The Volatility Structure of the Fixed Income Market under the HJM Framework: A Nonlinear Filtering Approach

Carl Chiarella [a], Hing Hung [a], Thuy-Duong Tô [b],*

[a]*School of Finance and Economics, University of Technology, Sydney, PO Box 123, Broadway NSW2007, Australia*

[b]*School of Banking and Finance, University of New South Wales, Sydney 2052, Australia*

Abstract

The dynamics for interest rate processes within the well-known multi-factor Heath, Jarrow and Morton (HJM) specification are considered. Despite the flexibility of and the notable advances in theoretical research about the HJM model, the number of empirical studies of it is still very sparse. This paucity is principally due to the difficulties in estimating models in this class, which are not only high-dimensional, but also nonlinear and involve latent state variables. The estimation of a fairly broad class of HJM models as a nonlinear filtering problem is undertaken by adopting the local linearization filter, 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 three-factor models 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

* Corresponding author
  Email addresses: `carl.chiarella@uts.edu.au` (Carl Chiarella),
`hing.hung@uts.edu.au` (Hing Hung), `t.to@unsw.edu.au` (Thuy-Duong Tô).

1 Introduction

Management of interest rate risk is of crucial importance to financial institutions and corporations. The volatility structure of the 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 centred 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 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 need to be 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 Eraker (2001), Elerian et al. (2001), Jacquier et al. (2007) and Golightly and Wilkinson (2008). 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).

Shoji and Ozaki (1997) investigate the finite sample performances of five different estimation methods for stochastic processes based on discrete data using GMM and

ML associated with various discretization systems. GMM is found to underperform all other methods, including the MLE based on a simple Euler discretization scheme. Faff and Gray (2006) study various different short rate models, including the Vasicek (1977), the CIR (1985a, 1985b), the Brennan-Schwartz (1980) and the CKLS (1992) models. They conclude that GMM has great difficulty estimating the speed of mean reversion in interest rates. Zhou (2001) provide further evidence on the finite sample properties of the ML 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 a univariate setting. Under a multivariate setting, this performance can deteriorate. Finally 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 all of the above-cited findings this paper pursues further the filtering approach. The linear filtering and prediction problem became well understood after the important work of Kalman (1960) and Kalman and Bucy (1961), as well as other important contribution of econometricians as summarised in Pollock (2003). Application of the Kalman filter is now widespread in finance and economics. Examples are the recent contributions of Bates (2006) and Jiang and Oomen (2007) who use the linear filter to extract latent variables from affine Markov processes. However, non-affine Markov processes are used frequently in various financial models. These processes require nonlinear filtering, which is still an active research area due to its technical difficulty. Various approximations for nonlinear filters have been proposed, such as the Extended Kalman filter, the Iterated Extended Kalman filter and 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 (that is 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 have chosen to adopt 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

$$r(t, x) = r(0, t + x) + \int_0^t \boldsymbol{\sigma}(s, t + x)' \left[\bar{\boldsymbol{\sigma}}(s, t + x) - \boldsymbol{\phi}(s) \right] ds \\ + \int_0^t \boldsymbol{\sigma}(s, t + x)' d\boldsymbol{W}(s), \tag{1}$$

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

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 1   *(i) For each $1 \leq i \leq I$, there exists $L_i \in \mathbb{N}$ such that the compo-*

4

nents, $\sigma_i(t, x)$, of the forward rate volatility process have the form

$$\sigma_i(t, x) = \sum_{l=1}^{L_i} c_{il}(t)\sigma_{il}(x), \tag{2}$$

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

(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

$$c_{il}(t) = \hat{c}_{il}(t, r(t, x_1), \ldots, r(t, x_M)), \tag{3}$$

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

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

(see Appendix A for a brief summary). This set of forward rates forms a Markov process. In terms of the real world measure, where $\boldsymbol{\phi} \equiv (\phi_1, \ldots, \phi_I)$ 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

$$\begin{aligned}
dr(t, x) =& [p_0(t, x, \tau_1, \ldots, \tau_N) + \boldsymbol{p}'(t, x, \tau_1, \ldots, \tau_N)\boldsymbol{r}(t, \tau_1, \ldots, \tau_N) \\
& - \boldsymbol{\phi}'\boldsymbol{\sigma}(t, t + x)]dt + \boldsymbol{\sigma}(t, t + x)'d\boldsymbol{W}(t),
\end{aligned} \tag{4}$$

and the full functional forms for the coefficient functions p_0 and $\boldsymbol{p}$ can be found in Appendix A.

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

$$y(t, x) = \frac{1}{x} \int_0^x r(t, u)du, \tag{5}$$

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

$$y(t, x) = q_0(t, x, \tau_1, \ldots, \tau_N) - \boldsymbol{q}'(t, x, \tau_1, \ldots, \tau_N)\boldsymbol{r}(t, \tau_1, \ldots, \tau_N), \tag{6}$$

where the $q_i(t, x, \tau_1, \ldots, \tau_N)$ is a set of deterministic functions (see Appendix A for definitions of the q_i). 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.

