Estimating Multivariate Exponential-Affine Term Structure Models from Coupon Bond Prices using Nonlinear Filtering

Mikkel Baadsgaard*, Jan Nygaard Nielsen†, and Henrik Madsen‡

May 22, 2000

Abstract

An econometric analysis of continuous-time models of the term structure of interest rates is presented. A panel of coupon bond prices with different maturities is used to estimate the embedded parameters of a continuous-discrete state space model of unobserved state variables: the spot interest rate, the central tendency and stochastic volatility. Emphasis is placed on the particular class of exponential-affine term structure models that permits solving the bond pricing PDE in terms of a system of ODEs. It is assumed that coupon bond prices are contaminated by additive white noise, where the stochastic noise term should account for model errors. A nonlinear filtering method is used to compute estimates of the state variables, and the model parameters are estimated by a quasi-maximum likelihood method provided that some assumptions are imposed on the model residuals. Both Monte Carlo simulation results and empirical results based on the Danish bond market are presented.

KEY WORDS: Nonlinear filtering, quasi maximum likelihood estimation, state space models, stochastic differential equations, stochastic volatility, term structure modelling.

*The Danish Ministry of Finance, Christiansborgs Slotsplads 1, DK-1218 København K. E-mail: mba@fm.dk
†E-mail: jnn@imm.dtu.dk
‡E-mail: hm@imm.dtu.dk

Contents

1 Introduction

The term structure of interest rates is, perhaps, the most important entity in finance as it describes the relationship between the yield on a default free discount bond and its maturity. It is a key concept in economic and financial theory, and in the risk-neutral valuation and hedging of interest rate contingent claims. Many models of the term structure are based on the assumption that all information about the economy is contained in a finite-dimensional vector of state variables whose dynamics are governed by stochastic processes[1]. The dynamics may be derived either by using absence of arbitrage arguments, obtained endogenously in a general equilibrium framework or identified from market data using econometric methods. The exact expression for the price of default free discount bond depends on the specification of the stochastic processes for the state variables and the associated market price of risk.

In the pioneering work by (Vasicek, 1977) a univariate diffusion process is proposed for modelling the unobservable instantaneous interest rate (spot rate). Cox, Ingersoll and Ross (1985) proposed the square-root process for the spot rate in a general equilibrium framework in order to introduce heteroscedasticity in the spot rate dynamics. All univariate models imply that the entire term structure is perfectly correlated, i.e. the fact that the entire term structure is inferred from the current short rate, and they do not allow for changes in the slope of the term structure. This is clearly at odds with numerous empirical findings, see the studies in e.g. (Dybvig, 1989) on US data and (Steeley, 1991) on UK data. Furthermore, Dybvig (1989) suggests that the short rate and the volatility of the short rate should be used as state variables, and Litterman and Scheinkman (1991) suggests that the spot rate volatility should be mean reverting. These empirical findings exclude the Ornstein-Uhlenbeck model and the log-normal model considered by (Stein and Stein, 1991; Heston, 1993). In order to exclude negative volatility the spot rate volatility may be modelled by a Cox-Ingersoll-Ross process. The ability of a term structure model to capture the stochastic feature of spot rate volatility is a direct measure of its hedging usefulness. Thus stochastic volatility is introduced in our model as a second state variable. A third state variable is introduced to model the mean level of the spot rate (the central tendency) following the findings in (Balduzzi, Das and Foresi, 1998). This yields the *special interest rate dynamics model* considered in (Chen, 1996) that also fits into the general "affine yield" setting considered by (Duffie and Kan, 1996). These models are also called *exponential-affine term structure models*, because bond yields are affine in the state variables within this model class. The methodology proposed here may, however, be applied to all term structure models for which a closed form expression for the price of a discount bond is available.

Estimating the term structure of interest rates is clearly a difficult problem to which a number of solutions have been proposed in the literature. Firstly, the Generalized Method of Moments (Hansen, 1982) is applied for estimating the parameters of a univariate model in (Chan, Karolyi, Longstaff and Sanders, 1992) using the one month Treasury bill as a proxy for the spot rate; Abken (1993) applies it to forward rates and the Efficient Method of Moments (Gallant and Tauchen, 1996) is applied in (Buraschi, 1996; Dai and Singleton, 1997). Pearson and Sun (1994) considers a two-dimensional CIR-model with explicit expressions for the bond prices and uses the probability distribution of the state variables in the likelihood function. Unfortunately, this method cannot be used when the number of simultaneously observed time series of bond prices exceeds the number of state variables.

