Parametric Estimation of Quadratic Term Structure Models of Interest Rate

Li Chen and H. Vincent Poor *
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544

Abstract

Nonlinear filtering techniques and the quasi maximum likelihood estimator (QMLE) are applied to the problem of estimating the parameters of quadratic models for the term structure of interest rates. It is assumed that zero coupon bond yields data have been contaminated by noise, which allows the application of nonlinear filtering techniques. Without the need of real time computation, we can instead apply the smoothing techniques. The asymptotic properties of the QMLE are also analyzed in two ways: the asymptotical optimality under Kullback-Leibler criterion and its consistent conditions in general. Finally Monte Carlo simulation results are presented to confirm the performance of this strategy.

1 Introduction

Affine term structure models (ATSMs) and quadratic term structure models (QTSMs) are two general classes of short rate models which have been widely

*Li Chen is a graduate student in electrical engineering department of Princeton University. H. Vincent Poor is a professor with electrical engineering department of Princeton University. Date: Oct. 2002 (This draft); Contact: lichen@princeton.edu

used to characterize the dynamics of the short rate. Much empirical work has been done to testify the effectiveness of these two class models. As early as Chan, Karolyi, Longstaff and Sanders (1992 [3]) who applied the Generalized Method of Moments (GMM)(Hansen,1982 [13]) to estimating a range of single-factor models, they used the one-month Treasury rate as a substitute for the short rate. Recently Gallant and Tauchen (1996 [11]) proposed the Efficient Method of Moments (EMM) which was applied in Dai and Singleton (2000 [7]) for the affine term structure models and in Ahn, Dittmar and Gallant (2002 [1]) for the quadratic term structure models. EMM uses the elements of the score vector of the auxiliary model as the moment functions, whose parameters are estimated by quasi-maximum likelihood, then applies these moment functions in the minimum chi-squared criterion. It approximates the true data generating process by a semi-nonparametric model (Gallant and Nychka, 1987 [10]). In Ahn, Dittmar and Gallant (2002 [1]), they justified that QTSMs are superior to the ATSMs in that they are able to provide a better goodness of fit of term structure dynamics than ATSMs. This is because the quadratic relationship between the short rate and state variables can import the nonlinearity into the interest rate dynamics.

However, due to market micro-structure issues, such as bid-ask spreads and asynchronous trading, the observable entities in the market are contaminated by noise. This consideration leads to another general approach to estimation methods for these two classes of short rate models: linear or nonlinear filtering techniques (e.g. the Kalman-Bucy filter). The extended Kalman filter was applied by Chen and Scott (1993 [4])(1995 [5])and by Duan and Simonato (1999 [8]) to exponential-affine models. Lund (1997 [16]) considered a nonlinear relationship between observations and the underlying state variables and proposed the Iterated Extended Kalman Filter (IEKF). Since the algorithm is relatively complicated, Lund implemented his method only for Gaussian diffusion models. Nielsen, Vestergaard and Madsen (2000 [19]) applied the truncated second order filter (TSOF) (Maybeck, 1982 [17]) to multi-factor affine models. This technique estimated the parameters by using the quasi maximum likelihood estimator (QMLE), but in their state-space model the measurement equation is still linear.

Motivated by all these works, in this paper, we will apply a nonlinear filtering technique and QMLE to quadratic models proposed by (Ahn, Dittmar and Gallant,2002 [1]). Two aspects of the asymptotical properties of the QMLE are discussed. In particular, it turns out that the QMLE we apply is asymptotically optimal under the Kullback-Leibler criterion. Moreover, we also derive the general consistent conditions for the QMLE we employ.

The remainder of this article is organized as follows. In Section 2, we provide a general framwork of quadratic term structure models (QTSMs), based on the work of Ahn, Dittmar and Gallant (2002 [1]). Section 3 introduces the state space-time model, nonlinear filtering, smoothing method and the quasi-likelihood estimator. Its asymptotic properties are discussed in Section 4. Section 5 presents results from Monte Carlo studies that verify the analytical results of Section 4. Finally, Section 6 contains our conclusions.

2 The General Quadratic Term Structure Model

Consider a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq \mathcal{T}_0}, \mathbb{P})$, let X_t denote an N-dimensional underlying state vector, which is assumed to satisfy the following SDE under the physical measure $\mathbb{P}$:

$$dX_t = [\mu + \xi X_t]dt + \Sigma dW_t \tag{1}$$

where μ is an N-dimensional vector of constants, ξ and Σ are $N \times N$ matrices, W_t is an N-dimensional vector of standard $(\mathcal{F}_t)$–Brownian motions that are mutually independent and the distribution of the initial state vector X_0 is assumed to be known.

QTSMs assume that, given a state of the form (1), the nominal instantaneous interest rate $R(t)$ is a quadratic function of the state variables:

$$R(t) = \alpha + \beta^{'} X_t + X_t^{'} \Phi X_t, \quad t \geq 0, \tag{2}$$

where α is a constant, β is a constant N-dimensional vector, and Φ is an $N \times N$ positive semi-definite matrix of constants. If we assume that $\alpha - \frac{1}{4}\beta^{'}\Phi^{-1}\beta \geq 0$, then $R(t)$ will be always nonnegative. To price securities based on this model, it is also necessary to specify the state-price density process $M(t)$, which defines the canonical valuation equation:

$$v(t) = E_t^{\mathbb{P}} \left\{ \frac{M(T)}{M(t)} v(T) \right\}, \tag{3}$$

where $v(t)$ is the price of a security at time t, and $E_t^{\mathbb{P}}\{\cdot\}$ the expectation conditioned on the filtration $\mathcal{F}_t$, under the probability measure $\mathbb{P}$. From arbitrage theory, it is assumed that there exists a unique risk neutral measure $\mathbb{Q}$ which is equivalent to $\mathbb{P}$. Under measure $\mathbb{Q}$, $\{\frac{v(t)}{B(t)}\}$ is a martingale, where $B(t)^{-1}$ is a discount process defined by

$$B(t) = \exp\left(\int_0^t R(s)ds\right). \tag{4}$$

In Ahn, Dittmar and Gallant (2002 [1]), the following assumption about the state-price density process is made to ensure that the state variables X_t still follow the Ornstein-Uhlenbeck process under measure $\mathbb{Q}$:

$$\begin{aligned}
\frac{dM(t)}{M(t)} &= -R(t)dt + 1_N^{'} diag[\eta_{0i} + \eta_{1i}^{'} X_t]_N dZ_N(t) && (5)\\
&= -R(t)dt + 1_N^{'}[(\eta_0 + \eta_1^{'} X_t)] \diamond dZ_N(t), && (6)
\end{aligned}$$

where

$$\eta_0 = (\eta_{01}, \eta_{02}, \eta_{03}, ..., \eta_{0N})^{'}, \eta_1 = (\eta_{11}, \eta_{12}, ...\eta_{1N})^{'}, \tag{7}$$

and $\diamond$ represents a Hadamard product. $Z_n(t)$ is an N-dimensional vector of standard Brownian motions which are mutually independent. We let Υ denote the covariance matrix between $dW_n(t)$ and $dZ_n(t)$.