5

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 (1). We shall specifically consider there volatility functions, namely

$$\sigma_1(t, x) = \gamma_1 r^\lambda(t), \tag{7}$$

$$\sigma_2(t, x) = \gamma_2 \, e^{-\kappa_2(x-t)}, \tag{8}$$

$$\sigma_3(t, x) = \gamma_3(x - t) \, e^{-\kappa_3(x-t)} \, . \tag{9}$$

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, where each volatility shock is transmitted along the yield curve, with the impact dying out as maturities increase. 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 forms that lead 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

$$d\phi_i = \alpha_i(\bar{\phi}_i - \phi_i)dt + \beta_i\sqrt{\phi_i(t)}dW_i(t). \tag{10}$$

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.

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) can be written as

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

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 (6) for the yields, into which they introduce an error term, then estimate their model via the Maximum Likelihood procedure. The main drawback of this approach is that the estimation procedure can only identify κ, as the relationship (6), which is actually an expression of the whole yield curve as an affine function of some particular yields rather than the forward rates, 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 the Bliss and Ritchken (1996) approach by using both the relationship (6) and the Markovian system (4) in their estimation procedure. They use the Kalman filtering method where (6) serves as the observation equation and (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 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

$$d\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t))dt + \sum_{i=1}^{m} \mathbf{g}_i(t, \mathbf{x}(t))dW_i(t), \tag{11}$$

and the discrete observation equation

$$\mathbf{z}_{t_j} = \mathbf{C}(t_j)\mathbf{x}(t_j) + \mathbf{e}_{t_j}, \text{ for } j = 0, 1, \ldots, J, \tag{12}$$

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 $\{\mathbf{e}_{t_j} : \mathbf{e}_{t_j} \sim \mathcal{N}(0, \boldsymbol{\Pi}), j = 0, \ldots, J\}$ is a sequence of i.i.d. random vectors.

It should be noted that a full (nonlinear) specification of the observation equation would be

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

where $\mathbf{h}$ and $\mathbf{p}_i$ are nonlinear functions, $\{\boldsymbol{\xi}_{t_j} : \boldsymbol{\xi}_{t_j} \sim \mathcal{N}(0, \boldsymbol{\Lambda}), \boldsymbol{\Lambda} = \text{diag}((\lambda_1, \ldots, \lambda_n)), j = 0, \ldots, J\}$ is a sequence of random vector i.i.d., and $\xi_{t_j}^i$ 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 $\boldsymbol{\xi}$. Henceforth we will work with the discrete observation equation (12).

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

$$
\begin{aligned}
\mathbf{f}(t, \mathbf{x}(t)) \approx & f(s, \mathbf{u}) + \left(\frac{\partial \mathbf{f}(s, \mathbf{u})}{\partial s} + \frac{1}{2} \sum_{k,l=1}^{d} [\mathbf{G}(s, \mathbf{u})\mathbf{G}'(s, \mathbf{u})]^{k,l} \frac{\partial^2 \mathbf{f}(s, \mathbf{u})}{\partial \mathbf{u}^k \partial \mathbf{u}^l} \right) (t - s) \\
& + \mathbf{J_f}(s, \mathbf{u})(\mathbf{x}(t) - \mathbf{u}),
\end{aligned}
\tag{13}
$$

where $(s, \mathbf{u}) \in \mathbb{R} \times \mathbb{R}^d$, $\mathbf{J_f}(s, \mathbf{u})$ 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 $\mathbf{G}(s, \mathbf{u}) \equiv (\mathbf{g}_1, \dots, \mathbf{g}_m)$. 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 following an Itô calculus expansion according to the Itô-Taylor formula will better take into account its stochastic nature.

Using such approximations for $\mathbf{f}$ and $\mathbf{g}_i$, the solution of the nonlinear state equation (11) can be approximated by the solution of the piecewise linear stochastic differential equation

$$
\begin{aligned}
d\mathbf{y}(t) = & \left(\mathbf{A}(t_j, \hat{\mathbf{y}}_{t_j|t_j})\mathbf{y}(t) + \mathbf{a}(t, t_j, \hat{\mathbf{y}}_{t_j|t_j}) \right) dt \\
& + \sum_{i=1}^{m} \left(\mathbf{B}_i(t_j, \hat{\mathbf{y}}_{t_j|t_j})\mathbf{y}(t) + \mathbf{b}_i(t, t_j, \hat{\mathbf{y}}_{t_j|t_j}) \right) dW_i(t)
\end{aligned}
\tag{14}
$$

for all $t \in [t_j, t_{j+1})$, starting at $\mathbf{y}(t_0) = \hat{\mathbf{y}}_{t_0|t_0} = \hat{\mathbf{x}}_{t_0|t_0}$. Here $\mathbf{y}(t)$ is used to denote the solution to the approximate system to distinguish it from the solution $\mathbf{x}(t)$ to the true system. The various quantities appearing in (14) are defined as

