Modelling International Bond Markets with Affine Term Structure Models*

Georg Mosburger† Paul Schneider‡

Abstract

This paper investigates the performance of international affine term structure models (ATSMs) that are driven by a mutual set of global state variables. We discuss which mixture of Gaussian and square root processes is best suited for modelling international bond markets. We derive necessary conditions for the correlation and volatility structure of mixture models to accommodate various empirical stylized facts such as the forward premium puzzle and differently shaped yield curves. Using UK-US data we estimate international ATSMs taking into account the joint transition density of yields and exchange rates without assuming normality. We find strong empirical evidence for negatively correlated global factors in international bond markets. Further, the empirical results do not support the existence of local factors in the UK-US setting, suggesting that diversification benefits from holding currency-hedged bond portfolios in these markets are likely to be small. Altogether, we find that mixture models greatly enhance the performance of ATSMs.

Keywords: Exchange rates, International affine term structure models, Estimation, Model Selection

JEL: C33, E43, F31, G12

*The authors wish to thank Manfred Frühwirth, Alois Geyer, Yihong Xia, and especially Engelbert Dockner and Helmut Elsinger for helpful comments and Yacine Aït-Sahalia and Bob Kimmel for helpful comments and examples of likelihood coefficients.

†Department of Finance, University of Vienna, Bruennerstrasse 72, 1210 Vienna, Austria, Phone.: +43-1-4277-38060, Fax: +43-1-4277-38054, E-mail: georg.mosburger@univie.ac.at

‡Department of Finance, Vienna University of Economics and Business Administration, Nordbergstrasse 15, 1090 Vienna, Austria, Phone.: +43-1-31336-4337, E-mail: paul.schneider@wu-wien.ac.at

1 Introduction

Affine term structure models (henceforth ATSMs) driven by Markovian latent factors have received a lot of attention in the literature that deals with the description of single economies, presumably due to their analytic tractability which is convenient for pricing and risk management. However, relatively little work has been done concerning their capabilities within the context of a mutual model for two economies. With the increased integration of global capital markets, there is a deeply-felt need to develop arbitrage free cross-country ATSMs that are (i) consistent with stylized empirical facts (ii) while maintaining tractability.

For international bond markets, these stylized empirical facts encompass differently shaped yield curves, time varying correlations across two countries' yields and the forward premium anomaly. Those empirical properties ought to be generated by the model *in addition* to the stylized empirical facts that have been investigated so thoroughly in the literature on single economies (see e.g. Litterman and Scheinkman, 1991; Duffee, 2002; Duarte, 2004; Dai and Singleton, 2003).

In the international context, Ahn (2004) and Dewachter and Maes (2001) present multi-national three factor, pure square root models in which both economies are driven by a local (country-specific) factor and a common (international) factor. Their models extend the earlier work of Nielsen and Saá-Requejo (1993) and provide important implications for the forward premium puzzle and international diversification effects within the framework of affine models. Further, Backus, Foresi, and Telmer (2001) provide an extensive analysis of the forward premium anomaly and analyze in a discrete-time setting whether standard ATSMs are consistent with the anomaly. An extension along these lines is provided by Han and Hammond (2003), who try to reconcile the forward premium anomaly with a multi-country pure square root model. Finally, to this end, Brennan and Xia (2004) propose a multi-country pure Gaussian term structure model.

Although all of the above mentioned models ask whether a specific type of model specification is able to reproduce the forward premium anomaly, none of the existing papers analyzes which specification is most suited for jointly fitting yields across countries and generating well-documented features in the international finance literature. Additionally, none of the existing models does make full use of the range of admissible distributional capabilities. ATSMs allow for a much richer parametrization while maintaining tractability and parameter identification. Every model specification type exhibits theoretical properties which, altogether, reflect a trade-off between modelling time-varying volatilities, correlations between factors and economic theory. On the grounds of pure models, a system specified entirely with Gaussian processes offers maximal flexibility with respect to magnitude and sign of conditional and unconditional correlations among the state variables. However, this advantage is at the cost of non-negativity of nominal interest rates (no-arbitrage) and time-varying conditional volatilities, the domain of correlated square root (CSR) processes, which in turn are only able to display zero conditional and non negative unconditional correlation.

Leaving the grounds of pure models implies further complications, since there is contradictory evidence even within the single economy term structure literature. On the one hand, Dai and Singleton (2000) state that across a wide variety of parameterizations of ATSMs, the data used in their study consistently called for negative conditional correlations among the state variables. Such type of correlations is precluded in multi-factor (pure) CSR models. On the other hand, using a different data sample and different market prices of risk, Duffee (2002) notices that the goodness-of-fit rises monotonically with the number of factors that affect conditional volatilities. Using the same market prices of risk as Duffee (2002), Tang and Xia (2005) favor a model class with one square root process and two Gaussian processes for a variety of data sets from different economies. The bottom line in this discussion concerning the specification of single economy term structure models seems to be that when conditional volatility is very pronounced in the data, models with more square root factors are more appropriate. However, when the data strongly calls for negative correlations among factors, models with more Gaussian factors should perform better. Joint modelling of exchange rates and yields imposes a final and heavy layer of difficulty on model specifications. Exchange rates exhibit a certain type of heteroskedastic variation, which, in an no-arbitrage setting, has to be compatible with the model implied variation that is generated as a function of market prices of risk and the Markovian latent state variables.

We work within the most general general setting that satisfies the admissibility conditions from Dai and Singleton (2000) extended to multiple countries. This setting is, in theory, flexible enough to produce the above mentioned empirical facts. We investigate both theoretically and empirically the tradeoffs arising from different specifications in an international context. In particular, we are interested in the performance of mixture models, models with both Gaussian *and* CSR processes. In addition, we explore whether there seem to exist local factors in international bond markets by assessing the performance of models in which all economies are driven by the same set of *common* factors relative to the performance of models in which some factors are *local* in the sense that that they have only impact on interest rates in one specific economy.

Using swap and LIBOR rates for the UK and the US and the corresponding exchange rate data, we estimate a series of models by means of maximum likelihood using the closed form likelihood expansions proposed in Aït-Sahalia (2002), Aït-Sahalia (2001) and Aït-Sahalia and Kimmel (2002). To the best of our knowledge, this study is the first one which estimates international ATSMs taking into account the joint distribution of yields and the exchange rate without assuming normality of the transition densities. Joint estimation gives us the opportunity to combine economic theory (no arbitrage) with time series properties. Further, an estimation that does not assume normality removes the bias introduced by a (false) normal assumption especially for high dimensional systems (see Frühwirth-Schnatter and Geyer, 1996). Representatives of the $A_0(3), A_1(3), A_2(3), A_3(3)$ are chosen according to the local factor specification as well as maximally parameterized common factor specifications. All in all we estimate eight models. All parameter estimates are admissible in the sense of Dai and Singleton (2000) and imply time series of the latent state variables that "*could have occurred*".

The best model according to its overall likelihood score is a model with two square root and one Gaussian process. This model tightly reproduces in sample yields and provides slightly better in sample forecasts of the signs of log exchange rate returns than a drift adjusted random walk. However, even though this model provides a tight fit of the yield data, the random walk has superior forecasting quality concerning levels of the exchange rate as well as yields for most maturities. Strikingly, the model most widely used in international settings, the pure CSR model, provides the worst fit to the data. This can probably be attributed to the strong negative correlation that seems to be present between the state variables that drive international economies. Concerning the forward premium puzzle only the representative from the $A_1(3)$ class generates risk premia that are volatile enough in order to generate a negative Fama coefficient. Further, we find no empirical support for the existence of local factors driving term structures and exchange rates across the US and UK, suggesting that diversification benefits from holding currency-hedged bond portfolios in these markets are likely to be small.

The remainder of this paper is structured as follows. In Section 2 we give a detailed presentation of our international affine term structure model. Section 3 discusses under which conditions differently specified admissible models are capable to reproduce several stylized empirical facts reported in the recent international finance literature. Section 4 describes the model estimation and presents the empirical results. Finally, in Section 5 we present concluding remarks.

2 Model Setup

2.1 Short Rates and Factors

We assume that the world economy consists of two countries, a domestic country d and a foreign country f, and is represented by a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}_t, \mathbb{P})$, where $\mathcal{F} = \{\mathbb{F}_t; 0 \leq t \leq T\}$. Short rates are modelled in nominal terms and are assumed to be affine functions of N unobserved state variables $Y(t) = (Y_1(t), Y_2(t), \ldots, Y_N(t))^\top$:

$$r^i(t) = \delta_0^i + \delta_Y^{i\top} Y(t), \quad i \in \{d, f\}, \tag{1}$$

where δ_0^i is a scalar and δ_Y^i is a $N \times 1$ vector that represents loadings on the latent factors $Y(t)$.[1] Further, under the objective probability measure $\mathbb{P}$, the vector of state variables is assumed to follow the affine diffusion

$$dY(t) = \mathcal{K}(\Theta - Y(t))dt + \Sigma\sqrt{S(t)}dW(t), \tag{2}$$

where $\mathcal{K}$, Σ and $S(t)$ are $N \times N$ matrices, Θ is an $N \times 1$ vector and $W(t)$ represents an N-dimensional independent standard Brownian motion under $\mathbb{P}$. Further, $S(t)$ is a diagonal $N \times N$ matrix with elements on the main diagonal given by:

$$S(t)_{ii} = \alpha_i + \beta_i^\top Y(t), \tag{3}$$

where α_i is a scalar and β_i is an $N \times 1$ vector given by i-th column of the matrix $\mathcal{B} = [\beta_1, \cdots, \beta_N]$.

To ensure admissibility and maximal flexibility, we work with the canonical models introduced by Dai and Singleton (2000) (henceforth DS).[2] They refer to N factor models, where the number of factors driving the conditional variance is $m \leq N$ as elements of the class $A_m(N)$. Further, they show that all admissible N factor models can uniquely be classified into $N+1$ non-nested subfamilies ($m = 0, 1, \ldots, N$) and that all of the extant ATSMs in the literature reside within some subfamily $A_m(N)$ and can be obtained from invariant transformations of the respective canonical model.[3] For

[1] Alternatively we could choose a structural modelling approach that takes into account price levels and consumption and make assumptions about the utility functions of representative agents as done by e.g. Constantinides (1992), but we are mainly concerned with model implied interactions between short rates, pricing kernels and exchange rate. For a study with pricing kernels in real terms see Brennan and Xia (2004).

[2] The admissibility conditions guarantee non-negativity of the conditional variances over the whole support of the state vector $Y(t) \in \mathbb{R}^N$. See Duffie and Kan (1996) and Dai and Singleton (2000).

[3] DS introduce *affine transformations* $\mathcal{T}_A Y(t) = LY(t) + \nu$, where L is a nonsingular $N \times N$ matrix and ν is an $N \times 1$

completeness, we refer to Appendix A where we report details about sufficient parameter restrictions and normalizations provided by DS that guarantee admissibility and identification of the canonical models.

We perform all analytical computations with the general specification in (1), for empirical investigations we will, however, put several restrictions on the canonical specification. Current literature puts a lot of emphasis on using local factors, i.e. factors that influence only the short rate of one specific country while having no impact on the other short rates (see Ahn, 2004; Dewachter and Maes, 2001; Brennan and Xia, 2004). In our model setup local factors can easily be accommodated by restricting some of the elements of δ_y^d and δ_y^f to take on values of zero. For example, if we let $Y_1(t)$ represent the common factor which affects all short rates in our world economy, then, restricting our attention to a three factor world, we could let $r^d(t) = \delta_1^d Y_1(t) + \delta_2^d Y_2(t)$ and $r^f(t) = \delta_1^f Y_1(t) + \delta_3^f Y_3(t)$. Local factors specific to one economy are modelled to be uncorrelated with the local factors specific to the other economy. Thus, we restrict the entries in the $\mathcal{K}$ matrix such that the drift of the factors specific to one economy is unaffected by the common state variables and the factors specific to the other economy. If the above example were taken from the $A_3(3)$ family, then starting from the canonical representation, we restrict $\mathcal{K}_{21}$, $\mathcal{K}_{23}$ and $\mathcal{K}_{31}$, $\mathcal{K}_{32}$ to be zero.

2.2 Bond Prices and Yields

Denote the time t price of a zero-coupon bond denominated in currency of country $i \in \{d, f\}$ with unit face value maturing at time $T = t + \tau$ by $P^i(Y(t), \tau)$. In the absence of arbitrage opportunities prices of zero-coupon bonds are given by

$$\begin{aligned} P^i(Y(t), \tau) &= \mathbb{E}_t^{\mathbb{Q}^i}\left[\exp\left(-\int_t^{t+\tau} r^i(s)ds\right)\right] \\ &= \mathbb{E}_t^{\mathbb{Q}^i}\left[\exp\left(-\int_t^{t+\tau} \delta_0^i + \delta_y^{i\top} Y(s)ds\right)\right], \end{aligned} \tag{4}$$

where $\mathbb{E}_t^{\mathbb{Q}^i}$ denotes expectation under the equivalent martingale measure of country i conditional on time t. Thus, in order to compute equation (4) we need to work with the factor dynamics under the equivalent probability measure $\mathbb{Q}^i$. Let dW^i denote the vector of $\mathbb{Q}^i$ Brownian motions. By applying

vector. *Diffusion rescaling* $\mathcal{T}_D$ affects the diffusion parameters and the market prices of risk. *Brownian motion rotation* $\mathcal{T}_O$ rotates unobserved independent Brownian motions into other unobserved Brownian motions and finally *permutation* $\mathcal{T}_P$ is a reordering of the state variables. All these transformations preserve admissibility of the model and leave short rates, bond prices and their distributions unchanged and are therefore termed "invariant transformations".

Girsanov's theorem we have $dW = dW^i - \Lambda^i(Y(t),t)dt$, where $\Lambda^i(Y(t),t)$ is an $N \times 1$ vector that represents the market prices of factor risk in the respective country i. In this paper, we adopt the market price of risk specification proposed by DS that is known as *completely affine*. In this specification $\Lambda^i(Y(t),t) = \sqrt{S(t)} \cdot \lambda_1^i$, where λ_1^i is a constant $N \times 1$ vector. From this, we can restate the dynamics of the state vector under the respective equivalent martingale measure of country $i \in \{d, f\}$ as

$$\begin{aligned} &= \mathcal{K}(\Theta - Y(t))dt - \Sigma\sqrt{S(t)}\Lambda^i(Y(t),t)dt + \Sigma\sqrt{S(t)}dW^i(t) \\ &= \mathcal{K}^i(\Theta^i - Y(t))dt + \Sigma\sqrt{S(t)}dW^i(t) \end{aligned} \tag{5}$$

where $\mathcal{K}^i$ and Θ^i denote the $\mathbb{Q}^i$ transformed mean reversion parameters that are given by

$$\widetilde{\mathcal{K}^i} = \left(\mathcal{K} + \Sigma\Phi^i\right), \qquad \widetilde{\Theta^i} = \left(\mathcal{K} + \Sigma\Phi^i\right)^{-1}\left(\mathcal{K}\Theta - \Sigma\Psi^i\right),$$

where the jth row of Φ^i is given by $\lambda_{1j}^i \cdot \beta_j^\top$ and Ψ^i is an $N \times 1$ vector whose jth element is given by $\lambda_{1j}^i \cdot \alpha_j$.

