widetilde{\boldsymbol{\theta}}_{LCGLS}^{(1)}$. Then, we obtain a second restricted estimate, $\widetilde{\boldsymbol{\theta}}_{LCGLS}^{(2)}$, by linearizing $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ around $\widetilde{\boldsymbol{\theta}}_{LCGLS}^{(1)}$. We repeat this process until the resulting constrained estimate satisfies the self-consistency restrictions, $\mathbf{r}(\widetilde{\boldsymbol{\theta}}_{LCGLS}^{(n)}) = \mathbf{0}$ within a given tolerance.

While the results in Diez de los Rios (2015a) suggest that only a few iterations of equation (29) might be required for this estimator to converge, Golinski and Spencer (2017) have recently noted that this estimator tends to diverge when the number of bond pricing factors is larger than three. This occurs because the GLS estimator, $\widehat{\boldsymbol{\theta}}_{GLS}$, by using the generalized inverse of $\mathbf{V}_g(\boldsymbol{\theta}^0)$ alone without the self-consistency restrictions, drops the S most informative linear combinations of $\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})$, and therefore there might not be not enough information on the reduced set of $K - S$ distance functions $\mathbf{T}_1'\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})$ to identify $\boldsymbol{\theta}$. This renders $\widehat{\boldsymbol{\theta}}_{GLS}$ numerically unstable and the algorithm to compute $\widetilde{\boldsymbol{\theta}}_{LCGLS}$ to diverge. This is a problem because the number of factors needed to adequately capture the cross-sectional variability of yields in more than one country is usually larger than three. In the appendix, we provide an alternative way of solving (28) that avoids this issue and allows us to estimate multi-country models with a large number of bond pricing factors. Specifically, our new method directly imposes the self-consistent restrictions implicit in $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ by reparameterizing the model in terms of $K - S$ free parameters and linearizing $\mathbf{r}(\boldsymbol{\theta})$ around $\widehat{\boldsymbol{\theta}}_{OLS}$.[7]

[7] The reader is referred to Diez de los Rios (2015a) for a discussion of several extensions of this regression framework, including (i) the estimation subject to equality constraints, (ii) the existence of unspanned macro risks, (iii) how to deal with situations where only a subset of bonds is available, and (iv) how to compute small-sample standard errors and implement bias corrections.

4 Relationship with maximum likelihood estimation

In this section, we now discuss the relationship of our ALS estimator to the ML approach. However, as a difference with the literature on the ML estimation of one-country GDTSMs, where the canonical representation of Joslin, Singleton and Zhu (2011) has substantially lessened many of the numerical challenges faced by researchers, there is no accepted canonical representation for multi-country models. For this reason, we start by deriving a canonical version of a multi-country GDTSM by adapting the methodology of Joslin, Singleton and Zhu (2001) to the international setup.

4.1 The canonical model

As noted in the previous sebsection, self-consistency of the model implies that not all the parameters of the generic representation of a multi-country GDTSM are free. For this reason, we now focus on providing normalizations for the general representation outlined above that ensure that the model-implied yields reproduce the bond pricing factors, $\mathbf{x}_{b,t}$.[8]

In particular, we follow Dai and Singleton (2000) and JSZ in employing the affine transformations of the state variables outlined in Appendix C to show that our generic representation of a multi-country term structure model above is observationally equivalent to a canonical model with latent state variables and restrictions on both the parameters that govern the dynamic evolution of the state variables under the risk-neutral measure and the loadings of the short rates across the different countries. We collect such a result in Lemma 1.

Lemma 1 *The generic representation of a multi-country term structure model in equations (1), (2), and (3) is observationally equivalent to a model where:* ***(i)*** *the short rates are linear in a set of latent "bond" factors* $\mathbf{z}_t$

$$\begin{pmatrix} r_{\$,t} \\ r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{J,t} \end{pmatrix} = \begin{pmatrix} r^{\mathbb{Q}}_{\$,\infty} \\ r^{\mathbb{Q}}_{1,\infty} \\ r^{\mathbb{Q}}_{2,\infty} \\ \vdots \\ r^{\mathbb{Q}}_{J,\infty} \end{pmatrix} + \begin{pmatrix} 1-\sum_{j=1}^{J}\gamma_{j,1} & 1-\sum_{j=1}^{J}\gamma_{j,2} & \cdots & 1-\sum_{j=1}^{J}\gamma_{j,F} \\ \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,F} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,F} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{J,1} & \gamma_{J,2} & \cdots & \gamma_{J,F} \end{pmatrix} \begin{pmatrix} z_{b,1,t} \\ z_{b,2,t} \\ \vdots \\ z_{b,F,t} \end{pmatrix}, \quad (30)$$

$$\mathbf{r}_t = \mathbf{r}^{\mathbb{Q}}_{\infty} + \mathbf{\Gamma}^{(b)}\mathbf{z}_{b,t},$$

[8]The results in this subsection originally appeared in Bauer and Diez de los Rios (2012).

where $\mathbf{r}_\infty^{\mathbb{Q}} = (r_{\$,\infty}^{\mathbb{Q}}, r_{1,\infty}^{\mathbb{Q}}, ..., r_{J,\infty}^{\mathbb{Q}})'$ *and* $\mathbf{\Gamma}^{(b)}$ *is a matrix that stacks the short-rate loadings on each of the factors and satisfies that the sum of each of the columns of* $\mathbf{\Gamma}^{(b)}$ *is equal to one;* ***(ii)*** *the joint dynamic evolution of the latent bond factors, and exchange rates,* $\mathbf{z}_t = (\mathbf{z}_{b,t}', \Delta\mathbf{s}_t')'$, *under the risk-neutral measure is given by the following VAR(1) process:*

$$\begin{pmatrix} \mathbf{z}_{b,t+1} \\ \Delta\mathbf{s}_{t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\theta}_s^{\mathbb{Q}} \end{pmatrix} + \begin{pmatrix} \mathbf{\Psi}_{bb}^{\mathbb{Q}} & \mathbf{0} \\ \mathbf{\Psi}_{sb}^{\mathbb{Q}} & \mathbf{\Psi}_{ss}^{\mathbb{Q}} \end{pmatrix} \begin{pmatrix} \mathbf{z}_{b,t} \\ \Delta\mathbf{s}_t \end{pmatrix} + \begin{pmatrix} \mathbf{u}_{b,t+1}^{\mathbb{Q}} \\ \mathbf{u}_{s,t+1}^{\mathbb{Q}} \end{pmatrix}, \tag{31}$$

which can be represented in compact form as $\mathbf{z}_{t+1} = \boldsymbol{\theta}^{\mathbb{Q}} + \mathbf{\Psi}^{\mathbb{Q}}\mathbf{z}_t + \mathbf{u}_{t+1}^{\mathbb{Q}}$, *where* $\mathbf{u}_t^{\mathbb{Q}} \sim$ *iid* $N(0, \mathbf{\Omega})$, *the matrix* $\mathbf{\Psi}_{bb}^{\mathbb{Q}}$ *is in ordered real Jordan form with relevant elements (i.e., eigenvalues) collected in the vector* $\boldsymbol{\psi}$, *and* $\boldsymbol{\theta}_s^{\mathbb{Q}}$ *and* $\mathbf{\Psi}_{s\bullet}^{\mathbb{Q}}$ *satisfy restrictions analogous to (5) and (6) which guarantee that uncovered interest parity holds under the risk-neutral measure; and* ***(iii)*** $\mathbf{z}_t$ *follows an unrestricted VAR(1) process under the historical measure:* $\mathbf{z}_{t+1} = \boldsymbol{\theta} + \mathbf{\Psi}\mathbf{z}_t + \mathbf{u}_{t+1}$, *where* $\mathbf{u}_t \sim$ *iid* $N(0, \mathbf{\Omega})$.

Proof. See Appendix D. ■

Remark 1 *When the eigenvalues in* $\mathbf{\Psi}_{bb}^{\mathbb{Q}}$ *are real and distinct,* $\mathbf{\Psi}_{bb}^{\mathbb{Q}}$ *is a diagonal matrix. Furthermore, as noted by Hamilton and Wu (2012), in such a case the elements of* $\mathbf{\Psi}_{bb}^{\mathbb{Q}}$ *have to be in descending order,* $\psi_{bb,1}^{\mathbb{Q}} > \psi_{bb,2}^{\mathbb{Q}} > \ldots \psi_{bb,F}^{\mathbb{Q}}$, *in order to have a globally identified structure.*

Remark 2 *Note that we could have alternatively normalized* $\mathbf{\Gamma}^{(1)}$ *such that the loadings of the U.S. short rate on the factors are all equal to one, which would then resemble the JSZ normalization for the domestic setup. However, such an approach is not maximal given that it does not allow the existence of (country-specific) factors that could drive the term structure of some of the countries without affecting the U.S. yield curve.*

Remark 3 *The representation in Lemma 1 nests the models proposed by Graveline and Joslin (2011) and Jotikasthira, Le and Lundblad (2015) in which the* j*th economy's short rate is driven by local factors (i.e.,* $r_{j,t} = r_{j,\infty}^{\mathbb{Q}} + \mathbf{1}'\mathbf{z}_{b,t}^{(j)}$ *where* $\mathbf{1}$ *is a conformable vector of ones and* $\mathbf{z}_{b,t}^{(j)}$ *collects country* j*'s local factors) under appropriate zero restrictions on* $\mathbf{\Gamma}^{(1)}$.