It is proposed to use the yield of discount bonds with different maturities as the observable entities in (Chen and Scott, 1993; Daves and Ehrhardt, 1993; Pearson and Sun, 1994; Duffie and Singleton, 1997) by "inversion of the yield curve", i.e. it is assumed that m maturities are observed without observation error and the associated bond pricing equation is inverted such that the yields are used as "instruments" for the state variables. Thus it is necessary to convert bond prices to yields as described in e.g. (Anderson, Breedon, Deacon, Derry and Murphy, 1996). In a series of recent papers, it is assumed that the yields (or

[1] A notable exception is the framework based on forward rates proposed in (Heath, Jarrow and Morton, 1992).

other observable entities) are contaminated by observation noise due to asynchronous trading, rounding off prices, bid-ask spreads, temporary deviations that are not arbitraged away and other market imperfections. This makes it convenient to cast the term structure model in state space form augmented by an observation equation that relates the observation to the underlying state variables. This leads to the application of Kalman filtering techniques, see (Jazwinski, 1970; Maybeck, 1982) for an introduction to such techniques. If the data consists of zero-coupon yields and the term structure model is Gaussian, the linear Kalman filter in combination with a maximum likelihood method may be applied to estimate the state variables and the model parameters. Pennacchi (1991) was the first to use this approach in financial econometrics. Exponential-affine models, in particular multifactor Gaussian and CIR models, are considered in (Chen and Scott, 1993; Chen and Scott, 1995; Claessens and Pennacchi, 1996; Lund, 1997*a*; Babbs and Nowman, 1999; Duan and Simonato, 1999). The Extended Kalman filter is applied by (Cumby and Evans, 1995; Claessens and Pennacchi, 1996), who considers defaultable bonds. The two approaches only differ in the update step. Frühwirth-Schnatter (1994) approximates the true update density by a Gaussian density with the mean and variance of the exact update density using numerical integration. However, the dimension of the integral is the same as the number of states implying that this approach is computationally demanding. Lund (1997*a*) considers a nonlinear observation equation and applies the Iterated Extended Kalman Filter (IEKF), but utilizes a Gaussian model and observed yields in the empirical part of the paper. However, it is argued in (Lund, 1997*a*) that bond yields contain less information than original bond prices so the application of coupon-bearing bonds potentially permits more powerful tests, because the prices of long-term bonds are relatively more sensitive to the model parameters. Using bond prices as observations lead to a nonlinear relation between the observations and the state variables[2], see e.g. (Nielsen, 1996).

The econometric method proposed in this paper may be applied to estimate parameters in multivariate, nonlinear state space models from observed bond prices and thus constitutes a considerable extension of the filtering techniques reported in the literature so far. The work reported here represents a generalization of the second order filter applied in (Nielsen, Vestergaard and Madsen, 2000) to stochastic volatility models in the sense that the proposed methodology allows for a nonlinear observation equation.

Following the work reported in (Chen and Scott, 1993; Chen and Scott, 1995; Jegadeesh and Pennacchi, 1996; Duffie and Singleton, 1997; Lund, 1997*a*; Lund, 1997*b*; Honoré, 1998; Duan and Simonato, 1999), a "panel data" approach is taken, i.e. the time-series information in the data and the cross-sectional information obtained by simultaneously observing prices of coupon-bearing bonds with different maturities are used to fully exploit the information in the data set for reasons of efficiency[3].

The rest of the paper is organized as follows. In Section 2 the term structure modelling framework is presented. In Section 3 the non-linear filtering method is presented, and Section 4 describes how the model parameters are estimated by a quasi maximum likelihood method. In Section 5 the properties of the estimates are examined by a Monte Carlo study, and some results from the Danish bond market are presented in Section 6. Section 7 concludes.

2 Term structure models

Let the stochastic process $\{\mathbf{X}\}$ describing the state of the economy be defined on the state space $\mathcal{S}$, which will, in general, be the d-dimensional Euclidean space $\mathcal{R}^d$ or a subset thereof. Assume that $\{\mathbf{X}\}$ solves the Itô Stochastic Differential Equation (SDE)

$$d\mathbf{X}_t = \mathbf{f}(t, \mathbf{X}_t; \boldsymbol{\psi})dt + \mathbf{g}(t, \mathbf{X}_t; \boldsymbol{\psi})d\mathbf{W}_t; \qquad \mathbf{X}_0 = \mathbf{X}_{t_0}, \tag{1}$$

[2] For exponential-affine models the relationship is obviously affine.