Given the affine structure of the factor dynamics under the equivalent martingale measure represented by equation (5) together with the affine structure of the short rates in equation (1), Duffie and Kan (1996) show that bond prices denominated in their respective home currency are given by

$$P^i(Y(t),\tau) = \exp\left(A^i(\tau) - B^i(\tau)^\top Y(t)\right), \tag{6}$$

where $A^i(\tau)$ and $B^i(\tau)$ are given by the solutions to the ODEs

$$\begin{aligned} \frac{dA^i(\tau)}{d\tau} &= -\Theta^{i\top}\mathcal{K}^{i\top}B^i(\tau) + \frac{1}{2}\sum_{j=1}^{N}\left[\Sigma^\top B^i(\tau)\right]_j^2\alpha_j - \delta_0^i \\ \frac{dB^i(\tau)}{d\tau} &= -\mathcal{K}^{i\top}B^i(\tau) - \frac{1}{2}\sum_{j=1}^{N}\left[\Sigma^\top B^i(\tau)\right]_j^2\beta_j + \delta_y^i \end{aligned} \tag{7}$$

with boundary conditions

$$A^i(0) = 0, \qquad B^i(0) = 0.$$

Here $A^i(\tau)$ is a scalar function and $B^i(\tau)$ is an $N \times 1$ vector valued function. From e.g. Fisher and Gilles (1996) we have that under the physical measure the instantaneous bond price dynamics in affine diffusion models are given by

$$\frac{dP^i(Y(t),\tau)}{P^i(Y(t),\tau)} = \left[r^i(t) + e^i(t,\tau)\right] dt - v(t,\tau) dW(t),$$

where $e^i(t,\tau) = B^i(\tau)^\top \Sigma \sqrt{S(t)}\, \Lambda^i$ denotes the instantaneous expected excess return to holding the bond and the instantaneous bond volatility is given by $v^i(t,\tau) = B^i(\tau)^\top \Sigma \sqrt{S(t)}$. Further, zero-coupon yields defined as $y^i(Y(t),\tau) = -\frac{1}{\tau} \ln P^i(Y(t),\tau)$ are affine in the state variables and given by

$$y^i(Y(t),\tau) = \frac{1}{\tau}\left[-A^i(\tau) + B^i(\tau)^\top Y(t)\right]. \tag{8}$$

2.3 Pricing Kernels and Exchange Rates

Given the assumption of no-arbitrage and complete markets, there exists a positive and unique pricing kernel (state-price density or state-price deflator) for each country i, denoted M^i, such that the product of the pricing kernel and any traded asset is a martingale under the physical measure $\mathbb{P}$ (see Harrison and Kreps (1979) and Harrison and Pliska (1981)). This yields the fundamental pricing equation:

$$x^i(t) = \mathbb{E}_t^{\mathbb{P}}\left[\frac{M^i(T)}{M^i(t)} \cdot x^i(T)\right] \qquad i \in \{d, f\}, \tag{9}$$

where $x^i(t)$ is the nominal value of a traded asset denominated in currency of country i which gives claim to the stochastic cash flow $x^i(T)$ denominated in currency of country i at time T. Equivalently, equation (9) can be reformulated as

$$1 = \mathbb{E}_t^{\mathbb{P}}\left[\frac{M^i(T)}{M^i(t)} \cdot R^i(t,T)\right] \qquad i \in \{d, f\}, \tag{10}$$

where $R^i(t,T) = x^i(T)/x^i(t)$ denotes the gross return from t to T generated by the asset in terms of country i's currency. As shown by Backus, Foresi, and Telmer (2001), in the absence of arbitrage, the exchange rate is tightly linked to the pricing kernels of the two countries. Define the exchange rate $X(t)$ as the number of units of domestic currency that have to be paid at time t in order to obtain one unit of foreign currency and consider two assets, one delivering a stochastic payoff in domestic currency the other one in foreign currency. Taking the asset denominated in domestic currency and using the fundamental asset pricing equation (10) the return $R^d(t,T)$ must satisfy

$$1 = \mathbb{E}_t^{\mathbb{P}} \left[\frac{M^d(T)}{M^d(t)} \cdot R^d(t,T) \right]. \tag{11}$$

However, we can also state the return on this asset in terms of the foreign currency since $R^f(t,T) = (X(t)/X(T)) \cdot R^d(t,T)$ and

$$1 = \mathbb{E}_t^{\mathbb{P}} \left[\frac{M^f(T)}{M^f(t)} \cdot \frac{X(t)}{X(T)} \cdot R^d(t,T) \right]. \tag{12}$$

Since the law of one price implies that both relations have to hold, we must have

$$\mathbb{E}_t^{\mathbb{P}} \left[\frac{M^d(T)}{M^d(t)} \cdot R^d(t,T) \right] = \mathbb{E}_t^{\mathbb{P}} \left[\frac{M^f(T)}{M^f(t)} \cdot \frac{X(t)}{X(T)} \cdot R^d(t,T) \right]. \tag{13}$$

By rearranging this equation and substituting T by $t+\tau$ we can see that the exchange rate is completely and endogenously determined by the dynamics of the two pricing kernels since now the following relation can be established

$$\frac{M^d(t+\tau)}{M^d(t)} \cdot \frac{X(t+\tau)}{X(t)} = \frac{M^f(t+\tau)}{M^f(t)}. \tag{14}$$

Apart from the tight link to the pricing kernels, the exchange rate also has distinct empirical features. Regressions of the exchange rate returns on the interest rate differential across countries have very low R^2 statistics, implying that the lion's share of the variation of exchange rate movements remains unexplained by the factor risks driving the term structure of the two countries. Therefore, we differentiate between risk factors that drive the pricing kernel dynamics and those that drive the term structure. Additionally, as many empirical investigations have shown (see also Section 4.1, especially Figure 2), exchange rate volatility is extremely high as compared to the volatility of the interest rate differential across two countries. In order to be able to account for this feature, we allow the pricing kernels to additionally be driven by a source of risk B_{N+1} that is orthogonal to any other of the term structure related risks $W_i(t)$. This decomposition of pricing kernel variation into "explainable" and "unexplainable" variation can also be found in Brandt and Santa-Clara (2002). They, however, attribute the unexplained pricing kernel changes to market incompleteness. We rather follow the point of view taken by Dewachter and Maes (2001) and accredit the unexplained variation to risk factors governing other types of assets than those of the bond market.

Equipped with such practical and theoretical considerations, we specify the dynamics of the pricing kernel of country i as

$$\frac{dM^i(t)}{M^i(t)} = -r^i(t)dt - \Lambda^i(Y(t),t)^\top \, dB(t) - \Phi^i \, dB_{N+1}(t) \tag{15}$$

where the pricing kernels are driven by a vector of N $\mathbb{P}$-Brownian motions $B(t) = (B_1(t), \ldots, B_N(t))^\top$ and an additional source of risk $B_{N+1}(t)$. $dB_i(t)$ is assumed to be independent of $dB_j(t)$ for $i \neq j$, i.e. $dB_i(t) \cdot dB_j(t) = 0$. The two innovation vectors W and B are also assumed to be mutually uncorrelated in order to reflect the difference between exchange rate risk and interest rate risk.[4]

Inspecting equation (14) as τ goes to zero together with the pricing kernel dynamics and an application of Ito's lemma yields the following dynamics for the exchange rate

$$\begin{aligned} d\log X(t) &= d\log M^f(t) - d\log M^d(t) \\ &= \left(r^d(t) - r^f(t) + \frac{1}{2}\left(\|\Lambda^d(Y(t),t)\|^2 - \|\Lambda^f(Y(t),t)\|^2 \right) \right) dt \\ &+ \frac{1}{2}((\Phi^d)^2 - (\Phi^f)^2)dt \\ &+ \left(\Lambda^d(Y(t),t) - \Lambda^f(Y(t),t) \right)^\top dB(t) + (\Phi^d - \Phi^f)dB_{N+1}, \end{aligned} \tag{16}$$

where $\|\cdot\|$ denotes the Euclidean norm. Equation (16) clearly shows that the uncovered interest rate parity does not hold under the physical measure $\mathbb{P}$. The expected rate is equal to the interest rate differential plus a risk premium that investors demand to compensate for exchange rate risk. This departure from the uncovered interest rate parity is solely due to differences in the market prices of the risk factors driving both economies. Thus, the uncovered interest rate parity is assumed to hold under the physical measure $\mathbb{P}$ only if each factor source of risk is compensated equally (in absolute terms) in the domestic country and the foreign country.

3 Implications of the Model

In this section we illustrate empirical features inherent to different model specifications. We discuss necessary conditions under which models are capable of reproducing negative correlations between short rates across countries (see Singleton, 1994) and the forward premium puzzle. Backus, Foresi, and Telmer

[4]Non perfect correlations are a prerequisite for our estimation procedure. With completely affine market prices of risk, the covariance matrix of the yield dynamics and the log exchange rate dynamics is singular for $\rho_i = 1, \quad i = 1, \ldots, N$ and $\Phi^i = 0$.

(2001) show that in affine models of the short rate the forward premium anomaly can be accounted for under two conditions. The first condition calls for a positive probability of negative interest rates. In the admissible framework, presented in this paper, this can only be accommodated by the inclusion of Gaussian factors. Alternatively, a way to generate the forward premium anomaly is to allow for asymmetric factor loadings (δ) across countries. The subsequent analysis is based on a general common factor framework (with δ free), however since we estimate our models using local factors, we also discuss the effect of restricting the model to a common factor - local factor setting.

3.1 Correlations

Although the Brownian motions driving the vector of state variables $Y(t)$ are independent, conditional and unconditional instantaneous correlations between the single factors can be different from zero due to interdependencies in the drift. This becomes apparent by inspecting the mean-reversion matrix $\mathcal{K}$ in (36). Unlike common specifications for square root models, where $\mathcal{K}$ is usually diagonal (e.g. Nielsen and Saá-Requejo, 1993; Ahn, 2004; Hodrick and Vassalou, 2002), the canonical form allows for off diagonal elements which implies that the drift of one factor will in general be a function of the other factors. This results in a rich unconditional correlation structure which is necessary for an affine model to being able to exhibit the empirical findings from Section 4.[5]

We can choose an invariant transformation of the canonical model that is suitable to eliminate feedback among Gaussian processes and between Gaussian and correlated square root (CSR) processes. The dependency structure is thereby transferred from $\mathcal{K}$ into the diffusion expression. To be more specific, the procedure can be performed by an invariant affine transformation of the latent factors $Y(t) = (Y_1(t), Y_2(t), \ldots, Y_N(t))^\top$ into $Z(t) = (Z_1(t), Z_2(t), \ldots, Z_N(t))^\top$ with $Z(t) = LY(t) + \nu$, where L is a nonsingular $N \times N$ matrix and ν is an $N \times 1$ vector. Such a transformation is possible because of the linear structure of affine term structure models and the fact that the factors are unobservable.[6] Under the physical measure, the dynamics of the transformed $Z(t)$ system are:

[5] In pure square root models the conditional volatility between the factor dynamics is zero due to admissibility.

[6] An example for the effect of such a transformation on the parameters can be seen by investigating the new factor loadings

$$r^i(t) = \delta_0^i + \delta_Y^{i\top} Y(t) = \delta_0^i + \delta_Y^{i\top} L^{-1}(Z(t) - \nu) = \underbrace{\delta_0^i - \delta_Y^{i\top} L^{-1}\nu}_{\delta_0^{i*}} + \underbrace{\delta_Y^{i\top} L^{-1}}_{\delta_Y^{i*\top}} Z(t),$$

The transformed Σ matrix allows for inspection of the correlation structure implied by the model. The transformation of the other parameters $(\mathcal{K}, \Theta, \Sigma, \alpha_i, \beta_i, \lambda)$ is equivalent.

$$
\begin{aligned}
dZ(t) &= L\mathcal{K}L^{-1}\left(\nu + L\Theta - Z(t)\right)dt + L\Sigma\sqrt{S^*(t)}dW(t) \\
&= \mathcal{K}^*\left(\Theta^* - Z(t)\right)dt + \Sigma^*\sqrt{S^*(t)}dW(t),
\end{aligned} \tag{17}
$$

where

$$
S^*(t)_{ii} = \alpha_i + \beta_i^\top L^{-1}(Z(t) - \nu), \quad \mathcal{K}^* = L\mathcal{K}L^{-1}, \quad \Theta^* = \nu + L\Theta, \quad \Sigma^* = L\Sigma
$$

The desired transformations can be done by finding a matrix L and a vector ν such that $\mathcal{K}^{DD}$ is diagonalized and $\mathcal{K}^{BD}$ is set to zero.

Denoting the transformed state variables by $Z(t)$ we can rewrite any canonical model as:

$$
r^i(t) = \delta_0^{i*} + \delta_y^{i*\top} Z(t), \quad i \in \{d, f\}, \tag{18}
$$

and

$$
dZ(t) = \mathcal{K}^*(\Theta^* - Z(t))dt + \Sigma^*\sqrt{S^*(t)}dW(t) \tag{19}
$$

with

$$
\mathcal{K}^* = \begin{pmatrix} \mathcal{K}^{BB}_{m\times m} & 0_{m\times(N-m)} \\ 0_{(N-m)\times m} & \mathcal{K}^{DD*}_{(N-m)\times(N-m)} \end{pmatrix}, \Sigma^* = \begin{pmatrix} I_{m\times m} & 0_{m\times(N-m)} \\ \Sigma^{DB}_{(N-m)\times m} & \Sigma^{DD}_{(N-m)\times(N-m)} \end{pmatrix},
$$

where $\mathcal{K}^{DD*}_{(N-m)\times(N-m)}$ is a diagonal matrix (and the diagonal elements of $\Sigma^{DD}_{(N-m)\times(N-m)}$ are equal to one). Since we now have moved some of the dependency structure from the drift to the Σ matrix, instantaneous conditional and unconditional covariances between the factors can be read off the Σ matrix.