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

9

The approximate stochastic differential equation (14) and the corresponding observation equation (see (12))

$$\mathbf{z}_{t_j} = \mathbf{C}(t_j)\mathbf{y}(t_j) + \mathbf{e}_{t_j}, \text{ for } j = 0, 1, \ldots, J, \tag{15}$$

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 $\rho \leq t$. 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 (15)). The mean and covariance matrix of this conditional distribution are given directly by the local linearization filter 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 (14) and (15), 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

$$\mathcal{L}_{\mathbf{Z}}(\boldsymbol{\theta}) = -\frac{rJ}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=1}^{J}\ln|\boldsymbol{\Sigma}_{t_j}| - \frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\nu}'_{t_j}\boldsymbol{\Sigma}_{t_j}^{-1}\boldsymbol{\nu}_{t_j}, \tag{16}$$

where the innovation equations are

$$\boldsymbol{\nu}_{t_j} = \mathbf{z}_{t_j} - \mathbf{C}(t_j)\hat{\mathbf{y}}_{t_j|t_{j-1}}, \tag{17}$$

$$\boldsymbol{\Sigma}_{t_j} = \mathbf{C}(t_j)\mathbf{P}_{t_j|t_{j-1}}\mathbf{C}'(t_j) + \boldsymbol{\Pi}. \tag{18}$$

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

$$\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}} \mathcal{L}_{\mathbf{Z}}(\boldsymbol{\theta}). \tag{19}$$

3.4 Econometric implementation

We now view our model as a continuous-discrete nonlinear state space system, where (4) and (10) serve as the nonlinear state equations, and (6) serves as the linear (affine) observation equation. Similar to the standard practice in the literature, we introduce

10

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 (7) - (9), of the market price of risk specification (10) 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. The application of genetic algorithm to HJM stochastic systems have been tested successfully in Chiarella and Tô (2006).

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 the 6 years from 7th July 1999 to 30th June 2005. The time to maturity of contracts are chosen partly because of the availability of data in Datastream and partly to match the range of maturities of bonds traded in market. The very short end of the yield curve is not included due to the fact that short maturity rates seem to undergo very little diffusion type movements so that their motion is not well captured by stochastic differential equations of the type we seek to estimate here. This is mostly due to the fact that short rates are driven very much by the interest rate policies of the central bank. Hence a model of short rates would also need to incorporate macro-economic factors and the central bank's reaction to them. This task is beyond the scope of the current study. The very long end of the yield curve is also not included as bonds at such maturities are not traded frequently enough to ensure data quality.

11

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 period,

Fig. 1. Zero yield curve, 2-year time to maturity
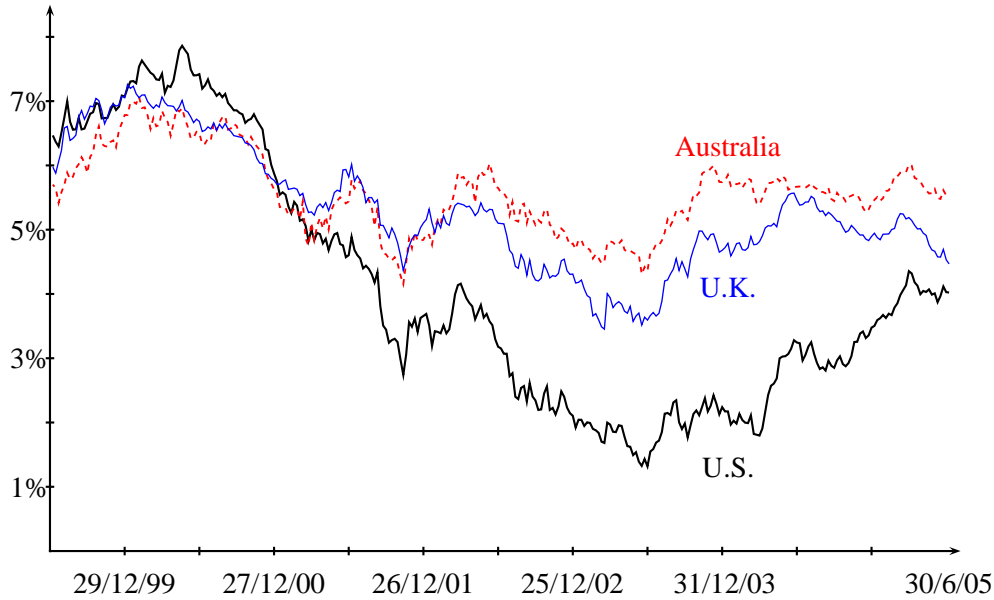
interest rates have changed significantly. The rates increased in all markets by around 1.5% from July 1999 to June 2000. The rates then decreased but at different rhythms across the markets. In the U.S., the rates dived sharply from 7.5% to 1% in the next 3 years, then started to pick up again in the second half of 2003, moved slightly around the 2% level, then followed a sharp upward trend to 4% by June 2005. 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 until June 2004, and slightly decreased afterwards. 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	4.10%	4.99%	5.88%	5.59%	5.97%	6.23%	5.31%	5.53%	5.52%
Stdev	1.92%	1.41%	0.99%	0.63%	0.60%	0.53%	0.94%	0.83%	0.56%
AC(1)	0.995	0.991	0.983	0.976	0.965	0.956	0.990	0.987	0.979