[3] Honoré (1998) imposes some linear restrictions on the observation errors to avoid using a filtering approach.

or, written componentwise as,

$$dX_t^i = f^i(t, \mathbf{X}_t; \boldsymbol{\psi})dt + \sum_{j=1}^{d} \sigma^{ij}(t, \mathbf{X}_t; \boldsymbol{\psi})dW_t^j; \qquad i = 1, \ldots, d, \tag{2}$$

where $\mathbf{f}\colon [t_0, T] \times \mathcal{R}^d \times \mathcal{R}^p \mapsto \mathcal{R}^d$ and $\mathbf{g}\colon [t_0, T] \times \mathcal{R}^d \times \mathcal{R}^p \mapsto \mathcal{R}^{d\times d}$ are assumed to satisfy sufficient regularity (Lipschitz and bounded growth) conditions to ensure the existence and uniqueness of solutions to (1), see eg. (Øksendal, 1995; Karatzas and Shreve, 1996); $\mathbf{X}_{t_0}$ is a stochastic initial condition satisfying $E[||\mathbf{X}_{t_0}||^2] < \infty$; $\boldsymbol{\psi}$ is a p-dimensional parameter vector belonging to $\boldsymbol{\psi}$, a subset of $\mathcal{R}^p$; and $\{\mathbf{W}_t\}$ is a d-dimensional Wiener process defined on the usual probability space $(\Omega, \mathcal{F}, P)$, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra, and P is the objective probability measure[4].

Remark 2.1 *The spot rate is typically expressed as a function $r_t(\mathbf{X}_t)$ of the state variables. In this paper it is assumed that the instantaneous spot rate is the first element of the state vector, $r_t = X_t^1$.*

The unique arbitrage-free price at time t, $P(t, T, \mathbf{X})$, of a zero-coupon bond maturing at time T ($t \le T$) can be obtained as the discounted expected value of the cash flow. The conditional expectation should be taken with respect to the equivalent martingale measure Q defined by the Radon-Nikodym derivative

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = \exp\left\{ -\int_{t_0}^{t} \lambda^T(u, \mathbf{X}_u; \boldsymbol{\psi})d\mathbf{W}_u - \frac{1}{2}\int_{t_0}^{t} ||\lambda(u, \mathbf{X}_u; \boldsymbol{\psi})||^2 du \right\}, \tag{3}$$

which is characterized by a vector $\lambda(t, \mathbf{X}_t; \boldsymbol{\psi})$ known as the *market price of risk*, where $\dim(\lambda(t, \mathbf{X}_t; \boldsymbol{\psi})) = \dim(\mathbf{X}_t)$. The i'th component of $\lambda(t, \mathbf{X}_t; \boldsymbol{\psi})$ measures the extent to which risk taken in the i'th factor is compensated through a higher expected return. In other words the price of a bond that pays out one unit-of-account at maturity T is given by

$$P(t, T, \mathbf{X}_t; \boldsymbol{\psi}) = E^Q\left[e^{-\int_t^T r_u(\mathbf{X}_u)du} | \mathcal{F}_t\right] = E^Q\left[e^{-\int_t^T r_u du} | \mathcal{F}_t\right]. \tag{4}$$

According to Girsanov's Theorem the stochastic process $\{\mathbf{X}\}$ satisfies the SDE

$$d\mathbf{X}_t = [\mathbf{f}(t, \mathbf{X}_t; \boldsymbol{\psi}) + \lambda(t, \mathbf{X}_t; \boldsymbol{\psi})\mathbf{g}(t, \mathbf{X}_t; \boldsymbol{\psi})]dt + \mathbf{g}(t, \mathbf{X}_t; \boldsymbol{\psi})d\mathbf{W}_t^Q \tag{5}$$

where $\mathbf{W}_t^Q$ is a Wiener process under the martingale measure Q.

Under mild regularity conditions the bond prices solves a partial differential equation (PDE), see e.g. (Duffie, 1996). The corresponding PDE is given by

$$\mathcal{D}P(t, T, \mathbf{X}_t; \boldsymbol{\psi}) - r_t P(t, T, \mathbf{X}_t; \boldsymbol{\psi}) = 0; \quad (t, \mathbf{X}_t) \in [0, T) \times \mathcal{S} \tag{6}$$

with the boundary condition

$$P(T, T, \mathbf{X}_T) = 1, \tag{7}$$

where

$$\begin{aligned} \mathcal{D}P(t, T, \mathbf{X}_t; \boldsymbol{\psi}) \quad &= \quad \frac{\partial P(t, T, \mathbf{X}_t; \boldsymbol{\psi})}{\partial t} + \frac{\partial P(t, T, \mathbf{X})}{\partial \mathbf{X}_t^T}(\mathbf{f}(t, \mathbf{X}_t; \boldsymbol{\psi}) + \lambda(t, \mathbf{X}_t; \boldsymbol{\psi})\mathbf{g}(t, \mathbf{X}_t; \boldsymbol{\psi})) \\ &\quad + \frac{1}{2}\mathrm{tr}\left[\mathbf{g}(t, \mathbf{X}_t; \boldsymbol{\psi})\mathbf{g}^T(t, \mathbf{X}_t; \boldsymbol{\psi})\frac{\partial P(t, T, \mathbf{X})}{\partial \mathbf{X}_t \partial \mathbf{X}_t^T}\right] \end{aligned} \tag{8}$$

[4]See (Protter, 1990) for definitions involving the theory of stochastic processes.