By using equation (18) and taking differences we obtain the dynamics of the two short rates:

$$
dr^d(t) = \delta_y^{d*\top} dZ(t) \qquad \text{and} \qquad dr^f(t) = \delta_y^{f*\top} dZ(t).
$$

The instantaneous covariance between $r^d(t)$ and $r^f(t)$ is given by

$$
\operatorname{Cov}(dr^d(t), dr^f(t)) = \sum_{k=1}^{N} \delta_k^{d*}\delta_k^{f*} \operatorname{Var}(dZ_k) + \sum_{1\le l<m\le N} (\delta_l^{d*}\delta_m^{f*} + \delta_l^{f*}\delta_m^{d*}) \operatorname{Cov}(dZ_l, dZ_m), \tag{20}
$$

and the instantaneous correlation is given by

$$\operatorname{Corr}(dr^d(t), dr^f(t)) = \frac{\operatorname{Cov}(dr^d(t), dr^f(t))}{\sqrt{\operatorname{Var}(dr^d(t)) \cdot \operatorname{Var}(dr^f(t))}}, \tag{21}$$

with

$$\operatorname{Var}(dr^i(t)) = \sum_{k=1}^{N} (\delta_k^{i*})^2 \operatorname{Var}(dZ_k) + 2 \sum_{1 \le l < m \le N}^{N} \delta_l^{i*} \delta_m^{i*} \operatorname{Cov}(dZ_l, dZ_m), \quad i \in \{d, f\}. \tag{22}$$

To inspect the properties of mixture/pure models, we fix the number of state variables to three ($N = 3$). Let us first consider a model in which all state variables follow CSR processes (m=3). The models proposed by Ahn (2004) and Dewachter and Maes (2001) fall into this class. In the maximal $A_3(3)$ model we have the following specification:

$$r^i(t) = \delta_0^i + \delta_1^i Y_1(t) + \delta_2^i Y_2(t) + \delta_3^i Y_3(t), \quad i \in \{d, f\},$$

$$\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11} & \mathcal{K}_{12} & \mathcal{K}_{13} \\ \mathcal{K}_{21} & \mathcal{K}_{22} & \mathcal{K}_{23} \\ \mathcal{K}_{31} & \mathcal{K}_{32} & \mathcal{K}_{33} \end{pmatrix} \left[\begin{pmatrix} \Theta_1 \\ \Theta_2 \\ \Theta_3 \end{pmatrix} - \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \end{pmatrix} \right] dt$$
$$+ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{Y_1(t)} & 0 & 0 \\ 0 & \sqrt{Y_2(t)} & 0 \\ 0 & 0 & \sqrt{Y_3(t)} \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \end{pmatrix}.$$

It can easily be seen that the state variables in this model are all conditionally uncorrelated with each other. By imposing that all delta weights are greater or equal zero in order to ensure positive short rates, we constrain the instantaneous correlation between two short rates to be nonnegative.[7]

In a next step, we now consider a specification in which only two factors drive the conditional volatilities of all factors, i.e. $m = 2$. After a suitable transformation

$$r^i(t) = \delta_0^{i*} + \delta_1^{i*} Z_1(t) + \delta_2^{i*} Z_2(t) + \delta_3^{i*} Z_3(t), \quad i \in \{d, f\},$$

[7] This is due to the fact that any two different (positive) linear combinations of uncorrelated random variables are positively correlated to each other. In the empirical section our representative of the $A_3(3)$ class is the only model where we had to drop the constraint of positive delta weights, since the data called for negative correlations.

$$\begin{pmatrix} dZ_1(t) \\ dZ_2(t) \\ dZ_3(t) \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11} & \mathcal{K}_{12} & 0 \\ \mathcal{K}_{21} & \mathcal{K}_{22} & 0 \\ 0 & 0 & \mathcal{K}_{33} \end{pmatrix} \left[\begin{pmatrix} \Theta_1 \\ \Theta_2 \\ 0 \end{pmatrix} - \begin{pmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \end{pmatrix} \right] dt$$
$$+ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \sigma_{31} & \sigma_{32} & 1 \end{pmatrix} \begin{pmatrix} \sqrt{Z_1(t)} & 0 & 0 \\ 0 & \sqrt{Z_2(t)} & 0 \\ 0 & 0 & \sqrt{\alpha_3 + Z_1(t) + Z_2(t)} \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \end{pmatrix}.$$

In $A_2(3)$ models $Z_3(t)$ is Gaussian and can therefore become negative. Thus, an inconvenient feature of all models in which $m < N$ is that there is a positive probability of generating negative short rates. However, by introducing a Gaussian process the model is now flexible enough to generate conditional correlations between Gaussian and CSR factors. In our example we are now free to determine σ_{31} and σ_{32} as to introduce non-zero correlations between $Z_3(t)$ and $Z_1(t)$ and between $Z_3(t)$ and $Z_2(t)$. The conditional correlations among the state variables driven by CSR processes, however, remain zero. Inclusion of Gaussian processes enables modelling correlations between Gaussian and any other state variables. This in turn implies that the correlation between any two short rates can now attain negative values. This can easily be seen by examining equation (20) and noting that we can now assign negative values to $\mathrm{Cov}(dZ_1, dZ_3)$ and $\mathrm{Cov}(dZ_2, dZ_3)$. Nevertheless, it should be clear that this flexibility comes at the price of limiting the volatility dynamics of the short rate. Thus, as already noted in Dai and Singleton (2000) there is an important tradeoff between modelling the structure of factor volatilities and admissible non-zero conditional correlations between the factors driving the short rate and thus between any two short rates.

Further as noted by Ahn (2004), common factor models, however, imply a lower bound on the correlation of the short rates strictly greater than -1. The reason why common factors cannot generate the full band of correlations is due to the fact that if either of the common factors increases, both, the covariance and the volatilities in the denominator of equation (20) increase. In local factor models, however, an increase in the local factor of country d raises the volatility of its short rate, but it does not affect the volatility of short rate f, nor the covariance between the short rates of countries d and f. Thus, when the local factor specific to country d explodes, the instantaneous correlation between country d and country f tends to zero.

3.2 Forward Premium Puzzle

Many empirical studies report that the changes in exchange rates and interest rate differential across countries are negatively correlated although theory would suggest a positive relation (see Bansal (1997), Bekaert (1996) and for a survey paper Engel (1996)). This finding has been entitled as "forward premium anomaly". In this section we show under which conditions affine models can reproduce this forward premium anomaly. Consider the regression equation

$$\log X(t+\Delta) - \log X(t) = a_1 + a_2(\log F(t,t+\Delta) - \log(X(t))) + \varepsilon(t+\Delta). \tag{23}$$

From covered interest rate parity $\log F(t,t+\Delta) - \log(X(t)) \approx (r^d - r^f)\Delta$ for Δ very small and the slope coefficient a_2 (also known as Fama coefficient) is given by

$$a_2 = \frac{\mathrm{Cov}\left(\log \frac{X(t+\Delta)}{X(t)}, (r^d(t) - r^f(t))\Delta\right)}{\mathrm{Var}((r^d(t) - r^f(t))\Delta)}. \tag{24}$$

The unbiased expectation hypothesis implies $a_1 = 0$ and $a_2 = 1$. However, assuming no arbitrage there is no reason for the unbiased expectation hypothesis to hold under the physical measure. Under no arbitrage a_1 and a_2 can be seen as affine "corrections" to account for the change in the drift of the exchange rate that renders equation (23) true under the expectation taken with respect to the physical probability measure. As mentioned above, a_2 is therefore often reported to be negative. In our model, the covariance term in a_2 can become negative for various reasons. Define

$$d = r^d(t) - r^f(t) \quad \text{and} \quad p = \frac{1}{2}\left(\|\Lambda^d(Y(t),t)\|^2 - \|\Lambda^f(Y(t),t)\|^2\right),$$

where d represents the interest differential across countries and p can be understood as exchange rate risk premium. In fact, the expected appreciation of the log exchange rate under the physical probability measure $\mathbb{P}$ in (16) is precisely $(d+p)dt$. Now, consider the covariance term in equation (24). This term can be rewritten as

$$\begin{aligned}
\mathrm{Cov}\left(\log \frac{X(t+\Delta)}{X(t)}, r^d(t) - r^f(t)\right) &= \mathrm{Cov}\left(d+p, d\right) \\
&= \mathrm{Var}(d) + \mathrm{Cov}(d,p).
\end{aligned}$$

Here we assume Δ to be sufficiently small, allowing us to use directly the infinitesimal dynamics in (16) without much error. Thus, in order to accommodate for the forward premium anomaly a model must be able to generate $\mathrm{Var}(d) + \mathrm{Cov}(d, p) < 0$. Fama (1984) gives the two necessary conditions. First, the covariance between d and p has to be negative, that is the interest rate differential has to covary negatively with the risk premium demanded by investors to compensate for exchange rate risk. Second, the variance of the exchange rate risk premium (p) has to be greater than the variance of the interest rate differential (d).

With the *completely affine* market price of risk specification the regression slope a_2 of our model is given by

$$a_2 = \frac{\mathrm{Var}(d) + \mathrm{Cov}(d,p)}{\mathrm{Var}(d)} = 1 + \frac{\mathrm{Cov}(d,p)}{\mathrm{Var}(d)} \tag{25}$$

with

$$\begin{aligned}
\mathrm{Var}(d) + \mathrm{Cov}(d,p) &= \frac{1}{2}\left(\sum_{k=1}^{N} \omega_k \gamma_k \,\mathrm{Var}(Y_k) + \sum_{1\leq l<m\leq N} \eta_{l,m} \,\mathrm{Cov}(Y_l, Y_m)\right) && (26)\\
\mathrm{Var}(d) &= \sum_{k=1}^{N} \gamma_k^2 \,\mathrm{Var}(Y_k) + 2 \sum_{1\leq l<m\leq N} \gamma_l \gamma_m \,\mathrm{Cov}(Y_l, Y_m) && (27)
\end{aligned}$$

where

$$\gamma_k = \delta_k^d - \delta_k^f \tag{28}$$

$$\omega_k = \sum_{j=1}^{N} \mathcal{B}_{kj}\left((\lambda_j^d)^2 - (\lambda_j^f)^2\right) + 2\gamma_k \tag{29}$$

$$\eta_{l,m} = \gamma_m \omega_l + \gamma_l \omega_m. \tag{30}$$

Since any term in equation (26) can become negative, our model is able to account for the forward premium puzzle. Clearly, the sign of the slope coefficient hinges greatly on γ and ω and the sign of the covariances between the state variables. In order to build some intuition for what information is contained in system (28) - (30), it is instructive to think of the short rate dynamics $dr(t)$ in terms of weights and factors since $dr(t) = \delta_y^\top dY(t)$. From this relation it can be seen that the δ-weights inflate (deflate) the variation in the factor dynamics. Hence, if an estimation puts a lot of weight on one factor, the variation of that factor most likely explains much of the variation in the short rate. In our model, in which an economy is made up of its nominal short rate, a natural way of paraphrasing "our estimation resulted in a high δ_1" is to say that an economy has high *exposure* to factor $Y_1(t)$. Using this terminology, the existence of the forward premium anomaly indicates a tendency for domestic (foreign) investors that are less exposed to a specific factor than foreign (domestic) investors to demand a higher risk premium in absolute terms for this factor and all other factors that are influenced by this factor, all other things equal. To explore the relation in more depth, we again focus on specific examples of three factor models.

Let us first consider the pure CSR specification, i.e. the $A_3(3)$ model. In this ATSM subfamily $\mathcal{B}$ is given by the identity matrix. Thus, the Fama slope coefficient a_2 becomes negative if

$$a\,\mathrm{Var}(Y_1) + b\,\mathrm{Var}(Y_2) + c\,\mathrm{Var}(Y_3) + d\,\mathrm{Cov}(Y_1, Y_2) + e\,\mathrm{Cov}(Y_1, Y_3) + f\,\mathrm{Cov}(Y_2, Y_3) < 0, \tag{31}$$

where

$$\begin{aligned}
a &= \Big((\lambda_1^d)^2 - (\lambda_1^f)^2 + 2(\delta_1^d - \delta_1^f)\Big)\,(\delta_1^d - \delta_1^f) \\
b &= \Big((\lambda_2^d)^2 - (\lambda_2^f)^2 + 2(\delta_2^d - \delta_2^f)\Big)\,(\delta_2^d - \delta_2^f) \\
c &= \Big((\lambda_3^d)^2 - (\lambda_3^f)^2 + 2(\delta_3^d - \delta_3^f)\Big)\,(\delta_3^d - \delta_3^f) \\
d &= \Big((\lambda_1^d)^2 - (\lambda_1^f)^2 + 2(\delta_1^d - \delta_1^f)\Big)\,(\delta_2^d - \delta_2^f) + \Big((\lambda_2^d)^2 - (\lambda_2^f)^2 + 2(\delta_2^d - \delta_2^f)\Big)\,(\delta_1^d - \delta_1^f) \\
e &= \Big((\lambda_1^d)^2 - (\lambda_1^f)^2 + 2(\delta_1^d - \delta_1^f)\Big)\,(\delta_3^d - \delta_3^f) + \Big((\lambda_3^d)^2 - (\lambda_3^f)^2 + 2(\delta_3^d - \delta_3^f)\Big)\,(\delta_1^d - \delta_1^f) \\
f &= \Big((\lambda_2^d)^2 - (\lambda_2^f)^2 + 2(\delta_2^d - \delta_2^f)\Big)\,(\delta_3^d - \delta_3^f) + \Big((\lambda_3^d)^2 - (\lambda_3^f)^2 + 2(\delta_3^d - \delta_3^f)\Big)\,(\delta_2^d - \delta_2^f).
\end{aligned}$$

Since all unconditional variances and covariances have to be positive in a pure CSR model in order to be admissible, it is clear that the sign of inequality (31) depends on the coefficients a to f. Further, we can see that if both economies' short rates are exposed equally to all factors, then (31) becomes zero and there is no way to account for the anomaly. Strikingly, one can show that if we move to a setting in which we also include local factors which only affect one short rate but not both, it is necessary that the two countries are not exposed in the same way to the common factor in order to generate the anomaly. This fact is documented in Backus, Foresi, and Telmer (2001).[8]

To find an example of how the mechanics work in an admissible CSR model, we can investigate under

[8] See Ahn (2004) for a model setup within this setting. Dewachter and Maes (2001) consider a similar setting, however they assign the same weight to the common factor in both economies.

what conditions the coefficient a in the above equation system becomes negative. For this exposure, we restrict all δs to be positive. All other things equal, with the domestic economy being less exposed to factor one than the foreign economy, that is $\delta_1^d < \delta_1^f$, we must have $(\lambda_1^d)^2 - (\lambda_1^f)^2 > 2(\delta_1^d - \delta_1^f)$. Thus, the magnitude of the risk premium demanded by domestic investors has to be higher than that demanded by foreign investors, in absolute terms. On the other hand, if $\delta_1^d > \delta_1^f$, i.e. the domestic economy has a higher exposure to factor one, we need $(\lambda_1^d)^2 - (\lambda_1^f)^2 < 2(\delta_1^d - \delta_1^f)$ for a to be negative. This relation of magnitude between the difference in the factor loadings and the difference in the respective squared market prices of risk is the only way to account for the forward premium puzzle in the $A_3(3)$ model.

Now, consider a mixture model in which the conditional volatility of the short rate is driven only by two of the three factors, i.e. one of the factor $(Y_3(t))$ is a Gaussian factor and compute again the coefficients in equation (26). In the $A_2(3)$ family these coefficients are

$$
\begin{aligned}
a &= \left((\lambda_1^d)^2 - (\lambda_1^f)^2 + \beta_{13}\left((\lambda_3^d)^2 - (\lambda_3^f)^2\right) + 2(\delta_1^d - \delta_1^f)\right)(\delta_1^d - \delta_1^f) \\
b &= \left((\lambda_2^d)^2 - (\lambda_2^f)^2 + \beta_{23}\left((\lambda_3^d)^2 - (\lambda_3^f)^2\right) + 2(\delta_2^d - \delta_2^f)\right)(\delta_2^d - \delta_2^f) \\
c &= 2(\delta_3^d - \delta_3^f)^2 \\
d &= \left((\lambda_1^d)^2 - (\lambda_1^f)^2 + \beta_{13}\left((\lambda_3^d)^2 - (\lambda_3^f)^2\right) + 2(\delta_1^d - \delta_1^f)\right)(\delta_2^d - \delta_2^f) \\
&\quad + \left((\lambda_2^d)^2 - (\lambda_2^f)^2 + \beta_{23}\left((\lambda_3^d)^2 - (\lambda_3^f)^2\right) + 2(\delta_2^d - \delta_2^f)\right)(\delta_1^d - \delta_1^f) \\
e &= \left((\lambda_1^d)^2 - (\lambda_1^f)^2 + \beta_{13}\left((\lambda_3^d)^2 - (\lambda_3^f)^2\right) + 2(\delta_1^d - \delta_1^f)\right)(\delta_3^d - \delta_3^f) \\
&\quad + \left((\lambda_3^d)^2 - (\lambda_3^f)^2 + 2(\delta_3^d - \delta_3^f)\right)(\delta_3^d - \delta_3^f) \\
f &= \left((\lambda_2^d)^2 - (\lambda_2^f)^2 + \beta_{23}\left((\lambda_3^d)^2 - (\lambda_3^f)^2\right) + 2(\delta_2^d - \delta_2^f)\right)(\delta_3^d - \delta_3^f) \\
&\quad + \left((\lambda_3^d)^2 - (\lambda_3^f)^2 + 2(\delta_3^d - \delta_3^f)\right)(\delta_2^d - \delta_2^f),
\end{aligned}
$$

where β_{13} and β_{23} are elements of the matrix $\mathcal{B}$:

$$
\mathcal{B} = \begin{pmatrix} 1 & 0 & \beta_{13} \\ 0 & 1 & \beta_{23} \\ 0 & 0 & 0 \end{pmatrix}.
$$

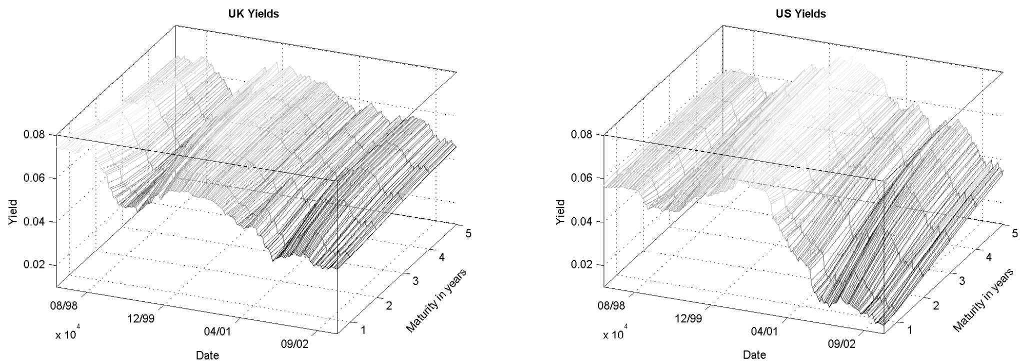
From the admissibility conditions we have $\beta_{13}, \beta_{23} \geq 0$.