Remark 4 *Global and country-specific factors can be accomodated in our setup by imposing appropriate zero restrictions on* $\mathbf{\Gamma}^{(b)}$ *and* $\mathbf{\Omega}_{bb}$ *so that the correlation between yields in two different countries is driven only by the global factors.*

Note, now, that the canonical model in Lemma 1 implies that yields on domestic and foreign zero-coupon bonds are affine in $\mathbf{z}_{b,t}$:

$$\mathbf{y}_t = \mathbf{a}_z + \mathbf{b}_z \mathbf{z}_{b,t}. \tag{32}$$

Thus, state variables that are linear combinations of the yields can simply be understood as invariant (affine) transformations of the latent factors $\mathbf{z}_{b,t}$:

$$\mathbf{x}_{b,t} = \mathbf{P}'\mathbf{y}_t = \mathbf{P}'(\mathbf{a}_z + \mathbf{b}_z \mathbf{z}_{b,t}) = \mathbf{c} + \mathbf{D}\mathbf{z}_{b,t},$$

which we can exploit to show the restrictions that parameters of the generic representation of the multi-country GDTSM above need to satisfy to be self-consistent.

Proposition 2 *The multi-country term structure model given by equations (2), (1) and (3), with state variables that are linear combinations of yields,* $\mathbf{x}_{b,t} = \mathbf{P}'\mathbf{y}_t$*, is self-consistent when*

$$\begin{aligned}
\boldsymbol{\Delta}^{(b)} &= \boldsymbol{\Gamma}^{(b)}\mathbf{D}^{-1}, \\
\boldsymbol{\Delta}^{(0)} &= \mathbf{r}_{\infty}^{\mathbb{Q}} - \boldsymbol{\Delta}^{(b)}\mathbf{c}, \\
\boldsymbol{\Phi}_{bb}^{\mathbb{Q}} &= \mathbf{D}\boldsymbol{\Psi}_{bb}^{\mathbb{Q}}\mathbf{D}^{-1}, \\
\boldsymbol{\mu}_{b}^{\mathbb{Q}} &= (\mathbf{I} - \boldsymbol{\Phi}_{bb}^{\mathbb{Q}})\mathbf{c},
\end{aligned}$$

where $\mathbf{c} = \mathbf{P}'\mathbf{a}_z$*,* $\mathbf{D} = \mathbf{P}'\mathbf{b}_z$ *and* $\mathbf{a}_z$*,* $\mathbf{b}_z$ *are implicitly defined in equation (32). The parameters under the physical measure remain unrestricted.*

Note that, as a result, the risk-neutral dynamics of the yield curve (and therefore, the cross-section of interest rates) is entirely determined by (a) $\mathbf{r}_{\infty}^{\mathbb{Q}}$, the long-run mean of the short rates under $\mathbb{Q}$; (b) the free elements in $\boldsymbol{\Gamma}^{(b)}$, i.e., the factor loadings, (c) $\boldsymbol{\psi}$, the speed of mean reversion of the state variables under $\mathbb{Q}$; and (d) $\boldsymbol{\Sigma}$, the covariance matrix of the innovations from the VAR. On the other hand, the VAR dynamics under $\mathbb{P}$ remain unrestricted.

Given this separation between risk-neutral and physical dynamics, and given the fact that the VAR dynamics remain unrestricted, one could use a two-step estimator similar to the one proposed by JSZ. In the first step, one would estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$ by OLS given that, since the VAR dynamics are unrestricted, OLS recovers the estimates of the conditional mean (Zellner, 1962). In the second step, one would estimate the remaining parameters of the model ($r^{\mathbb{Q}}_{\infty}$, $\mathbf{\Gamma}^{(b)}$, $\boldsymbol{\psi}$, $\mathbf{\Sigma}$) via numerical maximization of the likelihood function, taking as given the $\mathbb{P}$-dynamics estimates obtained in the first step.

Note, however, that such an ML estimator still implies a numerical search over a very large dimensional space when either the number of countries or the number of factors is moderately large. For example, in the case of a seven-country and 10-factor model, as in our empirical illustration below, the number of parameters is 213 (7 for $r^{\mathbb{Q}}_{\infty}$, 60 for $\mathbf{\Gamma}^{(b)}$, 10 for $\boldsymbol{\psi}$, and 136 for $\mathbf{\Sigma}$).[9] This renders the ML estimation un-implementable in such cases, leaving the LCGLS estimator proposed above as the only reliable alternative for the estimation of international term structure models with a large number of countries.[10]

4.2 Efficiency considerations

More importantly, it is possible to prove that the LCGLS estimates are asymptotically equivalent to MLE. In the standard case, Kodde, Palm and Pfann (1990) present the conditions under which the optimal ALS estimator is equivalent to the ML estimator. In particular, these authors note that if (i) the system of relationships $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ is complete, i.e., $G = H$ and the Jacobian $\partial\mathbf{g}/\partial\boldsymbol{\theta}'$ has full rank; and (ii) $\boldsymbol{\pi}$ is estimated by ML, or a method asymptotically equivalent to ML, then the optimal ALS estimator is asymptotically equivalent to the ML estimator of $\boldsymbol{\theta}$.

Diez de los Rios (2015b) extends the results in Kodde, Palm and Pfann (1990) to the case of optimal ALS estimation in a singular setup. In such a case, the optimal ALS estimator is still asymptotically equivalent to the ML estimator as long as $\widehat{\boldsymbol{\pi}}$ is estimated

[9] In addition, such an approach requires the analysis of several different subcases depending on whether all the eigenvalues $\mathbf{\Psi}^{\mathbb{Q}}_{bb}$ are real and distinct, there are repeated eigenvalues or such eigenvalues are complex. On the other hand, one does not need to a priori determine whether the eigenvalues are real and distinct when estimating the model using our linear regression approach given that our method will, in practice, numerically determine which subcase is most empirically relevant.

[10] Specifically, should one be interested in the parameters of the canonical representation, these can be recovered from the LCGLS estimates in the following way. First, note from Proposition 2 that $\mathbf{\Psi}^{\mathbb{Q}}_{bb}$ is related to the Jordan decomposition of $\mathbf{\Phi}^{\mathbb{Q}}_{bb}$. Therefore, an estimate of $\mathbf{\Psi}^{\mathbb{Q}}_{bb}$ can be obtained by finding the real Jordan normal form of $\widehat{\mathbf{\Phi}}^{\mathbb{Q}}_{bb}$. In particular, when the eigenvalues in $\mathbf{\Psi}^{\mathbb{Q}}_{bb}$ are real and distinct, $\widehat{\boldsymbol{\psi}}^{\mathbb{Q}}$ can be obtained by a simple spectral decomposition of $\widehat{\mathbf{\Phi}}^{\mathbb{Q}}_{bb} = \widehat{\mathbf{D}} diag(\widehat{\boldsymbol{\psi}}^{\mathbb{Q}})\widehat{\mathbf{D}}^{-1}_{bb}$. Second, given the estimate of $\widehat{\mathbf{D}}$ obtained in the previous step, an estimate of $\mathbf{\Gamma}^{(b)}$ is obtained as follows $\widehat{\mathbf{\Gamma}}^{(b)} = \left[\widehat{\mathbf{\Delta}}^{(b)}\widehat{\mathbf{D}}\right] / diag\left[\mathbf{1}'_J\widehat{\mathbf{\Delta}}^{(b)}\widehat{\mathbf{D}}\right]$. Note that our estimate of $\widehat{\mathbf{\Gamma}}^{(b)}$ satisfies that the sum of each of its columns is equal to one. Third, an estimate of the long-run mean of the short rate under $\mathbb{Q}$ can be obtained from $\widehat{\mathbf{r}}^{\mathbb{Q}}_{\infty} = \widehat{\mathbf{\Delta}}^{(0)} + \widehat{\mathbf{\Delta}}^{(b)\prime}(\mathbf{I} - \widehat{\mathbf{\Phi}}^{\mathbb{Q}}_{bb})^{-1}\widehat{\boldsymbol{\mu}}^{\mathbb{Q}}_{b}$. Fourth, given the structure of the optimization problems in (23) and (28), the estimates of the $\mathbb{P}$-dynamics parameters of the state variables implied by our linear framework also coincide with the OLS estimates of the VAR model in equation (1). Finally, standard errors for the coefficients of the canonical representation can be obtained using the Delta method and the results in Magnus (1985) regarding differentiation of eigenvalues and eigenvectors.

by a method that is asymptotically equivalent to constrained ML (i.e., $\widehat{\boldsymbol{\pi}}$ satisfies the self-consistency restrictions $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$). We note that the (linearized) CGLS estimator satisfies these two conditions, and, therefore, it is equivalent to the ML estimator.

5 Empirical application

In this section, we use the CGLS estimation method outlined above to estimate a seven-country, 10-factor model and decompose 10-year zero coupon bond yields into an expectations and a term premium component. This decomposition allows us to analyze the covariation of the term premia across yield curves denominated in different currencies within a unified framework.

Our data set consists of end-of-quarter observations over the period March 1988 (1988Q1) to March 2009 (2009Q1) of the U.S. dollar bilateral exchange rates against the British pound, the German Mark/Euro, the Canadian dollar, the Australian dollar, the Swiss Franc, and the Japanese Yen, along with the appropriate zero-coupon yield curves for these countries. Specifically, we consider the full spectrum of maturities from one quarter to 10 years.[11]

It is well documented that three principal components (labelled level, slope and curvature) are sufficient to explain over 95 per cent of the variation in U.S. government bond yields (Litterman and Scheinkman, 1991). This stylized fact also holds individually in the four countries examined here (Table 1). Panel A reports the variation in the levels of yields in each country explained by the first k principal components (PCs) from the cross-section of yields. In each country, three "domestic" PCs explain 99.9 per cent of the variation in the yield curve. In fact, given that we do not use data on the yields of bonds with maturities longer than 10 years, it can be argued that the seven domestic yield curves can be well approximated by only two PCs each (i.e., local level and slope) given that, in this case, two "domestic" PCs explain 99.8 per cent of their variation.