12

4.2 Empirical Results

We separately ran the estimation for 1-, 2- and 3-factor models, whose specifications are illustrated in Table 2

Table 2
Model specification

Model	$\sigma_1(t,x) =$	$\sigma_2(t,x) =$	$\sigma_3(t,x) =$
	$\gamma_1 r^\lambda(t)$	$\gamma_2\, e^{-\kappa_2(x-t)}$	$\gamma_3(x-t)\, e^{-\kappa_3(x-t)}$
Model 1	✓		
Model 2	✓	✓	
Model 3	✓	✓	✓

4.2.1 In-sample Estimation

We ran separate estimations for each model and chose the best one based on various information criteria, namely the Akaike (1974, 1976), the Schwarz-Bayesian (Schwarz (1978)) and the Hannan-Quinn (1979). Standard model specification tests, such as the significance of the estimates, and the properties of the errors, were also considered. In the U.S. market, as has been documented in the literature, the 3-factor model is the best, whereas in the Australian and U.K. markets, the 2-factor model is the preferred one. The estimated parameters of the preferred models for each market can be found in Table 3.

In the U.S. market, we find that there are three components to the volatility of the interest rate markets, namely

$$\sigma_1(t,x) = 2.7572r(t)^{1.7480},$$
$$\sigma_2(t,x) = 0.0249\, e^{-0.2965(x-t)},$$
$$\sigma_3(t,x) = 0.0038(x-t)\, e^{-0.3320(x-t)}\,.$$

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 1.7480, which is close to the value 1.5 found by Chan et al. (1992), and slightly above the range of 0.5 to 1.5 reported in Pagan et al. (1996) (depending on the interest rate series used). The second volatility factor depends on

13

Table 3
Estimated parameter values for the preferred model
This table reports the parameter estimates for the preferred model to each market and their corresponding standard errors. The notation xE-y stands for $x \times 10^{-y}$.

Parameter	U.S. market		U.K. market		Australian market	
	Est	Stderr	Est	Stderr	Est	Stderr
γ_1	2.7572	0.0219	0.1856	1.55E-05	0.0219	2.95E-09
γ_2	0.0249	0.0006	0.0214	1.49E-06	0.0327	2.08E-08
γ_3	0.0038	0.0000	-	-	-	-
λ	1.7480	0.0313	0.6864	1.64E-05	0.2113	4.45E-08
κ_2	0.2965	0.0057	0.2462	1.95E-06	0.4066	2.14E-07
κ_3	0.3320	0.0042	-	-	-	-
α_1	2.2796	0.0734	71.8748	6.71E-03	93.7003	7.88E-05
α_2	3.0075	0.0341	67.9437	6.13E-04	39.8334	2.04E-05
α_3	1.6307	0.0518	-	-	-	-
$\bar{\phi}_1$	2.0848	0.0379	0.2926	1.23E-05	0.2976	1.41E-07
$\bar{\phi}_2$	0.1535	0.0015	0.0860	2.75E-06	0.2100	4.56E-08
$\bar{\phi}_3$	4.7827	0.0301	-	-	-	-
β_1	0.3651	0.0127	0.9081	4.97E-05	0.2221	1.33E-07
β_2	0.0049	0.0001	0.1610	1.55E-05	0.0823	3.82E-08
β_3	1.0930	0.0231	-	-	-	-
r_0	0.0993	0.0010	0.0980	3.87E-06	0.0492	3.73E-08
σ_e^2	1.00E-06	8.92E-09	1.69E-07	4.12E-12	1.85E-07	2.06E-14

the slope of the yield curve, where each volatility shock is transmitted along the yield curve, and the impact dies out as maturities increase. The maturities of any two rates need to differ by around 2.34 years for this volatility factor to halve. The third volatility factor creates a hump in the volatility curve, which occurs at around 3 years to maturity (as implied by the value of κ_3).

The combination of these three volatility factors forms the instantaneous volatility for the forward rates. The overall volatility changes over time as the yield curve moves. The evolutions of the volatility for the spot rate, the 1-year, 6-year and 12-year forward rates are displayed in Figure 2. The short rate and other rates at the very short end of the

14

curve are much more volatile than longer term forward rates, but the volatility itself is less changing over time. Over the sample period, the volatility of the longer term forward rates fluctuates significantly around an increasing trend, and comes closer to the volatility level of shorter term rates at the end of the sample period.