Again, this model has the inconvenience that it generates negative short rates with a positive probability. Yet, the conditions that have to be fulfilled in order to generate the forward premium anomaly are not as restrictive as in the $A_3(3)$ model. To see this consider again the coefficients a to f. Clearly, coefficient c cannot become negative anymore. However, since the unconditional covariance between the Gaussian factor and the CSR factors is not bounded to be positive, this model offers more flexibility. That is, even if investors in the less exposed country do not demand a higher risk premium in absolute terms, it is still possible for the model to generate a negative slope coefficient a_2.

Next consider the $A_0(3)$ model. In this model class all state variables have constant variances implying constant risk premia over time and zero correlation between the interest rate differential (d) and the exchange rate risk premium (p). Thus, such model will never be able to generate a negative Fama coefficient with the completely affine market price of risk specification.

Altogether, as reported in several other studies, completely affine models are heavily restricted in their ability to generate the forward premium puzzle, since they need that the state variables driving the term structure exhibit at least some conditional volatility and that the market prices obey restrictive conditions.

Figure 1: Constructed UK and US zero coupon yields as implied by LIBOR and swap rates (06.01.1998 – 07.01.2003).



4 Empirical Analysis

4.1 Data and Empirical Facts

For our empirical analysis we use fixed-for-variable swap data and LIBOR rates. The choice to model the term structure by means of swap rates has recently been followed by many researchers, see for example Dai and Singleton (2000), Duffie and Singleton (1997), Collin-Dufresne and Goldstein (2002) or Dewachter and Maes (2001). This is done mainly for two reasons. First, swap rates are truly constant maturity yields, whereas in the Treasury market the maturities of constant maturity yields are only approximately constant. Second, they may be more relevant for pricing issues since most interest rate derivatives are priced by means of LIBOR and swap rates. One inconvenience of this approach is that these rates are not strictly without default risk. However, as Duffie and Huang (1996) and Collin-Dufresne and Solnik (2001) show, they are only minimally affected by credit risk because of their special netting features. Another problem encountered when analyzing swap rates is that the two-year contract is the shortest maturity available.[9] We therefore augment the data with short-term LIBOR rates which serve as a proxy for short-term swap rates that are not traded.

We retrieve LIBOR rates of 6 and 12 month maturities and swap rates for maturities of 2 to 5 years for the UK and the US. To avoid seasonality effects (see Piazzesi (2003)) we retrieve these data every Tuesday on a weekly basis from 06/01/1998 to 07/01/2003 (262 observations) from EcoWin. We then use these rates to bootstrap zero-coupon LIBOR and swap yields according to Piazzesi (2003).[10] The as such constructed yields are visualized in Figure 1. To complete the data we retrieve middle quote exchange rate data from Bloomberg.

Table 1: Summary statistics of the UK and US term structure
Means and standard deviations are reported in percentage points on an annual basis. Δs_{t+1} represents the annualized weekly log-returns of the exchange rate, i.e. the returns from period t to $t+1$.

	uk3m	uk6m	uk1yr	uk2yr	uk3yr	uk4yr	uk5yr	us3m	us6m	us1yr	us2yr	us3yr	us4yr	us5yr	Δs_{t+1}
mean	5.60	5.60	5.72	5.80	5.88	5.88	5.88	4.57	4.62	4.79	5.14	5.40	5.57	5.73	0.0037
std.	1.17	1.13	1.08	0.88	0.79	0.71	0.65	1.77	1.77	1.73	1.48	1.30	1.14	1.04	0.53
uk3m	**1**	0.99	0.96	0.87	0.79	0.75	0.73	0.82	0.80	0.77	0.71	0.65	0.61	0.56	0.06
uk6m		**1**	0.99	0.92	0.85	0.81	0.80	0.83	0.82	0.80	0.75	0.70	0.66	0.62	0.05
uk1yr			**1**	0.97	0.92	0.89	0.88	0.83	0.83	0.83	0.80	0.76	0.73	0.69	0.05
uk2yr				**1**	0.99	0.97	0.96	0.82	0.83	0.85	0.85	0.83	0.82	0.80	0.04
uk3yr					**1**	0.99	0.99	0.81	0.83	0.85	0.86	0.86	0.85	0.84	0.05
uk4yr						**1**	1.00	0.79	0.81	0.83	0.86	0.86	0.86	0.85	0.05
uk5yr							**1**	0.79	0.81	0.83	0.86	0.87	0.87	0.86	0.05
us3m								**1**	1.00	0.99	0.95	0.92	0.89	0.86	0.10
us6m									**1**	0.99	0.97	0.94	0.91	0.88	0.10
us1yr										**1**	0.99	0.97	0.94	0.92	0.10
us2yr											**1**	0.99	0.98	0.97	0.10
us3yr												**1**	1.00	0.99	0.10
us4yr													**1**	1.00	0.10
us5yr														**1**	0.10

As can be seen in Figure 1 the UK term structure is inverted at the beginning of the sample period.

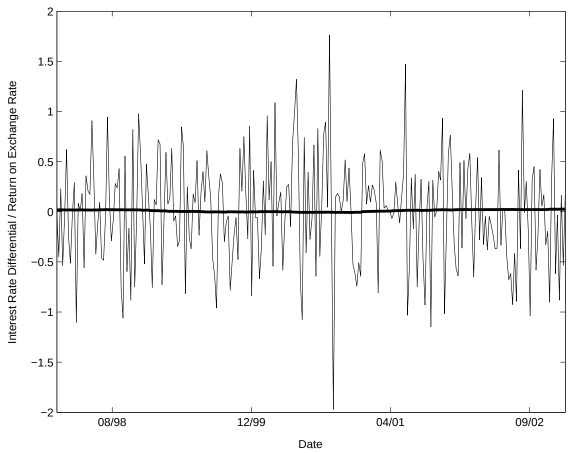
[9]One year swap rates started trading in 1997. Prior to this year the shortest available maturity for swap contracts was the two year contract.

[10]A practical problem when using swap and LIBOR rates together is that the data is recorded asynchronous since LIBOR data are recorded at 11 a.m. London time, while swap data are typically recorded at the end of day. Jones (2002) proposes a model to mitigate this problem. In our model we, however, ignore the problem of asynchronous recording.

Table 1 reports some descriptive statistics of the data. For both, the UK and the US, average yields are increasing, while their standard deviations are generally decreasing with maturity. Additionally, when comparing the average yield curves, we can infer that yields are generally lower in the UK and that the average yield curve in the UK is not as steep as in the US. Correlations within national bond markets are extremely high (ranging from 0.73 to almost 1) and monotonically decreasing with maturity. Across countries we also observe significant positive correlations ranging from 0.56 to 0.87, although to a lesser degree and without a clear pattern. All in all, the high correlations across countries as well as across maturities suggest that both term structures are driven by a common factor.

Another interesting fact is that the annualized log-returns of the exchange rate correlate positively with each of the yields, taking on correlation values from 0.04 to 0.10. However, the log-returns of the exchange rate are higher correlated to US yields than to UK yields, implying that the yield differentials ("UK minus US") are negatively correlated to exchange rate movements. This is clearly evidence against the uncovered interest rate parity, which would suggest that the exchange rate appreciates as the interest rate differential rises. Further, by inspecting the standard deviation relative to the mean of the data elements, we find that the exchange rate returns are excessively volatile compared to the yields. This statement also holds true as we compare the volatility of the yield differentials with exchange rate returns. This evidence is depicted in Figure 2, which plots the interest rate differential against annualized exchange rate returns.

Figure 2: Interest Rate Differential vs. Exchange Rate Returns
Comparison of the in-sample interest rate differential and annualized log exchange rate returns. The thick line represents the interest rate differential which is computed by subtracting the US 3 months yields from the UK 3 month yields. The thin line shows the annualized log returns of the GBP/USD exchange rate.



4.2 Estimation Procedure

Theoretically it is not necessary to include the exchange rate into the estimation, since it is endogenously determined by the pricing kernel dynamics in an arbitrage free setting. However, the functional form of the instantaneous drift and variance provides important information for the scale of the (differences of) market prices of risk. An estimation that does not take into account the exchange rate is likely to produce unrealistic implied exchange rate drifts and variances. To the best of our knowledge, we are the first who directly estimate the joint dynamics of yields and exchange rate taking into account the full distributional capabilities of the affine framework. In particular we do not assume the transition densities from one observation to be multivariate normal or χ^2 which is only the case for a very small, restricted subset of the $A_m(N)$ families.

In the preceding literature on affine term structure models in a two economy framework Quasi Maximum Likelihood (QML) has been the predominant estimation procedure (e.g. Han and Hammond, 2003; Dewachter and Maes, 2001; Brennan and Xia, 2004), presumably due to its ease of application. However, as pointed out in Frühwirth-Schnatter and Geyer (1996), the bias introduced by QML increases with the dimensionality of the model. Closed form transition densities for maximum likelihood estimation are only known for very few multivariate diffusion models. For example the transition densities for canonical ATSMs, except for restricted pure Gaussian and restricted pure square root models, are not known in closed form. In our application the dynamics of the exchange rate adds an additional layer of complication, since its drift and diffusion depends on the latent state variables.

Recent research in the field has sought to find suitable approximations to work around the problem of not having closed form transition densities. Apart from QML, which neglects the non-normality inherent to general diffusion models, a very intuitive and straightforward method is Simulated Maximum Likelihood (SML) (see Pedersen, 1995; Santa-Clara, 1995; Elerian, 1998; Durham and Gallant, 2001; Brandt and Santa-Clara, 2002), which already has found an application in international economics (Brandt and Santa-Clara, 2002). Unfortunately SML is a computationally intensive procedure. However, it can be greatly enhanced with respect to speed and precision with variance and bias reduction techniques such as control variates.

In order to being able to employ computationally intensive global optimization procedures for our maximum likelihood estimation that need many likelihood function evaluations, we employ the technique from Aït-Sahalia (2001), Aït-Sahalia (2002) and Aït-Sahalia and Kimmel (2002), who provide formulae for the calculation of closed form expansions of the likelihood function for discretely sampled diffusions that theoretically can be developed with arbitrary accuracy (depending on the order of expansion). These formulae are obtained from comparing terms of equal order from a proposed form of solution that is guessed from a Hermite expansion about the discretization Δ of dt with the Kolmogorov transition partial differential equations. For systems that cannot be reduced to unit diffusions, an additional expansion about the state variables is performed. Even though only the pure Gaussian model is reducible in the sense of Aït-Sahalia (2002) it is still possible to obtain the coefficients of the likelihood function from a linear system that can be evolved and solved order by order. Despite the fact that these equations are linear, for high dimensional systems like ours and high orders (higher than 2), solving these symbolic linear equations can become a non-trivial computational obstacle due to the sheer size of the coefficient expressions.

We assume that at each month t, $t = 1, \ldots, T$, N yields are observed without error. It is the same number N that denotes the number of latent state variables that drives both economies. These yields are for fixed times to maturity $\tau_1, \ldots, \tau_N$. The other k yields for the remaining maturities are assumed to be measured with serially and mutually uncorrelated, mean-zero measurement error. Denote the parameter vector by θ. Stack the N perfectly observed yields into a vector $y(t)$ and the k imperfectly observed yields into a vector $\widetilde{y}(t)$. Given an initial value of θ, equation (8) can be inverted in order to obtain an implied state vector $Y_0(t)$:

$$Y_0(t) = H_1^{-1}(y(t) - H_0). \tag{32}$$

In equation (32), H_0 is an $N \times 1$ vector with element i given by $A^j(\tau_i)/\tau_i$, and H_1 is a $N \times N$ matrix with row i given by $B^{j\top}(\tau_i)/\tau_i$. The superscript j indicates that the coefficients are computed under equivalent martingale measure $\mathbb{Q}^j$.

Given an implied state vector $Y_0(t)$, implied yields for the other k maturities can be computed. In order to do this it is necessary to compute G_0 and G_1, which contain the solutions to the differential equations (7) stacked in the same fashion as in H_0 and H_1. Stack these yields in a vector $\widehat{\widetilde{y}}(t) = -G_0 + G_1 Y_0(t)$. The measurement error is then given by $e_t = \widetilde{y}(t) - \widehat{\widetilde{y}}(t)$. We assume that $e_t \overset{iid}{\sim} MVN(0, C)$, where C is the time-invariant diagonal variance-covariance matrix of the measurement errors e_t. The associated log likelihood is denoted by l_e.

With observation times $t_0, \ldots, t_M$, at each time t_n we can evaluate the joint likelihood of the latent state variables and the log exchange rate conditional on the realizations at t_{n-1} using the likelihood approximation of order one.

$$l_x^{(1)}(x(t_n) \mid x(t_{n-1}), \Delta) = -2\log(2\pi\Delta) - D_v(x(t_n)) + \frac{C_x^{(-1)}(x(t_n) \mid x(t_{n-1})}{\Delta} + \sum_{k=0}^{1} C_x^{(k)}(x(t_n) \mid x(t_{n-1}))\frac{\Delta^k}{k!} \tag{33}$$

where $x(t_n) = \begin{pmatrix} Y_0(t_n) & \log X(t_n) \end{pmatrix}^\top$, $D_v(x(t)) = 1/2\log(\det(\mathrm{Var}(x(t))))$ and in our investigation $\Delta = 1/52$. The coefficients C_x are functions of the instantaneous drift and covariance matrix of latent state variables and the log exchange rate and are computed according to the formulae in Aït-Sahalia (2002).[11] We are interested in the joint likelihood of the log exchange rate with the yields rather than with the latent state variables. The transformation χ between the system of yields and the log exchange rate and the system of latent state variables and the log exchange rate is

$$\chi\begin{pmatrix} y \\ \log X \end{pmatrix} = \chi(w) = \begin{pmatrix} H_1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \cdot \left[\begin{pmatrix} y \\ \log X \end{pmatrix} - \begin{pmatrix} H_0 \\ 0 \end{pmatrix}\right]. \tag{34}$$

The determinant of the Jacobian is

$$\det J = \det\left(\frac{\partial \chi(w)}{\partial w}\right) = \det\begin{pmatrix} H_1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \frac{1}{\det H_1},$$

so that the joint likelihood of yields observed with error and without error with the log exchange rate becomes

$$\sum_{n=1}^{M}\left(l_x^{(1)}(x(t_n) \mid x(t_{n-1}), \Delta) - \log|\det H_1| + l_e(t_n)\right). \tag{35}$$

4.3 The Maximization Technique and Some Practical Considerations

The estimation procedure is subject to a number of complicating factors. First, a non-convex scalar valued function is optimized over roughly thirty parameters which makes it quite unlikely to actually find a global maximum. Second, the objective function, the likelihood function, is highly complex and it is extremely complicated to provide analytic gradients for gradient based solvers.[12] Third, even the

[11] The coefficients C_x are available from the authors upon request.