According to the Feynman-Kac representation theorems the bond price obtained by computing the expected value (4) or solving the PDE (8) is the same. It is only possible to obtain explicit solutions for a few particular models. However, some results may be obtained for the special class of *exponential-affine term structure models*.

2.1 Exponential-affine term structure models

Duffie and Kan (1996) provides the most general definition of the class of exponential-affine term structure models, i.e.

$$d\mathbf{X}_t = (\mathbf{a}\mathbf{X}_t + \mathbf{b})dt + \Sigma \begin{pmatrix} \sqrt{v_1(\mathbf{X}_t)} & 0 & \dots & 0 \\ 0 & \sqrt{v_2(\mathbf{X}_t)} & \dots & 0 \\ & & \ddots & \\ 0 & \dots & 0 & \sqrt{v_d(\mathbf{X}_t)} \end{pmatrix} d\mathbf{W}_t, \tag{9}$$

where $\mathbf{a} \in \mathcal{R}^{d \times d}$, $b \in \mathcal{R}^d$, $\Sigma \in \mathcal{R}^{d \times d}$, and

$$v_i(\mathbf{x}) = \alpha_i + \boldsymbol{\beta}_i^T \mathbf{x} \tag{10}$$

where, for each i, α_i is a scalar and $\boldsymbol{\beta}_i \in \mathcal{R}^d$. See (Duffie and Kan, 1996) for the coefficient restrictions that ensures the existence of unique solutions to (9).

Remark 2.2 *It is seen that in order to obtain an exponential-affine model the drift* $\mathbf{f}$ *and squared diffusion* $\mathbf{g}\mathbf{g}^T$ *should be affine in the state vector and time-homogenous. This also implies that* $P(t, T, \mathbf{X}_t; \boldsymbol{\psi})$ *need only be parametrized in the time-to-maturity* $\tau = T - t$.

Remark 2.3 *As argued in (Campbell, Lo and MacKinlay, 1997, Sec. 11.1.4) exponential-affine term structure models limit the way in which interest rate volatility can change with the level of interest rates.*

The term structure model considered in the empirical part of the paper is the three-factor *special interest rate dynamics* model proposed by (Chen, 1996), which fits into the framework (9)–(10) above[5]. The first factor is the instantaneous spot rate which is described by the following SDE

$$dr_t = \kappa_1(\theta_t - r_t)dt + \sqrt{v_t}dW_t^1 \tag{11}$$

where κ_1 is a constant parameter, θ_t is the stochastic central tendency towards which the spot rate mean reverts and $\sqrt{v_t}$ is the stochastic volatility of the spot rate.

The volatility of the spot rate is assumed to evolve according to a square-root process, cf. (Dybvig, 1989), i.e.

$$dv_t = \kappa_3(\bar{v} - v_t)dt + \eta\sqrt{v_t}dW_t^3 \tag{12}$$

where κ_3 $\bar{v}$ and η are parameters.

Following the findings in (Balduzzi et al., 1998), a third state variable is introduced to model the dynamics of the central tendency and it is described by a square-root process

$$d\theta_t = \kappa_2(\bar{\theta} - \theta_t)dt + \xi\sqrt{\theta_t}dW_t^2 \tag{13}$$

[5] Chen (1996) also considers a more general model that does not fit into the framework (9)–(10).

where κ_2, $\bar{\theta}$ and ξ are parameters[6].