Using the assumptions (5)-(7), the state vector X_t satisfies the following

SDE under the risk neutral measure $\mathbb{Q}$:

$$
\begin{aligned}
dX_t &= [\mu - \delta_0 + (\xi - \delta_1)X_t]dt + \Sigma d\widetilde{W}_t && (8)\\
&= [\widetilde{\mu} + \widetilde{\xi} X_t]dt + \Sigma d\widetilde{W}_t, && (9)
\end{aligned}
$$

where

$$
\delta_0 = -\Sigma \Upsilon \eta_0 \quad \text{and} \quad \delta_1 = -\Sigma \Upsilon \eta_1 \tag{10}
$$

are the market prices of risk, and $\widetilde{W}_t$ are N-dimensional standard Brownian motions under the risk neutral measure $\mathbb{Q}$.

Let $V(t, \tau)$ represent the price of a zero-coupon bond at time t with maturity $t+\tau$ $(\tau \geq 0)$. The following partial differential equation (PDE) is given directly by Feynman-Kac formula:

$$
\left[\frac{1}{2} tr \left(\Sigma \Sigma^{'} \frac{\partial^2 V(t,\tau)}{\partial X(t) \partial X(t)^{'}}\right) + \frac{\partial V(t,\tau)}{\partial X(t)^{'}} \times [\widetilde{\mu} + \widetilde{\xi} X(t)] + \frac{\partial V(t,\tau)}{\partial t}\right] \frac{1}{V(t,\tau)} = R(t), \tag{11}
$$

By using separation of variables, the solution $V(t, \tau)$ can be written as an exponential quadratic function as follows:

$$
V(t,\tau) = \exp[A(\tau) + B(\tau)^{'} X(t) + X(t)^{'} C(\tau) X(t)], \tag{12}
$$

with the initial condition

$$
V(t, 0) = 1, \tag{13}
$$

where $A(\tau), B(\tau)$, and $C(\tau)$ satisfy the following ordinary differential equations

(ODEs):

$$\frac{dC(\tau)}{d\tau} = 2C(\tau)\Sigma\Sigma^{'}C(\tau) + C(\tau)\widetilde{\xi} + \widetilde{\xi}^{'}C(\tau) - \Psi, \tag{14}$$

$$\frac{dB(\tau)}{d\tau} = 2C(\tau)\Sigma\Sigma^{'}B(\tau) + \widetilde{\xi}^{'}B(\tau) + 2C(\tau)\widetilde{\mu} - \beta, \tag{15}$$

and

$$\frac{dA(\tau)}{d\tau} = tr[\Sigma\Sigma^{'}C(\tau)] + \frac{1}{2}B(\tau)^{'}\Sigma\Sigma^{'}B(\tau) + B(\tau)^{'}\widetilde{\mu} - \alpha, \tag{16}$$

with initial conditions:

$$A(0) = 0, \quad B(0) = 0_{N\times 1} \quad \text{and} \quad C(0) = 0_{N\times N}. \tag{17}$$

Since the yield to maturity $y(t,\tau)$ is defined as $-\ln(V(t,\tau))/\tau$, we have

$$y(t,\tau) = -\frac{1}{\tau}[A(\tau) + B(\tau)^{'}X(t) + X(t)^{'}C(\tau)X(t)]. \tag{18}$$

Here we will use the separable QTSM that assumes no interactions in determination of the short rate (i.e., QTSM3 mentioned in Ahn, Dittmar and Gallant (2002 [1]), which means that ξ, δ_1, Ψ and Σ are all diagonal matrices:

$$\Sigma = diag[\sigma_1, \sigma_2, ..., \sigma_n] \tag{19}$$

$$\xi = diag[\xi_1, \xi_2, ..., \xi_n] \tag{20}$$

$$\delta_1 = diag[\delta_{11}, \delta_{12}, ..., \delta_{1n}] \tag{21}$$

and

$$\Psi = diag[\Psi_1, \Psi_2, ..., \Psi_n]. \tag{22}$$

Since identifying without restrictions the parameters from the observation is impossible, we impose the restrictions on the parameters according to the canonical form in Ahn, Dittmar and Gallant (2002[1]):

$$\Psi = I_N \tag{23}$$

$$\text{and} \quad \beta = 0_{N\times 1}, \tag{24}$$

where I_N denotes the $N \times N$ identity matrix and $0_{N\times 1}$ denotes the N-dimensional vector with all zero components.

3 The General Nonlinear Filtering Technique and Quasi Maximum Likelihood Estimator

3.1 Nonlinear Filtering

A general time-homogenous state-space model with the discrete observations $\{Y_{t_i}\}$ at an increasing squence of times $\{t_i\}_{0\leq i\leq n}$ is given by

$$dX(t) = F(X(t),\theta)dt + G(X(t),\theta)dW_t \tag{25}$$

$$Y_{t_i} = H(X(t_i),\theta) + n_i(\theta), \quad \text{for every} \quad t_i \geq 0, \tag{26}$$

where $H : \mathbb{R}^N \times \Theta \mapsto \mathbb{R}^m$, is an m-dimensional nonlinear function of the state vector. θ represents the parameter vector and Θ denotes the parameter space which is assumed to be compact. $\{n_i(\theta)\}$ is an independently and identically distributed (i.i.d.) Gaussian noise sequence with mean zero and covariance matrices $\{Q_i(\theta)\}$. Here we assume that $Q_i(\theta) = \varepsilon^2 I_m$ for all i, and that $t_0 = 0$.

For the specific case of a quadratic model, we have

$$
\begin{aligned}
F(X(t),\theta) &= \mu + \xi X(t), && (27)\\
G(X(t),\theta) &= \Sigma, && (28)\\
\text{and} \quad H^k(X(t),\theta) &= -\frac{1}{\tau_k}[A(\tau_k) + B(\tau_k)^{'} X(t) + X(t)^{'} C(\tau_k) X(t)] && (29)\\
\text{for} \quad & 1 \le k \le m,
\end{aligned}
$$

where $\theta = (\mu, \xi, \delta_0, \delta_1, \Sigma)$ and $H^k(X(t),\theta)$ represents the kth component of the function $H(X(t_i),\theta)$. Here we define $\{Y_{t_i}\}$ as a set of observed yields with different maturities $\{\tau_1, ..., \tau_m\}$ from the market.