[12] Recall that the likelihood function involves a matrix that contains the solutions to $N+1$ dimensional differential equations the parameters of which are non linear functions of the parameter vector θ. Additionally, the administrative and computational effort to calculate the derivatives of the likelihood coefficients with respect to the parameter vector would be enormous since

constraints are nonlinear, since stationarity imposes that the real part of the eigenvalues of the drift matrix $\mathcal{K}$ be positive. Finally we encountered difficulties in numerically solving the differential equation (7) for many admissible parameterizations. In this case we set the likelihood function to zero. These considerations let us apply the following procedure for our estimates:

Step 1 Generate J admissible, random starting parameter vectors within a reasonable range. Start J genetic optimization procedures with suitable penalty functions for the constraints, where for *each* call of the likelihood function the implied realizations of the latent state variables are updated as a function of the corresponding parametrization. Parameter vectors with implied state variables *that could not have occurred* are rejected.

Step 2 Take the best solutions from Step 1 according to their likelihood score and employ a gradient based solver (e.g. KNITRO, or donlp2) without updating the state variable vector.

Step 3 Update the state variables corresponding to the solution parameters from Step 2, discard parameters if the implied state variables are not admissible and go to Step 2 as long as the parameter vectors have not converged. Finally, compute the outer product of the gradients.

In our estimation we chose $J = 100$ and an order one approximation of the likelihood function. We found that genetic algorithms were the only tool capable of dealing with the discontinuities that arise when for each iteration the state variable vector is updated. It is noteworthy that the state variables implied by the maximizing parameterizations were all comparable in scale. Also, the achievable likelihood scores are very sensitive to the initial time series of latent state variables. The time series of Y_0 implied by the parametrization for all models can be found in Figure 4.

4.4 Empirical Results

For our empirical investigation, we fix the numbers of factors that describe the joint term structure in the US and the UK to three, i.e. $N = 3$. Dewachter and Maes (2001) give strong evidence that three "international" factors result in a high explanatory power and that the loss in explanatory power compared to a three factor model that models each market separately is rather insignificant.

For each of the four non-nested $A_m(3)$ subfamilies, i.e. $A_0(3)$, $A_1(3)$, $A_2(3)$, and $A_3(3)$, we estimate two representatives. The first representative is preselected following the local factor string in the literature (see Ahn (2004)). Specifically, these models are restricted such that there is one local UK

the coefficient expressions themselves are already quite large.

factor and one local US factor the marginal distributions of which are conditionally and unconditionally independent. Both of the local factors are allowed to affect the common factor by entering its drift, diffusion or both. The common factor on the other hand does not enter the local factors SDEs.[13] The second representative of each of the four subfamilies is constructed to be a pure common factor model. That is, in this second type of models, interest rates across both, the US and the UK, are modelled to be driven by the same (common) set of state variables. Although the local factor and the common factor model specification seem to differ largely, it has to be emphasized that local factor specifications merely represent a number of restrictions on the common factor specification, which is the more general specification. Therefore, each of the $A_m(3)$ models specified as a local factor model is nested in the respective more general $A_m(3)$ common factor model. In the following subsections we will present the results of the model estimations.

4.4.1 Common Factor Specification

The overall likelihoods of the estimated common factor models can be seen in Table 4.4.1.[14] The best model according to its likelihood score is the $A_2(3)$ model followed by the $A_1(3)$ model. The model with worst performance is the pure CSR $A_3(3)$ model. Even with δ unrestricted (as can be seen in Tables 5 and 9) the pure square root model achieved the lowest likelihood score of all models.

To grant a fair comparison of these four non-nested models we additionally compute Akaike Information Criteria (AIC) for all of these models. The ranking order, however, remains the same. Both, the $A_2(3)$ and the $A_1(3)$, are very successful in capturing the first two moments of the yield series (means and volatilities) and are closely reproducing in-sample yields of the US and UK term structure. This is documented in Figure 3 which plots the actual yields against the yields implied by the best model, the $A_2(3)$ model. Although in-sample implied pricing errors are low, the forecasting ability of the models remains to be questioned. We measure the forecasting ability of the models by means of Root Mean Squared Forecast Errors (RMSEs). Duffee (2002) reports that the completely affine market price of risk specification is unable to beat the random walk in forecasting future yields. The RMSEs reported in Table 13 confirm this finding. As for all of the estimated models, the RMSEs for random walk forecasts are, with just a few exceptions, lower than those implied by the estimated models suggesting a rather

[13]For the representative of the $A_1(3)$ class the dependency structure is exactly reversed in order to keep the common factor/local factor specification symmetric.

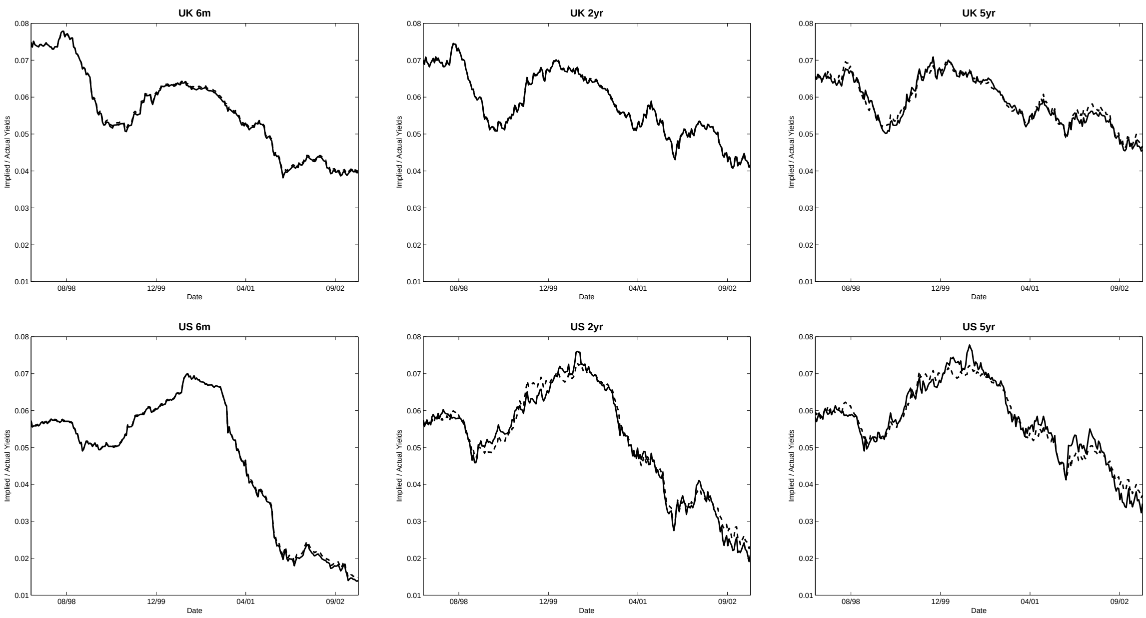
[14]For the specification of the estimated models we refer to Appendix B. The parameter estimates are reported in Tables 5 through 12 in Appendix C.

poor forecasting ability of the class of ATSMs.

Table 2: Comparison of Estimated Common Factor Models.
This table reports the log-likelihoods of all estimated common factor models along with the corresponding Akaike scores. Likelihoods are estimated with closed form likelihood expansions. From equation (35), the total likelihood of a model is given by the sum of three components. AIC denotes the Akaike information criterion. The smaller the AIC value, the "closer" the model is to reality.

Model Type	Free Parameters	$\sum_{n=1}^{M} l_x^{(1)}(t_n)$	$-\sum_{n=1}^{M} \log \lvert \det H_1 \rvert$	$\sum_{n=1}^{M} l_e(t_n)$	Total Log-Likelihood	AIC
$A_0(3)$ CF	32	1,610.4	2,767.2	15,138.5	**19,516.1**	-38,968.2
$A_1(3)$ CF	36	1,307.6	3,503.5	15,036.2	**19,847.3**	-39,622.6
$A_2(3)$ CF	37	2,057.6	3,800.5	14,949.7	**20,807.8**	-41,541.6
$A_3(3)$ CF	38	1,435.0	2,945.0	14,884.4	**19,264.4**	-38,452.8

Figure 3: Implied vs. Actual Yields.
Comparison of the in-sample implied and actual yields for maturities of 6 months, 2 years and 5 years (UK and US). The yields are implied by the parameter estimates of the best model specification, i.e. the $A_2(3)$ common factor model. The dashed line represents the model implied yields, whereas the solid line represents actual yields.

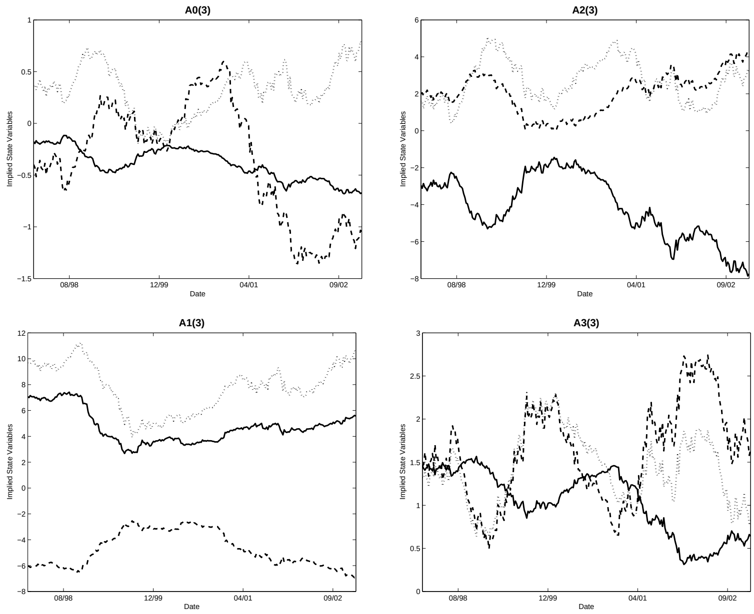


It remains to be answered why the model of choice for most of the previous studies, the $A_3(3)$ model, in which all of the factors exhibit conditional volatility performs the worst relative to all other affine model specifications. This fact can most likely be explained by its very restrictive correlation structure. As pointed out, factors that are governed by CSR processes are theoretically not able to display negative correlations, that is in pure CSR models all state variables are restricted to be positively correlated to each other. However, as a result of our estimation we can observe realizations of latent state variables that are negatively correlated, contradicting the theoretical specification. This further indicates that the $A_3(3)$ class is not the best choice for our data sample. As already Dai and Singleton (2000) have noticed in their single economy specification analysis on the US term structure, the data called for negative correlations among state variables. In Figure 4 we plot the dynamics of the implied state variables for each of the estimated models. With the bare eye it can be verified that the two models which perform best produce state variables that are negatively correlated. This provides strong evidence for negative correlations among the factors driving international bond markets.

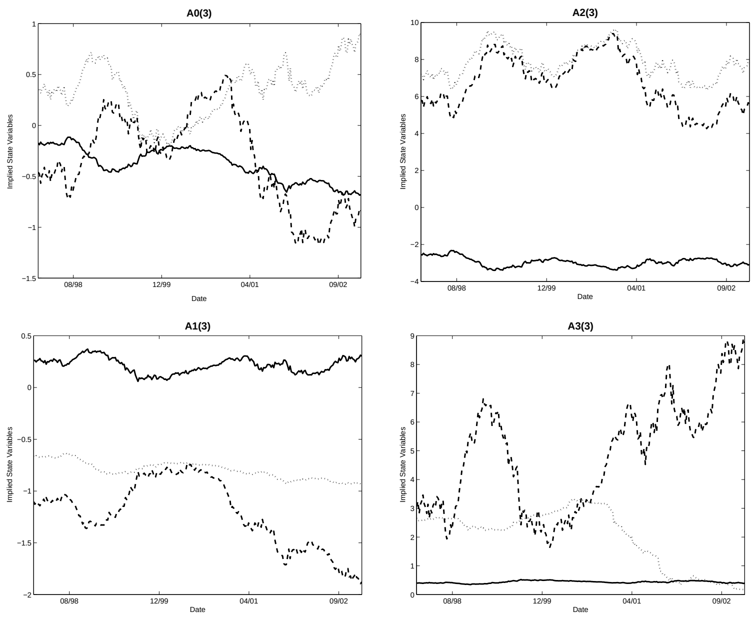
To assess the ability of the models to capture exchange rate movements, we first consider the implied Fama coefficients. Surprisingly, the only model that is able to account for the high unconditional volatility of the exchange rate risk premia is the $A_1(3)$ model. The Fama coefficient over the sample period generated by this model is -2.22, whereas the actual Fama coefficient computed by means of 1 month LIBOR rates amounts to -2.85. The implied coefficients of the other models, however, range from 0.63 for the $A_2(3)$ model to 1 (or close to 1) for the $A_3(3)$ and the $A_0(3)$ model. The ability of the $A_1(3)$ model to forecast exchange rates is again assessed by RMSEs. These are only slightly worse than those of the random walk. The RMSE for the in-sample 1 week ahead forecast of the exchange rate implied by the model is 0.024, whereas the error generated by a random walk is 0.023. For the 4 week ahead forecast the RSMEs are 0.026 and 0.020 for the model and the random walk, respectively. However, although the model is not able to generate smaller forecast errors than the random walk, it predicts slightly better whether the exchange rate is going to appreciate or depreciate in the future. For the 1 week ahead forecast the model is able to predict the right direction of change in 56% of the cases, the random walk is only right in 55%. Regarding the 4 week ahead forecast, the model succeeds in 58%, whereas the random walk only succeeds with a probability of 56%.

Figure 4: Implied State Vectors of the Common Factor Models.
Comparison of the model implied state vectors. Y_3 is represented by the solid line, Y_2 is shown by the dashed line and the trajectory of Y_1 is represented by the dotted line.

Panel A: Common Factor Models



Panel B: Local Factor Models



4.4.2 Local vs. Common Factor Models: Do There Exist Local Factors?

Next, we estimate each of the four affine subfamilies in its local factor specification. In all estimated models, except the $A_1(3)$ model, Y_1 represents the local UK factor, Y_2 is specific to the US and Y_3 is a common factor that influences both countries' interest rates. In the $A_2(3)$ model, we assign the Gaussian factor to represent the common factor for symmetry reasons.[15] In the $A_1(3)$ model we assign, due to symmetry reasons, Y_1 to represent the common factor, Y_2 to be the local UK factor and Y_3 to be local to the US. Further, for the model estimation, we have restricted market prices of risk for factors that are specific to the other country to zero. For example, if $Y_1(t)$ is the local UK factor and $Y_2(t)$ is specific to the US economy, we restrict the market prices of risk λ_2^{UK} and λ_1^{US} to zero.[16]

Table 3: Comparison of Estimated Local Factor Models
This table reports the log-likelihoods of all estimated local factor models along with the corresponding Akaike scores. Likelihoods are estimated with closed form likelihood expansions. From equation (35), the total likelihood of a model is given by the sum of three components. AIC denotes the Akaike information criterion. The smaller the AIC value, the "closer" the model is to reality.