Remark 2.4 *The model (11)–(13) corresponds to (9)–(10) with*

$$\mathbf{X}_t = \begin{pmatrix} r_t \\ \theta_t \\ v_t \end{pmatrix}; \mathbf{a} = \begin{pmatrix} -\kappa_1 & \kappa_1 & 0 \\ 0 & -\kappa_2 & 0 \\ 0 & 0 & -\kappa_3 \end{pmatrix}; \mathbf{b} = \begin{pmatrix} 0 \\ \kappa_2\bar{\theta} \\ \kappa_3\bar{v} \end{pmatrix}, \mathbf{\Sigma} = \mathbf{I},$$

$$\alpha_1 = \alpha_2 = \alpha_3 = 0,\ \boldsymbol{\beta}_1^T = \mathbf{0};\ \boldsymbol{\beta}_2^T = \begin{pmatrix} 0 & \xi^2 & 0 \end{pmatrix},\ \textit{and}\ \boldsymbol{\beta}_3^T = \begin{pmatrix} 0 & 0 & \eta^2 \end{pmatrix}.$$

For exponential-affine term structure models with three state variables the price of a zero-coupon bond is given by

$$P(\tau;\boldsymbol{\psi}) = P(t,T,\mathbf{X}_t;\boldsymbol{\psi}) = A(\tau)e^{-B(\tau)r_t - C(\tau)\theta_t - D(\tau)v_t}. \tag{14}$$

The functions $A(\tau)$, $B(\tau)$, $C(\tau)$ and $D(\tau)$ are determined by requiring that (14) be the solution to (6). When it is assumed that the Wiener processes are mutually independent, and the market price of risks are constant, the PDE for the three-factor model presented above becomes

$$\begin{aligned} &\frac{1}{2}v_t\frac{\partial^2 P(t,T,\mathbf{X}_t;\boldsymbol{\psi})}{\partial r_t^2} + \frac{1}{2}\eta^2 v_t\frac{\partial^2 P(t,T,\mathbf{X}_t;\boldsymbol{\psi})}{\partial v_t^2} + \frac{1}{2}\xi^2\theta_t\frac{\partial^2 P(t,T,\mathbf{X}_t;\boldsymbol{\psi})}{\partial \theta_t^2} \\ &+\left[\kappa_1(\theta_t - r_t) + \lambda_r v_t\right]\frac{\partial P(t,T,\mathbf{X}_t;\boldsymbol{\psi})}{\partial r_t} + \left[\kappa_2(\bar{\theta} - \theta_t) + \lambda_\theta\xi\theta_t\right]\frac{\partial P(t,T,\mathbf{X}_t;\boldsymbol{\psi})}{\partial \theta_t} \\ &+\left[\kappa_3(\bar{v} - v_t) + \lambda_v\eta v_t\right]\frac{\partial P(t,T,\mathbf{X}_t;\boldsymbol{\psi})}{\partial v_t} + \frac{\partial P(t,T,\mathbf{X}_t;\boldsymbol{\psi})}{\partial t} = r_t P(t,T,\mathbf{X}_t;\boldsymbol{\psi}), \end{aligned} \tag{15}$$

where λ_r, λ_θ and λ_v are the market prices of risk for the factors r_t, θ_t and v_t. By substitution of (14) into the PDE (15) the following system of Ordinary Differential Equations (ODEs) is obtained

$$1 = \kappa_1 B(\tau) + B'(\tau) \tag{16}$$

$$0 = -\kappa_1 B(\tau) + \frac{1}{2}\xi^2 C^2(\tau) + (\kappa_2 - \lambda_\theta\xi)C(\tau) + C'(\tau) \tag{17}$$

$$0 = \frac{1}{2}B^2(\tau) + \frac{1}{2}\eta^2 D^2(\tau) - \lambda_r B(\tau) + (\kappa_3 - \lambda_v\eta)D(\tau) + D'(\tau) \tag{18}$$

$$0 = \kappa_2\bar{\theta}C(\tau) + \kappa_3\bar{v}D(\tau) + \frac{A'(\tau)}{A(\tau)} \tag{19}$$

with the initial conditions $A(0) = 1$ and $B(0) = C(0) = D(0) = 0$, where, say, $B'(\tau) = \frac{\partial B(\tau)}{\partial \tau}$.

The solution to (15) is given in (Chen, 1996) in terms of the Bessel function (of the first and second kind), the Kummel function and the confluent hypergeometric function. However, it is computationally more convenient to solve (16)-(19) numerically using e.g. a Runge-Kutta method.

Remark 2.5 *Only (16) can be solved in closed form, i.e. without the need for special functions.*

[6] Although W_t^1, W_t^2 and W_t^3 are assumed independent, the spot rate r_t, its mean θ_t and its volatility v_t are correlated through (11).

The yield $R(\tau;\boldsymbol{\psi})$ is given by

$$R(\tau;\boldsymbol{\psi}) \;=\; -\frac{1}{\tau}\ln P(\tau;\boldsymbol{\psi}) \tag{20}$$

$$\;=\; -\frac{1}{\tau}\left[\ln A(\tau) - B(\tau)r - C(\tau)\theta - D(\tau)v\right] \tag{21}$$

provided that $P(\tau;\boldsymbol{\psi})$ is given by (14). Thus the functions $B(\tau), C(\tau)$, and $D(\tau)$ determine the sensitivity of a bond's yield to the factors r, θ and v, and can be called *factor loadings* for r, θ and v, respectively.