3.1.1 Time Propagation

Given the above model, for every i, the conditional density of $X(t)$ given $X(t_{i-1})$ and θ evolves according to the Kolmogorov forward equation:

$$
\begin{aligned}
\frac{\partial p(X(t)|X(t_{i-1},\theta))}{\partial t} &= -\sum_{k=1}^{N} \frac{\partial}{\partial x_t^k}\{p(X(t)|X(t_{i-1},\theta)) F_k(X(t),\theta)\} \\
+\frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} \frac{\partial^2}{\partial x_t^j \partial x_t^k} & \{p(X(t)|X(t_{i-1},\theta))[G(X(t),\theta)G^{'}(X(t),\theta)]_{jk}\},
\end{aligned}
\quad (30)
$$

for $t \in [t_{i-1}, t_i]$ and with the initial condition

$$
p(X|X(t_{i-1}),\theta) = \delta(X - X(t_{i-1})) \quad (31)
$$

where $\delta(\cdot)$ denotes the Dirac delta-function. This equation describes the propagation of the conditional density through the inter-sample interval $[t_{i-1}, t_i]$.

Since the computation of the entire density function is infeasible, we resort to the conditional moment as an estimator of the state vector. In particular, let $\hat{X}_{t|t_n}$ denote the estimator of X_t conditioned on $\{Y_{t_i}\}_{0\leq i\ n}$ and $\hat{\Sigma}_{t|t_n}$ denote the estimator of the covariance matrix of X_t conditioned on $\{Y_{t_i}\}_{0\leq i\leq n}$. Then, the propagation of the conditional mean and covariance over $[t_{i-1}, t_i]$ can be also derived from (30) (Maybeck, 1982 [17]):

$$\frac{d\hat{X}_{t|t_{i-1}}}{dt} = F(\widehat{X(t)}, \theta) \tag{32}$$

and

$$\begin{aligned} \frac{d\hat{\Sigma}_{t|t_{i-1}}}{dt} &= \{F(X\widehat{(t),\theta)}X'(t) - F(\widehat{X(t)}, \theta)\hat{X}'_{t|t_{i-1}}\} \\ &\quad + \{X(t)\widehat{F'(X}(t),\theta) - \hat{X}_{t|t_{i-1}}F'(\widehat{X(t)}, \theta)\} \\ &\quad + G(X(t),\widehat{\theta)G(}X(t),\theta)' \end{aligned} \tag{33}$$

where

$$F(\widehat{X(t)}, \theta) = E\left\{F(X(t), \theta)|\vec{Y}_{t_{i-1}}\right\} \tag{34}$$

and $\vec{Y}_{t_i} = \{Y_{t_k}\}_{0\leq k\leq i}$.

Applying the (32) and (33), we have the following proposition.

Proposition 1 *Given the quadratic model of (8) and the assumptions (19)-*

(22), we have the following time-updates:

$$
\begin{aligned}
\hat{X}_{t_i|t_{i-1}} &= e^{-\xi\Delta_i}\hat{X}_{t_{i-1}|t_{i-1}} + \xi^{-1}(I_N - e^{-\xi\Delta_i})\mu && (35)\\
\textit{and} \quad \hat{\Sigma}_{t_i|t_{i-1}} &= e^{-2\xi\Delta_i}\hat{\Sigma}_{t_{i-1}|t_{i-1}} + \frac{1}{2}\xi^{-1}(I_N - e^{-2\xi\Delta_i})\Sigma\Sigma' && (36)\\
\textit{for} \quad & 0 \le t_i \le t_n
\end{aligned}
$$

where $\Delta_i = |t_i - t_{i-1}|$.

3.1.2 Measurement Updating

As to the measurement update, the conditional mean and covariance are given by the following equations:

$$
\begin{aligned}
\hat{X}_{t_i|t_i} &= \int X p(X|\vec{Y}_{t_i})dX && (37)\\
\text{and} \quad \hat{\Sigma}_{t_i|t_i} &= \int XX' p(X|\vec{Y}_{t_i})dX - \hat{X}_{t_i|t_i}\hat{X}'_{t_i|t_i} && (38)
\end{aligned}
$$

According to Bayes' formula, we have

$$
\begin{aligned}
p(X(t_i)|\vec{Y}_{t_i}) &= \frac{p(Y_{t_i}|X(t_i), \vec{Y}_{t_{i-1}})p(X(t_i)|\vec{Y}_{t_{i-1}})}{p(Y_{t_i}|\vec{Y}_{t_{i-1}})} && (39)\\
&= \frac{p(Y_{t_i}|X(t_i))p(X(t_i)|\vec{Y}_{t_{i-1}})}{p(Y_{t_i}|\vec{Y}_{t_{i-1}})} && (40)
\end{aligned}
$$

and

$$
p(Y_{t_i}|\vec{Y}_{t_{i-1}}) = \int_\Omega p(Y_{t_i}|X(t_i))p(X(t_i)|\vec{Y}_{t_{i-1}})dX(t_i), \qquad (41)
$$

so that

$$\hat{X}_{t_i|t_i} = X(t_i)\widehat{p(\vec{Y}_{t_i}|X(t_i))}/\widehat{p(\vec{Y}_{t_i}|X(t_i))} \tag{42}$$

$$\text{and} \quad \hat{\Sigma}_{t_i|t_i} = XX'\widehat{p(\vec{Y}_{t_i}|X(t_i))}/\widehat{p(\vec{Y}_{t_i}|X(t_i))} - \hat{X}_{t_i|t_i}\hat{X}'_{t_i|t_i} \tag{43}$$

Maybeck (1982 [17]) pointed out that approximating the updating expectation by a series expansion of $p(X|\vec{Y}_{t_i})$ would incur a considerable measurement error. He assumes that the conditional mean and covariance can be expressed as power series of the innovations $\{Y_{t_i} - E[Y_{t_i}|\vec{Y}_{t_{i-1}}]\}$, instead, and uses a linear approximation, since the innovations are relatively small.