Applying a principal component analysis to the cross-section of global yields reveals, on the other hand, that more than 2 components are required to explain the cross-sectional

[11] Yield curve data are obtained from the Wright (2011) database, which consists of local currency zero-coupon government yield curves at the monthly (or higher) frequency for 10 industrialized countries. We drop New Zealand, Norway and Sweden from our empirical illustration, because for these countries, the data begin a bit later than March 1988. We choose to work with the 7 countries above as a trade-off between maximizing the sample size and keeping a balanced panel of yields. Exchange rate data are obtained from Bloomberg.

variation in the combined 40 interest rates. Panel B of Table 1 shows that 10 "global" PCs are needed to explain 99.8 per cent of the variation (the same amount as with two domestic PCs per country). This fact is confirmed by looking at the root-mean-squared pricing errors (RMSPE) from fitted values of a regression of the yield levels on k PCs, which are given in Panel C of Table 1. Two domestic PCs in each country deliver RMSPE close to 10 basis points in each of the four countries. To obtain a similar RMSPE we again need to use the first 10 global PCs. Against this backdrop, we use 10 PCs to capture the cross-sectional variation of our panel of international bond yields.

5.1 Fitting yields

Figure 1 presents both the estimated bond yield loadings implied by the affine term structure model, as well as the regression coefficients that one would obtain from projecting bond yields on the first 10 PCs (i.e., the loadings from a principal components analysis). The latter coefficients are from a linear factor model that minimizes the sum of the squared differences between model predictions and actual yields, and thus provide a natural benchmark to compare the pricing errors implied by our no-arbitrage model. Importantly, Figure 1 shows that the multi-country term structure model is flexible enough to replicate the shapes of the loadings on individual bond yields obtained from a principal component analysis.

We confirm the model's fit by providing RMSPE and mean-absolute pricing errors (MAPE) in Table 2. The column labelled "Affine" provides estimates of the goodness-of-fit measures for the affine term structure model; the column "Unrestricted" gives the results for an unrestricted regression of bond yields on the global PCs; while "Difference" characterizes the difference between the two quantities. The loss from imposing the no-arbitrage conditions is around 5 basis points at either the country or global level. While the loss is bigger than in one-country models (e.g., the loss in the Canadian yield curve illustration in Diez de los Rios (2015a) is less than one basis point), it is still economically small.[12]

In fact, we can use the fact that the minimized value of the ALS criterion function has an asymptotic χ^2 distribution to test the validity of the model. Specifically, we have that the dimensionality of the distance function is 3488 and the number of parameters

[12] While unreported for the sake of space, it is worth noting that OLS estimates of the no-arbitrage parameters do not deliver a good cross-sectional fit. Specifically, the loss from imposing the no-arbitrage conditions using the OLS estimates of the model is close to 17 bps.

of interest is 595. This leaves 2893 degrees of freedom. The 1% (5%) critical value for a $\chi^2(2893)$ is 3072.9 (3019.2), while the minimized value of the ALS criterion is 2202.6. Therefore, there is no evidence that the no-arbitrage restrictions imposed by the affine term structure model on the reduced-form model are inconsistent with the data.

5.2 Prices of risk

It is possible to show that the one-period expected excess return for holding an n-period bond is given by

$$E_t rx_{j,t+1}^{(n)} = E_t \left[\log \frac{P_{j,t+1}^{(n-1)}}{P_{j,t}^{(n)}} \right] - r_{j,t} = JIT + \mathbf{B}_j^{(n-1)\prime} (\boldsymbol{\lambda}_{b0} + \boldsymbol{\lambda}_{bb} \mathbf{x}_{b,t} + \boldsymbol{\lambda}_{bs} \Delta \mathbf{s}_t),$$

where JIT is a (constant) Jensen's inequality term and

$$\boldsymbol{\lambda}_{b0} = \boldsymbol{\mu}_b - \boldsymbol{\mu}_b^{\mathbb{Q}},$$
$$\boldsymbol{\lambda}_{bb} = \boldsymbol{\Phi}_{bb} - \boldsymbol{\Phi}_{bb}^{\mathbb{Q}},$$
$$\boldsymbol{\lambda}_{bs} = \boldsymbol{\Phi}_{bs}.$$

Thus, the risk premia on holding a bond for a period are linear in the state variables, $\mathbf{x}_t = (\mathbf{x}_{b,t}, \Delta \mathbf{s}_t)'$, and have three terms: (i) a Jensen's inequality term; (ii) a constant risk premium related to $\boldsymbol{\lambda}_{b0}$; and (iii) a time-varying risk-premium component where time variation is governed by the parameters in $\boldsymbol{\lambda}_b$ and $\boldsymbol{\lambda}_s$. Note that $\boldsymbol{\lambda}_{b,t} = \boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_b \mathbf{x}_{b,t} + \boldsymbol{\lambda}_s \Delta \mathbf{s}_t$ has the interpretation of the market price of bond risks, given that it captures how much expected bond holding returns must rise to compensate for exposure to the bond shocks, $\mathbf{v}_{b,t+1}$.In fact, when agents are risk neutral (i.e., $\boldsymbol{\mu}_b = \boldsymbol{\mu}_b^{\mathbb{Q}}$, $\boldsymbol{\Phi}_{bb} = \boldsymbol{\Phi}_{bb}^{\mathbb{Q}}$ and $\boldsymbol{\Phi}_{bs} = \boldsymbol{\Phi}_{bs}^{\mathbb{Q}} = \mathbf{0}$), we have that the market price of bond risk is equal to zero for all t.

Similarly, the one-period excess return earned by a domestic investor for holding a one-period zero-coupon bond from country j (i.e., the currency return) is:

$$E_t rs_{j,t+1}^{(n)} = E_t \left[\log \frac{S_{j,t+1}}{S_{j,t}} \right] + r_{j,t} - r_{\$,t} = JIT + \mathbf{e}_j' (\boldsymbol{\lambda}_{s0} + \boldsymbol{\lambda}_{sb} \mathbf{x}_{b,t} + \boldsymbol{\lambda}_{ss} \Delta \mathbf{s}_t),$$

where we have that

$$\mathbf{e}_j' \boldsymbol{\lambda}_{s0} = \mathbf{e}_j' \boldsymbol{\mu}_s + \delta_j^{(0)} - \delta_{\$}^{(0)},$$
$$\mathbf{e}_j' \boldsymbol{\lambda}_{sb} = \mathbf{e}_j' \boldsymbol{\Phi}_{sb} + \boldsymbol{\delta}_j^{(b)} - \boldsymbol{\delta}_{\$}^{(b)},$$
$$\mathbf{e}_j' \boldsymbol{\lambda}_{bs} = \mathbf{e}_j' \boldsymbol{\Phi}_{ss}.$$

Again, the currency risk premia are linear in the state variables, $\mathbf{x}_t = (\mathbf{x}_{b,t}, \Delta\mathbf{s}_t)'$, and have three terms: (i) a Jensen's inequality term; (ii) a constant risk premium; and (iii) a time-varying risk-premium component. As in the case of the bond prices of risk, we note that $\boldsymbol{\lambda}_{s,t} = \boldsymbol{\lambda}_{s0} + \boldsymbol{\lambda}_{sb}\mathbf{x}_{b,t} + \boldsymbol{\lambda}_{ss}\Delta\mathbf{s}_t$ has the interpretation of the market price of foreign exchange risks, given that it captures how much expected currency returns must rise to compensate for exposure to the currency shocks, $\mathbf{v}_{s,t+1}$. Finally, note that when agents are risk neutral, we have that the market price of foreign exchange rate risk is equal to zero for all t, and the uncovered interest parity hypothesis holds under both the physical and risk-neutral measures.

Table 3 presents Wald statistics for the hypothesis that the prices of risk are equal to zero (i.e., risk neutrality). Importantly, we cannot reject that neither of the bond factors are priced nor the exchange rate risks.[13]

5.3 Term premium estimates

In this section, we use the parameter estimates of our seven-country, 10-factor GDTSM to decompose long-term interest rates into expectations of future short-term rates and term premia. In particular,

$$y_{j,t}^{(n)} = \frac{1}{n}\sum_{h=1}^{n} E_t r_{j,t+h-1} + tp_{j,t}^{(n)}. \tag{33}$$

That is, the n-period interest rate at time t, $y_{j,t}^{(n)}$, is equal to the average path of the short-term rate over the following n periods and a risk-premium component, $tp_{j,t}^{(n)}$, usually called the term premium. This term premium is the expected return from holding an n-period bond to maturity while financing this investment by selling a sequence of one-period bonds.

Figure 2 plots the term premium on 10-year bond yields implied by our model for the seven countries in our sample. We find that the estimated term premium is countercyclical and rising during recessions (particularly during the early 1990s and 2000s). Figure 2 also

[13] When the dynamics of the state variables are left unrestricted, the estimates of $\mathbb{P}$-parameters coincide with the OLS estimates of a VAR(1) process for $\mathbf{f}_t$ and, therefore, suffer from the well-known problem that OLS estimates of autoregressive parameters tend to underestimate the persistence of the system in finite samples. For this reason, we replace the reduced-form OLS estimates of the VAR(1) equation in (1) with bias-corrected estimates as suggested by Bauer, Rudebusch and Wu (2012). As in Diez de los Rios (2015a), we use the analytical approximation for the mean bias in VARs presented in Pope (1990) with the adjustment suggested by Kilian (1998), in order to guarantee that the bias-corrected estimates are stationary.

shows that our term premia estimates for all the countries are highly correlated across countries. In fact, the first PC of the cross-section of term premia explain 75% of the variation in the cross-section of risk premia, while the first two PCs explain 92%. This might indicate that while one cannot statistically reject that all 10 factors are priced in the cross-section of interest rates, only 2 factors might be needed to explain most of the (economically interesting) variation in term premia. Interestingly, our finding that only 2 factors are priced in the cross-section of term premia is in line with the results in Duffee (2010) and Joslin Priebsch and Singleton (2014), while it differs from those in Cochrane and Piazzesi (2008), who find that only level risk is priced in the term structure of U.S. interest rates. However, we leave for further research understanding the drivers of these 2 term premia factors.

6 Final Remarks

In this paper, we extend the linear estimator of Diez de los Rios (2015a) to overcome the numerical challenges that plague multi-country term structure models. Specifically, we consider a novel linear regression approach to the estimation of multi-country Gaussian dynamic term structure models that can completely avoid numerical optimization methods whenever yields on adjacent maturities are directly observed, and that can be interpreted as an ALS estimator. Importantly, our estimator remains easy to compute and asymptotically efficient, even when the number of countries is relatively large: a situation in which other recently proposed approaches lose their tractability.