Remark 2.6 *Chen (1996) provides a number of illustrations and interpretations of the factor loadings, and concludes that the factor loadings implied by the model are similar in nature to those empirically identified by (Litterman and Scheinkman, 1991).*

3 Nonlinear filtering with discrete time observations

In this section the continuous-discrete nonlinear filtering problem will be described for a general stochastic state space model and the approximations made to obtain the second order filter will be discussed. The presentation follows (Maybeck, 1982) and (Nielsen et al., 2000).

Assume that observations are made available at discrete time instants $t_1 < \ldots < t_i < \ldots t_N$, where N denotes the number of observations. The relation between the state variables $\{\mathbf{X}\}$ and the observations is given by the observation equation:

$$\mathbf{Y}_{t_i} = \mathbf{h}(t_i, \mathbf{X}_{t_i};\boldsymbol{\psi}) + \mathbf{v}_{t_i} \tag{22}$$

where $\mathbf{h}\colon [t_0, T]\times\mathcal{R}^d\times\mathcal{R}^p \mapsto \mathcal{R}^m$ is a known function, which is assumed to be twice continuously differentiable with respect to $\mathbf{X}_t$. Finally $\{\mathbf{v}_{t_i}\}$ is a m-dimensional zero mean Gaussian white noise process with covariance $\boldsymbol{\Sigma}_{t_i}$. The stochastic entities $\mathbf{X}_0$, $\mathbf{W}_t$ and $\mathbf{v}_{t_i}$ are assumed to be mutually independent for all t and t_i.

The filtering problem consists of establishing the conditional density $p(\mathbf{X}_t|\mathcal{Y}_{t_i})$ of the state vector $\mathbf{X}_{t_i}$, conditioned on the observations up to and including time t_i, $\mathcal{Y}_{t_i}$ denotes this information-set. Having found this conditional density the optimal estimator of the state vector (with respect to some specified criterion like the Minimum Mean Square Error (MMSE)) can be determined.

Prior to deriving the socalled *truncated second order filter*, the basic principle behind filtering methods is described. The initial value of the state variables $\mathbf{X}_{t_0}$ is assumed to follow a parameterized a priori distribution where the parameters are to be estimated using a Quasi Maximum Likelihood method (QML). Given the dynamics of the state variables (1) the distribution of the state vector immediately prior to observing the first vector of bond prices may be computed. Using this distribution and the observation equation (22), the distribution of the predicted value of the bond prices are determined. Next, given the distribution of the predicted bond prices and the observed bond prices, the a posteriori distribution of the state variable may be computed. Again the system dynamics (1) are used to obtain the distribution of the state vector at the time just before the next vector of bond prices becomes available.

The conditional density $p(\mathbf{X}_t|\mathcal{Y}_{t_{i-1}})$ can be found in the following manner. First consider the prediction density, which is the distribution of the state vector $\mathbf{X}_{t_i}$ conditioned on the information-set $\mathcal{Y}_{t_{i-1}}$. Since the solution to (1) is a Markov process, the process is completely described by the transition densities $p(\mathbf{X}_t|\mathbf{X}_{t'})$ for $t > t'$.

The transition densities can in principle be found by solving the Kolmogorov forward equation

$$\begin{aligned}\frac{\partial p(\mathbf{X}_t|\mathbf{X}_{t_{i-1}})}{\partial t} &= -\sum_{j=1}^{d} \frac{\partial}{\partial x_t^j} \left\{ p(\mathbf{X}_t|\mathbf{X}_{t_{i-1}}) f^j(t, \mathbf{X}_t; \boldsymbol{\psi}) \right\} \\ &\quad + \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{d} \frac{\partial^2}{\partial x_t^j \partial x_t^k} \left\{ p(\mathbf{X}_t|\mathbf{X}_{t_{i-1}}) \left[\mathbf{g}(t, \mathbf{X}_t; \boldsymbol{\psi}) \mathbf{g}^T(t, \mathbf{X}_t; \boldsymbol{\psi}) \right]^{jk} \right\} \end{aligned} \tag{23}$$

for $t \in [t_{i-1}, t_i)$ with the initial condition $p(\boldsymbol{\xi}|\mathbf{X}_{t_{i-1}}) = \delta(\boldsymbol{\xi} - \mathbf{X}_{t_{i-1}})$, where $\delta(\cdot)$ is the Dirac delta-function, assuming the existence of continuous partial derivatives as indicated.