Fig. 2. The U.S. market. The instantaneous volatility of forward rates
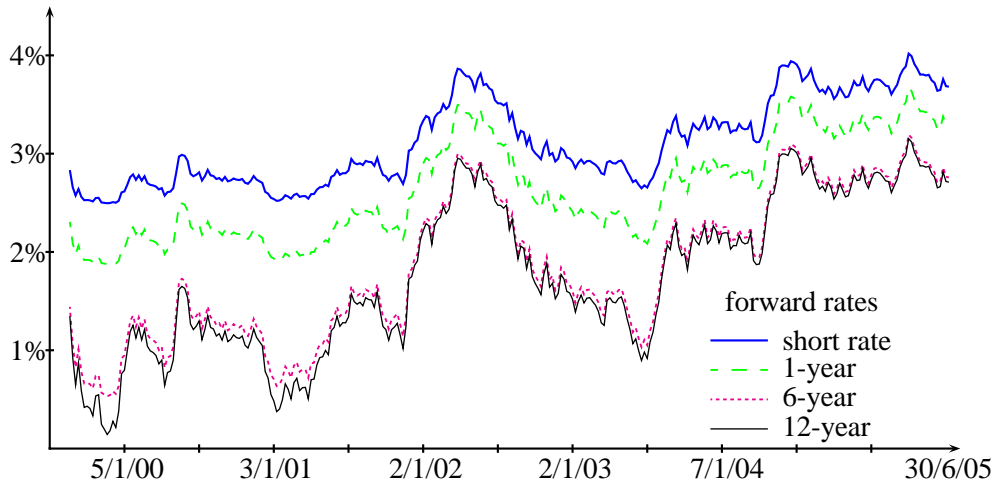
Each volatility factor contributes differently towards the total variation of the forward rates. The third volatility factor contributes very slightly, around 1.2% on average. The contribution of the other two factors varies according to the term of the forward rates and to the passage of time. Figure 3 shows the contribution of the level and slope factors toward the total short rate volatility. During the first one and a half years of the sample period, where interest rates in the U.S. fluctuated around a constant level, the contribution of the level factor towards the total short rate volatility showed no trend with a high level of fluctuation. During 2001 to mid-2002, interest rates dropped down sharply from 7% to 3%, the importance of the level of the rates to the short rate volatility increased significantly. The decrease in rates continued throughout the year until mid-2003, but switched from a high rate regime to a very low rate regime, diminishing the relative importance of short rate level towards total volatility. The relative behaviour between short and long ends of the yield curve became more important. In the final 2 years of the sample period, interest rates picked up again, and the level effect gained back its importance on interest rate volatility.

Given the fact that the third volatility factor plays a very negligible role in determining the overall forward rate variation, one could ask whether it should be included in the model specification. The answer is yes. Even though the 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 risk. The long run unit price $\bar{\phi}$ for the third volatility

15

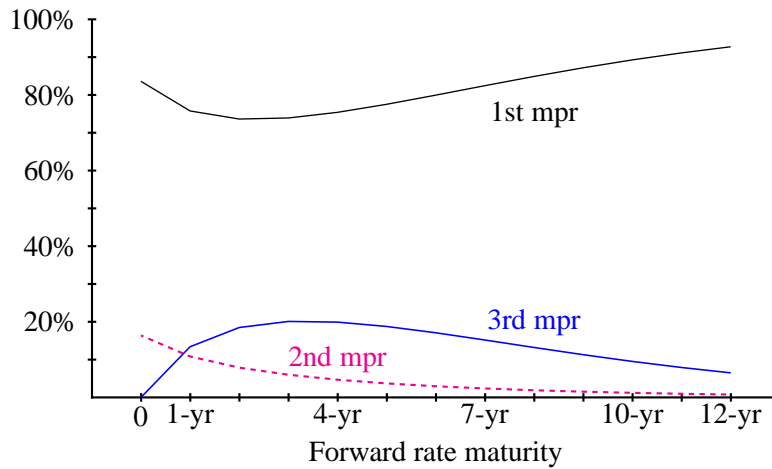Fig. 3. U.S. market. The contribution of the level and slope effects to the total short rate variation
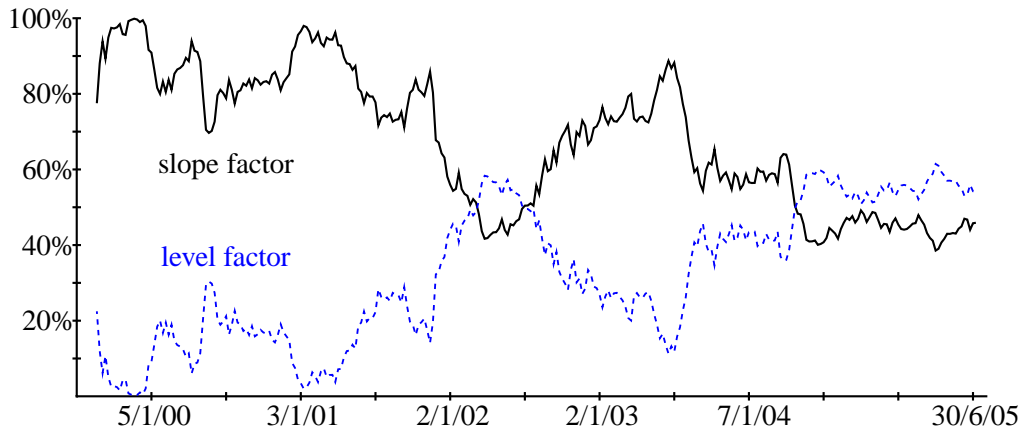
factor is 4.78, compared to the values 2.08 and 0.15 for the other two volatility factors. In addition, the speed of mean reversion of the price for the third volatility risk is only around 72% of that of the level factor risk and around half of that of the slope factor risk. On the way towards the long-run value, it takes 22 weeks for the price of the third volatility risk to halve, whereas it takes only 16 and 12 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 1.09 for β_3, compared to 0.37 for β_1 and 0.005 for β_2.