References

[1] Ang, A., and M. Piazzesi (2003): "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745-787.

[2] 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 Satff Working Paper No. 2012-5.

[3] Bauer, M.D. and J.D. Hamilton (2015): "Robust Risk Premia," Federal Reserve Bank of San Francisco Working Paper 2015-15.

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

[5] Bauer, M.D., G.D. Rudebusch and C. Wu (2012): "Correcting Estimation Bias in Dynamic Term Structure Models," *Journal of Business and Economic Statistics*, 30, 454-467.

[6] Bekaert, G. and R.J. Hodrick (2001): "Expectations Hypotheses Tests," *Journal of Finance*, 56, 4, 1357-1393.

[7] Brandt, M.W., and P. Santa-Clara (2002): "Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets," *Journal of Financial Economics* 63, 161-210.

[8] Cochrane, J. and M. Piazzesi (2005): "Bond Risk Premia," *American Economic Review*, 95, 138-60.

[9] Cochrane, J. and M. Piazzesi (2008): "Decomposing the Yield Curve," Mimeo, University of Chicago.

[10] Dai, Q. and K.J. Singleton (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943-1978.

[11] Dahlquist, M. and H. Hasseltoft (2011): "International Bond Risk Premia," *Journal of International Economics*, 90, 17-32.

[12] Diebold, F.X., C. Li, and V. Yue (2008): "Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach," *Journal of Econometrics* 146, 351-363.

[13] Diez de los Rios, A. (2015a): "A New Linear Estimator for Gaussian Dynamic Term Structure Models," *Journal of Business & Economic Statistics*, 33, 282-295.

[14] Diez de los Rios, A. (2015b): "Optimal Asymptotic Least Squares Estimation in a Singular Set-Up," *Economic Letters,* 128, 83-86.

[15] Duffee, G.R. (2010): "Sharpe Ratios in Term Structure Models," Mimeo, Johns Hopkins University.

[16] Duffee, G.R. (2011): "Information in (and not in) the Term Structure," *Review of Financial Studies* 24, 2895-2934.

[17] Duffee, G.R. and R. Stanton (2012): "Estimation of Dynamic Term Structure Models," *Quarterly Journal of Finance,* 2, 1-51.

[18] Golinski, A. and P.D. Spencer (2017): "Estimating the Term Structure with Linear Regressions: Getting to the Roots of the Problem," Mimeo, York University.

[19] Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models*, Cambridge University Press (Cambridge).

[20] Gourieroux, C., A. Monfort and A. Trognon (1982): "Nonlinear Asymptotic Least Squares," INSEE Document de travail no. 8207.

[21] Gourieroux, C., A. Monfort and A. Trognon (1985): "Moindres Carres Asymptotiques," Annales de l'INSEE 58, 91-122.

[22] Graveline, J. and S. Joslin (2011): "G10 Swap and Exchange Rates," MIT Mimeo.

[23] Hamilton, J.D. and J.C. Wu (2012): "Identification and Estimation of Gaussian Affine Term Structure Models," *Journal of Econometrics*, 168, 315-331.

[24] Jotikasthira C., A. Le and C. Lundblad (2010): "Why Do Term Structures in Different Currencies Comove?" *Journal of Financial Economics*, 2015, 115, 58-83.

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

[26] Joslin, S., K.J. Singleton and H. Zhu (2011): "A New Perspective on Gaussian DTSMs," *Review of Financial Studies,* 24, 926-970.

[27] Kodde, D.A., F.C. Palm and G.A. Pfann (1990): "Asymptotic Least-Squares Estimation Efficiency Considerations and Applications," *Journal of Applied Econometrics*, 5, 229-243.

[28] Litterman, R. and J.A. Scheinkman (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, June, 54-61.

[29] Lütkepohl, H. (1989): "A Note on the Asymptotic Distribution of Impulse Response Functions of Estimated VAR Models with Orthogonal Residuals," *Journal of Econometrics*, 42, 371-376.

[30] Magnus, J. (1985): "On Differentiating Eigenvalues and Eigenvectors," *Econometric Theory*, 1, 179-191.

[31] Meldrum, A., M. Razcko, and P. Spencer (2016): "Overseas Unspanned Factors and Domestic Bond Returns," Bank of England Staff Working Paper No. 618.

[32] Monfort, A. and F. Pegoraro (2012): "Asset Pricing with Second-Order Esscher Transforms," *Journal of Banking and Finance*, 1678-1687.

[33] Newey, W.K. and D.L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in R.F. Engle and D.L. McFadden (eds), *Handbook of Econometrics: Vol. 4*, Elsevier Science Press (Amsterdam), 2111-2245.

[34] Peñaranda, F. and E. Sentana (2012): "Spanning Tests in Return and Stochastic Discount Factor Mean-Variance Frontiers: a Unifying Approach," *Journal of Econometrics,* 170, 303-324.

[35] Sarno, L., P. Schneider and C. Wagner (2012): "Properties of Foreign Exchange Risk Premiums," *Journal of Financial Economics*, 105, 279-310.

[36] Sentana, E. (2002): "Did the EMS Reduce the Cost of Capital?" *Economic Journal*, 112, 786-809.

[37] Wright, J. H. (2011): "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset," *American Economic Review* 101, 1514-1534.

Appendix

A Bond Pricing

A.1 Domestic bonds

We start by assuming (to then verify that this guess is right) that the price of a U.S. zero-coupon bond of maturity n periods at time t is exponentially affine in the factors:

$$P_{\$,t}^{(n)} = \exp\left[A_{\$}^{(1)} + \mathbf{B}_{\$}^{(1)\prime}\mathbf{x}_{b,t}\right]. \tag{34}$$

Substituting (34) into (8) in the main text of the paper, we have that:

$$\begin{aligned} P_{\$,t}^{(n+1)} &= E_t^{\mathbb{Q}}\left[\exp\left(-r_{\$,t} + A_{\$}^{(n)} + \mathbf{B}_{\$}^{(n)\prime}\mathbf{x}_{b,t+1}\right)\right], \\ &= E_t^{\mathbb{Q}}\left\{\exp\left[A_{\$}^{(1)} + \mathbf{B}_{\$}^{(1)\prime}\mathbf{x}_{b,t} + A_{\$}^{(n)} + \mathbf{B}_{\$}^{(n)\prime}(\boldsymbol{\mu}_b^{\mathbb{Q}} + \boldsymbol{\Phi}_{bb}^{\mathbb{Q}}\mathbf{x}_{b,t} + \mathbf{v}_{b,t+1})\right]\right\}, \\ &= E_t^{\mathbb{Q}}\left\{\exp\left[A_{\$}^{(1)} + A_{\$}^{(n)} + \mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\mu}_b^{\mathbb{Q}} + \left(\mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\Phi}_{bb}^{\mathbb{Q}} + \mathbf{B}_{\$}^{(1)\prime}\right)\mathbf{x}_{b,t} + \mathbf{B}_{\$}^{(n)\prime}\mathbf{v}_{b,t+1}\right]\right\}. \end{aligned}$$

Note that the last term in the previous equation satisfies

$$E_t^{\mathbb{Q}}\left[\exp\left(\mathbf{B}_{\$}^{(n)\prime}\mathbf{v}_{b,t+1}\right)\right] = \exp\left[\frac{1}{2}\mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_{\$}^{(n)}\right].$$

Thus we have that

$$A_{\$}^{(n+1)} + \mathbf{B}_{\$}^{(n+1)\prime}\mathbf{x}_{b,t} = \left(A_{\$}^{(n)} + \mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\mu}_{bb}^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_{\$}^{(n)} + A_{\$}^{(1)}\right) + \left(\mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\Phi}_{bb}^{\mathbb{Q}} + \mathbf{B}_{\$}^{(1)\prime}\right)\mathbf{x}_{b,t}.$$

And matching coefficients we arrive at the following pricing recursions:

$$\mathbf{B}_{\$}^{(n+1)\prime} = \mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\Phi}_{bb}^{\mathbb{Q}} + \mathbf{B}_{\$}^{(1)\prime}, \tag{35}$$

$$A_{\$}^{(n+1)} = A_{\$}^{(n)} + \mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\mu}_b^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}_{\$}^{(n)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_{\$}^{(n)} + A_{\$}^{(1)}. \tag{36}$$

Furthermore, the recursion is started by exploiting the fact that the affine pricing relationship is trivially satisfied for domestic one-period bonds (i.e., $y_{\$,t}^{(1)} = r_{\$,t}$):

$$\log P_{\$,t}^{(1)} = -y_{\$,t}^{(1)} = -r_{\$,t} = -\delta_{\$}^{(0)} - \boldsymbol{\delta}_{\$}^{(1)\prime}\mathbf{x}_{b,t}.$$

In particular, matching coefficients, we have that $A_{\$}^{(1)} = -\delta_{\$}^{(0)}$, and $\mathbf{B}_{\$}^{(1)} = -\boldsymbol{\delta}_{\$}^{(b)}$.

A.2 Foreign bonds

In a similar fashion to the case of domestic bonds, we also start by assuming that the price of a country j bond of maturity n periods at time t is exponentially affine in the factors:

$$P_{j,t}^{(n)} = \exp\left[A_j^{(1)} + \mathbf{B}_j^{(1)\prime}\mathbf{x}_{b,t}\right]. \tag{37}$$

with $A_j^{(1)} = -\delta_j^{(0)}$, and $\mathbf{B}_j^{(1)} = -\boldsymbol{\delta}_j^{(b)}$ for one-period bonds.