The conditional density $p(\mathbf{X}_t|\mathcal{Y}_{t_{i-1}})$ may then be found as

$$p(\mathbf{X}_t|\mathcal{Y}_{t_{i-1}}) = \int_{\mathcal{S}} p(\mathbf{X}_t|\mathbf{X}_{t_{i-1}}) p(\mathbf{X}_{t_{i-1}}|\mathcal{Y}_{t_{i-1}}) d\mathbf{X}_{t_{i-1}} \tag{24}$$

where $p(\mathbf{X}_{t_{i-1}}|\mathcal{Y}_{t_{i-1}})$ is the conditional density for the previous observation update, which can be calculated as follows according to *Bayes rule*

$$\begin{aligned} p(\mathbf{X}_{t_i}|\mathcal{Y}_{t_i}) &= \frac{p(\mathbf{Y}_{t_i}|\mathbf{X}_{t_i}, \mathcal{Y}_{t_{i-1}}) p(\mathbf{X}_{t_i}|\mathcal{Y}_{t_{i-1}})}{p(\mathbf{Y}_{t_i}|\mathcal{Y}_{t_{i-1}})} \\ &= \frac{p(\mathbf{Y}_{t_i}|\mathbf{X}_{t_i}) p(\mathbf{X}_{t_i}|\mathcal{Y}_{t_{i-1}})}{p(\mathbf{Y}_{t_i}|\mathcal{Y}_{t_{i-1}})} \end{aligned} \tag{25}$$

The denominator is given by

$$p(\mathbf{Y}_{t_i}|\mathcal{Y}_{t_{i-1}}) = \int_{\mathcal{S}} p(\mathbf{Y}_{t_i}|\mathbf{X}_{t_i}) p(\mathbf{X}_{t_i}|\mathcal{Y}_{t_{i-1}}) d\mathbf{X}_{t_i} \tag{26}$$

Equations (23)–(26) constitute the general continuous-discrete time filtering problem. Unfortunately, except for a few special cases (e.g. narrow-sense linear systems), closed form solutions to these equations are not available. The computation of the entire density function $p(\mathbf{X}_t|\mathcal{Y}_{t_{i-1}})$, which provides the connection between the evolution of the state variable and the observations, requires the solution of partial integro-differential equations (derived by means of the Kolmogorov forward equation) and observation updates involve solving functional integral difference equations (derived by means of the Bayes' formula). This implies that the general optimal nonlinear filter will be infinite dimensional. For practical purposes expansions truncated to some low order are required both in the time propagation and observation update of the nonlinear filter. One possible approach is to consider expansions of some of the conditional moments, and this will be pursued in the following. Other approaches are described in (Maybeck, 1982).

3.1 Conditional moments estimator

Let $\hat{\mathbf{X}}_{t|t_{i-1}}$ denote the conditional mean of $\mathbf{X}_t$ given the information set $\mathcal{Y}_{t_{i-1}}$, i.e. $\hat{\mathbf{X}}_{t|t_{i-1}} = E[\mathbf{X}_t|\mathcal{Y}_{t_{i-1}}] = E_{i-1}[\mathbf{X}_t]$ and let $\mathbf{V}_{t|t_{i-1}} = E[(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t_{i-1}})(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t_{i-1}})^T|\mathcal{Y}_{t_{i-1}}]$ denote the conditional variance of the state estimate for $t \in [t_{i-1}, t_i)$. Explicit expressions for the time evolution of $\mathbf{X}_{t|t_{i-1}}$ and $\mathbf{V}_{t|t_{i-1}}$ may be derived using the Kolmogorov forward equation (23), which results in differential equations for these conditional moments expressed in terms of expectations of $\mathbf{f}(t, \mathbf{X}_t; \boldsymbol{\psi})$ and $\mathbf{g}(t, \mathbf{X}_t; \boldsymbol{\psi})$. However, these differential equations cannot be solved explicitly because the appropriate densities are not available in closed form. An approximate filter is obtained by writing down *time propagation equations* that describes the evolution of the state variables between sampling instants, and *updating equations* that relates the conditional mean and conditional variance of the state variables to the observations at the sampling instants. However, a Taylor expansion of $\mathbf{f}(t, \mathbf{X}_t; \boldsymbol{\psi})$ and $\mathbf{g}(t, \mathbf{X}_t; \boldsymbol{\psi})$ truncated after the second order terms followed by taking expectations give rise to the following approximate time propagation equations, see (Maybeck, 1982) for the details,