The unit price of volatility risk, when scaled by the volatility, will determine the overall compensation to 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 to 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 considerable role in the total compensation, especially for yields with medium maturities. The market price of the first (the level) volatility factor risk is the dominant one overall, whereas the market price of risk for the slope volatility factor is minimal for the longer end of the yield curve.

The Australian market and the U.K. market, on the other hand, are found to have only two main factors that influence the volatility of interest rates, namely the interest rate level factor and the slope factor. The influence of the level factor towards the total volatility of interest rates increases with the tenor of the rates, as shown in Figure 5. However, in Australia, the level of interest rates does not affect the volatility much at the short end of the curve. The significance of this level effect increases quickly over the medium maturity segment and becomes the dominating factor at the longer end of

16

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

the yield curve.

Fig. 5. Australian and U.K. markets. The contribution of the level factor to the total volatility function
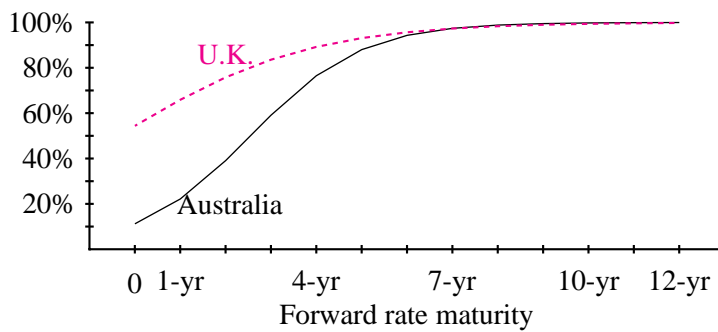
Unlike the U.S. market, the unit price of risk in Australian and U.K. markets has very high speed of mean reversion, reflected by the value of α. In order to halve its value, it takes only half a day for the price of risk in the U.K. market, a day for the price of the slope risk in the Australian market and more than a third of a day for the price of the level risk in the Australian market. Given that the volatility of the prices of risk is not high, the risk compensation per unit of risk is quite stable over time in these two markets. The long run price for the level risk is higher than that for the slope risk.

Given the dominating effect of the level factor towards the total volatility and the higher unit price of risk, it is clear that the total compensation for bearing this risk has a major impact on the drift value of the interest rate evolution. At the long end of the yield curve the impact of volatility on the drift is totally due to this level effect. The slope effect has a considerable effect only at the short end of the curve.

Fig. 6. Australian and U.K. markets. The contribution of each market price of risk to the total risk compensation
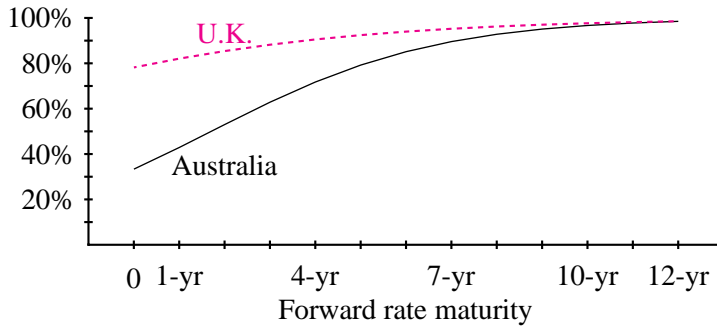

Table 4
Forecast errors (in basis points) from the preferred models

	U.S. market	U.K. market	Australian market
Mean error	7.26	2.90	-1.34
Standard deviation of error	12.68	7.69	7.87
Mean absolute error	12.47	6.43	6.35
Stdev (absolute error)	10.09	5.09	4.90

4.2.2   Forecasting

The selected models are used to forecast future movement of the yield curve. One-step ahead prediction is performed on 30 weeks ahead and the results of the prediction are reported in Table 4. Prediction performance for the U.K. and the Australian markets is excellent with mean absolute errors smaller than 6.5 basis points. Prediction performance for the U.S. market is reasonably good at mean absolute errors of 12.5 basis points. Compared to the Australian and the U.K. markets, yield curve movements in the U.S. market are influenced by one more factor, namely the curvature of the curve, and therefore are harder to predict. However, all models give satisfactory results.

