THE VOLATILITY STRUCTURE OF THE FIXED INCOME MARKET UNDER THE HJM FRAMEWORK: A NONLINEAR FILTERING APPROACH

CARL CHIARELLA[+], HING HUNG[‡], AND THUY-DUONG TÔ*

[+,‡]*School of Finance and Economics*
University of Technology, Sydney
PO Box 123, Broadway NSW2007
Australia
[+]carl.chiarella@uts.edu.au
[‡]hing.hung@uts.edu.au

**School of Commerce*
University of Adelaide
Adelaide, SA 5005
Australia
td.to@adelaide.edu.au

ABSTRACT. This paper considers the dynamics for interest rate processes within a multi-factor Heath, Jarrow and Morton (1992) specification. Despite the flexibility of and the notable advances in theoretical research about the HJM model, the number of empirical studies is still inadequate. This paucity is principally because of the difficulties in estimating models in this class, which are not only high-dimensional, but also nonlinear and involve latent state variables. This paper treats the estimation of a fairly broad class of HJM models as a nonlinear filtering problem, and adopts the local linearization filter of Jimenez and Ozaki (2003), which is known to have some desirable statistical and numerical features, to estimate the model via the maximum likelihood method. The estimator is then applied to the U.S., the U.K. and the Australian markets. Different two- and there-factor models are are found to be the best for each market, with the factors being the level, the slope and the "twist" effect. The contribution of each factor towards overall variability of the interest rates and the financial reward each factor claims are found to differ considerably from one market to another.

Key words: Term structure; Heath-Jarrow-Morton; Multifactor; Filtering; Local Linearization;

JEL classifications: C51, E43, G12

1. INTRODUCTION

Management of interest rate risk is of crucial importance to financial institutions and corporations. The volatility structure of this interest rate market plays a crucial role in assessing and managing the value as well as the risk of bond and interest rate derivative portfolios. Various interest rate models have been considered, amongst which the Heath-Jarrow-Morton (1992) (hereafter HJM) framework provides a very flexible framework for interest rate modelling. Despite its nice theoretical flexibility, the application of the HJM class of models to practical problems is hindered by the difficulty of model estimation. This is principally due to the fact that the underlying state variables of the HJM model are un-observable quantities, and the dynamics are usually non-Markovian and non-linear in their (latent) state variables.

Theoretical research on HJM models has shown that for a fairly broad family of volatility functions, the underlying stochastic system can be Markovianized, thereby easing the computational complexity involved. However, the problems of nonlinearity and the existence of latent variables still exist, and the empirical analysis of HJM models has centered around certain volatility functions that lead to convenient properties for the system, for example, the class of affine or square root affine volatilities.

It should also be noted that the estimation of stochastic models is already a challenging task for systems with affine or square root affine volatilities. The estimation techniques rely on the three basic tools: maximum likelihood, the method of moments and filtering techniques. The maximum likelihood estimator (MLE) is a method of choice for models whose likelihood is tractable, and was first applied by Chen and Scott (1993) and Pearson and Sun (1994). In many cases of interest the likelihood function is not available, and various approximation techniques are used. These include the Hermite expansion technique by Aït-Sahalia (1999, 2002, 2003), the simulated maximum likelihood by Brandt and Santa-Clara (2002), Brandt and He (2002), and the related Markov Chain Monte Carlo (MCMC) method by Jacquier et al. (1994), Kim et al. (1998), Eraker (2001) and Elerian et al. (2001). Research that uses the method of moments principle include the generalized method of moments (GMM) by Ho et al. (1996), the simulated method of moments (SMM) by Duffie and Singleton (1993), the indirect inference by Broze et al. (1998), the efficient method of moments (EMM) by Gallant and Tauchen (1996, 1997, 1998), the robust GMM by Dell'Aquila et al. (2003), and the GMM based on conditional characteristics functions by Singleton (2001). Filtering techniques, such as the Kalman filter, have recently been applied to estimate linear term structure models, such as in Jegadeesh and Pennacchi (1996), Geyer and Pichler (1999) and Rossi (2004).

Zhou (2001) study the finite sample properties of the maximum likelihood and the method of moments estimators for square-root interest rate diffusion models. The performance of the EMM method is found to be mixed even under an univariate setting. Under a multivariate setting, this performance can deteriorate. Recently Duffee and Stanton (2004) also analyze the performance of different estimation methods for dynamic term structure models. They find that the standard MLE does a very poor job of estimating the parameters that determine expected changes in interest rates. Furthermore they find that the EMM estimator is an unacceptable alternative, even where the MLE performs well. They conclude that the Kalman filter is a reasonable choice, even in the non-Gaussian setting where the filter is not exact. In that case, they advocate the use of a variant of the Kalman filter, where the updating equation for the state variables is a linearized version of the drift using its first derivative.

In light of the findings of Zhou (2001) and Duffee and Stanton (2004) this paper pursues further the filtering approach. Even though the linear filtering and prediction problem has well been understood after the important work of Kalman (1960) and Kalman and Bucy (1961), nonlinear filtering is still an active research area. Various approximation for nonlinear filters have been proposed, such as the Extended Kalman filter, the Iterated Extended Kalman filter, the Modified Gaussian filter. As these filters are quite computationally unstable, Ozaki (1993) introduced a Local Linearization filter, which was later developed further by Jimenez and Ozaki (2002, 2003) for systems whose volatility structure is dependent on the state variables (i.e. systems with multiplicative noise). The main idea is to linearize the system dynamics according to the Itô formula, utilizing both the drift and the diffusion terms, to better take into account the stochastic behaviour of the system, and then to apply the (readily available) optimal linear filter. We advocate the use of this filter as it has been shown by Shoji (1998) to have good bias properties and by Jimenez et al. (1999) to have a number of computational advantages. The estimation method is able to exploit both the time series and cross sectional information of the yield curve.

We empirically investigate different multi-factor interest rate models and apply the local linearization filter to analyze the volatility structure of the U.S., the U.K, and the Australian markets. These markets have been chosen to represent different regions in the world. The rest of the paper is organized as follows. Section 2 introduces the model. The econometric implication of the model and the proposed estimation method are discussed in Section 3. Empirical results are then presented in Section 4, and Section 5 concludes the paper.

2. MODEL FRAMEWORK

The general framework for the interest rate models considered in this paper is introduced in Heath, Jarrow and Morton (1992), where the instantaneous forward rates $r(t,x)$ (the rate that can be contracted at time t for instantaneous borrowing/lending at future time $t+x$) are assumed to satisfy SDEs of the form[1]

$$\begin{aligned} r(t,x) = r(0,t+x) + \int_0^t \boldsymbol{\sigma}(s,t+x)' \left[\bar{\boldsymbol{\sigma}}(s,t+x) - \boldsymbol{\phi}(s) \right] \, ds \\ + \int_0^t \boldsymbol{\sigma}(s,t+x)' d\boldsymbol{W}(s), \end{aligned} \tag{2.1}$$

where

$$\bar{\boldsymbol{\sigma}}(s,t+x) = \int_s^{t+x} \boldsymbol{\sigma}(s,u) du,$$

and $\boldsymbol{\sigma}(t,x)$, $\boldsymbol{\phi}(t)$ are I-dimensional processes and $\boldsymbol{W}(t)$ is a standard I-dimensional vector of independent Wiener processes under the market measure $\mathcal{P}$, $I \in \mathbb{N}_+$ and the superscript $'$ represents matrix transposition. The vector $\boldsymbol{\phi}(t)$ can be interpreted as the market price of interest rate risk vector associated with $d\boldsymbol{W}(t)$. In general, $\boldsymbol{\sigma}$ and $\boldsymbol{\phi}$ may depend on a number of forward rates $r(t,x)$.[2]

The HJM model framework is chosen as it yields arbitrage-free models that fit the initial yield curve by construction. The subclass of HJM models which are particularly suited to practical implementation are those which can be Markovianized. Carverhill (1994), Ritchken and Sankarasubramanian (1995), Bhar and Chiarella (1997a), Inui and Kijima (1998), de Jong and Santa-Clara (1999) and Björk and Svensson (2001) discuss various specifications of the forward rate volatilities $\sigma(t,x)$ that lead to Markovian representations of the forward rate dynamics. Chiarella and Kwon (2001b, 2003) introduce a specification that leads to a fairly broad and convenient class of models. The models in this class satisfy the assumption:

Assumption 2.1. *(i) For each $1 \leq i \leq I$, there exists $L_i \in \mathbb{N}$ such that the components, $\sigma_i(t,x)$, of the forward rate volatility process have the form*

$$\sigma_i(t,x) = \sum_{l=1}^{L_i} c_{il}(t) \sigma_{il}(x) \tag{2.2}$$

where $c_{il}(t)$ are stochastic processes and $\sigma_{ij}(x)$ are deterministic functions.

---

[1]We are in fact using the Brace et al. (1997) implementation of the HJM model. This is more appropriate to capture the dynamics of LIBOR and various other market quoted rates.

[2]In this notation, $r(t,0)$ denotes the instantaneous rate of interest that we henceforth write as $r(t)$.

(ii) There exist $M \in \mathbb{N}$ and a sequence $x_1 < \cdots < x_M \in \mathbb{R}_+$ such that the processes $c_{il}(t)$ have the form

$$c_{il}(t) = \hat{c}_{il}(t, r(t, x_1), \dots, r(t, x_M)), \tag{2.3}$$

where $\hat{c}$ is deterministic in its arguments.

Chiarella and Kwon (2003) then prove that the forward curve can be expressed as an affine function of a set of N discrete tenor forward rates

$$\boldsymbol{r}(t, \tau_1, \dots, \tau_N) = [r(t, \tau_1), \dots, r(t, \tau_N)]'$$

(see Appendix A for a brief summary). This set of forward rates forms a Markov process. In terms of the real world measure, where $\boldsymbol{\phi} \equiv (\phi_1, \dots, \phi_I)$ is the vector of market prices of risk associated with the Wiener process $\boldsymbol{W}$, the system of stochastic differential equations for the instantaneous forward rates becomes[3]

$$\begin{aligned} dr(t,x) =& [p_0(t, x, \tau_1, \dots, \tau_N) + \boldsymbol{p}_1'(t, x, \tau_1, \dots, \tau_N)\boldsymbol{r}(t, \tau_1, \dots, \tau_N) \\ &- \boldsymbol{\phi}'\boldsymbol{\sigma}(t, t+x)]dt + \boldsymbol{\sigma}(t, t+x)'d\boldsymbol{W}(t). \end{aligned} \tag{2.4}$$

The yield $y(t, x)$ on the $(t + x)$-maturity zero coupon bond can be calculated from the instantaneous forward rates via

$$y(t, x) = \frac{1}{x} \int_0^x r(t, u) du, \tag{2.5}$$

and can also be expressed as an affine function of the forward rates, that we write in the form

$$y(t, x) = q_0(t, x, \tau_1, \dots, \tau_N) - \boldsymbol{q}'(t, x, \tau_1, \dots, \tau_N)\boldsymbol{r}(t, \tau_1, \dots, \tau_N), \tag{2.6}$$

where the $q_i(t, x, \tau_1, \dots, \tau_N)$ is a set of deterministic functions[4]. We therefore have an affine term structure model. This model is not nested inside the popular affine model class considered in Duffie and Kan (1996), even though there will be occasions when the two classes overlap.

3. Estimation Framework

3.1. **The model specification.**

The empirical work of Litterman and Scheinkman (1991), Chen and Scott (1993), Knez et al. (1994), Singh (1995), who use principal component analysis, suggests that there are at most three factors affecting the volatility of interest rates. Guided by this insight we propose to use a three-dimensional Wiener process in the specification (2.1).