Here we restate the final results for the updating step according to Maybeck (1982 [17]):

$$\hat{X}_{t_i|t_i} = \hat{X}_{t_i|t_{i-1}} + K_{t_i}[Y_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta) - \gamma_{t_i|t_{i-1}}(\theta)] \tag{44}$$

$$\text{and} \quad \hat{\Sigma}_{t_i|t_i} = \hat{\Sigma}_{t_i|t_{i-1}} - K_{t_i}h(\hat{X}_{t_i|t_{i-1}}, \theta)\hat{\Sigma}_{t_i|t_{i-1}}, \tag{45}$$

where

$$K_{t_i} = \hat{\Sigma}_{t_i|t_{i-1}}h(\hat{X}_{t_i|t_{i-1}}, \theta)'A_{t_i}^{-1} \tag{46}$$

with

$$
\begin{aligned}
A_{t_i} &= h(\hat{X}_{t_i|t_{i-1}},\theta)\hat{\Sigma}_{t_i|t_{i-1}}h(\hat{X}_{t_i|t_{i-1}},\theta)' \\
&- \gamma_{t_i|t_{i-1}}(\theta)\gamma_{t_i|t_{i-1}}(\theta)' + Q_i(\theta) \qquad (47) \\
h(\hat{X}_{t_i|t_{i-1}},\theta) &= \frac{\partial H(\hat{X}_{t_i|t_{i-1}},\theta)}{\partial X} \qquad (48) \\
\text{and} \quad \gamma^k_{t_i|t_{i-1}}(\theta) &= \frac{1}{2}tr[\frac{\partial^2 H^k(\hat{X}_{t_i|t_{i-1}},\theta)}{\partial X^2}\hat{\Sigma}_{t_i|t_{i-1}}] \qquad (49) \\
\text{for} \qquad & 1 \le k \le m.
\end{aligned}
$$

3.1.3 Nonlinear Smoothing

If there is no requirement for real time estimation, we can resort to a smoother for better performance. Smoothing is an extension of the filtering problem and filtering is the basis of smoothing. Here we will apply the nonlinear smoothing strategy proposed by Leondes, Peller and Stear (1970 [15]). The main results are stated as follows:

$$
\begin{aligned}
\frac{d\hat{X}_{t_i|t_n}}{dt} &= E\{F(X(t_i),\theta)|\vec{Y}_n\} \\
&- E\{G(X(t_i),\theta)G'(X(t_i),\theta)\}E\left\{\frac{1}{q(X(t_i))}\frac{\partial q(X(t_i))}{\partial X}|\vec{Y}_n\right\}, (50)
\end{aligned}
$$

where

$$
q(X(t_i)) = p(X(t_i)|\vec{Y}_i) \quad \text{for} \quad i = 1, 2, ..., n. \qquad (51)
$$

As mentioned in Leondes, Peller and Stear (1970 [15]), we can apply the obtained nonlinear filter results $\{\hat{X}_{t_i|t_i}\}_{0\le i\le n}$ and $\{\hat{\Sigma}_{t_i|t_i}\}_{0\le i\le n}$ to approximate

$\{q(X(t_i))\}_{0\leq i\leq n}$ by Gaussian density functions. Then we can obtain a nonlinear smoother as follows:

$$\begin{aligned}
\frac{d\hat{X}_{t|t_n}}{dt} &= \mu + \xi \hat{X}_{t|t_n} + \Sigma\Sigma' \hat{\Sigma}_{t|t}^{-1}[\hat{X}_{t|t_n} - \hat{X}_{t|t}] &(52)\\
\text{and} \quad \frac{d\hat{\Sigma}_{t|t_n}}{dt} &= (\xi + \Sigma\Sigma'\hat{\Sigma}_{t|t}^{-1})\hat{\Sigma}_{t|t_n} + \hat{\Sigma}_{t|t_n}(\xi + \Sigma\hat{\Sigma}_{t|t}^{-1}) - \Sigma\Sigma'. &(53)
\end{aligned}$$

3.2 Quasi-Maximum Likelihood Estimator(QMLE)

Given θ, using the above filtering technique, we can calculate the conditional moment estimator of the state vector $\{\hat{X}_{t_i|t_i}\}_{0\leq i\leq n}$. In order to estimate the parameter vector θ, we will give a quasi-maximum likelihood estimator. The asymptotic properties of this estimator are analyzed in the following section. Let $\{\hat{n}_i(\theta)\}$ denote the one-step prediction error defined by

$$\hat{n}_i(\theta) = Y_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \tag{54}$$

and let $l_{\hat{n}_i}(\hat{n}_i, \theta)$ denote the true log-likelihood function for $\hat{n}_i(\theta)$. Since it is quite difficult to obtain analytical expression of $l_{\hat{n}_i}(\hat{n}_i, \theta)$, instead we give a quasi log-likelihood function $\hat{l}_{\hat{n}_i}(\hat{n}_i, \theta)$ as follows:

$$\begin{aligned}
\log p(\{\hat{n}_i(\theta)\}) &= \sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(\hat{n}_i, \theta) &(55)\\
&= \sum_{i=1}^{n} -\frac{1}{2}(\log|M_i(\theta)| + (\hat{n}_i(\theta) - \gamma_{t_i|t_{i-1}}(\theta))' M_i(\theta)^{-1} \\
&\quad (\hat{n}_i(\theta) - \gamma_{t_i|t_{i-1}}(\theta)) + m\log(2\pi)) &(56)
\end{aligned}$$

where

$$M_i(\theta) = h(\hat{X}_{t_i|t_{i-1}}, \theta)\hat{\Sigma}_{t_i|t_{i-1}} h(\hat{X}_{t_i|t_{i-1}}, \theta)' + Q_i(\theta). \tag{57}$$

Here we assume $\{\hat{n}_i(\theta)\}$ are mutually independent Gaussian random variable with the mean $\{\gamma_{t_i|t_{i-1}}(\theta)\}$ and covariance $\{M_i(\theta)\}$. The rationale for this assumption will be discussed in the following section.