Note that, substituting (37) into (12) in the main text of the paper, we have that:

$$
\begin{aligned}
P_{j,t}^{(n+1)} &= E_t^{\mathbb{Q}}\left[\exp\left(-r_{\$,t} + \Delta s_{j,t+1} + A_j^{(n)} + \mathbf{B}_j^{(n)\prime}\mathbf{x}_{b,t+1}\right)\right], \\
&= E_t^{\mathbb{Q}}\left\{\exp\left[A_{\$}^{(1)} + \mathbf{B}_{\$}^{(1)\prime}\mathbf{x}_{b,t} + \mathbf{e}_j'(\boldsymbol{\mu}_s^{\mathbb{Q}} + \boldsymbol{\Phi}_{sb}^{\mathbb{Q}}\mathbf{x}_{b,t} + \mathbf{v}_{s,t+1}) + \dots\right.\right. \\
&\quad \left.\left. + A_j^{(n)} + \mathbf{B}_j^{(n)\prime}(\boldsymbol{\mu}_b^{\mathbb{Q}} + \boldsymbol{\Phi}_{bb}^{\mathbb{Q}}\mathbf{x}_{b,t} + \mathbf{v}_{b,t+1})\right]\right\}, \\
&= E_t^{\mathbb{Q}}\left\{\exp\left[A_{\$}^{(1)} + \mathbf{e}_j'\boldsymbol{\mu}_s^{\mathbb{Q}} + A_j^{(n)} + \mathbf{B}_j^{(n)\prime}\boldsymbol{\mu}_b^{\mathbb{Q}} + \dots\right.\right. \\
&\quad \left.\left. + \left(\mathbf{B}_j^{(n)\prime}\boldsymbol{\Phi}_{bb}^{\mathbb{Q}} + \mathbf{B}_{\$}^{(1)\prime} + \mathbf{e}_j'\boldsymbol{\Phi}_{sb}^{\mathbb{Q}}\right)\mathbf{x}_{b,t} + \mathbf{B}_j^{(n)\prime}\mathbf{v}_{b,t+1} + \mathbf{e}_j'\mathbf{v}_{s,t+1}\right]\right\}, \\
&= E_t^{\mathbb{Q}}\left\{\exp\left[-\frac{1}{2}\mathbf{e}_j'\boldsymbol{\Sigma}_{ss}\mathbf{e}_j + A_j^{(1)} + A_j^{(n)} + \mathbf{B}_j^{(n)\prime}\boldsymbol{\mu}_b^{\mathbb{Q}} + \dots\right.\right. \\
&\quad \left.\left. + \left(\mathbf{B}_j^{(n)\prime}\boldsymbol{\Phi}_{bb}^{\mathbb{Q}} + \mathbf{B}_j^{(1)\prime}\right)\mathbf{f}_t + \mathbf{B}_j^{(n)\prime}\mathbf{v}_{b,t+1} + \mathbf{e}_j'\mathbf{v}_{s,t+1}\right]\right\},
\end{aligned}
$$

where, for the last equality, we have used the fact that the uncovered interest parity holds under the risk-neutral measure.

Once again, note that the last term in the previous equation satisfies:

$$
\begin{aligned}
E_t^{\mathbb{Q}}\left\{\exp\left[\begin{pmatrix} \mathbf{B}_j^{(n)\prime} & \mathbf{e}_j' \end{pmatrix}\begin{pmatrix} \mathbf{v}_{b,t+1} \\ \mathbf{v}_{s,t+1} \end{pmatrix}\right]\right\} &= \exp\left[\frac{1}{2}\begin{pmatrix} \mathbf{B}_j^{(n)\prime} & \mathbf{e}_j' \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}_{bb} & \boldsymbol{\Sigma}_{sb}' \\ \boldsymbol{\Sigma}_{sb} & \boldsymbol{\Sigma}_{ss} \end{pmatrix}\begin{pmatrix} \mathbf{B}_{\$}^{(n)} \\ \mathbf{e}_j \end{pmatrix}\right] \\
&= \exp\left[\frac{1}{2}\mathbf{B}_j^{(n)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_j^{(n)} + \frac{1}{2}\mathbf{e}_j'\boldsymbol{\Sigma}_{ss}\mathbf{e}_j + \mathbf{B}_j^{(n)\prime}\boldsymbol{\Sigma}_{sb}\mathbf{e}_j\right].
\end{aligned}
$$

Thus we have that

$$
A_j^{(n+1)} + \mathbf{B}_j^{(n+1)\prime}\mathbf{x}_{b,t} = \left[A_j^{(n)} + \mathbf{B}_j^{(n)\prime}\left(\boldsymbol{\mu}_{bb}^{\mathbb{Q}} + \boldsymbol{\Sigma}_{sb}\mathbf{e}_j\right) + \frac{1}{2}\mathbf{B}_j^{(n)\prime}\boldsymbol{\Sigma}_{bb}\mathbf{B}_j^{(n)} + A_j^{(1)}\right] + \left(\mathbf{B}_j^{(n)\prime}\boldsymbol{\Phi}_{bb}^{\mathbb{Q}} + \mathbf{B}_j^{(1)\prime}\right)\mathbf{x}_{b,t}.
$$

And matching coefficients we arrive at the following pricing recursions:

$$
\mathbf{B}_j^{(n+1)\prime} = \mathbf{B}_j^{(n)\prime}\boldsymbol{\Phi}_{11}^{\mathbb{Q}} + \mathbf{B}_j^{(1)\prime}, \tag{38}
$$

$$
A_j^{(n+1)} = A_j^{(n)} + \mathbf{B}_j^{(n)\prime}\boldsymbol{\mu}_1^{\mathbb{Q}} + \frac{1}{2}\mathbf{B}_j^{(n)\prime}\boldsymbol{\Sigma}_{11}\mathbf{B}_j^{(n)} + A_j^{(1)}. \tag{39}
$$

B Details on computation of the CGLS estimator

Specifically, we start by linearizing $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ around the unconstrained OLS estimate of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{OLS}$, described above. Let $\widetilde{\mathbf{r}}(\boldsymbol{\theta}) = \mathbf{0}$ be the linearized version $\mathbf{r}(\boldsymbol{\theta})$ around $\widehat{\boldsymbol{\theta}}_{OLS}$:

$$
\widetilde{\mathbf{r}}(\boldsymbol{\theta}) = \left[\mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS}) - \frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS})}{\partial \boldsymbol{\theta}'}\widehat{\boldsymbol{\theta}}_{OLS}\right] + \frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS})}{\partial \boldsymbol{\theta}'}\boldsymbol{\theta} = \mathbf{a} + \mathbf{A}\boldsymbol{\theta},
$$

with $\mathbf{A} = \frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS})}{\partial \boldsymbol{\theta}'}$ and $\mathbf{a} = \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS}) - \frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}}_{OLS})}{\partial \boldsymbol{\theta}'}\widehat{\boldsymbol{\theta}}_{OLS}$.

Then, we reparameterize the parameter space into the alternative K parameters $\boldsymbol{\alpha}$ $(S \times 1)$ and $\boldsymbol{\beta}$ $((K-S) \times 1)$ such that $\boldsymbol{\alpha} = \widetilde{\mathbf{r}}(\theta)$. Specifically, we can choose

$$\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{A} \\ \mathbf{A}_{\perp} \end{pmatrix} \boldsymbol{\theta}, \tag{40}$$

where $\mathbf{A}'_{\perp}$ is a basis for the orthogonal component of the row span of $\mathbf{A}$. This transformation allows us to impose the parametric restrictions $\widetilde{\mathbf{r}}(\boldsymbol{\theta}) = \boldsymbol{\alpha} = \mathbf{0}$ by inverting (40):

$$\boldsymbol{\theta} = \widetilde{\mathbf{A}}^{-1}(\mathbf{E}_2\boldsymbol{\beta} - \widetilde{\mathbf{a}}), \tag{41}$$

where $\mathbf{E}_2 = [\mathbf{0}, \mathbf{I}]'$ and substituting $\boldsymbol{\theta}$ into the distance function $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \boldsymbol{\gamma}(\boldsymbol{\pi}) - \boldsymbol{\Gamma}(\boldsymbol{\pi})\boldsymbol{\theta}$ to obtain a new distance function in terms of the smaller set of parameters $\boldsymbol{\beta}$:

$$\mathbf{h}(\boldsymbol{\pi}, \boldsymbol{\beta}) = \widetilde{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{\beta},$$

with $\widetilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + \boldsymbol{\Gamma}\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{a}}$, and $\widetilde{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}\widetilde{\mathbf{A}}^{-1}\mathbf{E}_2$.

Thus, the optimal ALS estimator of $\boldsymbol{\beta}$ can be obtained as

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{LCGLS} &= \arg\min_{\boldsymbol{\theta}} T \left[\widetilde{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{\beta}\right]' \widehat{\mathbf{V}}_g^{+} \left[\widetilde{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{\beta}\right], \\ &= \left(\widetilde{\boldsymbol{\Gamma}}'\widehat{\mathbf{V}}_g^{+}\widetilde{\boldsymbol{\Gamma}}\right)^{-1} \left(\widetilde{\boldsymbol{\Gamma}}'\widehat{\mathbf{V}}_g^{+}\widetilde{\boldsymbol{\gamma}}\right), \end{aligned}$$

and the optimal estimate of $\boldsymbol{\theta}$ can be obtained using (41):

$$\widehat{\boldsymbol{\theta}}_{LCGLS} = \widetilde{\mathbf{A}}^{-1}(\mathbf{E}_2\widehat{\boldsymbol{\beta}}_{GLS} - \widetilde{\mathbf{a}}). \tag{42}$$