Model Type	Free Parameters	$\sum_{n=1}^{M} l_x^{(1)}(t_n)$	$-\sum_{n=1}^{M} \log\lvert\det H_1\rvert$	$\sum_{n=1}^{M} l_e(t_n)$	Total Log-Likelihood	AIC
$A_0(3)$ LF	27	1,617.1	2,729.6	14,840.1	**19,186.8**	-38,319.6
$A_1(3)$ LF	30	1,798.2	2,055.4	14,953.8	**18,807.4**	-37,554.8
$A_2(3)$ LF	31	932.5	3,098.0	14,844.8	**18,875.3**	-37,688.6
$A_3(3)$ LF	30	1,366.6	2,989.7	13,874.2	**18,230.5**	-36,401.0

As shown in Table 3 the model that performs best according to both, its likelihood and its AIC, is the pure Gaussian model $A_0(3)$ followed by the $A_2(3)$ model. Again, as in the common factor specification, the $A_3(3)$ model has the lowest likelihood value and also ranks last according to its AIC.

In order to compare the common factor specification with its nested local factor counterpart, we compute likelihood ratios (LR). The LRs, reported in Table 4, are exceeding by far the 99% critical values, implying that the common factor specifications are by far better suited to capture dynamics

[15] Remember that in $A_m(N)$ models the $m < N$ factors that are driving the conditional volatility conventionally make up the first m factors, i.e. the factors $Y_1, \ldots, Y_m$ are CSR factors and the remaining factors $Y_{m+1}, \ldots Y_N$ are Gaussian. See Appendix A.

[16] For further details concerning the specification of the local factor models refer to Appendix B.

in the joint term structure and the exchange rate than local factor specifications. Together with the analysis of the common factor specification above, this result provides conclusive evidence against local factors in the joint UK-US term structure and the exchange rate. By definition, a local factor impacts only one economy, has negligible effects on the other and is marginally uncorrelated. The state variables implied by our common factor models have different impacts on the UK and US economies, however none of them are insignificant as can be seen in Tables 9 to 12 in Appendix C. Further, as highlighted above, the results in the common factor specification show the importance of flexible correlation structures among the state variables, that allows some of the factors to be negatively correlated. Altogether, this strongly indicates that local factors play a subordinated role. Similar results, although in another setting are found in Inci and Lu (2004).

Table 4: Log-Likelihood Ratios.
This table reports the log-likelihoods ratios (LR) between the estimated common factor models and their nested local factor counterpart. The likelihood ratios are χ^2 distributed with degrees of freedom corresponding to the difference between the number of free parameters in the common factor specification and the number of free parameters in the respective nested local factor specification. The degrees of freedom are given in the column labelled “df” and the critical value corresponding to the 99% confidence interval is given in the last column.

Model Type	LR	df	Critical Value (99%)
$A_0(3)$	658.6	5	15.09
$A_1(3)$	2079.8	6	16.81
$A_2(3)$	4000.8	7	18.48
$A_3(3)$	2067.8	8	20.09

This issue has important implications for portfolio diversification across international bond markets. Consider a UK-investor who currently holds only UK bonds and considers to additionally invest in currency-hedged US bonds. Since both term structures and the GBP/USD exchange rate seems to be driven by a set of common of factors rather than local factors, the return uncertainties of a currency-hedged bond portfolio across those two countries would have the same sources of risk as his initially undiversified position in UK bonds only. The evidence against local factors, thus, suggests that the investor would not greatly enhance the mean-variance characteristics of his portfolio by additionally investing in a currency-hedged portfolio of US bonds. If there would, however, exist local factors in the US bond market, the investor could achieve significant diversification benefits from holding the currency-hedged bond portfolio in these markets.

5 Conclusion

We investigate the theoretical properties and the empirical performance of international canonical affine term structure models that are driven by a common set of latent state variables. We derive necessary conditions for the correlation and volatility structure of mixture models to accommodate the empirical stylized facts concerning the forward premium puzzle and yield curves and show the tradeoff that is inherent in the specification of ATSMs. Although models with Gaussian processes have the inconvenience of negative interest rates with positive probability and restricting conditional volatility, it seems that they are nevertheless – at least in theory – better suited to capture empirical stylized facts of joint term structure dynamics since they allow for a more flexible correlation structure among the driving state variables.

Using UK and US LIBOR and swap rate data, as well as GBP/USD exchange rate data we estimate common factor, as well as local factor representatives from the $A_0(3), A_1(3), A_2(3), A_3(3)$ models by means of maximum likelihood. We take into account the joint distribution of yields and the exchange rate without assuming normality of the transition densities. Strikingly, the model most widely used in international settings, the $A_3(3)$ provides the worst fit to the data, in the local factor, as well as the common factor setting. This can probably be attributed to the strong negative correlation that seems to be present between the latent factors that drive international economies. The best model overall comes from the common factor $A_2(3)$ class. Forecasts of the log exchange rate with this model and the common factor $A_1(3)$ models are in the range of a drift adjusted random walk, forecasts for the direction of the appreciation/depreciation of the log exchange rate are slightly better than a drift adjusted random walk. Even though this model provides a tight fit of the yield data, we can confirm the finding from Duffee (2002) that yield forecasts with completely affine market prices of risk are not able to outperform a simple random walk forecast. Concerning the forward premium puzzle only the representative from the $A_1(3)$ generates risk premia that are variable enough relative to the short rate differential to generate a negative Fama coefficient. Further, we find strong evidence against the existence of local factors inherent in the UK-US term structure and the exchange rate, indicating that diversification effects are likely to be small when diversifying bond portfolios across these countries.

An interesting question that is left for further research is the modelling with asymmetric factors, where the local factors are modelled with different kinds of processes as well as modelling the joint term structure dynamics with multiple (possibly correlated) common factors. Another open question is whether there is evidence for local factors in the joint term structure and the exchange rate across emerging markets.

References

Ahn, D.-H., 2004, "Common Factors and Local Factors: Implications for Term Structures and Exchange Rates," *Journal of Financial and Quantitative Analysis*, 39, 69–102.

Aït-Sahalia, Y., 2001, "Closed-Form Likelihood Expansions for Multivariate Diffusions," Working paper, Princeton University and NBER.

Aït-Sahalia, Y., 2002, "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.

Aït-Sahalia, Y., and R. Kimmel, 2002, "Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions," Working paper, Princeton University and NBER.

Backus, D. K., S. Foresi, and C. I. Telmer, 2001, "Affine Term Structure Models and the Forward Premium Anomaly," *Journal of Finance*, 56, 279–304.

Bansal, R., 1997, "An Exploration of the Forward Premium Puzzle in Currency Markets," *Review of Financial Studies*, 10, 369–403.

Bekaert, G., 1996, "The Time-Variation of Risk and Return in Foreign Exchange Markets: A General Equilibrium Perspective," *Review of Financial Studies*, 9, 427–470.

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.

Brennan, M. J., and Y. Xia, 2004, "International Capital Markets and Exchange Risk," Working Paper, UCLA, Wharton.

Collin-Dufresne, P., and R. Goldstein, 2002, "Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility," *Journal of Finance*, 57, 1685–1730.

Collin-Dufresne, P., and B. Solnik, 2001, "On the Term Structure of Default Premia in the Swap and LIBOR Markets," *Journal of Finance*, 56, 1095–1115.

Constantinides, G. M., 1992, "A Theory of the Nominal Term Structure of Interest Rates," *Review of Financial Studies*, 5, 531–552.

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

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

Dewachter, H., and K. Maes, 2001, "An Admissible Affine Model for Joint Term Structure Dynamics of Interest Rates," Working paper, KULeuven.

Duarte, J., 2004, "Evaluating an Alternative Risk Preference in Affine Term Structure Models," *Review of Financial Studies*, 17, 379–404.

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

Duffie, D., and M. Huang, 1996, "Swap Rates and Credit Quality," *Journal of Finance*, 51, 921–949.

Duffie, D., and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379–406.

Duffie, D., and K. Singleton, 1997, "An Econometric Model of the Term Structure of Interest Rate Swap Yields," *Journal of Finance*, 52, 1287–1323.

Durham, G. B., and R. A. Gallant, 2001, "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes," Working Paper, University of North Carolina.

Elerian, O., 1998, "A Note on the Existence of Closed Form Conditional Transition Density for the Milstein Scheme," Working Paper, Nuffield College, Oxford University.

Engel, C., 1996, "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence," *Journal of Empirical Finance*, 3, 123–192.

Fama, E., 1984, "Forward and spot exchange rates," *Journal of Monetary Economics*, 14, 319–338.

Fisher, M., and C. Gilles, 1996, "Estimating Exponential-Affine Models of the Term Structure," Working Paper, Federal Reserve Board.

Frühwirth-Schnatter, S., and A. Geyer, 1996, "Bayesian Estimation of Economemtric Multi-Factor Cox-Ingersoll-Ross-Models of the Term Structure of Interest Rates Via MCMC Methods," Working Paper, Vienna University of Economics and BA.

Han, B., and P. Hammond, 2003, "Affine Models of the Joint Dynamics of Exchange Rates and Interest Rates," Working paper, University of Calgary and Stanford University.

Harrison, M., and D. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Security Markets," *Journal of Economic Theory*, 20, 381–408.

Harrison, M., and S. Pliska, 1981, "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and Their Applications*, 11, 215–260.

Hodrick, R., and M. Vassalou, 2002, "Do we need multi-country models to explain exchange rate and interest rate and bond return dynamics?," *Journal of Economic Dynamics & Control*, 26, 1275–1299.

Inci, A. C., and B. Lu, 2004, "Exchange Rates and Interest Rates: Can Term Structure Models Explain Currency Movements?," *Journal of Economic Dynamcis & Control*, 28, 1595–1624.

Jones, C. S., 2002, "Estimating Yield Curves From Asynchronous LIBOR and Swap Quotes," Working paper, University of Southern California.

Litterman, R., and J. A. Scheinkman, 1991, "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1, 54–61.

Nielsen, L. T., and J. Saá-Requejo, 1993, "Exchange Rate and Term Structure Dynamics and the Pricing of Derivative Securities," Working Paper, INSEAD.

Pedersen, A., 1995, "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22.

Piazzesi, M., 2003, "Affine Term Structure Models," Working Paper, UCLA and NBER.

Santa-Clara, P., 1995, "Simulated Likelihood Estimation of Diffusions with an Application to the Short Term Interest Rate," Ph.D Dissertation, INSEAD.

Singleton, K. J., 1994, "Persistence of International Interest Rate Correlation," Working Paper, Prepared for the Berkeley Program in Finance.

Tang, H., and Y. Xia, 2005, "An International Examination of Affine Term Structure Models and the Expectations Hypothesis," Working Paper, Wharton School.

Appendix A: Admissibility and Identification Conditions for Canonical Models

In the canonical models proposed by DS, the m factors that drive the conditional volatility conventionally make up the first block in the factor vector, such that $Y(t) = \left(Y^B_{m\times 1}, Y^D_{(N-m)\times 1}\right)^\top$. Here, block B denotes the square root part of the vector of state variables and D denotes the Gaussian part. The coefficient matrices of the factor dynamics in equation (2) are:

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}^{BB}_{m\times m} & 0_{m\times(N-m)} \\ \mathcal{K}^{DB}_{(N-m)\times m} & \mathcal{K}^{DD}_{(N-m)\times(N-m)} \end{pmatrix} \tag{36}$$

for $m > 0$, $\mathcal{K}$ upper or lower triangular for $m = 0$

$$\Theta = \begin{pmatrix} \Theta^B_{m\times 1} \\ 0_{(N-m)\times 1} \end{pmatrix} \tag{37}$$

$$\Sigma = I_{N\times N} \tag{38}$$

$$\alpha = \begin{pmatrix} 0_{m\times 1} \\ 1_{(N-m)\times 1} \end{pmatrix} \tag{39}$$

$$\mathcal{B} = \begin{pmatrix} I_{m\times m} & \mathcal{B}^{BD}_{m\times(N-m)} \\ 0_{(N-m)\times m} & 0_{(N-m)\times(N-m)} \end{pmatrix} \tag{40}$$

$$S(t)_{ii} = \alpha_i + \beta_i^\top Y(t), \tag{41}$$

where β_i represents the i-th column of $\mathcal{B}$ and $S(t)$ is diagonal. Further, the coefficients in equation (1) and in equations (36) - (41) are subject to the following admissibility conditions in DS:

$$[\delta^d_Y]_j \geq 0, \quad [\delta^f_Y]_j \geq 0, \quad m+1 \leq j \leq N$$

$$\mathcal{K}_i\Theta = \sum_{j=1}^{m} \mathcal{K}_{ij}\Theta_j > 0, \quad 1 \leq i \leq m,$$

$$\mathcal{K}_{ij} \leq 0, \quad 1 \leq j \leq m,$$

$$\Theta_i \geq 0, \quad 1 \leq i \leq m,$$

$$\mathcal{B}_{ij} \geq 0, \quad 1 \leq i \leq m, \quad m+1 \leq j \leq N.$$

Appendix B: Model Descriptions

Local Factor Models

In the subsequent model descriptions we denote for notational convenience $(\lambda_i^{UK})^2 - (\lambda_i^{US})^2 = \lambda_i^{\blacksquare}$ and $(\lambda_i^{UK} - \lambda_i^{US}) = \lambda_i^{\bigstar}$. In all estimated models, except the $A_1(3)$ model, Y_1 represents the local UK factor, Y_2 is specific to the US and Y_3 is a common factor that influences both countries' interest rates. In the $A_2(3)$ model, we assign the Gaussian factor to represent the common factor for symmetry reasons. In the $A_1(3)$ model we assign, due to symmetry reasons, Y_1 to represent the common factor, Y_2 to be the local UK factor and Y_3 to be local to the US. Further, for the model estimation, we have restricted market prices of risk for factors that are specific to the other country to zero. For example, if $Y_1(t)$ is the local UK factor and $Y_2(t)$ is specific to the US economy, we restrict the market prices of risk λ_2^{UK} and λ_1^{US} to zero.