Since the Jacobian transfer matrix

$$\frac{\partial(\hat{n}_1, \hat{n}_2, ..., \hat{n}_n)}{\partial(Y_{t_1}, Y_{t_2}, ..., Y_{t_n})} = I_n \tag{58}$$

we have

$$\begin{aligned}
\log p(Y_{t_1}, Y_{t_2}, ..., Y_{t_n}; \theta) &= \sum_{i=1}^{n} \log p_i(\vec{Y}_{t_i}, \theta) && (59)\\
&= \sum_{i=1}^{n} l_{\hat{n}_i}(Y_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \theta) && (60)\\
&\cong \sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(Y_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \theta) && (61)
\end{aligned}$$

where

$$p_i(\vec{Y}_{t_i}, \theta) = \begin{cases} p(Y_{t_i}|\vec{Y}_{i-1}, \theta), & \text{for} \quad i > 1 \\ p(Y_{t_1}, \theta), & \text{for} \quad i = 1 \end{cases} \tag{62}$$

After filtering $\{\hat{X}_{t_i|t_{i-1}}\}_{0 \le i \le n}$, by using this quasi log-likelihood function, we can obtain the QMLE from the nonlinear optimization:

$$\hat{\theta}_n = \arg\max_{\theta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(Y_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \theta) \right\} \tag{63}$$

For implementing the QMLE, since $Q_i(\theta) = \varepsilon^2 I_m$, by applying Woodbury's formula, M_i^{-1} can be simplified as

$$M_i^{-1} = \varepsilon^{-2}[I_m - h_i(\varepsilon^2 \hat{\Sigma}_{t_i|t_{i-1}}^{-1} + h_i h_i^{'})^{-1} h_i^{'}]. \tag{64}$$

In the same way, we can simplify the $|M_i|$ as:

$$|M_i| = \varepsilon^{2(m-N)} |\hat{\Sigma}_{t_i|t_{i-1}}| |\varepsilon^2 \hat{\Sigma}_{t_i|t_{i-1}}^{-1} + h(t_i, \hat{X}_{t_i|t_{i-1}}, \theta)^{'} h(t_i, \hat{X}_{t_i|t_{i-1}}, \theta)|. \tag{65}$$

4 Asymptotic Properties of the QMLE

Since we apply the quasi log-likelihood function instead of the true one to implement the QMLE, we need to verify its validity. We now consider this issue.

First we give a lemma, a proof of which is given in the appendix.

Lemma 1 *Given a sequence of random variables $\{Y_i\}_{i\geq 1}$, suppose $\{g_n(\vec{Y}_n, \theta)\}$ and $\{k_n(\vec{Y}_n, \theta)\}$ are two sequences of measurable functions (vec$Y_n = \{Y_i\}_{1\geq i \geq n}$) that satisfy the following conditions*

1)

$$g_n(\vec{Y}_n,\theta) \quad \to \quad s(\theta) \quad i.p. \tag{66}$$

$$and \quad k_n(\vec{Y}_n,\theta) \quad \to \quad s(\theta) \quad i.p., \tag{67}$$

$$as \qquad n \to \infty.$$

2)$\{g_n(\vec{Y}_n,\theta)\}$, $\{k_n(\vec{Y}_n,\theta)\}$ *and* $s(\theta)$ *are identifiable and smooth on* Θ. *(Note: the appropriate definitions of "identifiable" and "smooth" can be found in Peracchi (2000, [20])*

3)The sequences $\{\hat{\theta}_n^g\}$ *and* $\{\hat{\theta}_n^k\}$ *defined as follows*

$$\hat{\theta}_n^g \quad = \quad \arg\max_{\theta\in\Theta}\{g_n(\vec{Y}_n,\theta)\}, n=1,2,... \tag{68}$$

$$and \quad \hat{\theta}_n^k \quad = \quad \arg\max_{\theta\in\Theta}\{k_n(\vec{Y}_n,\theta)\}, n=1,2,..., \tag{69}$$

are convergent.

Then

$$\lim_{n\to\infty}(\hat{\theta}_n^g - \hat{\theta}_n^k) = 0 \quad i.p. \tag{70}$$

4.1 Optimality under the Kullback-Leibler Criterion

From (59), we know that $p(\vec{Y}_n,\theta_0)$ is the true likelihood function of the observation vector $\vec{Y}_n = (Y_{t_1}, Y_{t_2}, ..., Y_{t_n})$. From (62), we can write

$$p(\vec{Y}_n,\theta_0) = \prod_{i=1}^{n} p_i(\vec{Y}_{t_i},\theta_0). \tag{71}$$

A natural way to analyze the consistency problem is to try to evaluate the divergence between the true likelihood functions and the likelihood function used in the QMLE, as $n \to \infty$. Here we will apply the Kullback-Leibler divergence defined for a probability density p_1 with respect to a probability density p_2 as

$$D_{KL}(p_1, p_2) = \int \log \frac{p_1}{p_2} p_1(x) dx. \tag{72}$$

An estimator $\hat{\theta}_n$ is said to satisfy the Kullback-Leibler criterion (Kullback-Leibler,1951 [14]) if

$$\bar{\theta}_n = \arg\min_{\theta} \left\{ \frac{1}{n} D_{KL}^n [p(\vec{Y_n}, \theta_0), \hat{p}(\vec{Y_n}; \theta)] \right\}, \tag{73}$$

where

$$\log \hat{p}(\vec{Y_n}; \theta) = \sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(Y_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \theta). \tag{74}$$

Following (73), we have

$$\begin{aligned}
\bar{\theta}_n &= \arg\min_{\theta}\left\{\frac{1}{n}D_{KL}\left\{p(\vec{Y},\theta_0),\hat{p}(\vec{Y};\theta)\right\}\right\} && (75)\\
&= \arg\min_{\theta}\left\{\frac{1}{n}E\left\{\log\frac{\prod_{i=1}^{n}p_i(\vec{Y}_{t_i},\theta_0)}{\prod_{i=1}^{n}\hat{p}_i(\vec{Y}_{t_i},\theta)}\right\}\right\} && (76)\\
&= \arg\min_{\theta}\left\{\frac{1}{n}E\left\{\sum_{i=1}^{n}\log p_i(\vec{Y}_{t_i},\theta_0)-\frac{1}{n}\sum_{i=1}^{n}\hat{l}_i(\vec{Y}_{t_i},\theta)\right\}\right\} && (77)\\
&= \arg\max_{\theta}\left\{\frac{1}{n}E\left\{\sum_{i=1}^{n}\hat{l}_i(\vec{Y}_{t_i},\theta)\right\}\right\} && (78)\\
&= \arg\max_{\theta}\left\{\frac{1}{n}\sum_{i=1}^{n}\bar{l}_i(\theta)\right\} && (79)\\
\text{where}\quad \bar{l}_i(\theta) &= E\left\{\hat{l}_i(\vec{Y}_{t_i},\theta)\right\}=\int\hat{l}_i(\vec{Y}_{t_i},\theta)p(\vec{Y}_{t_i})d\vec{Y}_{t_i} && (80)\\
\text{with}\quad \log\hat{p}_i(\vec{Y}_{t_i},\theta) &= \hat{l}_i(\vec{Y}_{t_i},\theta)=\hat{l}_{\hat{n}_i}(Y_{t_i}-H(\hat{X}_{t_i|t_{i-1}},\theta),\theta). && (81)
\end{aligned}$$

The third step of the above deduction is true because $E\{\sum_{i=1}^{n}\log p_i(\vec{Y}_{t_i},\theta_0)\}$ is unconditional on θ.