C Invariant transformations of multi-country term structure models

Assume the following multi-country term structure model:

$$\begin{aligned} \mathbf{r}_t &= \boldsymbol{\Delta}_0 + \boldsymbol{\Delta}_1\mathbf{x}_t, \\ \mathbf{x}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1}, \\ \mathbf{x}_{t+1} &= \boldsymbol{\mu}^{\mathbb{Q}} + \boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{x}_t + \mathbf{v}_{t+1}^{\mathbb{Q}}, \end{aligned}$$

where both $\mathbf{v}_t$ and $\mathbf{v}_t^{\mathbb{Q}}$ are *i.i.d.* $N(0, \boldsymbol{\Sigma})$, and $\mathbf{x}_t = (\mathbf{x}'_{1,t}, \mathbf{x}'_{2,t})'$ being $\mathbf{x}_{1,t}$ a latent set of factors, and $\mathbf{x}_{2,t}$ observable. As in Dai and Singleton (2000), we are interested in applying invariant transformations, $\widehat{\mathbf{x}}_t = \mathbf{c} + \mathbf{D}\mathbf{x}_t$. We then have that the model above is observationally equivalent to:

$$\begin{aligned} \mathbf{r}_t &= \widehat{\boldsymbol{\Delta}}_0 + \widehat{\boldsymbol{\Delta}}_1\mathbf{x}_t, \\ \mathbf{x}_{t+1} &= \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\Phi}}\mathbf{x}_t + \widehat{\mathbf{v}}_{t+1}, \\ \mathbf{x}_{t+1} &= \widehat{\boldsymbol{\mu}}^{\mathbb{Q}} + \widehat{\boldsymbol{\Phi}}^{\mathbb{Q}}\mathbf{x}_t + \widehat{\mathbf{v}}_{t+1}^{\mathbb{Q}}, \end{aligned}$$

where now both $\widehat{\mathbf{v}}_t$ and $\widehat{\mathbf{v}}_t^{\mathbb{Q}}$ are *i.i.d.* $N(0, \widehat{\boldsymbol{\Sigma}})$ and

$$\begin{aligned}
\widehat{\boldsymbol{\Delta}}_0 &= \boldsymbol{\Delta}_0 - \boldsymbol{\Delta}_1 \mathbf{D}^{-1}\mathbf{c}, \\
\widehat{\boldsymbol{\Delta}}_1 &= \boldsymbol{\Delta}_1 \mathbf{D}^{-1}, \\
\widehat{\boldsymbol{\mu}} &= (\mathbf{I} - \mathbf{D}\boldsymbol{\Phi}\mathbf{D}^{-1})\mathbf{c} + \mathbf{D}\boldsymbol{\mu}, \\
\widehat{\boldsymbol{\Phi}} &= \mathbf{D}\boldsymbol{\Phi}\mathbf{D}^{-1}, \\
\widehat{\boldsymbol{\mu}}^{\mathbb{Q}} &= (\mathbf{I} - \mathbf{D}\boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{D}^{-1})\mathbf{c} + \mathbf{D}\boldsymbol{\mu}^{\mathbb{Q}}, \\
\widehat{\boldsymbol{\Phi}}^{\mathbb{Q}} &= \mathbf{D}\boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{D}^{-1}, \\
\widehat{\boldsymbol{\Sigma}} &= \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}'.
\end{aligned}$$

Of special interest to us are those invariant transformations that leave the set of observable variables, $\mathbf{x}_{2,t}$, unchanged. Such transformations can be expressed the following way:

$$\begin{pmatrix} \widehat{\mathbf{x}}_{1,t} \\ \widehat{\mathbf{x}}_{2,t} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{1,t} \\ \mathbf{x}_{2,t} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 + \mathbf{D}_1\mathbf{x}_{1,t} \\ \mathbf{x}_{2,t} \end{pmatrix}.$$

D Proof of Lemma 1

To proof this lemma, we use the invariant transformations of multi-country term structure models above as in Joslin, Singleton and Zhu (2011). In particular, we need to focus on invariant transformations that leave the set of exchange rates unchanged:

$$\begin{pmatrix} \widehat{\mathbf{f}}_t \\ \Delta\mathbf{s}_t \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_J \end{pmatrix} \begin{pmatrix} \mathbf{f}_t \\ \Delta\mathbf{s}_t \end{pmatrix}.$$

For simplicity, we assume that $\boldsymbol{\Phi}_{11}^{Q}$ can be diagonalized, that is $\boldsymbol{\Phi}_{11}^{Q} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{-1}$ where $\boldsymbol{\Lambda}$ is a diagonal matrix that contains the eigenvalues of $\boldsymbol{\Phi}_{11}^{Q}$, and $\mathbf{P}$ is a matrix that contains the corresponding eigenvectors.[14] The following two invariant transformations deliver the model in Lemma 1. First, we apply:

$$\begin{pmatrix} \widehat{\mathbf{f}}_t \\ \Delta\mathbf{s}_t \end{pmatrix} = \begin{pmatrix} -(\mathbf{I} - \boldsymbol{\Lambda})^{-1}\mathbf{T}^{-1}\boldsymbol{\mu}_1^{Q} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{T}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_J \end{pmatrix} \begin{pmatrix} \mathbf{f}_t \\ \Delta\mathbf{s}_t \end{pmatrix}.$$

Second, we exploit that for a given diagonal matrix such as $\boldsymbol{\Lambda}$, we can pre- and post-multiply it by another diagonal matrix, $\mathbf{B}$, and leave it unchanged it: $\boldsymbol{\Lambda} = \mathbf{L}\boldsymbol{\Lambda}\mathbf{L}^{-1}$. In particular, using

$$\begin{pmatrix} \widetilde{\mathbf{f}}_t \\ \Delta\mathbf{s}_t \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_J \end{pmatrix} \begin{pmatrix} \widehat{\mathbf{f}}_t \\ \Delta\mathbf{s}_t \end{pmatrix},$$

where

$$\mathbf{L} = \begin{pmatrix} \sum_{j=0}^{J} \widehat{\delta}_{1j} & 0 & \dots & 0 \\ 0 & \sum_{j=0}^{J} \widehat{\delta}_{2j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sum_{j=0}^{J} \widehat{\delta}_{Fj} \end{pmatrix},$$

[14] See appendix of Joslin, Singleton and Zhu (2011) for the case of non-diagonalizable matrices.

and $\widehat{\delta}_{ij}$ is the i-th element of vector $\widehat{\boldsymbol{\delta}}_j^{(1)}$, the vector of factor loadings of the short rate obtained from the first invariant transformation. Under such transformation, the factor loadings for the short rate will sum up to one, and thus the model can be expressed in the canonical form of Lemma 1.

Figure 1: Bond factor loadings

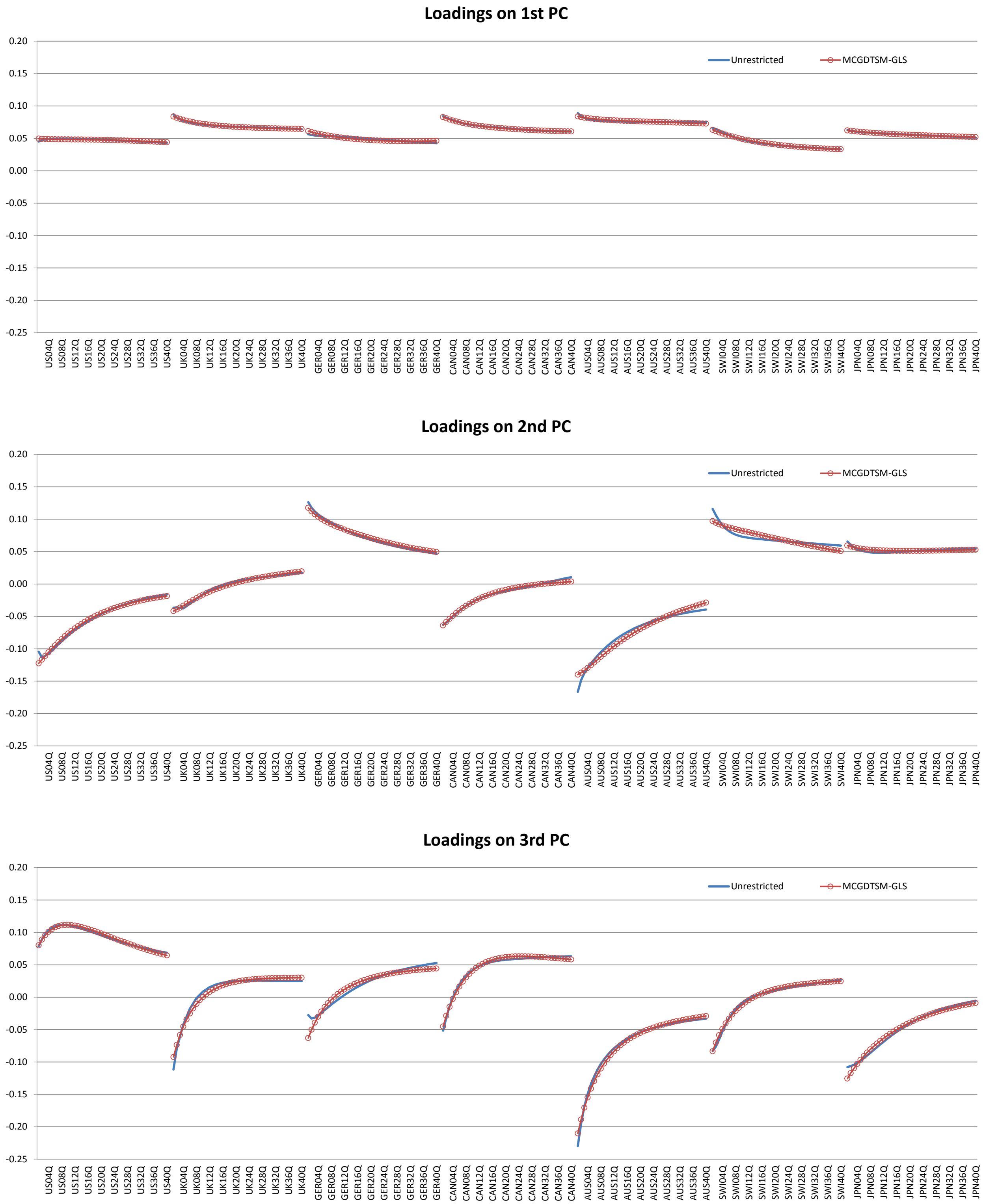
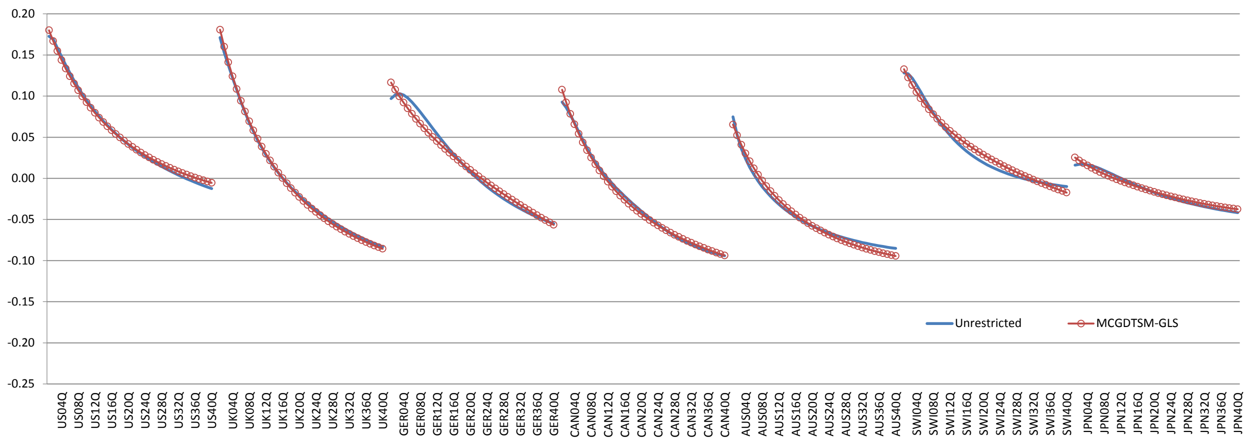
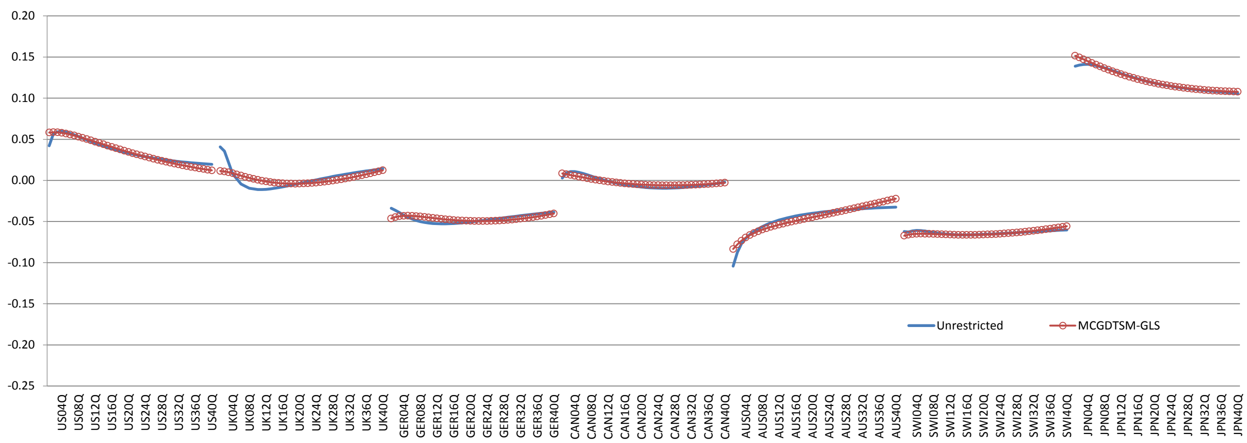