[3]For definition of the coefficient functions p_0 and $\boldsymbol{p}$, see Appendix A.

[4]Again see Appendix A for definitions of the q_i.

We shall specifically consider four volatility functions, namely

$$\sigma_1(t,x) = \gamma_1 r^{\lambda}(t), \tag{3.1}$$

$$\sigma_{2a}(t,x) = \gamma_2\,\mathrm{e}^{-\kappa_2(x-t)}, \tag{3.2}$$

$$\sigma_{2b}(t,x) = \gamma_2(r(t,\tau) - r(t)), \tag{3.3}$$

$$\sigma_3(t,x) = \gamma_3(x-t)\,\mathrm{e}^{-\kappa_3(x-t)}\,. \tag{3.4}$$

The first volatility function $\sigma_1(t,x)$ reflects the level factor, where the volatility is dependent on the level of the short rate. If $\lambda = 0.5$ we would obtain a Cox-Ingersoll-Ross (1985a) type of volatility. The second volatility function can be called the slope factor, which is modelled here in two different ways. The volatility can be dependent on the actual slope, as modelled by σ_{2b}, or the volatility can be thought of as a simple decreasing function of maturity as modelled by $\sigma_{2a}(t,x)$ which allows the shock to have much less impact on the yield curve as the yield maturity increases. The function $\sigma_3(t,x)$ creates a hump in the volatility function, which is a typical pattern found in swap markets. This can be thought of as a "twist" in the yield curve.

To model the market prices of risk ϕ_1, ϕ_2, ϕ_3, the current literature has assumed that they are dependent on the underlying interest rate. Since there is no guidance on what the functional form for this dependence should be, modellers have chosen those functional form that leads to nice model properties. As the underlying interest rate follows an Itô process, if the market prices of risk are dependent on the interest rate, they should also follow Itô processes. Instead of specifying a dependence structure as in the literature, the market prices of risk here are assumed to follow a stochastic differential equation

$$d\phi_i = \alpha_i(\bar{\phi}_i - \phi_i)dt + \beta_i\sqrt{\phi_i(t)}dW_i(t). \tag{3.5}$$

Intuitively, the specification suggests that the market prices of different interest rate risks are always positive and tend to converge to their long run equilibria.

3.2. **Econometric implication of the model.**

Some similar and other specialized models of the HJM class considered here have been empirically analyzed. Bliss and Ritchken (1996) consider the case where the volatility function in (2.2) can be written as[5]

$$\sigma(t,x) = c(t)\,\mathrm{e}^{-\kappa x}\,.$$

This specification covers our single-factor model, as each of our volatility functions can be written in the above form. For example, with $\sigma_1(t,x) = \gamma_1 r^{\lambda}(t)$, the value of κ is zero and $c(t) = \gamma_1 r^{\lambda}(t)$. The key idea of their approach is to exploit the relationship

[5]With this volatility function, the model can be Markovianized using two state variables.

(2.6) for the yields, into which they introduce an error term, then estimate their model via the Maximum Likelihood procedure [6]. The main drawback of this approach is that the estimation procedure can only identify κ, as the relationship (2.6) does not depend on the parameters characterizing function $c(t)$ (γ_1 and λ in our example). However, all of the parameters in the models are important in practical work, such as the determination of the price of a derivative contract.

de Jong and Santa-Clara (1999) also empirically study two-state variable HJM models where the volatility function of the system is proportional to the square root of the state variables. However, they overcome the disadvantages of Bliss and Ritchken (1996) approach by using both the relationship (2.6) and the Markovian system (2.4) in their estimation procedure. They use the Kalman filtering method where (2.6) serves as the observation equation and (2.4) is discretized into a state transition equation. In a more general setting, it is not clear how to discretize the structural stochastic system, and the behaviour of the estimator is clearly dependent on the method used in this discretization. In a more recent study, Rossi (2004) also uses the Kalman filter to estimate a HJM model. However, the HJM model is Gaussian, and only a linear Kalman filter is needed, therefore the issue of a discretization does not arise.

In this paper, we advocate the local linearization filter (hereafter the LL filter) of Jimenez and Ozaki (2002, 2003). This approach is still based on the Kalman filter for a discrete linear system. However, Jimenez and Ozaki do not discretize the nonlinear system directly, but rather approximate it by a system linear in both its drift and its diffusion terms, for which a linear Kalman filter turns out to be readily applicable. The approximation is not based on the first order Taylor approximation used in the standard extended Kalman filter framework, but is instead based on a second order approximation using the Itô formula to better take into account the stochastic behaviour of the underlying state variables.

In his comparative study, Shoji (1998) analyzed the performance of the maximum likelihood estimator based on the LL filter and the one based on the extended Kalman filter for a system with additive noise (i.e. the volatility function is not dependent on the state variables). Shoji used Monte Carlo simulation to show that the LL filter provided estimates with smaller bias, particularly in estimation of the coefficient of the drift term. Jimenez et al. (1999) compared the LL scheme with other linearization schemes for systems with either additive or multiplicative noise (the latter means that the volatility function is dependent on the state variables). They also reported a number

[6] The relationship Bliss and Ritchken use is actually an expression of the whole yield curve as an affine function of some particular yields rather than the forward rates. This can be derived very simply from the model here.

of numerical advantages of the LL filter, including numerical stability, better accuracy and the order of strong convergence.

3.3. The local linearization filter and the maximum likelihood estimator.

Consider the state space model defined by the continuous state equation

$$d\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t))dt + \sum_{i=1}^{m} \mathbf{g}_i(t, \mathbf{x}(t))dW_i(t), \tag{3.6}$$

and the discrete observation equation[7]

$$\mathbf{z}_{t_j} = \mathbf{C}(t_j)\mathbf{x}(t_j) + \mathbf{e}_{t_j}, \text{ for } j = 0, 1, \ldots, J, \tag{3.7}$$

where $\mathbf{f}$ and $\mathbf{g}_i$ are nonlinear functions, $\mathbf{x}(t) \in \mathbb{R}^d$ is the state vector at the instant of time t, $\mathbf{z}_{t_j} \in \mathbb{R}^r$ is the observation vector at the instant of time t_j, $\boldsymbol{W}$ is an m-dimensional Wiener process, and $\{\mathbf{e}_{t_j} : \mathbf{e}_{t_j} \sim \mathcal{N}(0, \mathbf{\Pi}), j = 0, \ldots, J\}$ is a sequence of i.i.d. random vectors.

The system functions $\mathbf{f}$ and $\mathbf{g}_i$ can be linearly approximated. Jimenez and Ozaki (2003) proposed to approximate them via a truncated Ito-Taylor expansion, for example, the approximation for $\mathbf{f}$ is

$$\begin{aligned}\mathbf{f}(t, \mathbf{x}(t)) \approx & f(s, \mathbf{u}) + \left(\frac{\partial \mathbf{f}(s, \mathbf{u})}{\partial s} + \frac{1}{2} \sum_{k,l=1}^{d} [\mathbf{G}(s, \mathbf{u})\mathbf{G}'(s, \mathbf{u})]^{k,l} \frac{\partial^2 \mathbf{f}(s, \mathbf{u})}{\partial \mathbf{u}^k \partial \mathbf{u}^l} \right) (t - s) \\ & + \mathbf{J}_{\mathbf{f}}(s, \mathbf{u})(\mathbf{x}(t) - \mathbf{u}),\end{aligned} \tag{3.8}$$

where $(s, \mathbf{u}) \in \mathbb{R} \times \mathbb{R}^d$, $\mathbf{J}_{\mathbf{f}}(s, \mathbf{u})$ is the Jacobian of $\mathbf{f}$ evaluated at the point $(s, \mathbf{u})$ and $\mathbf{G}(s, \mathbf{u})$ is the $d \times m$ matrix defined by $\mathbf{G}(s, \mathbf{u}) \equiv (\mathbf{g}_1, \ldots, \mathbf{g}_m)$. The presence of the volatility function $\mathbf{g}_i$ in the linearization of both the drift and the diffusion terms differentiates this linearization scheme from the simple second order Euler/Taylor expansion. It is because the underlying state system is stochastic, and follows an Itô calculus, expansion according to Itô-Taylor formula will better take into account their stochastic nature.

Using such approximations for $\mathbf{f}$ and $\mathbf{g}_i$, the solution of the nonlinear state equation (3.6) can be approximated by the solution of the piecewise linear stochastic differential

[7] A full (nonlinear) specification of the observation equation would be

$$\mathbf{z}_{t_j} = \mathbf{h}(t_j, \mathbf{x}(t_j)) + \sum_{i=1}^{n} \mathbf{p}_i(t_j, \mathbf{x}(t_j))\xi^i_{t_j} + \mathbf{e}_{t_j}, \text{ for } j = 0, 1, \ldots, J,$$

where $\mathbf{h}$ and $\mathbf{p}_i$ are nonlinear functions, $\{\boldsymbol{\xi}_{t_j} : \boldsymbol{\xi}_{t_j} \sim \mathcal{N}(0, \mathbf{\Lambda}), \mathbf{\Lambda} = \mathrm{diag}((\lambda_1, \ldots, \lambda_n)), j = 0, \ldots, J\}$ is a sequence of random vector i.i.d., and $\xi^i_{t_j}$ and $\mathbf{e}_{t_j}$ are uncorrelated for all i and j. However, in most finance applications, including ours, a linear specification for $\mathbf{h}$ is all that is required and there is no need to include the extra noise term $\boldsymbol{\xi}$.

equation[8]

$$
\begin{aligned}
d\mathbf{y}(t) = &\Big(\mathbf{A}(t_j, \hat{\mathbf{y}}_{t_j|t_j})\mathbf{y}(t) + \mathbf{a}(t, t_j, \hat{\mathbf{y}}_{t_j|t_j})\Big)\, dt \\
&+ \sum_{i=1}^{m} \Big(\mathbf{B}_i(t_j, \hat{\mathbf{y}}_{t_j|t_j})\mathbf{y}(t) + \mathbf{b}_i(t, t_j, \hat{\mathbf{y}}_{t_j|t_j})\Big)\, dW_i(t)
\end{aligned}
\tag{3.9}
$$

for all $t \in [t_j, t_{j+1})$, starting at $\mathbf{y}(t_0) = \hat{\mathbf{y}}_{t_0|t_0} = \hat{\mathbf{x}}_{t_0|t_0}$. The various quantities appearing in (3.9) are defined as

$$
\begin{aligned}
\hat{\mathbf{x}}_{t|\rho} &= \mathbb{E}\left(\mathbf{x}(t)|Z_\rho\right), \quad Z_\rho = \{\mathbf{z}_{t_j} : t_j \le \rho\}, \\
\hat{\mathbf{y}}_{t|\rho} &= \mathbb{E}\left(\mathbf{y}(t)|Z_\rho\right), \\
\mathbf{A}(s, \mathbf{u}) &= \mathbf{J}_{\mathbf{f}}(s, \mathbf{u}), \\
\mathbf{B}_i(s, \mathbf{u}) &= \mathbf{J}_{\mathbf{g_i}}(s, \mathbf{u}), \\
\mathbf{a}(t, s, \mathbf{u}) &= \mathbf{f}(s, \mathbf{u}) - \mathbf{J}_{\mathbf{f}}(s, \mathbf{u})\mathbf{u} + \frac{\partial \mathbf{f}(s, \mathbf{u})}{\partial s}(t - s) \\
&\quad + \frac{1}{2} \sum_{k,l=1}^{d} [\mathbf{G}(s, \mathbf{u})\mathbf{G}'(s, \mathbf{u})]^{k,l} \frac{\partial^2 \mathbf{f}(s, \mathbf{u})}{\partial \mathbf{u}^k \partial \mathbf{u}^l}(t - s), \\
\mathbf{b}_i(t, s, \mathbf{u}) &= \mathbf{g}_i(s, \mathbf{u}) - \mathbf{J}_{\mathbf{g}_i}(s, \mathbf{u})\mathbf{u} + \frac{\partial \mathbf{g}_i(s, \mathbf{u})}{\partial s}(t - s) \\
&\quad + \frac{1}{2} \sum_{k,l=1}^{d} [\mathbf{G}(s, \mathbf{u})\mathbf{G}'(s, \mathbf{u})]^{k,l} \frac{\partial^2 \mathbf{g}_i(s, \mathbf{u})}{\partial \mathbf{u}^k \partial \mathbf{u}^l}(t - s).
\end{aligned}
$$