$A_3(3)$

$$r^{UK}(t) = \delta_0^{\text{UK}} + \delta_1^{\text{UK}} Y_1(t) + \delta_3^{\text{UK}} Y_3(t), \quad r^{\text{US}}(t) = \delta_0^{\text{US}} + \delta_2^{\text{US}} Y_2(t) + \delta_3^{\text{US}} Y_3(t)$$

$$\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \\ d\log X(t) \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11}(\theta_1 - Y_1(t)) \\ \mathcal{K}_{22}(\theta_2 - Y_2(t)) \\ \mathcal{K}_{31}(\theta_1 - Y_1(t)) + \mathcal{K}_{32}(\theta_2 - Y_2(t)) + \mathcal{K}_{33}(\theta_3 - Y_3(t)) \\ r^{UK}(t) - r^{US}(t) + \frac{1}{2} \sum_{i=1}^{3} \lambda_i^{\blacksquare} Y_i(t) + (\Phi^{UK})^2 - (\Phi^{US})^2 \end{pmatrix} dt$$

$$+ \begin{pmatrix} \sqrt{Y_1(t)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{Y_2(t)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{Y_3(t)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^{\bigstar}\sqrt{Y_1(t)} & \lambda_2^{\bigstar}\sqrt{Y_2(t)} & \lambda_3^{\bigstar}\sqrt{Y_3(t)} & \Phi^{UK} - \Phi^{US} \end{pmatrix} \cdot \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dB_1(t) \\ dB_2(t) \\ dB_3(t) \\ dB_4(t) \end{pmatrix}$$

$A_2(3)$

$$r^{\text{UK}}(t) = \delta_0^{\text{UK}} + \delta_1^{\text{UK}} Y_1(t) + \delta_3^{\text{UK}} Y_3(t), \quad r^{\text{US}}(t) = \delta_0^{\text{US}} + \delta_2^{\text{US}} Y_2(t) + \delta_3^{\text{US}} Y_3(t)$$

$$\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \\ d\log X(t) \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11}(\theta_1 - Y_1(t)) \\ \mathcal{K}_{22}(\theta_2 - Y_2(t)) \\ \mathcal{K}_{31}(\theta_1 - Y_1(t)) + \mathcal{K}_{32}(\theta_2 - Y_2(t)) - \mathcal{K}_{33} Y_3(t) \\ r^{UK}(t) - r^{US}(t) + \frac{1}{2} \sum_{i=1}^{2} \lambda_i^{\blacksquare} Y_i(t) + \lambda_3^{\blacksquare}\phi(t) + (\Phi^{UK})^2 - (\Phi^{US})^2 \end{pmatrix} dt$$

$$+ \begin{pmatrix} \sqrt{Y_1(t)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{Y_2(t)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\phi(t)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^{\bigstar}\sqrt{Y_1(t)} & \lambda_2^{\bigstar}\sqrt{Y_2(t)} & \lambda_3^{\bigstar}\sqrt{\phi(t)} & \Phi^{UK} - \Phi^{US} \end{pmatrix} \cdot \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dB_1(t) \\ dB_2(t) \\ dB_3(t) \\ dB_4(t) \end{pmatrix}$$

with $\phi(t) = 1 + \beta_{13} Y_1(t) + \beta_{23} Y_2(t)$.

$A_1(3)$

$$r^{\mathrm{UK}}(t) = \delta_0^{\mathrm{UK}} + \delta_1^{\mathrm{UK}} Y_1(t) + \delta_2^{\mathrm{UK}} Y_2(t), \quad r^{\mathrm{US}}(t) = \delta_0^{\mathrm{US}} + \delta_1^{\mathrm{US}} Y_1(t) + \delta_3^{\mathrm{US}} Y_3(t)$$

$$\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \\ d\log X(t) \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11}(\theta_1 - Y_1(t)) \\ -\mathcal{K}_{22} Y_2(t) \\ -\mathcal{K}_{33} Y_3(t) \\ r^{UK}(t) - r^{US}(t) + \frac{1}{2} \quad \lambda_1^{\blacksquare} Y_1(t) + \lambda_2^{\blacksquare} \phi_1(t) + \lambda_3^{\blacksquare} \phi_2(t) + (\Phi^{UK})^2 - (\Phi^{US})^2 \end{pmatrix} dt$$

$$+ \begin{pmatrix} \sqrt{Y_1(t)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\phi_1(t)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\phi_2(t)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^{\bigstar} \sqrt{Y_1(t)} & \lambda_2^{\bigstar} \sqrt{\phi_1(t)} & \lambda_3^{\bigstar} \sqrt{\phi_2(t)} & \Phi^{UK} - \Phi^{US} \end{pmatrix} \cdot \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dB_1(t) \\ dB_2(t) \\ dB_3(t) \\ dB_4(t) \end{pmatrix}$$

with $\phi_1(t) = 1 + \beta_{12}$ and $\phi_2(t) = 1 + \beta_{13}$.

$A_0(3)$

$$r^{\mathrm{UK}}(t) = \delta_0^{\mathrm{UK}} + \delta_1^{\mathrm{UK}} Y_1(t) + \delta_3^{\mathrm{UK}} Y_3(t), \quad r^{\mathrm{US}}(t) = \delta_0^{\mathrm{US}} + \delta_2^{\mathrm{US}} Y_2(t) + \delta_3^{\mathrm{US}} Y_3(t)$$

$$\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \\ d\log X(t) \end{pmatrix} = \begin{pmatrix} -\mathcal{K}_{11} Y_1(t) \\ -\mathcal{K}_{22} Y_2(t) \\ -(\mathcal{K}_{31} Y_1(t) + \mathcal{K}_{32} Y_2(t) + \mathcal{K}_{33} Y_3(t)) \\ r^{UK}(t) - r^{US}(t) + \frac{1}{2} \quad \sum_{i=1}^{3} \lambda_i^{\blacksquare} + (\Phi^{UK})^2 - (\Phi^{US})^2 \end{pmatrix} dt$$

$$+ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^{\bigstar} & \lambda_2^{\bigstar} & \lambda_3^{\bigstar} & \Phi^{UK} - \Phi^{US} \end{pmatrix} \cdot \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dB_1(t) \\ dB_2(t) \\ dB_3(t) \\ dB_4(t) \end{pmatrix}$$

Common Factor Models

For all common factor models we have: $r^i(t) = \delta_0^i + \delta_1^i Y_1(t) + \delta_2^i Y_2(t) + \delta_3^i Y_3(t), i \in \{\text{UK}, \text{US}\}$.

$A_3(3)$

$$\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \\ d\log X(t) \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11}(\theta_1 - Y_1(t)) + \mathcal{K}_{12}(\theta_2 - Y_2(t)) + \mathcal{K}_{13}(\theta_3 - Y_3(t)) \\ \mathcal{K}_{21}(\theta_1 - Y_1(t)) + \mathcal{K}_{22}(\theta_2 - Y_2(t)) + \mathcal{K}_{23}(\theta_3 - Y_3(t)) \\ \mathcal{K}_{31}(\theta_1 - Y_1(t)) + \mathcal{K}_{32}(\theta_2 - Y_2(t)) + \mathcal{K}_{33}(\theta_3 - Y_3(t)) \\ r^{UK}(t) - r^{US}(t) + \frac{1}{2} \quad \sum_{i=1}^{3} \lambda_i^{\blacksquare} Y_i(t) + (\Phi^{UK})^2 - (\Phi^{US})^2 \end{pmatrix} dt$$

$$+ \begin{pmatrix} \sqrt{Y_1(t)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{Y_2(t)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{Y_3(t)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^{\bigstar}\sqrt{Y_1(t)} & \lambda_2^{\bigstar}\sqrt{Y_2(t)} & \lambda_3^{\bigstar}\sqrt{Y_3(t)} & \Phi^{UK} - \Phi^{US} \end{pmatrix} \cdot \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dB_1(t) \\ dB_2(t) \\ dB_3(t) \\ dB_4(t) \end{pmatrix}$$

$A_2(3)$

$$\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \\ d\log X(t) \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11}(\theta_1 - Y_1(t)) + \mathcal{K}_{12}(\theta_2 - Y_2(t)) \\ \mathcal{K}_{21}(\theta_1 - Y_1(t)) + \mathcal{K}_{22}(\theta_2 - Y_2(t)) \\ \mathcal{K}_{31}(\theta_1 - Y_1(t)) + \mathcal{K}_{32}(\theta_2 - Y_2(t)) - \mathcal{K}_{33} Y_3(t) \\ r^{UK}(t) - r^{US}(t) + \frac{1}{2} \quad \sum_{i=1}^{2} \lambda_i^{\blacksquare} Y_i(t) + \lambda_3^{\blacksquare}\phi(t) + (\Phi^{UK})^2 - (\Phi^{US})^2 \end{pmatrix} dt$$

$$+ \begin{pmatrix} \sqrt{Y_1(t)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{Y_2(t)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\phi(t)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^{\bigstar}\sqrt{Y_1(t)} & \lambda_2^{\bigstar}\sqrt{Y_2(t)} & \lambda_3^{\bigstar}\sqrt{\phi(t)} & \Phi^{UK} - \Phi^{US} \end{pmatrix} \cdot \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dB_1(t) \\ dB_2(t) \\ dB_3(t) \\ dB_4(t) \end{pmatrix}$$

with $\phi(t) = 1 + \beta_{13} Y_1(t) + \beta_{23} Y_2(t)$.

$A_1(3)$

$$\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \\ d\log X(t) \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11}(\theta_1 - Y_1(t)) \\ -(\mathcal{K}_{21} Y_1(t) + \mathcal{K}_{22} Y_2(t) + \mathcal{K}_{23} Y_3(t)) \\ -(\mathcal{K}_{31} Y_1(t) + \mathcal{K}_{32} Y_2(t) + \mathcal{K}_{33} Y_3(t)) \\ r^{UK}(t) - r^{US}(t) + \frac{1}{2} \quad \lambda_1^{\blacksquare} Y_1(t) + \lambda_2^{\blacksquare}\phi_1(t) + \lambda_3^{\blacksquare}\phi_2(t) + (\Phi^{UK})^2 - (\Phi^{US})^2 \end{pmatrix} dt$$

$$+ \begin{pmatrix} \sqrt{Y_1(t)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\phi_1(t)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\phi_2(t)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^{\bigstar}\sqrt{Y_1(t)} & \lambda_2^{\bigstar}\sqrt{\phi_1(t)} & \lambda_3^{\bigstar}\sqrt{\phi_2(t)} & \Phi^{UK} - \Phi^{US} \end{pmatrix} \cdot \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dB_1(t) \\ dB_2(t) \\ dB_3(t) \\ dB_4(t) \end{pmatrix}$$

with $\phi_1(t) = 1 + \beta_{12}$ and $\phi_2(t) = 1 + \beta_{13}$.

$A_0(3)$

$$
\begin{pmatrix} dY_1(t) \\ dY_2(t) \\ dY_3(t) \\ d\log X(t) \end{pmatrix} = \begin{pmatrix} -(\mathcal{K}_{21} Y_1(t)) \\ -(\mathcal{K}_{21} Y_1(t) + \mathcal{K}_{22} Y_2(t)) \\ -(\mathcal{K}_{31} Y_1(t) + \mathcal{K}_{32} Y_2(t) + \mathcal{K}_{33} Y_3(t)) \\ r^{UK}(t) - r^{US}(t) + \frac{1}{2} \quad \sum_{i=1}^{3} \lambda_i^{\blacksquare} + (\Phi^{UK})^2 - (\Phi^{US})^2 \end{pmatrix} dt
$$

$$
+ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^{\bigstar} & \lambda_2^{\bigstar} & \lambda_3^{\bigstar} & \Phi^{UK} - \Phi^{US} \end{pmatrix} \cdot \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dB_1(t) \\ dB_2(t) \\ dB_3(t) \\ dB_4(t) \end{pmatrix}
$$

Appendix C: Estimated Model Parameters

Table 5: Parameter Estimates of the $A_3(3)$ Local Factor Model.
This table reports the parameter estimates of the local factor $A_3(3)$ model. Parameters are estimated with closed form likelihood expansions. Asymptotic standard errors for the parameters are given below the respective parameter value in parentheses. On the left side of the table, parameters that are restricted to zero by the local factor specification are marked by —. The right-hand side of the table gives the standard deviation of the yields' measurement error and the corresponding standard error in parentheses. Yields that are assumed to be observed exactly are marked with "fixed".

	UK	US
δ_0	0.2336 (0.0031)	-0.0054 (0.0020)

	Index (i)				Country	
	1	2	3		UK	US
$\mathcal{K}_{1i}$	0.2364 (0.0439)	—	—	$\sigma(0.25)$	0.0006 (3e-05)	0.0016 (0.0002)
$\mathcal{K}_{2i}$	—	0.2459 (0.0259)	—	$\sigma(0.5)$	0 (fixed)	0 (fixed)
$\mathcal{K}_{3i}$	-0.0002 (0.0050)	-0.0168 (0.0010)	2.3819 (8.4527)	$\sigma(1)$	0.0013 (0.0002)	0.0020 (0.0003)
Θ_i	4.0019 (0.3791)	4.1985 (0.3791)	0.4480 (1.5883)	$\sigma(2)$	0 (fixed)	0.0030 (0.0007)
δ_i^{UK}	-0.0073 (0.0002)	—	-0.3335 (0.0066)	$\sigma(3)$	0.0008 (2e-05)	0.0034 (0.0010)
$\lambda_{1i}^{\mathrm{UK}}$	0.0203 (0.0202)	—	0.1283 (8.4437)	$\sigma(4)$	0.0009 (7e-05)	0.0037 (0.0018)
δ_i^{US}	—	0.0186 (0.0002)	0.0291 (0.0046)	$\sigma(5)$	0.0011 (1e-06)	0.0044 (0.0027)
$\lambda_{1i}^{\mathrm{US}}$	—	0.0355 (0.0186)	0.0799 (8.4150)			
$(\Phi^{\mathrm{UK}})^2 - (\Phi^{\mathrm{US}})^2$	-0.0380745 (0.366694)					
$(\Phi^{\mathrm{UK}} - \Phi^{\mathrm{US}})^2$	0 (0.014411)					

Table 6: Parameter Estimates of the $A_2(3)$ Local Factor Model.
This table reports the parameter estimates of the local factor $A_2(3)$ model. Parameters are estimated with closed form likelihood expansions. Asymptotic standard errors for the parameters are given below the respective parameter value in parentheses. On the left side of the table, parameters that are restricted to zero by the local factor specification are marked by —. The right-hand side of the table gives the standard deviation of the yields' measurement error and the corresponding standard error in parentheses. Yields that are assumed to be observed exactly are marked with "fixed".

	UK	US
δ_0	0.1425 (0.0020)	0.1462 (0.0021)

	Index (i)				Country	
	1	2	3		UK	US
$\mathcal{K}_{1i}$	0.3646 (0.0324)	—	—	$\sigma(0.25)$	0.0006 (3e-05)	0.0011 (0.0001)
$\mathcal{K}_{2i}$	—	0.3448 (0.0457)	—	$\sigma(0.5)$	0 (fixed)	0 (fixed)
$\mathcal{K}_{3i}$	0.6026 (0.5318)	-0.2375 (0.5940)	1.0800 (0.0234)	$\sigma(1)$	0.0015 (0.0002)	0.0014 (0.0002)
Θ_i	4.6751 (0.9712)	4.5911 (1.3566)	—	$\sigma(2)$	0 (fixed)	0.0021 (0.0004)
β_{1i}	—	—	0 (0.3340)	$\sigma(3)$	0.0008 (2e-05)	0.0023 (0.0006)
β_{2i}	—	—	0 (0.3730)	$\sigma(4)$	0.0009 (6e-05)	0.0025 (0.0008)
δ_i^{UK}	0.0271 (0.0003)	—	0.1014 (0.0011)	$\sigma(5)$	0.0012 (0.0001)	0.0029 (0.0013)
$\lambda_{1i}^{\mathrm{UK}}$	-0.0609 (0.0166)	—	1.5745 (0.8164)			
δ_i^{US}	—	0.0198 (0.0003)	0.0790 (0.0009)			
$\lambda_{1i}^{\mathrm{US}}$	—	-0.0152 (0.0192)	1.6197 (0.8689)			
$(\Phi^{\mathrm{UK}})^2 - (\Phi^{\mathrm{US}})^2$	-0.178905 (1.26083)					
$(\Phi^{\mathrm{UK}} - \Phi^{\mathrm{US}})^2$	0 (0.0232314)					

Table 7: Parameter Estimates of the $A_1(3)$ Local Factor Model.
This table reports the parameter estimates of the local factor $A_1(3)$ model. Parameters are estimated with closed form likelihood expansions. Asymptotic standard errors for the parameters are given below the respective parameter value in parentheses. On the left side of the table, parameters that are restricted to zero by the local factor specification are marked by —. The right-hand side of the table gives the standard deviation of the yields' measurement error and the corresponding standard error in parentheses. Yields that are assumed to be observed exactly are marked with "fixed".