5   Conclusion

The HJM framework provides a very flexible tool for interest rate modelling. Even though theoretical research has advanced quickly in this area, 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

18

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. 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 differently 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 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. In the U.S. market, the slope factor is much more important than the level factor for nearly all of the sample period.

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 somewhat homogeneous across maturities in the U.S. and U.K markets. In the Australian market, the slope factor contributes more at the very short end of the curve, but this contribution diminishes quickly along the 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,

examples include the particle filter in Creal (2008), the HodrickPrescott filter in Maravall and del Río (2007) and the median hybrid filter in Fried et al. (2006). 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.

Acknowledgments

A Markovianization of the interest rate dynamics

Assuming that the forward rate $r(t, x)$ defined in (1) has a volatility function $\boldsymbol{\sigma}(t, x)$ that satisfies Assumption 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.

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

where

$$
\bar{\sigma}_{il}(x) = \int_0^x \sigma_{il}(s) \, ds,
$$

$$
\varphi_{ll^*}^i(t) = \int_0^t c_{il}(s) c_{il^*}(s) \, ds,
$$

$$
\psi_l^i(t) = \int_0^t c_{il}(s) \, d\widetilde{W}_i(s) - \sum_{l^*=1}^{d_i} \int_0^t c_{il}(s) c_{il^*}(s) \bar{\sigma}_{il^*}(s) \, ds,
$$

$$
\epsilon_{ll^*} = \begin{cases} 1, & \text{if } l \neq l^*, \\ 0, & \text{if } l = l^*. \end{cases}
$$

and $\widetilde{W}_i, (i = 1, \dots, I)$ 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 2 *Suppose that the conditions of Assumption 1 are satisfied. If there exist $\tau_1, \tau_2, \ldots, \tau_N \in \mathbb{R}_+$ such that the matrix*

$$\boldsymbol{A}(t, \tau_1, \ldots, \tau_N) = \begin{bmatrix} a_1(t, \tau_1) & a_2(t, \tau_1) & \cdots & a_N(t, \tau_1) \\ a_1(t, \tau_2) & a_2(t, \tau_2) & \cdots & a_N(t, \tau_2) \\ \hdotsfor{4} \\ a_1(t, \tau_N) & a_2(t, \tau_N) & \cdots & a_N(t, \tau_N) \end{bmatrix} \tag{A.3}$$

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

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

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

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

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 differ-

ential equation

$$dr(t, x) = \left[\frac{\partial a_0(t, x)}{\partial x} - \frac{\partial \boldsymbol{a}(t, x)'}{\partial x} \boldsymbol{A}(t, \tau_1, \ldots, \tau_N)^{-1} \boldsymbol{a}_0(t, \tau_1, \ldots, \tau_N) \right.$$
$$\left. + \frac{\partial \boldsymbol{a}(t, x)'}{\partial x} \boldsymbol{A}(t, \tau_1, \ldots, \tau_N)^{-1} \boldsymbol{r}(t, \tau_1, \ldots, \tau_N) + \boldsymbol{\sigma}(t, t + x)' \bar{\boldsymbol{\sigma}}(t, t + x) \right] dt$$
$$+ \boldsymbol{\sigma}(t, t + x)' d\widetilde{\boldsymbol{W}}(t).$$

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

$$dr(t, x) = \left[\frac{\partial a_0(t, x)}{\partial x} - \frac{\partial \boldsymbol{a}(t, x)'}{\partial x} \boldsymbol{A}(t, \tau_1, \ldots, \tau_N)^{-1} \boldsymbol{a}_0(t, \tau_1, \ldots, \tau_N) \right.$$
$$+ \frac{\partial \boldsymbol{a}(t, x)'}{\partial x} \boldsymbol{A}(t, \tau_1, \ldots, \tau_N)^{-1} \boldsymbol{r}(t, \tau_1, \ldots, \tau_N) + \boldsymbol{\sigma}(t, t + x)' \bar{\boldsymbol{\sigma}}(t, t + x)$$
$$\left. - \boldsymbol{\phi}' \boldsymbol{\sigma}(t, t + x) \right] dt + \boldsymbol{\sigma}(t, t + x)' d\boldsymbol{W}(t),$$

which is (4) in the main text. Thus in terms of the notation in (4)

$$p_0(t, x, \tau_1, \ldots, \tau_N) = \frac{\partial a_0(t, x)}{\partial x} - \frac{\partial \boldsymbol{a}(t, x)'}{\partial x} \boldsymbol{A}(t, \tau_1, \ldots, \tau_N)^{-1} \boldsymbol{a}_0(t, \tau_1, \ldots, \tau_N)$$
$$+ \boldsymbol{\sigma}(t, t + x)' \bar{\boldsymbol{\sigma}}(t, t + x), \text{ and}$$
$$\boldsymbol{p}'(t, x, \tau_1, \ldots, \tau_N) = \frac{\partial \boldsymbol{a}(t, x)'}{\partial x} \boldsymbol{A}(t, \tau_1, \ldots, \tau_N)^{-1}.$$