The approximate stochastic differential equation (3.9) and the corresponding observation equation (see (3.7))

$$
\mathbf{z}_{t_j} = \mathbf{C}(t_j)\mathbf{y}(t_j) + \mathbf{e}_{t_j}, \text{ for } j = 0, 1, \ldots, J, \tag{3.10}
$$

form a linear state space system. The optimal linear filter proposed by Jimenez and Ozaki (2002) can be applied (see Appendix B for its definition) to determine the conditional mean $\hat{\mathbf{y}}_{t|\rho}$ and conditional covariance matrix $\mathbf{P}_{t|\rho} = \mathbb{E}((\mathbf{y}(t) - \hat{\mathbf{y}}_{t|\rho})(\mathbf{y}(t) - \hat{\mathbf{y}}_{t|\rho})'|Z_\rho)$ for all $\rho \le t$. The difference with the standard Kalman filter is that the volatility function here is also dependent of the state variables, albeit only via a linear function.

Due to the assumption of multivariate normality of the disturbances $\mathbf{e}_{t_j}$ (and if the initial state vector also has a proper multivariate normal distribution), the distribution of $\mathbf{z}_{t_{j+1}}$ conditional on Z_{t_j} is itself normal (see (3.10)). The mean and covariance matrix of this conditional distribution are given directly by the local linearization filter

[8] We use $\mathbf{y}(t)$ to denote the solution to the approximate system to distinguish it from $\mathbf{x}(t)$ the solution to the true system.

above. Therefore, a maximum likelihood estimator for the model parameters can be easily derived.

Let $\boldsymbol{\theta}$ be the vector of parameters of interest, which include all parameters specifying the state space model (3.9) and (3.10), plus the initial state values of $\hat{\mathbf{x}}_{t_0|t_0}$ and $\mathbf{P}_{t_0|t_0}$. The log likelihood function for $\mathbf{Z}$ is

$$\mathcal{L}_{\mathbf{Z}}(\boldsymbol{\theta}) = -\frac{rJ}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=1}^{J}\ln|\boldsymbol{\Sigma}_{t_j}| - \frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\nu}'_{t_j}\boldsymbol{\Sigma}_{t_j}^{-1}\boldsymbol{\nu}_{t_j} \tag{3.11}$$

where the innovation equations are

$$\boldsymbol{\nu}_{t_j} = \mathbf{z}_{t_j} - \mathbf{C}(t_j)\hat{\mathbf{y}}_{t_j|t_{j-1}}, \tag{3.12}$$

$$\boldsymbol{\Sigma}_{t_j} = \mathbf{C}(t_j)\mathbf{P}_{t_j|t_{j-1}}\mathbf{C}'(t_j) + \boldsymbol{\Pi}. \tag{3.13}$$

The maximum likelihood estimator of $\boldsymbol{\theta}$ is then

$$\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}} \mathcal{L}_{\mathbf{Z}}(\boldsymbol{\theta}). \tag{3.14}$$

3.4. **Econometric implementation.**

We now view our model as a continuous-discrete nonlinear state space system, where (2.4) and (3.5) serve as the nonlinear state equations, and (2.6) serves as the linear (affine) observation equation. Similar to the standard practice in the literature, we introduce into the observation equation a measurement error, which reflects the fact that the model cannot fit all observed yields simultaneously. This measurement error is assumed to follow a multivariate normal distribution. The local linearization filter can be readily applied to yield the maximum likelihood estimator of $\boldsymbol{\theta}$, the vector of parameters of interest, which includes all of the parameters of the volatility functions (4.1) - (4.4), of the market price of risk specification (3.5) and the initial conditional mean vector $\hat{\mathbf{x}}_{t_0|t_0}$ and conditional variance matrix $\mathbf{P}_{t_0|t_0}$.

The numerical difficulties associated with any estimation procedures for stochastic systems are well-known. Amongst them, system stability, matrix inversion to calculate the likelihood function, convergence of the optimization routine and significance of the estimates are the main problems. To partly overcome these problems, we maximize the likelihood function using a genetic algorithm (Holland (1975), Mitchell (1996), Vose (1999), Michalewicz (1999)). Genetic algorithms use the evolutionary principle to solve difficult problems with objective functions that do not possess “nice” properties such as continuity and differentiability. The algorithms search the solution space of a function, and implement a “survival of the fittest” strategy to improve the solutions.

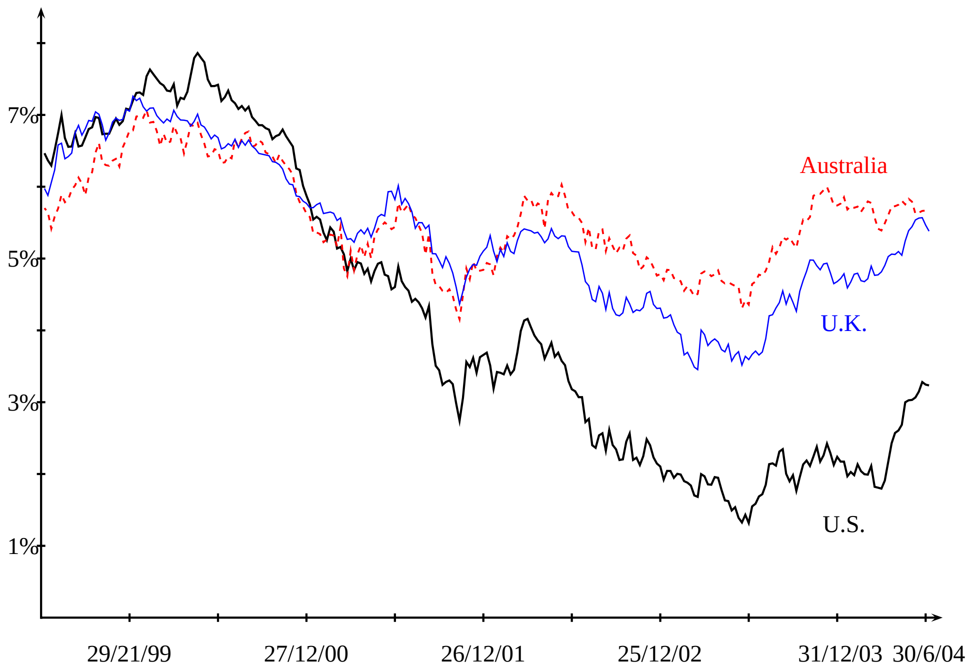
4. Empirical Analysis

4.1. **The Data.**

We estimate the model using the zero yield data in the U.S, U.K and Australian markets downloaded from Datastream®. The data consists of weekly observations for contracts with maturity of 2, 3, 4, 5, 7, 10 and 12 years, spanning 5 years from 7th July 1999 to 30th June 2004.

Figure 1 shows the 2-year zero rates for the different markets. Evolutions for rates at other maturities have similar shapes, though at different levels. Over the 5-year

FIGURE 1. Zero yield curve, 2-year time to maturity



period, interest rates have changed significantly. The rates increased in all markets by around 1.5% from July 1999 to June 2000. The rates then decreased but with different paces across the markets. In the U.S., the rates dived sharply from 7.5% to 1% in the next 3 years, then started to pick up again in the second half of 2003, moved slightly around the 2% level, then rose to 3% by June 2004. The U.K market also experienced a period of decreasing rates during the 3-year period of June 2000 - June 2003, but to a much lesser extent than the U.S. market. Then the rates picked up again as part of a global trend. The Australian market had a much more stable interest rate movement compared to the other two, around 6% in 1999 and 2000, and around 5% for the rest of the sample data. All of the rates display a high level of autocorrelation, as can be seen in Table 1.

TABLE 1. Summary statistics for the zero yield curve

	U.S.			Australia			U.K.		
	2-yr	5-yr	12-yr	2-yr	5-yr	12-yr	2-yr	5-yr	12-yr
Mean	4.21%	5.15%	6.06%	5.58%	6.01%	6.29%	5.36%	5.63%	5.61%
Stdev	2.07%	1.49%	0.99%	0.69%	0.65%	0.56%	1.02%	0.87%	0.56%
AC(1)	0.995	0.992	0.985	0.977	0.966	0.957	0.992	0.990	0.983

We also analyzed the principal components of the zero yield curve. In all of the markets, three components are able to explain nearly 100% of the variation in the yields. The last component plays a negligible role, only explaining 0.06% of the total variation in the U.S market, and less than 0.2% of the total variation in the U.K and the Australian markets, as reported in Table 2.

TABLE 2. Principal component analysis of zero yield curves

% variation explained	U.S.	Australia	U.K.
Principal component 1	99.21	97.56	99.14
Principal component 2	0.73	2.18	0.72
Principal component 3	0.06	0.17	0.13
Total of the 3 components	100	99.91	99.99

4.2. **Empirical Results.**

We separately ran the estimation for different combinations of the 4 volatility functions

$$\sigma_1(t,x) = \gamma_1 r^{\lambda}(t), \tag{4.1}$$

$$\sigma_{2a}(t,x) = \gamma_2\, \mathrm{e}^{-\kappa_2(x-t)}, \tag{4.2}$$

$$\sigma_{2b}(t,x) = \gamma_2(r(t,\tau) - r(t)), \tag{4.3}$$

$$\sigma_3(t,x) = \gamma_3(x-t)\, \mathrm{e}^{-\kappa_3(x-t)}\ . \tag{4.4}$$

The models considered and their code can be found in Table 3.

TABLE 3. Models considered

This table reports different models considered in the empirical analysis by combining various volatility functions.

Model Code	$\sigma_1(t,x) = \gamma_1 r^{\lambda}(t)$	$\sigma_{2a}(t,x) = \gamma_2\, e^{-\kappa_2(x-t)}$	$\sigma_{2b}(t,x) = \gamma_2(r(t,\tau)-r(t))$	$\sigma_3(t,x) = \gamma_3(x-t)\, e^{-\kappa_3(x-t)}$
Model 1	✓			
Model 2	✓	✓		
Model 2B	✓		✓	
Model 3	✓	✓		✓
Model 3B	✓		✓	✓
Model 3C	✓	✓	✓	

4.2.1. *The U.S. market.*

Among the six models estimated for the U.S. market, we find that model 3 is the best one. Except for model 3B, where we failed to find a combination of parameters that resulted in positive interest rates, we were able to maximize the likelihood functions for all of the other model specifications. Table 4 reports the likelihood values and various information criteria for each estimated model. Model 3 has a lower information criterion calculated based on different schemes, namely the Akaike, the Schwarz-Bayesian and the Hannan-Quinn, and therefore is the preferred model.