	UK	US
δ_0	0.1634 (0.0012)	0.1004 (0.0008)

	Index (i)				Country	
	1	2	3		UK	US
$\mathcal{K}_{1i}$	3.7371 (3.6164)	0.9195 (0.0814)	—	$\sigma(0.25)$	0.0007 (4e-05)	0.0011 (0.0001)
$\mathcal{K}_{2i}$	—	0.4209 (0.0087)	—	$\sigma(0.5)$	0 (fixed)	0 (fixed)
$\mathcal{K}_{3i}$	1.1311 (0.1167)	—	0.3470 (0.0124)	$\sigma(1)$	0.0016 (0.0002)	0.0019 (0.0003)
Θ_i	0.0191 (0.0186)	—	—	$\sigma(2)$	0 (fixed)	0.0021 (0.0003)
β_{1i}	—	0 (0.5185)	0 (2.0438)	$\sigma(3)$	0.0011 (6e-05)	0.0022 (0.0004)
δ_i^{UK}	0.0577 (0.0015)	0.1518 (0.0015)	—	$\sigma(4)$	0.0009 (7e-05)	0.0024 (0.0007)
$\lambda_{1i}^{\mathrm{UK}}$	-2.7699 (3.6359)	0.0739 (0.0189)	—	$\sigma(5)$	0.0014 (0.0002)	0.0025 (0.0007)
δ_i^{US}	0.1137 (0.0029)	—	0.0650 (0.0010)			
$\lambda_{1i}^{\mathrm{US}}$	-2.9704 (3.6270)	—	-0.0005 (0.0245)			
$(\Phi^{\mathrm{UK}})^2 - (\Phi^{\mathrm{US}})^2$	0.204691 (0.319961)					
$(\Phi^{\mathrm{UK}} - \Phi^{\mathrm{US}})^2$	0 (0.00799437)					

Table 8: Parameter Estimates of the $A_0(3)$ Local Factor Model.
This table reports the parameter estimates of the local factor $A_0(3)$ model. Parameters are estimated with closed form likelihood expansions. Asymptotic standard errors for the parameters are given below the respective parameter value in parentheses. On the left side of the table, parameters that are restricted to zero by the local factor specification are marked by —. The right-hand side of the table gives the standard deviation of the yields' measurement error and the corresponding standard error in parentheses. Yields that are assumed to be observed exactly are marked with "fixed".

	UK	US
δ_0	0.0844 (0.0004)	0.0797 (0.0004)

	Index (i)				Country	
	1	2	3		UK	US
$\mathcal{K}_{1i}$	0.3604 (0.0412)	—	—	$\sigma(0.25)$	0.0008 (5e-05)	0.0013 (0.0001)
$\mathcal{K}_{2i}$	—	0.3266 (0.0200)	—	$\sigma(0.5)$	0 (fixed)	0 (fixed)
$\mathcal{K}_{3i}$	0.4870 (0.0163)	0.0680 (0.0024)	0.7595 (0.0147)	$\sigma(1)$	0.0012 (0.0001)	0.0016 (0.0002)
Θ_i	—	—	—	$\sigma(2)$	0 (fixed)	0.0020 (0.0004)
δ_i^{UK}	0.0303 (0.0008)	—	0.1059 (0.0012)	$\sigma(3)$	0.0007 (2e-05)	0.0023 (0.0005)
$\lambda_{1i}^{\mathrm{UK}}$	-0.0604 (0.0206)	—	0.0420 (0.0065)	$\sigma(4)$	0.0009 (7e-05)	0.0025 (0.0010)
δ_i^{US}	—	0.0212 (0.0006)	0.0759 (0.0009)	$\sigma(5)$	0.0011 (0.0001)	0.0027 (0.0011)
$\lambda_{1i}^{\mathrm{US}}$	—	-0.0450 (0.0277)	0.0341 (0.0075)			
$(\Phi^{\mathrm{UK}})^2 - (\Phi^{\mathrm{US}})^2$	0 (27.913961)					
$(\Phi^{\mathrm{UK}} - \Phi^{\mathrm{US}})^2$	0 (0.26785)					

Table 9: Parameter Estimates of the $A_3(3)$ Common Factor Model.
This table reports the parameter estimates of the common factor $A_3(3)$ model. Parameters are estimated with closed form likelihood expansions. Asymptotic standard errors for the parameters are given below the respective parameter value in parentheses. The right-hand side of the table gives the standard deviation of the yields' measurement error and the corresponding standard error in parentheses. Yields that are assumed to be observed exactly are marked with "fixed".

	UK	US
δ_0	-0.0316	-0.0031
	(0.0013)	(0.0024)

	Index (i)				Country	
	1	2	3		UK	US
$\mathcal{K}_{1i}$	0.5141	0	-0.1086	$\sigma(0.25)$	0.0006	0.0011
	(0.0439)	(0.0166)	(0.0192)		(3e-05)	(0.0001)
$\mathcal{K}_{2i}$	0	0.5960	0	$\sigma(0.5)$	0	0
	(0.0190)	(0.0259)	(0.0120)		(fixed)	(fixed)
$\mathcal{K}_{3i}$	-0.4174	0	0.7643	$\sigma(1)$	0.0010	0.0018
	(0.0117)	(0.0137)	(0.0197)		(9e-05)	(0.0003)
Θ_i	1.1406	1.5224	0.7954	$\sigma(2)$	0	0.0022
	(0.0302)	(0.0406)	(0.0320)		(fixed)	(0.0004)
δ_i^{UK}	-0.0123	0.0260	0.0597	$\sigma(3)$	0.0008	0.0023
	(0.0004)		(0.0006)		(2e-05)	(0.0006)
$\lambda_{1i}^{\mathrm{UK}}$	-0.1108	0.0131	-0.2017	$\sigma(4)$	0.0009	0.0026
	(0.0147)		(0.0259)		(8e-05)	(0.0010)
δ_i^{US}	0.0332	-0.0208	0.0313	$\sigma(5)$	0.0011	0.0027
	(0.0009)	(0.0012)	(0.0015)		(0.0001)	(0.0011)
$\lambda_{1i}^{\mathrm{US}}$	-0.1896	-0.0028	-0.2302			
	(0.0189)	(0.0286)	(0.0328)			
$(\Phi^{\mathrm{UK}})^2 - (\Phi^{\mathrm{US}})^2$	0.0249739					
	(0.0805578)					
$(\Phi^{\mathrm{UK}} - \Phi^{\mathrm{US}})^2$	0					
	(0.00321977)					

Table 10: Parameter Estimates of the $A_2(3)$ Common Factor Model.
This table reports the parameter estimates of the common factor $A_2(3)$ model. Parameters are estimated with closed form likelihood expansions. Asymptotic standard errors for the parameters are given below the respective parameter value in parentheses. The right-hand side of the table gives the standard deviation of the yields' measurement error and the corresponding standard error in parentheses. Yields that are assumed to be observed exactly are marked with "fixed".

	UK	US
δ_0	0.0854	0.0743
	(0.0006)	(0.0009)

	Index (i)				Country	
	1	2	3		UK	US
$\mathcal{K}_{1i}$	0.7051	-0.0002	—	$\sigma(0.25)$	0.0007	0.0012
	(0.4205)	(0.0543)			(5e-05)	(0.0002)
$\mathcal{K}_{2i}$	0	0.7499	—	$\sigma(0.5)$	0	0
	(0.0343)	(1.2743)			(fixed)	(fixed)
$\mathcal{K}_{3i}$	0.0998	0.7409	1.0372	$\sigma(1)$	0.0010	0.0017
	(0.1374)	(0.3526)	(0.0155)		(9e-05)	(0.0003)
Θ_i	2.2788	0.6668	—	$\sigma(2)$	0	0.0020
	(1.4343)	(1.1490)			(fixed)	(0.0004)
β_{1i}	—	—	0.1088	$\sigma(3)$	0.0007	0.0021
			(0.1242)		(2e-05)	(0.0004)
β_{2i}	—	—	0.5543	$\sigma(4)$	0.0009	0.0023
			(0.2050)		(9e-05)	(0.0008)
δ_i^{UK}	0.0001	0.0259	0.0202	$\sigma(5)$	0.0012	0.0024
	(0.0001)	(0.0004)	(0.0002)		(0.0001)	(0.0008)
$\lambda_{1i}^{\mathrm{UK}}$	-0.3455	-0.3967	1.0832			
	(0.4235)	(1.2513)	(0.5216)			
δ_i^{US}	0.0067	0.0139	0.0185			
	(0.0002)	(0.0005)	(0.0003)			
$\lambda_{1i}^{\mathrm{US}}$	-0.3638	-0.3708	1.1181			
	(0.4235)	(1.2860)	(0.4764)			
$(\Phi^{\mathrm{UK}})^2 - (\Phi^{\mathrm{US}})^2$	0.117586					
	(0.219478)					
$(\Phi^{\mathrm{UK}} - \Phi^{\mathrm{US}})^2$	0.00129949					
	(0.00470785)					

Table 11: Parameter Estimates of the $A_1(3)$ Common Model.
This table reports the parameter estimates of the common factor $A_1(3)$ model. Parameters are estimated with closed form likelihood expansions. Asymptotic standard errors for the parameters are given below the respective parameter value in parentheses. The right-hand side of the table gives the standard deviation of the yields' measurement error and the corresponding standard error in parentheses. Yields that are assumed to be observed exactly are marked with "fixed".

	UK	US
δ_0	0.0383	0.0583
	(0.0005)	(0.0009)

	Index (i)				Country	
	1	2	3		UK	US
$\mathcal{K}_{1i}$	0.5641	—	—	$\sigma(0.25)$	0.0007	0.0010
	(0.1867)				(5e-05)	(0.0001)
$\mathcal{K}_{2i}$	0.0062	0.3602	-0.0944	$\sigma(0.5)$	0	0
	(0.0433)	(0.0275)	(0.0318)		(fixed)	(fixed)
$\mathcal{K}_{3i}$	0.5796	0.4092	0.3570	$\sigma(1)$	0.0001	0.0015
	(0.0331)	(0.0206)	(0.0184)		(0.0001)	(0.0002)
Θ_i	7.0338	—	—	$\sigma(2)$	0	0.0019
	(1.0847)				(fixed)	(0.0004)
β_{1i}	—	0	0	$\sigma(3)$	0.0008	0.0022
		(0.0182)	(0.0170)		(2e-05)	(0.0005)
δ_i^{UK}	0.0006	0.0146	0.0172	$\sigma(4)$	0.0008	0.0024
	(0.0003)	(0.0003)	(0.0003)		(7e-05)	(0.0008)
λ_{1i}^{UK}	0.0104	1.5792	0.0846	$\sigma(5)$	0.0011	0.0023
	(0.1872)	(1.1406)	(1.1709)		(0.0001)	(0.0009)
δ_i^{US}	0.0045	0.0275	0.0172			
	(0.0003)	(0.004)	(0.0004)			
λ_{1i}^{US}	0.0310	1.6265	0.0653			
	(0.1867)	(1.0449)	(1.5442)			
$(\Phi^{UK})^2 - (\Phi^{US})^2$	0.555552					
	(1.04452)					
$(\Phi^{UK} - \Phi^{US})^2$	0					
	(0.0171855)					

Table 12: Parameter Estimates of the $A_0(3)$ Common Factor Model.
This table reports the parameter estimates of the common factor $A_0(3)$ model. Parameters are estimated with closed form likelihood expansions. Asymptotic standard errors for the parameters are given below the respective parameter value in parentheses. The right-hand side of the table gives the standard deviation of the yields' measurement error and the corresponding standard error in parentheses. Yields that are assumed to be observed exactly are marked with "fixed".

	UK	US
δ_0	0.0838	0.0765
	(0.0004)	(0.0005)

	Index (i)				Country	
	1	2	3		UK	US
$\mathcal{K}_{1i}$	0.5474	—	—	$\sigma(0.25)$	0.0006	0.0010
	(0.0530)				(3e-05)	(0.0001)
$\mathcal{K}_{2i}$	0.6993	0.4084	—	$\sigma(0.5)$	0	0
	(0.0630)	(0.0170)			(fixed)	(fixed)
$\mathcal{K}_{3i}$	0.4900	0.0371	0.6296	$\sigma(1)$	0.0009	0.0015
	(0.0138)	(0.0023)	(0.0106)		(6e-05)	(0.0002)
Θ_i	—	—	—	$\sigma(2)$	0	0.0019
					(fixed)	(0.0004)
δ_i^{UK}	0.0295	0	0.0991	$\sigma(3)$	0.0007	0.0022
	(0.0005)	(0.0002)	(0.0010)		(2e-05)	(0.0005)
$\lambda_{1i}^{\mathrm{UK}}$	-0.0905	-0.3062	0.0005	$\sigma(4)$	0.0008	0.0024
	(0.1574)	(1.4582)	(0.0388)		(6e-05)	(0.0009)
δ_i^{US}	0.0027	0.0207	0.0665	$\sigma(5)$	0.0011	0.0024
	(0.0008)	(0.0004)	(0.0012)		(0.0001)	(0.0010)
$\lambda_{1i}^{\mathrm{US}}$	-0.0233	-0.2982	0.0371			
	(1.0654)	(20.4093)	(6.3749)			
$(\Phi^{\mathrm{UK}})^2 - (\Phi^{\mathrm{US}})^2$	-0.0282269					
	(12.2683)					
$(\Phi^{\mathrm{UK}} - \Phi^{\mathrm{US}})^2$	0					
	(0.00498988)					

Table 13: In Sample RMSEs.
For yield and log exchange rate forecasts the system of latent state variables is simulated with the Euler discretization scheme from equation (2). The starting values for the simulation are the state variables implied by the same parameter vector that governs the evolution of the system.

Bond Maturity	Forecast horizon	CF $A_2(3)$	RW
UK 0.5	3m	0.000277	0.000449
UK 0.5	6m	0.000333	0.000636
UK 2	3m	0.000474	0.000485
UK 2	6m	0.000697	0.000580
US 0.5	3m	0.000606	0.000687
US 0.5	6m	0.000925	0.00106
US 2	3m	0.000728	0.000606
US 2	6m	0.000941	0.000809
UK 0.25	3m	0.000273	0.000431
UK 0.25	6m	0.000297	0.000630
UK 1	3m	0.000376	0.000537
UK 1	6m	0.000455	0.000693
UK 3	3m	0.000545	0.000427
UK 3	6m	0.000875	0.000520
UK 4	3m	0.000609	0.000383
UK 4	6m	0.000995	0.000462
UK 5	3m	0.000588	0.000336
UK 5	6m	0.00101	0.000403
US 0.25	3m	0.000465	0.000688
US 0.25	6m	0.000817	0.00107
US 1	3m	0.000734	0.000687
US 1	6m	0.000994	0.00102
US 3	3m	0.000724	0.000529
US 3	6m	0.000916	0.000659
US 4	3m	0.000712	0.000476
US 4	6m	0.000900	0.000563
US 5	3m	0.000603	0.000441
US 5	6m	0.000838	0.000508