Figure 1: Bond factor loadings (cont.)

Loadings on 4th PC



Loadings on 5th PC



Loadings on 6th PC

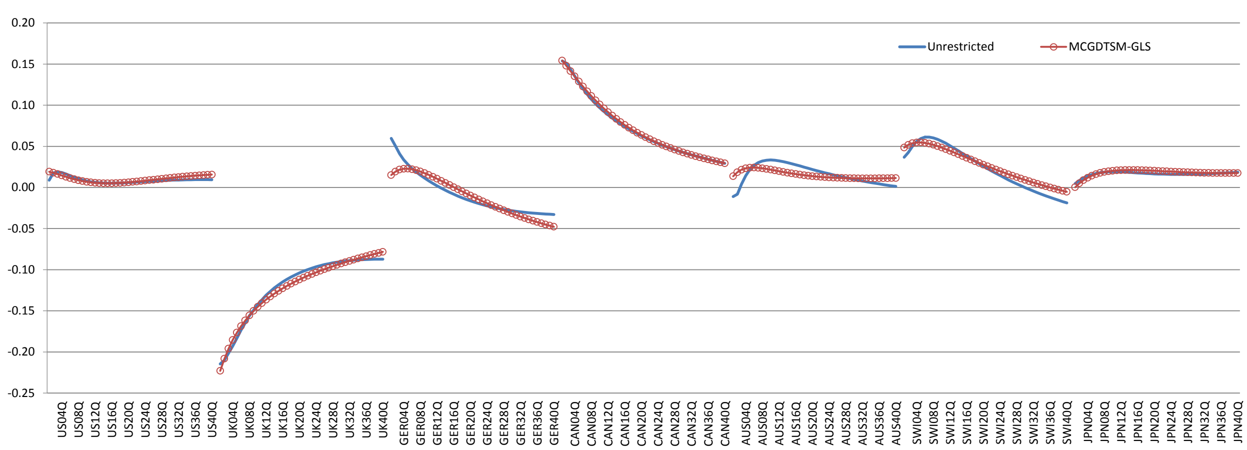
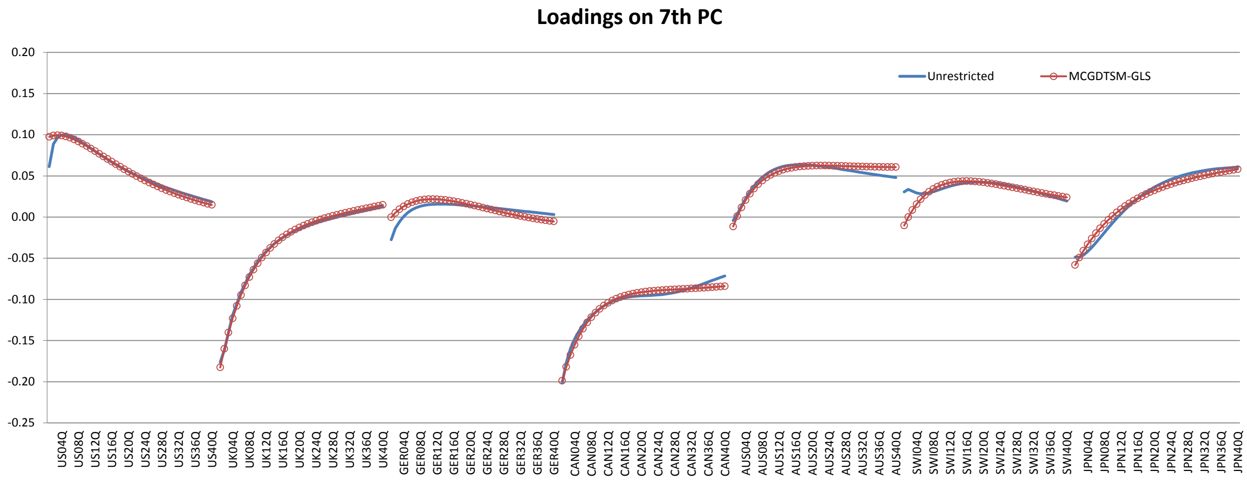
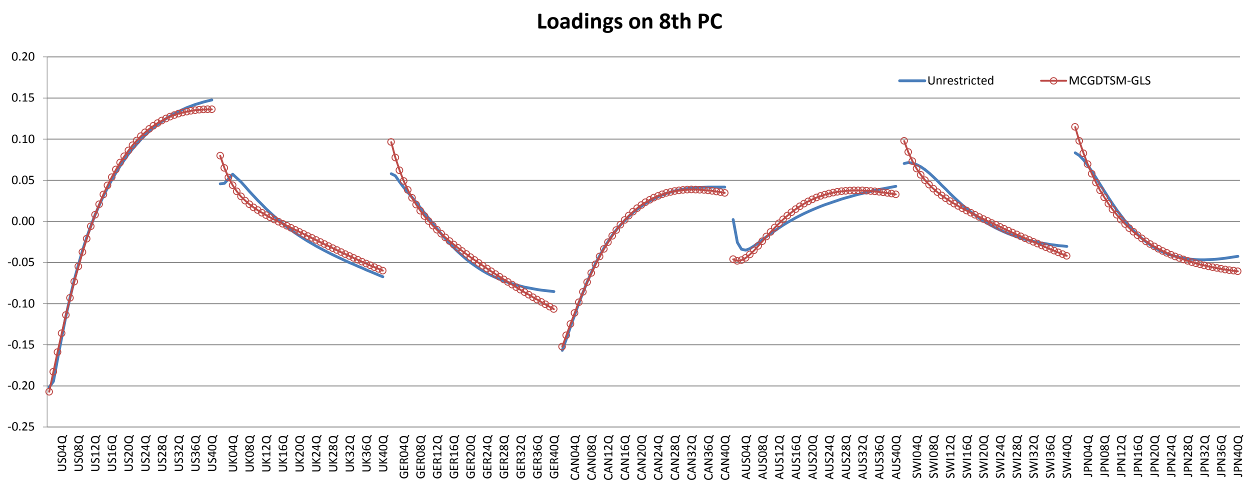


Figure 1: Bond factor loadings (cont.)