TABLE 4. U.S market. Information criteria.

	Model 1	Model 2	Model 2B	Model 3	Model 3C
Log likelihood	14647.68	22680.14	15780.21	23227.7	23112.27
Number of parameters	5	11	9	16	15
Akaike (AIC)	-8.016	-12.408	-8.632	-12.705	-12.642
Schwarz-Baysian (BIC)	-8.007	-12.391	-8.619	-12.681	-12.620
Hannan-Quinn (HIC)	-8.012	-12.402	-8.627	-12.696	-12.634

The estimated parameters for model 3 can be found in Table 5. The 3 volatility functions and the corresponding market prices of risk evolutions are

$$\sigma_1(t,x) = 0.0457 r(t)^{0.5624}$$

$$\sigma_2(t,x) = 0.0240\, e^{-0.0155(x-t)}$$

$$\sigma_3(t,x) = 0.0012(x-t)\, e^{-1.0446(x-t)}$$

$$d\phi_1 = 40.4048(1.4393 - \bar{\phi}_1)dt + 0.2928\sqrt{\phi_1}dW_1(t)$$

$$d\phi_2 = 49.9942(0.0133 - \bar{\phi}_2)dt + 0.0506\sqrt{\phi_2}dW_2(t)$$

$$d\phi_3 = 22.2553(39.780 - \bar{\phi}_3)dt + 9.9216\sqrt{\phi_3}dW_3(t)$$

TABLE 5. U.S. market. Estimated parameters for model 3

This table reports the parameter estimates for model 3, the preferred model among those estimated. Their corresponding standard errors are reported in square parenthesis. The notation xe-y stands for $x \times 10^{-y}$.

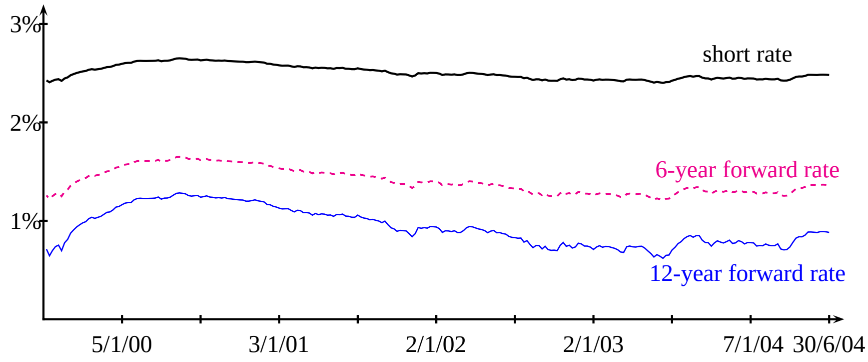
Par	Est	Par	Est	Par	Est
γ_1	0.0457	γ_2	0.0240	γ_3	0.0012
	[3.3e-9]		[2.3e-8]		[1.1e-9]
λ	0.5624	κ_2	0.1155	κ_3	1.0446
	[1.6e-7]		[1.2e-8]		[1.4e-7]
α_1	40.4048	α_2	49.9942	α_3	22.2553
	[2.5e-5]		[5.0e-5]		[4.6e-4]
$\bar{\phi}_1$	1.4393	$\bar{\phi}_2$	0.0133	$\bar{\phi}_3$	39.780
	[1.6e-6]		[2.7e-7]		[9.2e-4]
β_1	0.2928	β_2	0.0506	β_3	9.9216
	[4.7e-7]		[0.0796]		[1.1176]

The first volatility factor depends on the level of interest rate via the functional form $\gamma_1 r^{\lambda}$. The estimated value of λ is 0.56, which is quite close to the square root volatility specification usually used in empirical work. This value is much lower than the value 1.5 found by Chan et al. (1992), but is within the range of 0.5 to 1.5 reported in Pagan et al. (1996) (dependent on the interest rate series used). The second volatility factor allows a shock in the market to impact on the short end of the curve more than to yields at longer maturities. The value of κ_2 implies that it requires the two yield maturities to be 6 years apart for the impact of the same shock to halve. The third volatility factor creates a hump in the volatility curve, which occurs at around 1 year to maturity (as implied by the value of κ_3).

The combination between these three volatility factors forms the instantaneous volatility for the forward rates. This overall volatility changes over time as the yield curve moves. The volatility evolutions for the spot rate, the 6-year forward rate and the 12-year forward rate are displayed in Figure 2. The total variation for the forward rates decreases as the time to maturity increases. Moreover, the short maturity forward rates have much less variation over time than the longer maturity ones.

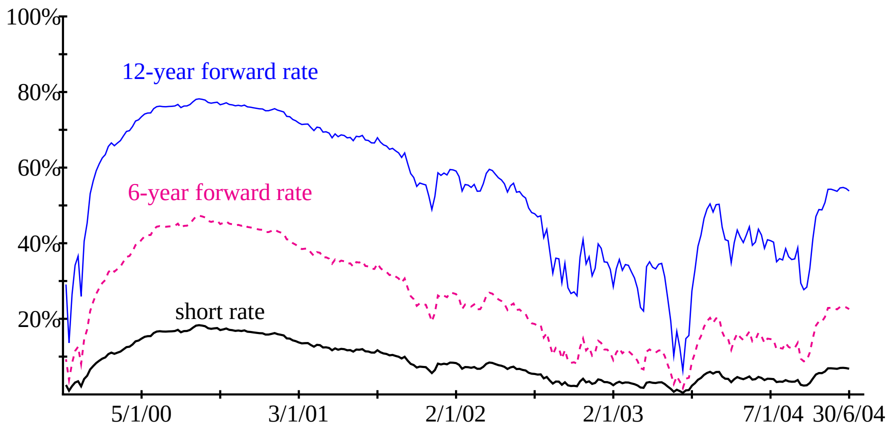
Each volatility factor contributes differently towards the total variation of the forward rates. The third volatility factor contributes very slightly, almost to a negligible extent (less than half a percentage point). The contribution of the other two factors vary according to the yield maturity and to the passage of time. Figure 3 shows the contribution of the first factor (the level factor), the contribution of the second factor is just a mirror image. The level effect is more dominant for the forward rate with longer time to maturity than the shorter ones, averaging at 54% for the 12-year forward rate,

FIGURE 2. The U.S. market. The instantaneous volatility of forward rates



25% for the 6-year forward rate and only 8.5% for the short rate. The dominance of the level effect on the overall forward rates volatility increased during the first year of the sample period, then declined steadily for the next 3 years, and finally levelled off at a slightly higher value in the last year.

FIGURE 3. U.S. market. The contribution of the level effect to the total forward rate variation



Given the fact that the third volatility factor plays a very negligible role in determining the overall forward rate variation, one would ask the question whether it should be included in the model specification. The answer is yes. Even though magnitude of this volatility factor is small, each unit of this volatility risk commands much higher financial reward than one unit of the other volatility risks. The long run unit price $\bar{\phi}$ for the third volatility factor is 39.8, compared to the values 1.45 and 0.01 for the other two volatility factors. In addition, the speed of mean reversion of the price for the third volatility risk is only half those of the other two volatility factors risk. On the way towards the long-run value, it takes 1.6 weeks for the level of the price of the third volatility risk to halve, whereas it takes only 0.9 and 0.7 weeks for the prices of the other two risk to halve. The intensity of this third unit price movement is also much higher, measured by a value of 9.9 for β_3, compared to 0.3 for β_1 and 0.05 for β_2.

This unit price, when scaled by the volatility, will determine the overall compensation for investors for bearing volatility risk. It is reflected on a discount to the drift of the forward curve, and consequently a premium to the drift of the bond equation. The discounts to compensate for bearing each volatility factor risk (called the market price of risk) are additive. Figure 4 graphs the contribution of each market price of risk into the total compensation investors require to bear the volatility risk. As can be seen from the graph, even though the third volatility factor is very small in magnitude, the corresponding market price of this risk plays a very significant role in the total compensation, especially for yields with short and medium maturities. The market price of the first volatility factor risk is the dominant one overall, but that role is somewhat diminished for yields from 1-3 year maturity.

FIGURE 4. The U.S. market. The contribution of each market price of risk to the total risk compensation

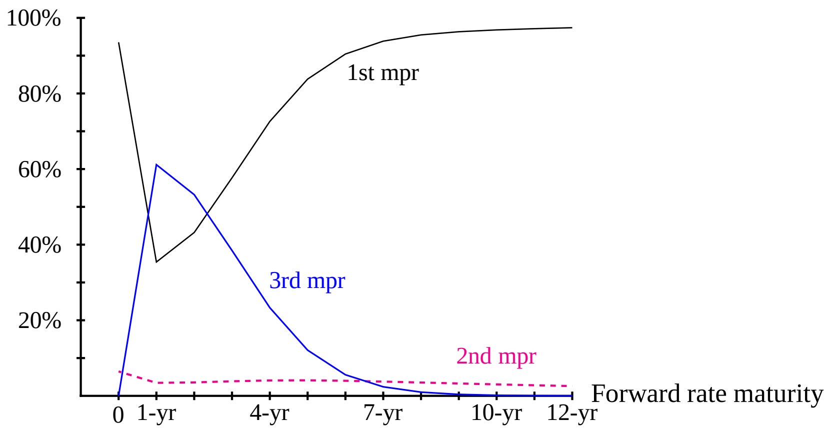


Table 6, Panel A reports the prediction errors obtained by the model. The average absolute prediction error for zero yield series is 14 basis points, whereas the mean of the prediction error is 3 basis points, which indicate a good in-sample prediction power.

In order to check the power of our model and the estimation, we used the parameter values to calculate the implied LIBOR rates, then compared these implied values with the market values. It should be noted that the LIBOR rates were not used in our

TABLE 6. U.S. market. Prediction errors.

Panel A. Prediction of the zero yields

This panel reports the prediction errors for all of the yield series used in the estimation. All values are reported as basis points.

Maturity (in years)	2	3	4	5	7	10	12	All
Mean error	2.16	5.34	2.21	2.53	2.48	1.89	3.29	2.84
Standard deviation error	20.4	19.5	18.3	18.7	18.0	16.7	16.6	18.3
Mean absolute error	13.7	14.7	13.5	14.3	13.8	12.8	13.0	13.7
Stdev absolute error	15.2	13.8	12.5	12.3	11.7	10.8	10.9	12.5

Panel B. Prediction of LIBOR rates

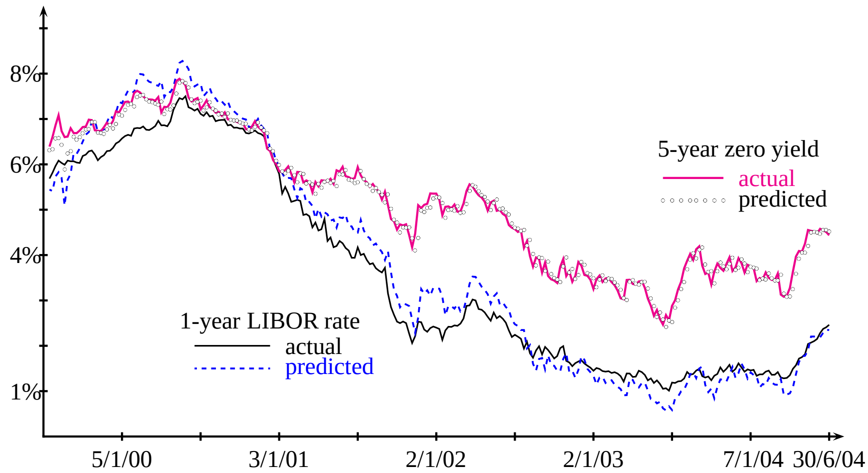
Parameters estimated from zero yield data are used to predict the actual LIBOR rates. All values in the table are in basis points. "Stdev" stands for standard deviation, "Correl" stands for correlation, "pred." stands for prediction, and "Abs. Err." stands for absolute error.