As shown in (Gallant and White,1988 [12]), under regularity conditions, we have the generalized version of the uniform law of large numbers:

$$\frac{1}{n}\sum_{i=1}^{n}\hat{l}_i(\vec{Y}_{t_i},\theta)-\frac{1}{n}\sum_{i=1}^{n}\bar{l}_i(\theta)\to 0\quad\text{a.s.} \tag{82}$$

Thus according to Lemma 1, we have

$$(\hat{\theta}_n \quad - \quad \bar{\theta}_n)\to 0\quad\text{i.p.} \tag{83}$$

This means that the QMLE $\hat{\theta}_n$ will asymptotically minimize the K-L divergence between $p(\vec{Y_n})$ and $\hat{p}(\vec{Y_n})$; i.e., it is asymptotically optimal under the K-L criterion.

4.2 General Consistency and Asymptotical Normality

As mentioned in Bollerslev and Wooldridge (1992 [2]), since the score of the normal log-likelihood has the martingale difference property when the first two conditional moments are correctly specified, the Gaussian distributed QMLE is generally consistent and asymptotically normally distributed. In particular we have the following result from (1992 [2]):

Lemma 2 *Given a quasi log-likelihood function, if the first two conditional moments are correctly specified, under regularity conditions (for the sake of readability, they are listed in the appendix) we have*

$$A_n^{-1} B_n A_n^{-1} \sqrt{n}(\theta_n^* - \theta_0) \to \mathcal{N}(0, I) \tag{84}$$

where

$$\begin{aligned} A_n &= \frac{1}{n} \sum_{i=1}^{n} E\{ \left(\frac{\partial \mu_i(\theta_0)}{\partial \theta} \right)' \Omega_i^{-1}(\theta_0) \frac{\partial \mu_i(\theta_0)}{\partial \theta} \\ &+ \frac{1}{2} \frac{\partial \Omega_i(\theta_0)'}{\partial \theta} [\Omega_i^{-1}(\theta_0) \otimes \Omega_i^{-1}(\theta_0)] \frac{\partial \Omega_i(\theta_0)}{\partial \theta} \} \end{aligned} \tag{85}$$

and

$$B_n = \frac{1}{n} \sum_{i=1}^{n} E \left\{ \frac{\partial \hat{l}_i(\theta_0)'}{\partial \theta} \frac{\partial \hat{l}_i(\theta_0)}{\partial \theta} \right\} \tag{86}$$

where $\mu_i(\theta)$ and $\Omega_i(\theta_0)$ represent the conditional mean and covariance given the observations $\{Y_k\}_{1\leq k\leq i}$, respectively. $\hat{l}_i(\theta)$ is quasi-likelihood function given the observations $\{Y_k\}_{1\leq k\leq i}$ and $\otimes$ represents the Kronecker product.

The validity of the assumed quasi-likelihood function (defined as (56)) is shown by the following deduction. According to (25) and (54), we have

$$\hat{n}_i(\theta) = n_i(\theta) + [H(X(t_i),\theta) - H(\hat{X}_{t_i|t_{i-1}},\theta)]. \tag{87}$$

Because, in our case, H is a quadratic function of the state vector, we can rewrite the above equation as

$$\begin{aligned} \hat{n}_i^k(\theta) = n_i^k(\theta) + h^k(\hat{X}_{t_i|t_{i-1}},\theta)'(X(t_i) - \hat{X}_{t_i|t_{i-1}}) \\ +\frac{1}{2}(X(t_i) - \hat{X}_{t_i|t_{i-1}})'h_2^k(\hat{X}_{t_i|t_{i-1}},\theta)(X(t_i) - \hat{X}_{t_i|t_{i-1}}) \\ \text{for} \quad 1 \leq k \leq m. \end{aligned} \tag{88}$$

where

$$\begin{aligned} h_2^k(\hat{X}_{t_i|t_{i-1}},\theta) \quad &= \quad \frac{\partial^2 H^k(\hat{X}_{t_i|t_{i-1}},\theta)}{\partial X^2} \\ \text{for} \qquad & 1 \leq k \leq m. \end{aligned} \tag{89}$$

Now we can prove that $\{\hat{n}_i(\theta)\}$ are mutually independent random variables that satisfy:

$$\begin{aligned} E\{\hat{n}_i(\theta)\} &= \gamma_{t_i|t_{i-1}}(\theta) \qquad (90) \\ \text{and} \quad Var(\hat{n}_i(\theta)) &= M_i(\theta) + \varphi_i(\theta) \qquad (91) \end{aligned}$$

where

$$\varphi_i(\theta) = E\{(\hat{n}_i(\theta) - n_i(\theta))(\hat{n}_i(\theta) - n_i(\theta))'\} - \{\gamma_{t_i|t_{i-1}}(\theta)\gamma_{t_i|t_{i-1}}(\theta)'\}. \qquad (92)$$

Now we define $\{l'_{\hat{n}_i}(\vec{Y}_i, \theta)\}$ to be a sequence of Gaussian log-likelihood functions with means and variances defined in (90) and (91). Thus, according to Lemma 2, if we apply $\sum_{i=1}^{n} l'_{\hat{n}_i}(\vec{Y}_i, \theta)$ to the QMLE, the obtained estimator $\hat{\theta}_n$ is generally consistent and asymptotical normally distributed. Here we assume that $l'_{\hat{n}_i}(\vec{Y}_i, \theta)$ satisfy the regularity conditions in Lemma 2. We can prove if the dimension of the state vector is Less than 3. But generally it is hard to say.

However, because we are unable to calculate $\{\varphi_n(\theta)\}$, instead as shown in (56), we take $Var(\hat{n}_i(\theta))$ as $M_i(\theta)$ which means that the second moment is misspecified. But we can still achieve the consistency under certain conditions according to Lemma 1.

Proposition 2 *Suppose $l'_{\hat{n}_i}(\vec{Y}_i, \theta)$ and $\hat{l}_{\hat{n}_i}(\vec{Y}_i, \theta)$ satisfy the conditions in Lemma*

1 and for any $\theta \in \Theta$*,*

$$\lim_{n\to\infty} E[||X(t_n) - \hat{X}_{t_n|t_{n-1}}||^3] = 0 \tag{93}$$

$$\textit{and} \quad \lim_{n\to\infty} E[||X(t_n) - \hat{X}_{t_n|t_{n-1}}||^4] = 0, \tag{94}$$

where $||\cdot||$ *denote the sup-norm on* Θ*. Then*

$$\hat{\theta}_n \to \theta_0 \quad i.p.. \tag{95}$$

Since generally we cannot guarantee the conditions (93) and (94) , in the next section, we will use Monte Carlo analysis to confirm its performance.