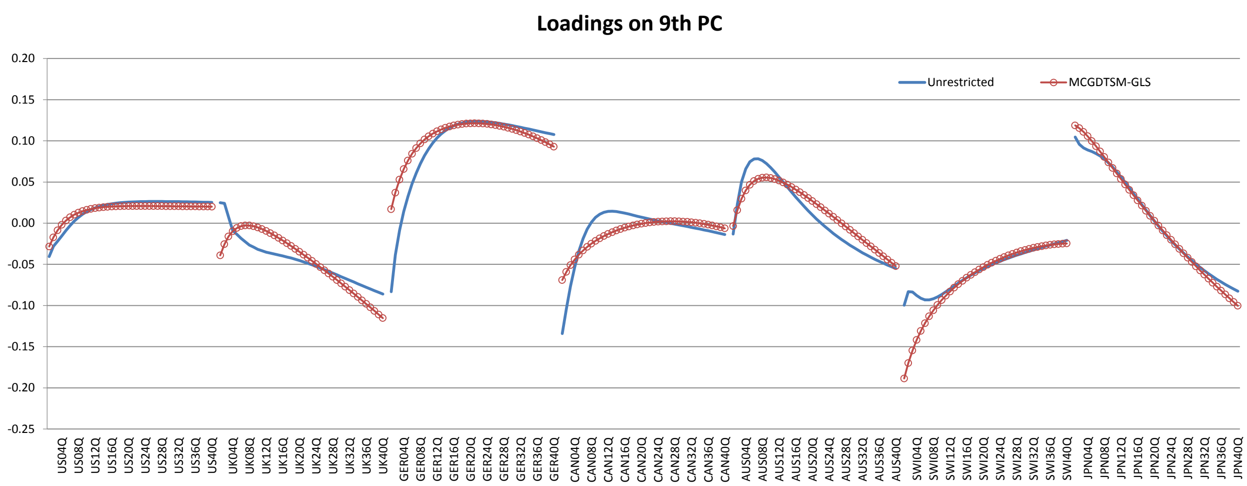


Figure 1: Bond factor loadings (cont.)

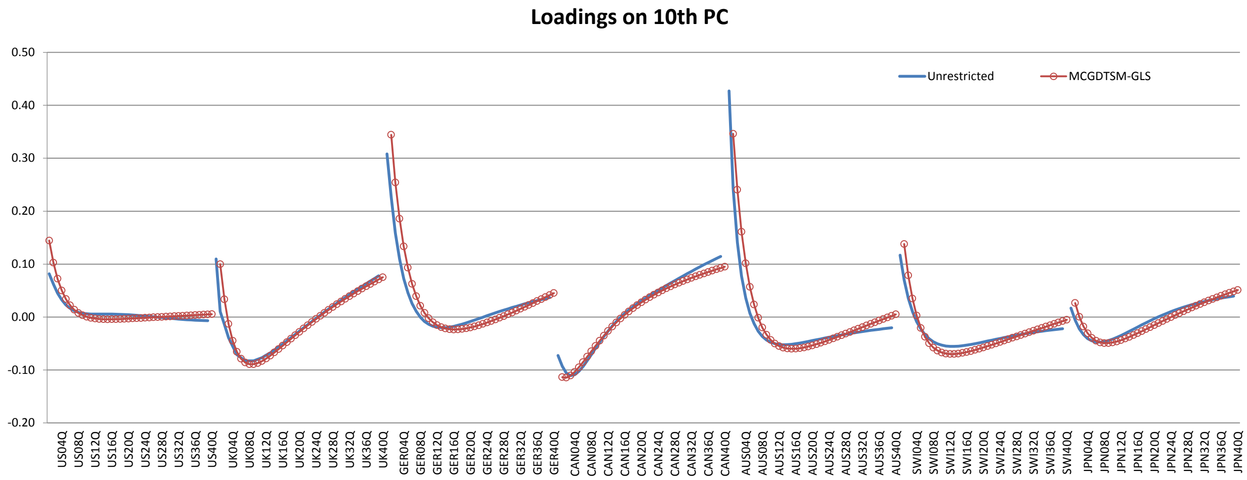


Figure 2: Estimated term premium on international 10-year yields

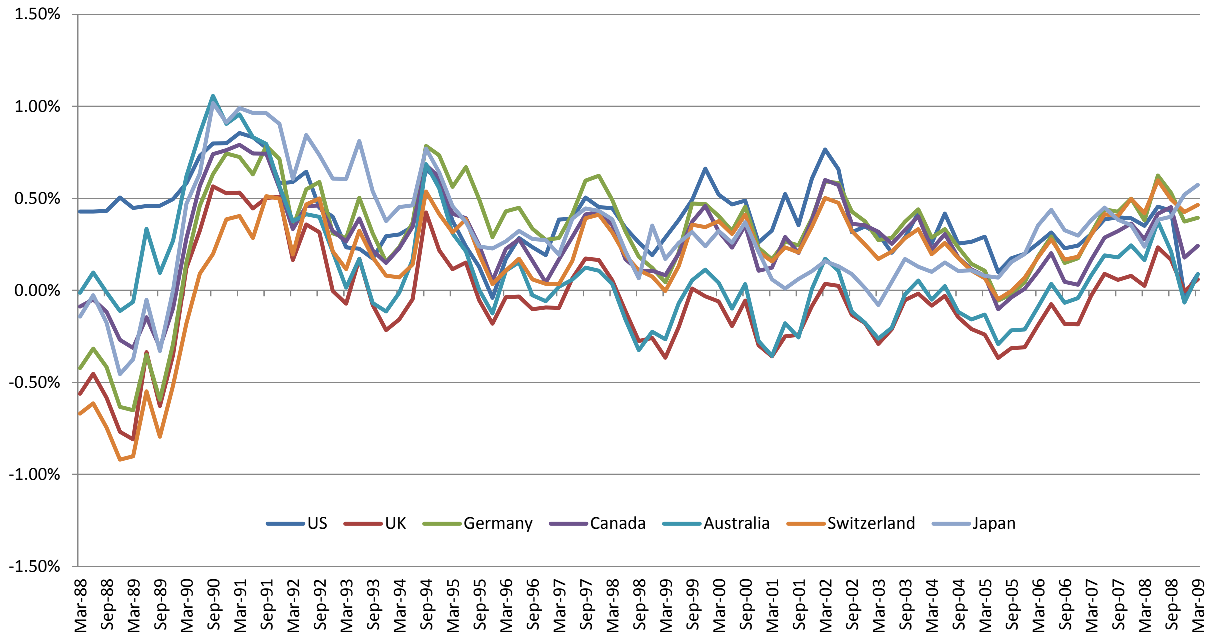


Table 1
Principal components analysis

Panel A: Per cent variation in yield curves explained by the first k domestic PCs

k	U.S.	U.K.	Germany	Canada	Australia	Switzerland	Japan
1	95.8	96.9	96.4	97.4	97.6	97.5	98.4
2	99.8	99.7	99.7	99.8	99.8	99.7	99.9
3	100.0	100.0	100.0	100.0	100.0	99.9	100.0

Panel B: Per cent variation in yield curves explained by the first k global PCs

k	per cent	k	per cent	k	per cent
1	88.6	6	99.0	11	99.8
2	93.9	7	99.4	12	99.9
3	96.4	8	99.6	13	99.9
4	97.8	9	99.7	14	99.9
5	98.5	10	99.8	15	99.9

Panel C: RMSE (in basis points) of a regression of yields on the first k PCs

k	U.S.	U.K.	Germany	Canada	Australia	Switzerland	Japan	Global
Domestic PCs								
1	37.7	43.8	35.4	38.9	43.7	27.0	25.6	36.7
2	8.0	13.6	10.1	10.8	14.1	10.2	6.4	10.8
3	3.4	4.9	3.3	4.0	5.1	4.1	2.3	4.0
Global PCs								
8	8.9	14.0	18.7	12.7	17.1	14.9	11.9	14.0
9	8.4	12.6	13.9	12.1	16.2	12.9	9.5	12.2
10	8.0	11.4	12.1	9.7	13.7	12.1	9.1	10.8
11	7.9	10.7	7.7	9.3	12.1	10.3	8.6	9.4
12	7.3	8.6	6.9	8.3	9.5	10.1	8.4	8.4

Note: Data are sampled quarterly March 1988 (1988Q1) to March 2009 (2009Q1).

Table 2
Model fit in basis points

	Affine	Unrestricted	Difference
U.S.	10.95	7.97	2.98
U.K.	20.1	13.6	6.5
Germany	16.9	10.07	6.83
Canada	10.81	10.79	0.02
Australia	21.12	14.14	6.98
Switzerland	15.76	10.17	5.59
Japan	10.35	6.39	3.96

Note: Affine model fit in basis points (1 = 0.01 per cent). RMSPE gives the root-mean-squared pricing error, and MAPE gives mean-absolute pricing error. "Affine" provides the fit of the multi-country term structure model, while "Unrestricted" provides the model fit of a regression of yields on the first 10 global principal components. "Difference" provides the loss of fit in basis points of estimating an affine term structure model instead of unrestricted OLS regressions.

Table 3
Wald statistics for the prices of risk being equal to zero

Panel A: Bond Prices of Risk ($H_0 : \mathbf{e}_j' \boldsymbol{\lambda}_b = \mathbf{0}$)

	Wald Test	p-value
PC1	51.58	[< 0.001]
PC2	59.52	[< 0.001]
PC3	39.48	0.001
PC4	48.88	[< 0.001]
PC5	43.40	[< 0.001]
PC6	44.89	[< 0.001]
PC7	45.24	[< 0.001]
PC8	44.03	[< 0.001]
PC9	44.30	[< 0.001]
PC10	43.98	[< 0.001]

Panel B: Foreign Exchange Prices of Risk ($H_0 : \mathbf{e}_j' \boldsymbol{\lambda}_s = \mathbf{0}$)

	Wald Test	p-value
GBP	10864.14	[< 0.001]
EUR	16259.49	[< 0.001]
CAD	15725.58	[< 0.001]
AUD	17866.31	[< 0.001]
CHF	14754.80	[< 0.001]
JPY	13412.00	[< 0.001]

Note: Data are sampled quarterly March 1988 (1988Q1) to March 2009 (2009Q1).