Maturity	Correl. (libor,pred.)	Stdev. libor	Stdev. pred.	Mean Error	Stdev. Error	Mean Abs. Err.	Stdev. Abs. Err.
1 month	0.9110	217.18	268.06	-5.61	113.78	89.62	70.11
2 months	0.9376	218.55	262.89	-8.96	95.62	77.67	56.28
3 months	0.9466	220.76	261.20	-9.72	88.27	72.66	50.86
4 months	0.9548	221.48	259.74	-11.30	81.63	67.52	47.06
5 months	0.9614	222.06	258.42	-13.04	75.87	62.87	44.27
6 months	0.9683	222.08	257.18	-15.23	69.64	58.69	40.32
7 months	0.9742	222.02	255.95	-16.71	63.93	54.57	37.13
8 months	0.9788	221.88	254.71	-18.21	58.92	50.92	34.66
9 months	0.9825	221.73	253.43	-19.66	54.54	47.75	32.76
10 months	0.9852	221.34	252.09	-20.40	50.94	44.92	31.42
11 months	0.9874	220.91	250.69	-21.07	47.82	42.45	30.39
12 months	0.9891	220.37	249.23	-21.65	45.11	40.18	29.75

estimation at all, and the maturities of the zero yield rate used in the estimation are from 2-12 years, whereas the maturities for the LIBOR rates are only less than 1 year. As can be seen from Panel B of Table 6, there is a very high correlation between the predicted LIBOR and the actual LIBOR rates, though the predicted series have somewhat higher variation than the actual series. Understandably the lowest correlation of 91.1% is for the 1 month rate LIBOR, as it is much outside the maturity range used in the estimation. The correlation increases as the time-to-maturity comes closer to the maturity range used in the estimation. The correlation between the predicted and the actual 1 year LIBOR rate is 98.9%. Across all LIBOR maturities, the average prediction error is from 5-22 basis points, whereas the absolute errors lie in the range of 40-90 basis point. In a different study, Jegadeesh and Pennacchi (1996) used the Kalman filter to estimate a linear 2-factor model with constant volatilities using futures data (with 3-month LIBOR as the underlying rate) and used the parameters to predict the actual LIBOR. They reported mean errors (not mean absolute error) in the range of 23-48 basis points for different maturities.

Figure 5 illustrates the predictive power of the model. The model gives excellent prediction for zero yield series which were used in the estimation. For the LIBOR series, whose values as well as maturity range were not used in the estimation, the predicted series matches the actual series well in the trend, whereas the value deviation is small.

FIGURE 5. The U.S. market. Actual and predicted interest rates.



At this point one would question whether the model predictive power is better than a simple random walk approach. However, it should be kept in mind that practitioners regularly need to price over-the-counter instruments whose underlying interest rate is not traded in the market, or to price illiquid instruments whose underlying interest rate's quote is not always available. Under such circumstances, the random walk prediction method is not feasible, and the ability of the model to predict accurately a totally different interest rate series in a totally different maturity range is very important. One of the reasons the implied volatility approach is popular among practitioners is that a less actively traded security can be priced consistently with other more liquid securities. Nevertheless, this implied volatility approach implies a continual change of parameter values, which is not desirable. Using our estimation procedure, consistency of security prices remains while parameters are kept constant and there is a clear indication of the confidence interval for the parameter values.

4.2.2. *The Australian market.*

Unlike the U.S. market, we find that model 2B is the preferred model based on various information criteria, as shown in Table 7. It should be noted that even though in theory some models encompass others (e.g. model 1 is nested inside model 2), in practice the likelihood value of the more general model might not be higher than that of the restricted one, as the restricting parameters have to be kept different from zero for computational purposes. The estimated parameters can be found in Table 8. The volatility functions for the estimated model and the corresponding market prices of risk are

$$\begin{aligned}
\sigma_1(t,x) &= 0.9166r(t)^{1.7080}\\
\sigma_2(t,x) &= 0.0824(r(t,\tau)-r(t))\\
d\phi_1 &= 48.0407(18.7346-\bar{\phi}_1)dt + 5.3123\sqrt{\phi_1}dW_1(t)\\
d\phi_2 &= 49.2477(15.1458-\bar{\phi}_2)dt + 8.4740\sqrt{\phi_2}dW_2(t).
\end{aligned}$$

TABLE 7. Australian market. Information criteria.

	Model 1	Model 2	Model 2B	Model 3	Model 3C
Log likelihood	74609.74	23351.56	90387.4	23354.29	23317.48
Number of parameters	5	11	9	16	15
Akaike (AIC)	-40.835	-12.775	-49.279	-12.774	-12.755
Schwarz-Baysian (BIC)	-40.827	-12.759	-49.266	-12.750	-12.732
Hannan-Quinn (HIC)	-40.832	-12.769	-49.274	-12.765	-12.746

Under this model specification, the volatility of different forward rates does not depend on the maturity of the rates, but rather depends explicitly on the level of the short rate and the slope of the yield curve. The estimated value for λ in the first volatility function is 1.7. A direct comparison with other studies is not available, as previous studies only focus on the U.S. market. For the U.S. market, Bhar et al. (2005) have employed a Bayesian updating algorithm to estimate the distribution for the parameter λ in a one factor HJM model implied by LIBOR rates of various maturities. They find that the distribution lies in the interval [0.5,4]. In light of this finding a value of 1.7 for the Australian market seems plausible.

The changes of the two volatility factors over time are illustrated in Figure 6. The overall variation of interest rates follows a declining trend. It can be seen that the first volatility factor plays a dominant role in determining the variation of interest rates. The contribution of the second volatility factor increases with time as the yield curve becomes more and more steep. On average for the whole sample period, the level

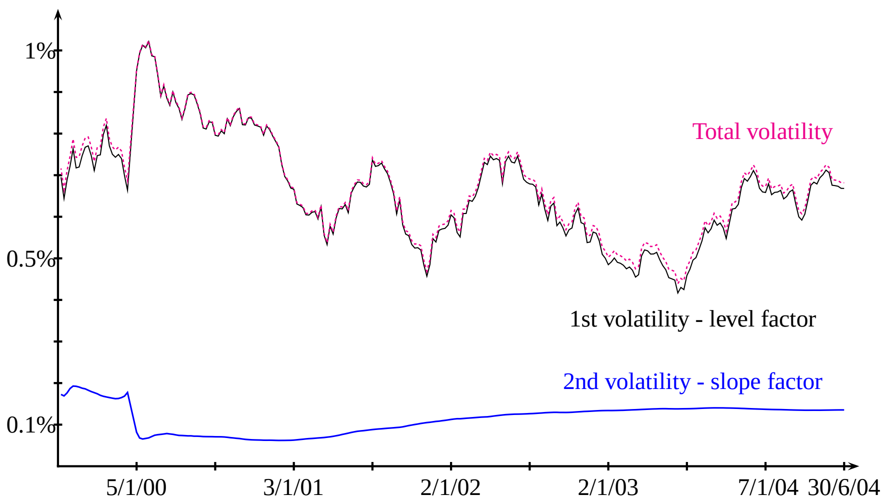
TABLE 8. Australian market. Estimated parameters for model 2B

This table reports the parameter estimates for model 2B, the preferred model among those estimated for the Australian market. Their corresponding standard errors are reported in square parenthesis. The notation xe-y stands for $x \times 10^{-y}$.

Par.	γ_1	λ	α_1	$\bar{\phi}_1$	β_1
Est.	0.9166	1.7080	48.041	18.735	5.312
Stderr.	[4.8e-7]	[1.6e-9]	[0.0052]	[0.0040]	[0.0004]
Par.	γ_2		α_2	$\bar{\phi}_2$	β_2
Est.	0.0824		49.248	15.146	8.474
Stderr.	[4.0e-6]		[0.0010]	[0.0107]	[0.0005]

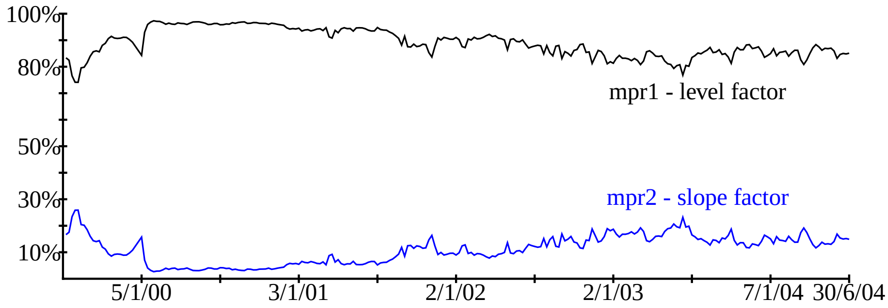
factor explains 96% of the variation in the yield curve and the slope factor explains the other 4%.

FIGURE 6. The instantaneous volatility of forward rates



The unit prices for bearing these two different volatility risk are similar, with an estimate of 18.7 for $\bar{\phi}_1$ and 15.1 for $\bar{\phi}_2$. Both of them have quite high rate of mean reversion, i.e. it takes only 0.75 weeks for the level of the price of the volatility risk to halve. Due to the similarity in the unit prices of the two different risk, and the domination of the first factor risk, 89% of the overall risk compensation is contributed by the first factor risk, and only 11% is contributed by the second factor risk. Figure 7 graphs the changes of this contribution overtime. The ability of the second factor to command financial reward increases in the second half of the sample period due to the steeper yield curve.

FIGURE 7. Australian market. The contribution of each market price of risk to the total risk compensation



The in-sample predictive power of this model for the Australian market is not as good as the model for the U.S market. There is one instance where the filter behaves really poorly. It is at the last observation of 1999, when the whole yield curve suddenly shifted up after a trend of downward movements. The filter did not predict this change, which resulted in a large under-prediction for the level of the rates. However, after this large error, the filter adapted to the new information and subsequently did reasonably well. Panel A of Table 9 reports the summary statistics for the prediction error. Without the large error, the overall absolute prediction error is 14.5 basis points. There is no clear pattern of prediction errors across different maturities.

Panel B of Table 9 reports the power of the model in predicting the LIBOR rates, which have a different maturity range and were not used in the estimation task. The average prediction error lies in the range of 37-46 basis points, whereas the mean absolute error lies in the range of 38-56 basis points. The correlation of the predicted series and the actual series increases steadily as the maturity of the LIBOR rates gets closer to the zero yield maturity range used in the estimation. The correlation is at a high level of 89.2% for the LIBOR rate of 1 year maturity, which is still 1 year lower than the lowest maturity yield used in the estimation. It is reasonable to expect that the model will predict well those interest rates with maturity lying within the estimation range.

TABLE 9. Australian market. Prediction errors.

Panel A. Prediction of the zero yields

This table reports the prediction errors for all of the Australian yield series used in the estimation. All values are reported as basis points.