5 Monte Carlo Studies

First we do the simulation for 2-factor quadratic model, the result is relatively poor. It turns out that only 68% of the movement of the yield curves can be captured by this model.

Therefore we consider QTSM with state dimension equal to 3 ($N = 3$). Thus we have 16 parameters, $\theta = (\alpha, \mu, \xi, \delta_0, \delta_1, \Sigma)$.

We set the true values of the parameters as shown in Table 1 based on the results of Ahn, Dittmar and Gallant (2002 [1]), and then we use Euler's method with step length as 0.0002 and sampling interval as $\Delta = 1/50$ to obtain 1500 observations for 1, 5 and 10 year zero-coupon bond yields for each time series. In this case, we set the observation noise sequence $\{n_i(\theta)\}$ to be a sequence of i.i.d. Gaussian random variables with mean zero and the variance

$$Q_i(\theta) = \varepsilon^2 I_m, \tag{96}$$

$$\text{where} \quad \varepsilon = 0.001. \tag{97}$$

Averaged results for filtering from 30 independent simulations are shown in Table [1]:

Table 1: Parameter Estimation Summary by Nonlinear Filtering

Parameter	Mean from Simulation	Stdev	True Value	$\|t\|$-stat
α	0.01581	0.002153	0.0176	4.553755
μ_1	0.03771	0.028936	0.0436	1.114904
μ_2	0.00571	0.003705	0.0070	1.907050
μ_3	0.02684	0.011793	0.0300	1.467653
ξ_{11}	-1.2957	0.899732	-1.5412	1.494328
ξ_{22}	-0.0027	0.001838	-0.0030	0.893998
ξ_{33}	-2.4966	0.561667	-2.8718	3.659240
δ_{01}	0.08291	0.013232	0.1100	11.213576
δ_{02}	0.15233	0.137974	0.1200	1.283421
δ_{03}	-0.09503	0.00973	-0.1000	2.797720
δ_{11}	-2.67994	1.268479	-3.1000	1.813797
δ_{22}	-0.03742	0.009608	-0.0400	1.470779
δ_{33}	-3.07813	0.939454	-3.5000	2.459596
σ_{11}	0.47951	0.07121	0.5000	1.576020
σ_{22}	0.48491	0.103293	0.5000	0.800164
σ_{33}	0.18996	0.043782	0.2000	1.256026

It is noticed that all the null hypotheses of the parameters can not be rejected at 95% level except for μ, ξ_{33} and δ_{01}. Averaged results for smoother from 30 independent simulations are shown in Table 2:

From the t-test, it can be easily seen that the results by the smoother are better than the results given by the filtering. All the null hypotheses can not be reject at 60% level except for α and ξ_{22}.

We also give the comparison of the simulated and filtered 5,10-year yields(shown

Table 2: Parameter Estimation Summary by Smoothing

Parameter	Mean from Simulation	Stdev	True Value	$\|t\|$-stat
α	0.01701	0.00139	0.0176	2.32283
μ_1	0.04181	0.02116	0.0436	0.46337
μ_2	0.00590	0.00333	0.0070	1.80921
μ_3	0.02898	0.01954	0.0300	0.28591
ξ_{11}	-1.5653	0.77941	-1.5412	0.16936
ξ_{22}	-0.0031	0.00021	-0.0030	2.55754
ξ_{33}	-2.8916	0.91125	-2.8718	0.11901
δ_{01}	0.1089	0.23700	0.1100	0.02542
δ_{02}	0.1183	0.59784	0.1200	0.01557
δ_{03}	-0.09986	0.05003	-0.1000	0.01533
δ_{11}	-3.24243	1.33080	-3.1000	0.58620
δ_{22}	-0.04128	0.22928	-0.0400	0.03058
δ_{33}	-3.4382	1.04870	-3.5000	0.32277
σ_{11}	0.5283	0.23259	0.5000	0.66643
σ_{22}	0.4493	0.27205	0.5000	1.02074
σ_{33}	0.1946	0.38127	0.2000	0.07758

in Fig.1), which confirms the effectiveness of our QMLE estimator. It turns out that 3-factor quadratic model can explain 95% of the movement of the yield curves.

6 Conclusion

Nonlinear Filtering Technique and Quasi Maximum Likelihood estimator are applied to estimate the quadratic models of the term structure of the interest rate. And it turns out that the 3-factor model can capture at least 95% of the movements of the yield curves. In this paper we also analyze the asymptotical properties of the QMLE which, to our knowledge, hasn't been done in this settings. Therefore we confirm the effectiveness of this method from both theoretical and empirical aspects.

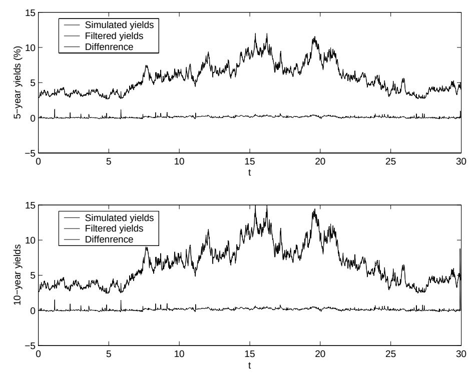

Figure 1: The comparison of simulated and filtered yields

7 Appendix

1) Proof of Lemma 1:

Since we have

$$p\lim_{n\to\infty} g_n(\vec{Y}_n,\theta) = p\lim_{n\to\infty} k_n(\vec{Y}_n,\theta) = s(\theta), \tag{98}$$

where $p\lim$ means the limit in probability, then it is easy to prove that

$$|\max_{\theta} g_n(\vec{Y}_n,\theta) - \max_{\theta} k_n(\vec{Y}_n,\theta)| \to 0 \quad \text{i.p..} \tag{99}$$

Using the definitions (68) and (69), we can rewrite (99) as

$$|g_n(\vec{Y}_n,\hat{\theta}_n^g) - k_n(\vec{Y}_n,\hat{\theta}_n^k)| \to 0 \quad \text{i.p..} \tag{100}$$

Because

$$|s(\hat{\theta}_n^g)-s(\hat{\theta}_n^k)| \le |s(\hat{\theta}_n^g)-g_n(\vec{Y}_n,\hat{\theta}_n^g)|+|g_n(\vec{Y}_n,\hat{\theta}_n^g)-k_n(\vec{Y}_n,\hat{\theta}_n^k)|+|s(\hat{\theta}_n^k)-k_n(\vec{Y}_n,\hat{\theta}_n^k)| \tag{101}$$

and all of the three items on the right-side of the above equation converge to zero in probability, as n goes to infinity, we have

$$|s(\hat{\theta}_n^g) - s(\hat{\theta}_n^k)| \to 0 \quad \text{i.p.} \quad \text{as} \quad n \to \infty. \tag{102}$$

This is equivalent to

$$\hat{\theta}_n^g - \hat{\theta}_n^k \to 0 \quad \text{i.p.} \quad \text{as} \quad n \to \infty \tag{103}$$

if $s(\theta)$ is identifiable on Θ.