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

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

where

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

This affine yield expression is equation (6) in the main text, once we identify

$$q_0(t, x, \tau_1, \ldots, \tau_N) = b_0(t, x) - \boldsymbol{b}(t, x)' \boldsymbol{A}(t, \tau_1, \ldots, \tau_N)^{-1} \boldsymbol{a}_0(t, \tau_1, \ldots, \tau_N),$$
$$\boldsymbol{q}'(t, x, \tau_1, \ldots, \tau_N) = \boldsymbol{b}(t, x)' \boldsymbol{A}(t, \tau_1, \ldots, \tau_N)^{-1}.$$

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

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

and the discrete observation equation

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

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.

It should be noted that Jimenez and Ozaki (2002) original specification is

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

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

Define $\hat{\mathbf{x}}_{t|\rho} = \mathbb{E}(\mathbf{x}(t)|Z_\rho)$ 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 $\mathbb{E}(\mathbf{W}(t)\mathbf{W}'(t)) = \mathbf{I}$, $\hat{\mathbf{x}}_{t_0|t_0} < \infty$ and $\mathbf{P}_{t_0|t_0} < \infty$.

Theorem 3 *(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*

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

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

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

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

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 $\rho < t$.

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.

Akaike, H. (1974), 'A New Look at the Statistical Model Identification', *I.E.E.E. Transactions on Automatic Control* **AC 19**, 716–723.

Akaike, H. (1976), Canonical Correlation Analysis of Time Series and the Use of an Information Criterion, *in* R. K. Mehra and D. G. Lainotis, eds, 'System Identification: Advances and Case Studies', Academic Press, New York, pp. 52–107.

Bates, D. S. (2006), 'Maximum Likelihood Estimation of Latent Affine Processes', *The Review of Financial Studies* **19**(3), 909–965.

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

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.

Brennan, M. J. and Schwartz, E. S. (1980), 'Analyzing Convertible Bonds', *Journal of Financial and Quantitative Analysis* **31**, 85–107.

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.

Chiarella, C. and Tô, T.-D. (2006), 'The Multifactor Nature of the Volatility of Futures Markets', *Computational Economics* **27**(2-3), 163–183.

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

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985b), 'A Theory of the Term Structure of Interest Rates', *Econometrica* **53**(2), 385–407.

Creal, D. D. (2008), 'Analysis of Filtering and Smoothing Algorithms for Lévy-Driven Stochastic Volatility Models', *Computational Statistics & Data Analysis* **52**(6), 2863 2876.

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.

Faff, R. and Gray, P. (2006), 'On the Estimation and Comparison of Short-Rate Models using the Generalised Method of Moments', *Journal of Banking and Finance* **30**, 3131–3146.

Fried, R., Bernholt, T. and Gather, U. (2006), 'Repeated Median and Hybrid Filters', *Computational Statistics & Data Analysis* **50**(9), 2313–2338.

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.

Golightly, A. and Wilkinson, D. (2008), 'Bayesian Inference for Nonlinear Multivariate Diffusion Models Observed with Error', *Computational Statistics & Data Analysis* **52**(3), 1674–1693.

Hannan, E. J. and Quinn, B. J. (1979), 'The Determination of the Order of an Autoregression', *Journal of the Royal Statistical Society* **B 41**, 190–195.

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., Johannes, M. and Polson, N. (2007), 'MCMC Maximum Likelihood for Latent State Models', *Journal of Econometrics* **137**(2), 615–640.

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.

Jiang, G. J. and Oomen, R. C. A. (2007), 'Estimating Latent Variables and Jump Diffusion Models Using High-Frequency Data', *Journal of Financial Econometrics* **5**(1), 1–30.

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.

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.

Maravall, A. and del Río, A. (2007), 'Temporal Aggregation, Systematic Sampling, and the HodrickPrescott Filter', *Computational Statistics & Data Analysis* **52**(2), 975–998.

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.

Pollock, D. (2003), 'Recursive Estimation in Econometrics', *Computational Statistics & Data Analysis* **44**(1), 37–75.

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.

Schwarz, G. (1978), 'Estimating the Dimension of a Model', *Annals of Statistics* **6**, 461–464.

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

Shoji, I. and Ozaki, T. (1997), 'Comparative Study of Estimation Methods for Continuous Time Stochastic Processes', *Journal of Time Series Analysis* **18**(5), 485–506.

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.

Vasicek, O. (1977), 'An Equilibrium Characterization of the Term Structure', *Journal of Financial Economics* **5**(2), 177–188.

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.