Maturity (in years)	2	3	4	5	7	10	12	All
Mean error	-19.8	-8.1	1.7	5.8	6.4	0.9	-6.0	-2.8
Stdev. error	165.5	163.8	162.5	161.5	159.5	156.7	154.9	160.6
Mean absolute error	28.2	20.5	22.4	25.8	25.9	22.5	22.8	24.0
Stdev absolute error	164.3	162.7	161.0	159.5	157.5	155.0	153.3	159.0
Without the large error at the last observation in 1999								
Mean error	-9.7	1.9	11.6	15.7	16.2	10.2	3.4	7.1
Stdev. error	29.5	26.9	26.0	25.5	24.0	23.9	25.1	25.9
Mean absolute error	18.1	10.5	12.6	16.0	16.3	13.0	13.4	14.3
Stdev absolute error	25.3	24.8	25.6	25.4	24.0	22.5	21.5	24.2

Panel B. Prediction of LIBOR rates

Parameters estimated from zero yield data are used to predict the actual LIBOR rates. All values in the table are in basis points. “Stdev” stands for standard deviation, “Correl” stands for correlation, “pred.” stands for prediction, and “Abs. Err.” stands for absolute error.

Maturity	Correl. (libor,pred.)	Stdev. libor	Stdev. pred.	Mean Error	Stdev. Error	Mean Abs. Err.	Stdev. Abs. Err.
1 month	0.5974	58.00	62.35	-46.45	54.13	55.30	45.01
2 months	0.6508	59.32	62.54	-45.81	51.00	53.22	43.18
3 months	0.7003	61.07	62.78	-44.86	47.97	51.06	41.27
4 months	0.7391	62.29	63.02	-44.39	45.27	49.42	39.69
5 months	0.7728	63.65	63.26	-43.88	42.78	47.81	38.32
6 months	0.7983	65.05	63.51	-43.47	40.85	46.42	37.45
7 months	0.8189	66.29	63.75	-42.68	39.20	45.14	36.33
8 months	0.8373	67.59	64.00	-41.81	37.69	43.92	35.20
9 months	0.8535	68.99	64.25	-40.75	36.35	42.61	34.14
10 months	0.8674	69.82	64.50	-39.73	34.97	41.45	32.91
11 months	0.8796	70.61	64.76	-38.67	33.69	40.27	31.74
12 months	0.8915	71.47	65.01	-37.35	32.41	38.90	30.52

4.2.3. *The U.K. market.*

Similar to the Australian market, the model 2B is found to be the preferred model for the U.K. market according to various information criteria reported in Table 10. The preferred model has two different volatility sources, their specification and their corresponding market prices of risk evolution are

$$\begin{aligned}\sigma_1(t,x) &= 0.07155r(t)^{0.7263}\\ \sigma_2(t,x) &= 5.4235(r(t,\tau)-r(t))\\ d\phi_1 &= 15.7712(17.9088-\bar{\phi}_1)dt + 1.8736\sqrt{\phi_1}dW_1(t)\\ d\phi_2 &= 42.9665(6.4041-\bar{\phi}_2)dt + 2.6493\sqrt{\phi_2}dW_2(t).\end{aligned}$$

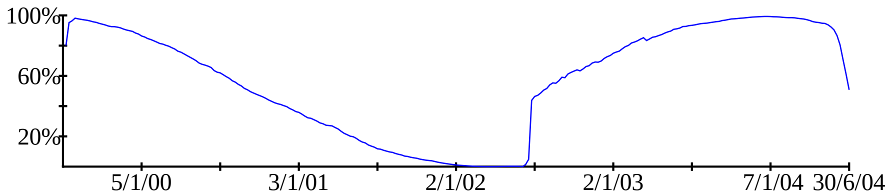
TABLE 10. The U.K. market. Information criteria.

	Model 1	Model 2	Model 2B	Model 3	Model 3C
Log likelihood	64139.58	23445.61	88576.99	23553.42	23446.83
Number of parameters	5	11	9	16	15
Akaike (AIC)	-35.104	-12.827	-48.477	-12.883	-12.825
Schwarz-Baysian (BIC)	-35.096	-12.810	-48.464	-12.859	-12.803
Hannan-Quinn (HIC)	-35.101	-12.821	-48.472	-12.874	-12.817

Even though the yield curve variation in the U.K. market is also determined by the factor and the slope effects as in the Australian market, the slope effect plays a much more significant role in the U.K. market. Figure 8 shows the changes over time of this slope effect contribution to the overall variation of the U.K yield curve. The erratic behaviour comes from the special feature of the U.K. interest rate market. At the beginning of the sample period (i.e. June 1999), the U.K. yield curve had a humped shape with the long maturity yields much smaller than the short maturity yields. The short maturity yield then gradually declined whereas the long maturity yields increased, reducing the the negative spread. The hump feature was still observed for the next two years, until March 2001. The trend of declining short rates and increasing long rates continued until the end of the sample period (June 2004), which resulted in an increasing trend for the spread and the new normal yield curve. The larger the absolute value of the spread, the higher the contribution of the slope factor towards the overall variation of the yield curve.

Unlike the Australian market, where the two volatility factors claim similar financial reward for investors, the level factor in the U.K. market claims an almost 3 times higher financial reward per unit of risk than the slope factor. In addition, the price of the level factor risk has much lower mean reversion rate than that of the slope factor. It takes

FIGURE 8. U.K. market. The contribution of the slope effect to the total forward rate variation



3.3 weeks for the price of the level factor to revert to the mean, compared to only 1.2 weeks for the price of the slope factor.

The in-sample predictive power of this model is similar to that of the Australian market. There is also one particular point of time (the two last observations in 2002) where the filter significantly underestimates the downward movement of the yield curve. On that particular week the yield curve shift down by more than 15 basis points (from 15-19 basis points, depending on maturities), whereas the usual movement of the yield curve is only of few basis points. It can be seen from Panel A, Table 11 that excluding this particular instance reduces the standard deviation of the prediction errors significantly. The mean of the absolute errors then varies between 11-20 basis points, which is reasonable.

Re the ability to predict totally different interest rate series (LIBOR v.s. zero yield rate) with a totally different maturity range, the model does a very good job in predicting the trend, as evidenced by the correlation level of up to 96.6%. However, the value prediction is not good, with mean absolute error lying in the range of 60-80 basis points. In this U.K. market, the zero yield series used in the estimation only have a standard deviation of 18-100 basis points, whereas the short term LIBOR rates have a much higher standard deviation of 93-104 basis points. Therefore, the prediction of LIBOR rates implied by the parameter values estimated from zero yield rates underestimates the variation of the actual LIBOR rates, which leads to the not so small errors.

TABLE 11. U.K. market. Prediction errors.

Panel A. Prediction of the zero yields

This table reports the prediction errors for all of the U.K. yield series used in the estimation. All values are reported as basis points.

Maturity (in years)	2	3	4	5	7	10	12	All
Mean error	-25.0	-9.2	-2.7	-0.3	0.6	-2.1	-4.6	-6.2
Stdev. error	176.8	176.5	176.2	176.1	176.2	176.9	177.5	176.6
Mean absolute error	31.1	24.4	22.7	21.7	23.1	29.2	33.1	26.5
Stdev absolute error	175.8	175.0	174.7	174.8	174.7	174.5	174.4	174.8
Without the large error at the last observation in 2002								
Mean error	-15.2	0.7	7.1	9.6	10.5	7.8	5.2	3.7
Stdev. error	20.1	17.1	13.7	11.5	12.5	21.3	26.0	17.5
Mean absolute error	19.5	12.6	11.0	10.0	11.4	17.5	21.4	14.8
Stdev absolute error	15.9	11.6	10.9	11.1	11.8	14.4	15.6	13.0

Panel B. Prediction of LIBOR rates

Parameters estimated from zero yield data are used to predict the actual LIBOR rates. All values in the table are in basis points. "Stdev" stands for standard deviation, "Correl" stands for correlation, "pred." stands for prediction, and "Abs. Err." stands for absolute error.

Maturity	Correl. (libor,pred.)	Stdev. libor	Stdev. pred.	Mean Error	Stdev. Error	Mean Abs. Err.	Stdev. Abs. Err.
1 month	0.8803	92.52	93.43	-79.76	45.50	80.79	43.65
2 months	0.9003	94.85	93.60	-76.46	42.09	77.01	41.08
3 months	0.9157	96.66	93.90	-73.82	39.22	74.02	38.83
4 months	0.9281	97.93	94.19	-72.23	36.60	72.25	36.55
5 months	0.9375	98.95	94.49	-70.81	34.47	70.81	34.47
6 months	0.9436	99.80	94.79	-69.44	33.05	69.45	33.05
7 months	0.9484	100.45	95.10	-67.88	31.85	67.94	31.72
8 months	0.9528	101.23	95.40	-66.20	30.74	66.32	30.49
9 months	0.9563	101.98	95.70	-64.56	29.89	64.73	29.52
10 months	0.9600	102.62	96.01	-62.65	28.84	62.86	28.37
11 months	0.9632	103.30	96.31	-60.69	27.94	60.95	27.37
12 months	0.9658	104.00	96.62	-58.70	27.25	59.00	26.59

5. Conclusion

The HJM framework provides a very flexible tool for interest rate modelling. Even though theoretical research has advanced quickly, the advantages of HJM models have not been fully realized in practical applications due to the lack of empirical work. More research needs to be done on the challenging task of HJM model estimation in order to obtain a better understanding of interest rate volatility that is much needed in the process of assessing and managing risk as well as pricing derivative securities. This paper has attempted to contribute to the empirical literature by proposing an estimation framework that can be applied to a broad class of nonlinear HJM models.

The paper uses the local linearization filter to build up a maximum likelihood estimator which is able to identify all parameters of the model, and to exploit both time series and cross-sectional data. The local linearization scheme is based on an Itô-Taylor expansion of the nonlinear drift and diffusion terms of the driving dynamics to better take into account the stochastic behaviour of the interest rate system, and an optimal linear filter is subsequently applied. This filter has been chosen because of its advantages over other filters as shown by Shoji (1998) and its better numerical and stability properties as demonstrated by Jimenez et al. (1999).

The estimator is then used to estimate the interest rate volatility structure in the U.S, the U.K. and the Australian markets, using zero coupon bond yields. A range of models was proposed and various information criteria were used to select the best model. In the U.S. market, a 3-factor model is found to be the best, whereas a 2-factor model is the preferred one in the Australian and the U.K. markets. The two factors driving the yield curve evolution in the U.K. and the Australian markets are the level and the slope factors. In the U.S. market, apart from the level factor and the slope factor (which is modelled slightly different than in the other two markets), there is a third significant factor that creates the hump feature on the yield curve.

Unlike previous research, we find that the level factor is a clearly dominant factor in determining the overall yield curve variation only in the Australian market. In the U.K. market, the slope factor is much more important for more than half of the sample period. In the U.S. market, the level factor is not as important as the slope factor for yields with short maturities. The role of the level factor increases as the yield to maturity lengthens.

From an investor point of view, a financial reward is required to bear volatility risk. In the U.S. and the U.K. market, the unit price of the volatility risk coming from the level factor is higher than that of the risk coming from the slope factor, whereas the two types of risk are priced similarly in the Australian market. In all of the markets, the level factor contributes mostly toward the overall risk compensation. This contribution is homogeneous across maturities in the Australian and the U.K. markets. In the U.S. market, the contribution from the level factor diminishes at medium-short maturity of 1-3 years, where the contribution of the third volatility factor is highest. This third volatility factor is the one that creates the hump feature for the U.S. yield curve. A knowledge of how each factor contributes to the overall volatility and the rewards for bearing the risk will help investors manage the risk of interest rate portfolios.

The filter adopted here is certainly not the only nonlinear filter available to modellers. It is left for future research to explore other filters, so as to find a good trade-off between reduction in computational requirements, increase in accuracy and better statistical reliability, all of which are crucial if financial managers are to re-assess their models frequently.