2) The Regularity Conditions needed for Lemma 2:

(i) Θ is compact and has nonempty interior. $\theta \in \text{int } \Theta$.

(ii) The conditional mean and covariance are measurable for all $\theta \in \Theta$. They are twice differentiable on int Θ.

(iii) $\{l_i(\theta) - l_i(\theta_0) : i = 1, 2, ...\}$ satisfies the uniform WLLN (UWLLN). (Wooldridge, 1990 [22]) and the sequences of first and second derivatives of $\{l_i(\theta) - l_i(\theta_0) : i = 1, 2, ...\}$ satisfy the UWLLN.

(iv) θ_0 is the identifiably unique maximizer of $\frac{1}{T}\sum_{i=1}^{n} E\{l_i(\theta) - l_i(\theta_0)\}$.

(v) Let $\{a_n(\theta)\}$ be the difference sequence of $\{A_n(\theta)\}$. Then, the sequence $\{a_n(\theta) - a_n(\theta_0)\}$ and the sequence $\{a_n(\theta_0)\}$ satisfy the UWLLN.

3)Proof of Proposition 2:

Since $l'_{\hat{n}_i}(\vec{Y}_i, \theta)$ is a Gaussian log-likelihood function whose first two conditional moments are correctly specified, according to Lemma 1 and Lemma 2, we only need to prove that

$$\frac{1}{n}\sum_{i=1}^{n} l'_{\hat{n}_i}(\vec{Y}_i, \theta) - \frac{1}{n}\sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(\vec{Y}_i, \theta) \to 0 \quad \text{i.p.}, \tag{104}$$

uniformly in Θ. This is equivalent to showing that $l'_{\hat{n}_i}(\theta) - \hat{l}_{\hat{n}_i}(\theta)$ converges uniformly to zero in probability as i goes to infinity. Since the only difference between $l'_{\hat{n}_i}(\theta)$ and $\hat{l}_{\hat{n}_i}(\theta)$ is their covariances, we need to show that the difference of their covariances will vanish uniformly as n goes to infinity. So what we need to prove is that

$$\varphi_i(\theta) \to 0 \quad \text{as} \quad i \to \infty, \quad \text{uniformly on} \quad \Theta \tag{105}$$

where $\varphi_i(\theta)$ is defined in (92).

According to (88) and (92), we have that

$$||\varphi_n(\theta)|| \to 0, \tag{106}$$

where $||\cdot||$ is the sup-norm on Θ, if,

$$\lim_{i\to\infty} E\{||X(t_i) - \hat{X}_{t_i|t_{i-1}}||^3\} = 0 \tag{107}$$

$$\text{and} \quad \lim_{i\to\infty} E\{||X(t_i) - \hat{X}_{t_i|t_{i-1}}||^4\} = 0. \tag{108}$$

This completes the proof.

References

[1] Ahn, D. H., Dittmar, R. F. and Gallant, A. R. (2002), "Quadratic Term Structure Models: Theory and Evidence," *The Review of Financial Studies* **15**, 243-288.

[2] Bollerslev, T. and Wooldridge, J. M. (1992), "Quasi-Maximum Likelihood Estimation of Dynamic Models with Time-Varying Covariances," *Econometric Reviews* **11**, 143-172.

[3] Chan, K. C., Karolyi, G. A., Longstaff, F. A. and Sander, A. B. (1992), "An Empirical Comparison of Alternative Models of the Short-term Interest Rate," *Journal of Finance* **47**, 1209-1227.

[4] Chen, R. R. and Scott, L.(1993), "Maximum likelihood estimation for multifactor equilibrium model of the term structure of interest rates," *Journal of Fixed Income* **3**, 14-31

[5] Chen, R. R. and Scott, L. (1995), "Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model," Technical report, University of Georgia. Manuscript.

[6] Constantinides, G. (1992), "A Theory of the Term Structure of Interest Rates," *Econometrica* **53**, 385-406.

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

[8] Duan, J. C. and Simonato, J. G. (1999), "Estimating Exponential-Affine Term Structure Models by Kalman Filter," *Review of Quantitative Finance and Accounting* **13(2)**, 111-135.

[9] Duncan, D. B. and Horn, S. D. (1972), "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis," *Journal of the American Statistical Association* **67**, 815-821.

[10] Gallant, A. R. and Nychka, D. W.(1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* **55**, 363-390.

[11] Gallant, A. R. and Tauchen, G. E.(1996), "Which Moments to Match?," *Econometric Theory* **12(4)**, 657-681.

[12] Gallant, A. R. and White, H. (1988), *A Unifed Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, Oxford.

[13] Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* **50**, 1029-1054.

[14] Kullback, S. and Leibler, R. A. (1951), "On Information and Sufficiency," *Annals of Mathematical Statistics* **22**, 79-86.

[15] Leondes, C. T., Peller, J. B. and Stear, E. B. (1970), "Nonlinear Smoothing Theory," *IEEE Trans. on System Science and Cybernetics.* **6(1)**, 63-71.

[16] Lund, J. (1997b), "Nonlinear Kalman Filtering Techniques for Term Structure Modes," working paper, Department of Finance, The Aarhus School of Business.

[17] Maybeck, P. S. (1982), *Stochastic Models, Estimation and Control*, Academic Press, London.

[18] Neyman, J. and Pearson, E. S. (1928), "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," *Biometrika* **20A**, 175-240, 263-294.

[19] Nielsen, J. N., Vestergaard, M. and Madsen, H. (2000), "Estimating Multivariate Exponential-Affine Term Structure Models from Coupon Bond Prices Using Nonlinear Filtering," Working paper, The Technical University of Denmark.

[20] Peracchi, F. (2000), *Econometrics*, John Wiley and Sos Ltd.

[21] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* **50**, 1-26.

[22] Wooldridge, J. M. (1990), "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory* **6**, 17-43.