APPENDIX A. MARKOVIANIZATION OF THE INTEREST RATE DYNAMICS

Assuming that the forward rate $r(t, x)$ defined in (2.1) has a volatility function $\boldsymbol{\sigma}(t, x)$ that satisfies Assumption 2.1. Proposition 3.4 in Chiarella and Kwon (2003) states that the forward rate curve can be expressed as an affine function of some state variables, i.e.

$$\begin{aligned} r(t,x) = r(0,t+x) &+ \sum_{i=1}^{I}\sum_{l=1}^{L_i}\sigma_{il}(t+x)\psi_l^i(t) \\ &+ \sum_{i=1}^{I}\sum_{\substack{l,l^*=1 \\ l\le l^*}}^{L_i}[\sigma_{il}(t+x)\bar{\sigma}_{il^*}(t+x)+\epsilon_{ll^*}\sigma_{il^*}(t+x)\bar{\sigma}_{il}(t+x)]\varphi_{ll^*}^i(t), \end{aligned} \tag{A.1}$$

where

$$\begin{aligned} \bar{\sigma}_{il}(x) &= \int_0^x \sigma_{il}(s)\,ds, \\ \varphi_{ll^*}^i(t) &= \int_0^t c_{il}(s)c_{il^*}(s)\,ds, \\ \psi_l^i(t) &= \int_0^t c_{il}(s)\,d\widetilde{W}_i(s) - \sum_{l^*=1}^{d_i}\int_0^t c_{il}(s)c_{il^*}(s)\bar{\sigma}_{il^*}(s)\,ds, \\ \epsilon_{ll^*} &= \begin{cases} 1, & \text{if } l \neq l^*, \\ 0, & \text{if } l = l^*. \end{cases} \end{aligned}$$

and $\widetilde{W}_i, (i = 1, \ldots, I)$ are standard Wiener processes under the equivalent measure $\tilde{\mathcal{P}}$.

Under this setting, the economic meaning of the state variables φ and ψ is not clear. The next step is to use the forward rates themselves as the state variables.

Let $\mathscr{S} = \{\psi_l^i(t), \varphi_{lk}^i(t)\}$. Define $N = |\mathscr{S}|$, choose an ordering for $\mathscr{S}$ and write $\chi_n(t)$ for the elements of $\mathscr{S}$ so that $\mathscr{S} = \{\chi_1(t), \ldots, \chi_N(t)\}$. Then (A.1) can be written

$$r(t,x) = a_0(t,x) + \sum_{n=1}^{N} a_n(t,x)\,\chi_n(t), \tag{A.2}$$

for suitable deterministic functions $a_0(t,x)$ and $a_n(t,x)$.

Corollary A.1. *Suppose that the conditions of Assumption 2.1 are satisfied. If there exist $\tau_1, \tau_2, \ldots, \tau_N \in \mathbb{R}_+$ such that the matrix*

$$\boldsymbol{A}(t,\tau_1,\ldots,\tau_N) = \begin{bmatrix} a_1(t,\tau_1) & a_2(t,\tau_1) & \cdots & a_N(t,\tau_1) \\ a_1(t,\tau_2) & a_2(t,\tau_2) & \cdots & a_N(t,\tau_2) \\ \cdots & \cdots & \cdots & \cdots \\ a_1(t,\tau_N) & a_2(t,\tau_N) & \cdots & a_N(t,\tau_N) \end{bmatrix} \tag{A.3}$$

is invertible for all $t \in \mathbb{R}_+$, then the variables $\chi_n(t)$ can be expressed in the form

$$\boldsymbol{\chi}(t) = \boldsymbol{A}(t,\tau_1,\ldots,\tau_N)^{-1}\left[\boldsymbol{a}_0(t,\tau_1,\ldots,\tau_N) - \boldsymbol{r}(t,\tau_1,\ldots,\tau_N)\right], \tag{A.4}$$

where

$$\begin{aligned} \boldsymbol{\chi}(t) &= [\chi_1(t), \ldots, \chi_N(t)]', \\ \boldsymbol{a}_0(t,\tau_1,\ldots,\tau_N) &= [a_0(t,\tau_1), \ldots, a_0(t,\tau_N)]', \\ \boldsymbol{r}(t,\tau_1,\ldots,\tau_N) &= [r(t,\tau_1), \ldots, r(t,\tau_N)]'. \end{aligned}$$

The whole forward curve then can be written in terms of these new economically meaningful state variables

$$\begin{aligned} r(t,x) = a_0(t,x) &- \boldsymbol{a}(t,x)'\boldsymbol{A}(t,\tau_1,\ldots,\tau_N)^{-1}\boldsymbol{a}_0(t,\tau_1,\ldots,\tau_N) \\ &+ \boldsymbol{a}(t,x)'\boldsymbol{A}(t,\tau_1,\ldots,\tau_N)^{-1}\boldsymbol{r}(t,\tau_1,\ldots,\tau_N), \end{aligned} \tag{A.5}$$

where

$$\boldsymbol{a}(t,x) = [a_1(t,x), \ldots, a_N(t,x)]'.$$

Therefore, the HJM models admits a N-dimensional affine realization in terms of the set of discrete tenor forward rates $\boldsymbol{r}(t,\tau_1,\ldots,\tau_N)$. This set of forward rates forms a

Markov process, and under $\widetilde{\mathcal{P}}$ each forward rate $r(t,x)$ satisfies the stochastic differential equation

$$\begin{aligned} dr(t,x) = \Bigg[& \frac{\partial a_0(t,x)}{\partial x} - \frac{\partial \boldsymbol{a}(t,x)'}{\partial x} \boldsymbol{A}(t,\tau_1,\dots,\tau_N)^{-1} \boldsymbol{a}_0(t,\tau_1,\dots,\tau_N) \\ & + \frac{\partial \boldsymbol{a}(t,x)'}{\partial x} \boldsymbol{A}(t,\tau_1,\dots,\tau_N)^{-1} \boldsymbol{r}(t,\tau_1,\dots,\tau_N) + \boldsymbol{\sigma}(t,t+x)'\bar{\boldsymbol{\sigma}}(t,t+x) \Bigg] dt \\ & + \boldsymbol{\sigma}(t,t+x)' d\widetilde{\boldsymbol{W}}(t). \end{aligned}$$

In terms of the real world measure, where $\boldsymbol{\phi} \equiv (\phi_1,\dots,\phi_I)$ is the vector of market prices of risk associating with the Wiener process $\boldsymbol{W}$, the system becomes

$$\begin{aligned} dr(t,x) = \Bigg[& \frac{\partial a_0(t,x)}{\partial x} - \frac{\partial \boldsymbol{a}(t,x)'}{\partial x} \boldsymbol{A}(t,\tau_1,\dots,\tau_N)^{-1} \boldsymbol{a}_0(t,\tau_1,\dots,\tau_N) \\ & + \frac{\partial \boldsymbol{a}(t,x)'}{\partial x} \boldsymbol{A}(t,\tau_1,\dots,\tau_N)^{-1} \boldsymbol{r}(t,\tau_1,\dots,\tau_N) + \boldsymbol{\sigma}(t,t+x)'\bar{\boldsymbol{\sigma}}(t,t+x) \\ & - \boldsymbol{\phi}'\boldsymbol{\sigma}(t,t+x) \Bigg] dt + \boldsymbol{\sigma}(t,t+x)' d\boldsymbol{W}(t), \end{aligned}$$

which is (2.4) in the main text.

The yield $y(t,x)$ can also be expressed as an affine function of forward rates

$$\begin{aligned} y(t,x) = & b_0(t,x) - \boldsymbol{b}(t,x)'\boldsymbol{A}(t,\tau_1,\dots,\tau_N)^{-1}\boldsymbol{a}_0(t,\tau_1,\dots,\tau_N) \\ & + \boldsymbol{b}(t,x)'\boldsymbol{A}(t,\tau_1,\dots,\tau_N)^{-1}\boldsymbol{r}(t,\tau_1,\dots,\tau_N), \end{aligned}$$

where

$$\begin{aligned} b_0(t,x) &= \frac{1}{x}\int_0^x a_0(t,u)du, \\ \boldsymbol{b}(t,x) &= \frac{1}{x}\int_0^x \boldsymbol{a}(t,u)du. \end{aligned}$$

This affine yield expression is equation (2.6) in the main text.

Appendix B. Local Linearization Filter for Linear Continuous-Discrete State Space Models

Jimenez and Ozaki (2002) analyzed a linear state space model defined by the continuous state equation

$$d\mathbf{x}(t) = (\mathbf{A}(t)\mathbf{x}(t) + a(t))\,dt + \sum_{i=1}^{m} (\mathbf{B}_i(t)\mathbf{x}(t) + \mathbf{b}_i(t))\,d\mathbf{W}_i(t), \tag{B.1}$$

and the discrete observation equation[9]

$$\mathbf{z}_{t_j} = \mathbf{C}(t_j)\mathbf{x}(t_j) + \mathbf{e}_{t_j}, \text{ for } j = 0, 1, \ldots, J, \tag{B.2}$$

where $\mathbf{x}(t) \in \mathbb{R}^d$ is the state vector at the instant of time t, $\mathbf{z}_{t_j} \in \mathbb{R}^r$ is the observation vector at the instant of time t_j, $\mathbf{W}$ is a m-dimensional vector of independent Wiener processes, and $\{\mathbf{e}_{t_j} : \mathbf{e}_{t_j} \sim \mathcal{N}(0, \mathbf{\Pi}), j = 0, \ldots, J\}$ is a sequence of random vector i.i.d.

Define $\hat{\mathbf{x}}_{t|\rho} = \mathbb{E}\left(\mathbf{x}(t)|Z_\rho\right)$ and $\mathbf{P}_{t|\rho} = \mathbb{E}((\mathbf{x}(t) - \hat{\mathbf{x}}_{t|\rho})(\mathbf{x}(t) - \hat{\mathbf{x}}_{t|\rho})'|Z_\rho)$ for all $\rho \leq t$, where $Z_\rho = \{\mathbf{z}_{t_j} : t_j \leq \rho\}$.

Suppose that $\mathbb{E}\left(\mathbf{W}(t)\mathbf{W}'(t)\right) = \mathbf{I}$, $\hat{\mathbf{x}}_{t_0|t_0} < \infty$ and $\mathbf{P}_{t_0|t_0} < \infty$.

Theorem B.1. *(Jimenez and Ozaki (2002)) The optimal (minimum variance) linear filter for the linear model (B.1)- (B.2) consists of equations of evolution for the conditional mean $\hat{\mathbf{x}}_{t|t}$ and the covariance matrix $\mathbf{P}_{t|t}$. Between observations, these satisfy the ordinary differential equation*

$$d\hat{\mathbf{x}}_{t|t} = \left(\mathbf{A}(t)\hat{\mathbf{x}}_{t|t} + \mathbf{a}(t)\right) dt, \tag{B.3}$$

$$\begin{aligned} d\mathbf{P}_{t|t} = \Bigg(& \mathbf{A}(t)\mathbf{P}_{t|t} + \mathbf{P}_{t|t}\mathbf{A}'(t) + \sum_{i=1}^{m} \mathbf{B}_i(t)\left(\mathbf{P}_{t|t} + \hat{\mathbf{x}}_{t|t}\hat{\mathbf{x}}'_{t|t}\right)\mathbf{B}'_i(t) \\ & + \sum_{i=1}^{m}\left(\mathbf{B}_i(t)\hat{\mathbf{x}}_{t|t}\mathbf{b}'_i(t) + \mathbf{b}_i(t)\hat{\mathbf{x}}'_{t|t}\mathbf{B}'_i(t) + \mathbf{b}_i(t)\mathbf{b}'_i(t)\right)\Bigg) dt, \end{aligned} \tag{B.4}$$

for all $t \in [t_j, t_{j+1})$. At an observation at t_j, they satisfy the difference equation

$$\hat{\mathbf{x}}_{t_{j+1}|t_{j+1}} = \hat{\mathbf{x}}_{t_{j+1}|t_j} + \mathbf{K}_{t_{j+1}}\left(\mathbf{z}_{t_{j+1}} - \mathbf{C}(t_{j+1})\hat{\mathbf{x}}_{t_{j+1}|t_j}\right), \tag{B.5}$$

$$\mathbf{P}_{t_{j+1}|t_{j+1}} = \mathbf{P}_{t_{j+1}|t_j} - \mathbf{K}_{t_{j+1}}\mathbf{C}(t_{j+1})\mathbf{P}_{t_{j+1}|t_j}, \tag{B.6}$$

where

$$\mathbf{K}_{t_{j+1}} = \mathbf{P}_{t_{j+1}|t_j}\mathbf{C}'(t_{j+1})\left(\mathbf{C}(t_{j+1})\mathbf{P}_{t_{j+1}|t_j}\mathbf{C}'(t_{j+1}) + \mathbf{\Pi}\right)^{-1} \tag{B.7}$$

is the filter gain. The prediction $\hat{\mathbf{x}}_{t|\rho}$ and $\mathbf{P}_{t|\rho}$ are accomplished, respectively, via expressions (B.3) and (B.4) with initial conditions $\hat{\mathbf{x}}_{t_0|t_0}$ and $\mathbf{P}_{t_0|t_0}$ and $\rho < t$.

[9]Their original specification is

$$\mathbf{z}_{t_j} = \mathbf{C}(t_j)\mathbf{x}(t_j) + \sum_{i=1}^{n}\mathbf{D}_i(t_j)\mathbf{x}(t_j)\boldsymbol{\xi}^i_{t_j} + \mathbf{e}_{t_j}, \text{ for } j = 0, 1, \ldots, J,$$

where $\{\boldsymbol{\xi}_{t_j} : \boldsymbol{\xi}_{t_j} \sim \mathcal{N}(0, \mathbf{\Lambda}), \mathbf{\Lambda} = \operatorname{diag}((\lambda_1, \ldots, \lambda_n)), j = 0, \ldots, J\}$ is a sequence of random vector i.i.d., and $\mathbb{E}(\boldsymbol{\xi}^i_{t_j}, \mathbf{e}_{t_j}) = \boldsymbol{\vartheta}^i(t_j)$. However, in most finance applications, the noise term $\boldsymbol{\xi}$ is not required.

The analytical solution for this system of equations can be easily found, for details see Jimenez and Ozaki (2003). They also provide some equivalent expressions that are easier to implement via computer programs.

REFERENCES

Aït-Sahalia, Y. (1999), 'Transition Densities for Interest Rate and Other Nonlinear Diffusions', *Journal of Finance* **54**(4), 1361–1395.

Aït-Sahalia, Y. (2002), 'Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach', *Econometrica* **70**, 223–263.

Aït-Sahalia, Y. (2003), Closed-Form Likelihood Expansions for Multivariate Distibutions, Working paper, Princeton University and NBER.

Bhar, R. and Chiarella, C. (1997a), 'Transformation of Heath-Jarrow-Morton models to Markovian Systems', *European Journal of Finance* **3**, 1–26.

Bhar, R., Chiarella, C., Hung, H. and Runggaldier, W. (2005), 'The Volatility of the Instantaneous Spot Interest Rate Implied by Arbitrage Pricing - A Dynamic Bayesian Approach', *Automatica* . Forthcoming.

Björk, T. and Svensson, L. (2001), 'On the Existence of Finite Dimensional Realizations for Nonlinear Forward Rate Models', *Mathematical Finance* **11**, 205–243.

Bliss, R. R. and Ritchken, P. (1996), 'Empirical Tests of Two State-Variable Heath-Jarrow-Morton Models', *Journal of Money, Credit and Banking* **28**(3), 452–476.

Brace, A., Gatarek, D. and Musiela, M. (1997), 'The Market Model of Interest Rate Dynamics', *Mathematical Finance* **7**(2), 127–155.

Brandt, M. W. and He, P. (2002), Simulated Likelihood Estimation of Affine Term Structure Models from Panel Data, Working paper, University of Pennsylvania.

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

Broze, L., Scaillet, O. and Zakoian, J.-M. (1998), 'Quasi-Indirect Inference for Diffusion Processes', *Econometric Theory* **14**(2), 161–186.

Carverhill, A. (1994), 'When is the Short Rate Markovian?', *Mathematical Finance* **4**(4), 305–312.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. and Sanders, A. B. (1992), 'An Empirical Comparison of Alternative Models of the Short-Term Interest Rate', *Journal of Finance* **47**, 1209–1221.

Chen, R.-R. and Scott, L. (1993), 'Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rate', *Journal of Fixed Income* **3**(3), 14–31.

Chiarella, C. and Kwon, O. K. (2001b), 'Forward Rate Dependent Markovian Transformations of the Heath-Jarrow-Morton Term Structure Model', *Finance and Stochastics* **5**(2), 237–257.

Chiarella, C. and Kwon, O. K. (2003), 'Finite Dimensional Affine Realisations of HJM Models in Terms of Forward Rates and Yields', *Review of Derivatives Research* **6**(2), 129–155.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985a), 'An Intertemporal General Equilibrium Model of Asset Prices', *Econometrica* **53**(2), 363–384.

de Jong, F. and Santa-Clara, P. (1999), 'The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables', *Journal of Financial and Quantitative Analysis* **34**(1), 131–157.

Dell'Aquila, R., Ronchetti, E. and Trojani, F. (2003), 'Robust GMM Analysis of Models for the Short Rate Process', *Journal of Empirical Finance* **10**(3), 373–397.

Duffee, G. R. and Stanton, R. H. (2004), Estimation of Dynamic Term Structure Models, Working paper, Haas School of Business, U.C. Berkeley.

Duffie, D. and Kan, R. (1996), 'A Yield-Factor Model of Interest Rates', *Mathematical Finance* **6**(4), 379–406.

Duffie, D. and Singleton, K. (1993), 'Simulated Moments Estimation of Markov Models of Asset Prices', *Econometrica* **61**(4), 929–952.

Elerian, O., Chib, S. and Shephard, N. (2001), 'Likelihood Inference for Discretely Observed Nonlinear Diffusions', *Econometrica* **69**(4), 959–993.

Eraker, B. (2001), 'MCMC Analysis of Diffusion Models with Application to Finance', *Journal of Business and Economic Statistics* **19**(2), 177–191.

Gallant, A. R. and Tauchen, G. (1996), 'Which Moments to Match', *Econometric Theory* **12**(4), 657–681.

Gallant, A. R. and Tauchen, G. (1997), 'Estimation of Continuous Time Models for Stock Returns and Interest Rates', *Macroeconomic Dynamics* **1**, 135–168.

Gallant, A. R. and Tauchen, G. (1998), 'Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions', *Journal of the American Statistical Association* **93**(441), 10–24.

Geyer, A. L. J. and Pichler, S. (1999), 'A State-Space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure', *Journal of Financial Research* **22**(1), 107–130.

Heath, D., Jarrow, R. and Morton, A. (1992), 'Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation', *Econometrica* **60**(1), 77–105.

Ho, M., Perraudin, W. and Sørensen, B. (1996), 'A continuous-time arbitrage-pricing model with stochastic volatility and jump', *Journal of Business and Economic Statistics* **14**(1), 31–43.

Holland, J. (1975), *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor. Extended New Eddition, MIT Press, Cambridge, 1992.

Inui, K. and Kijima, M. (1998), 'A Markovian Framework in Multi-Factor Heath-Jarrow-Morton Models', *Journal of Financial and Quantitative Analysis* **33**(3), 423–440.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994), 'Bayesian Analysis of Stochastic Volatility Models', *Journal of Business and Economic Statistics* **12**(4), 371–417.

Jegadeesh, N. and Pennacchi, G. (1996), 'The Behavior of Interest Rates Implied by the Term Structure of Eurodollar Futures', *Journal of Money, Credit and Banking* **28**(3), 426–446.

Jimenez, J. C. and Ozaki, T. (2002), 'Linear Estimation of Continuous-Discrete Linear State Space Models with Multiplicative Noise', *Systems and Control Letters* **47**, 91–101. Erratum appears in V49, page 161, 2003.

Jimenez, J. C. and Ozaki, T. (2003), 'Local Linearization Filters for Non-Linear Continuous-Discrete State Space Models with Multiplicative Noise', *International Journal of Control* **76**(12), 1159–1170.

Jimenez, J. C., Shoji, I. and Ozaki, T. (1999), 'Simulation of Stochastic Differential Equations through the Local Linearization Method. A Comparative Study', *Journal of Statistical Physics* **94**, 587–602.

Kalman, R. E. (1960), 'A New Approach to Linear Filtering and Prediction Problems', *Journal of Basic Engineering, Transactions ASME, Series D* **82**, 35–45.

Kalman, R. E. and Bucy, R. S. (1961), 'New Results in Linear Filtering and Prediction Theory', *Journal of Basic Engineering, Transactions ASME, Series D* **83**, 95–108.

Kim, S., Shephard, N. and Chib, S. (1998), 'Stochatic Volatility: Likelihood Inference and Comparison with ARCH Models', *Review of Economic Studies* **65**(224), 361–393.

Knez, P. J., Litterman, R. and Scheinkman, J. (1994), 'Exploration into Factors Explaining Money Market Returns', *Journal of Finance* **49**(5), 1861–1882.

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

Michalewicz, Z. (1999), *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag.

Mitchell, M. (1996), *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA.

Ozaki, T. (1993), 'A Local Linearization Approach to Nonliner Filtering', *International Journal of Control* **57**, 75–96.

Pagan, A. R., Hall, A. D. and Martin, V. (1996), Modeling the Term Structure, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics', Vol. 14, Elsevier Science, chapter 4, pp. 91–118.

Pearson, N. D. and Sun, T.-S. (1994), 'Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model', *Journal of Finance* **49**(4), 1279–1304.

Ritchken, P. and Sankarasubramanian, L. (1995), 'The Importance of Forward Rate Volatility Structures in Pricng Interest Rate-Sensitive Claims', *Journal of Derivatives* **3**(1), 25–41.

Rossi, G. D. (2004), 'Kalman Filtering of Consistent Forward Rate Curves: A tool to Estimate and Model Dynamically the Term Structure', *Journal of Empirical Finance* **11**(2), 277–308.

Shoji, I. (1998), 'A Comparative Study of Maximum Likelihood Estimators for Nonlinear Dynamical System Models', *International Journal of Control* **71**(3), 391–404.

Singh, M. K. (1995), 'Estimation of Multifactor Cox, Ingersoll, and Ross Term Structure Model: Evidence on Volatility Structure and Parameter Stability', *Journal of Fixed Income* **5**(2), 8–28.

Singleton, K. (2001), 'Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function', *Journal of Econometrics* **102**(1), 111–141.

Vose, M. D. (1999), *The Simple Genetic Algorithm: Foundations and Theory*, MIT Press, Cambridge, MA.

Zhou, H. (2001), 'Finite Sample Properties of EMM, GMM, QMLE and MLE for a Square-Root Interest Rate Diffusion Model', *Journal of Computational Finance* **5**, 89–122.