$$\begin{aligned}
\frac{d\hat{\mathbf{X}}_{t|t_{i-1}}}{dt} &= \mathbf{f}(t, \hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\psi}) + E_{i-1}[\mathbf{B}_{t|t_{i-1}}] && (27)\\
\frac{d\mathbf{V}_{t|t_{i-1}}}{dt} &= \mathbf{F}(t, \hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\psi})\mathbf{V}_{t|t_{i-1}} + \mathbf{V}_{t|t_{i-1}}\mathbf{F}^T(t, \hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\psi}) \\
&\quad + E_{i-1}\left[\mathbf{g}(t, \hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\psi})\mathbf{g}^T(t, \hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\psi})\right] && (28)
\end{aligned}$$

with the initial conditions $\hat{\mathbf{X}}_{t_{i-1}|t_{i-1}}$ and $\mathbf{V}_{t_{i-1}|t_{i-1}}$.
The bias-correction term $E_{i-1}[\mathbf{B}_{t|t_{i-1}}]$ is a n-dimensional vector with the kth component

$$E^k_{i-1}[\mathbf{B}_{t|t_{i-1}}] = \frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 \mathbf{f}^k(t, \mathbf{x}; \boldsymbol{\psi})}{\partial \mathbf{x}^2}\mathbf{V}_{t|t_{i-1}}\right\}\Bigg|_{\mathbf{x}=\hat{\mathbf{X}}_{t|t_{i-1}}} \tag{29}$$

and $\mathbf{F}(t, \hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\psi})$ is given by the $n \times n$ matrix

$$\mathbf{F}(t, \hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\psi}) = \frac{\partial \mathbf{f}(t, \mathbf{x}; \boldsymbol{\psi})}{\partial \mathbf{x}}\Bigg|_{\mathbf{x}=\hat{\mathbf{X}}_{t|t_{i-1}}} \tag{30}$$

The last term in (28) is a $d \times d$ symmetric matrix with element ij given by (where the dependence on $\hat{\mathbf{X}}_{t|t_{i-1}}$, $t|t_{i-1}$, and $\boldsymbol{\psi}$ have been dropped for convenience)

$$\begin{aligned}
E^{jk}_{i-1}[\mathbf{g}\mathbf{g}^T] &= \sum_{\nu=1}^{d}\sum_{l=1}^{d} \mathbf{g}^{j\nu}(\mathbf{g}^T)^{l\nu} + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{g}^{j\nu}}{\partial \mathbf{x}}^T \frac{\partial (\mathbf{g}^T)^{lk}}{\partial \mathbf{x}}\right)\mathbf{V}\right\} \\
&\quad + \frac{1}{2}\mathbf{g}^{j\nu}\mathrm{tr}\left\{\frac{\partial^2 (\mathbf{g}^T)^{lk}}{\partial \mathbf{x}^2}\mathbf{V}\right\} + \frac{1}{2}\mathrm{tr}\left\{\mathbf{V}\frac{\partial^2 \mathbf{g}^{j\nu}}{\partial \mathbf{x}^2}\right\}(\mathbf{g}^T)^{lk} && (31)
\end{aligned}$$

Remark 3.1 *Notice that* $\mathbf{g}^{j\nu}$ *denotes element* $j\nu$ *of* $\mathbf{g}$*, whereas* $(\mathbf{g}^T)^{lj}$ *denotes element* lj *of the transpose of* $\mathbf{g}$*. Also notice that the partial derivative of a scalar with respect to a vector yields a row vector such that, say,* $\frac{\partial (\mathbf{g}^T)^{lj}}{\partial \mathbf{x}}$ *is a row vector, and* $\frac{\partial \mathbf{g}^{j\nu}}{\partial \mathbf{x}}^T$ *is a column vector.*

The observation update of the mean and the covariance is approximated by a power series in the residual, which for computational tractability is truncated at first order terms.

$$\begin{aligned}
\hat{\mathbf{X}}_{t_i|t_i} &= \mathbf{a}_0 + \mathbf{a}_1(\mathbf{Y}_{t_i} - \hat{\mathbf{Y}}_{t_i|t_{i-1}}) && (32)\\
\mathbf{V}_{t_i|t_i} &= \mathbf{b}_0 + \mathbf{b}_1(\mathbf{Y}_{t_i} - \hat{\mathbf{Y}}_{t_i|t_{i-1}}) && (33)
\end{aligned}$$

where $\mathbf{a}_0$, $\mathbf{a}_1$, $\mathbf{b}_0$ and $\mathbf{b}_1$ are random variables as a function of $\mathbf{Y}_{t_{i-1}}$. In order to avoid negative computed values of $\hat{\mathbf{V}}_{t_i|t_i}$ the $\mathbf{b}_1$ is set identically to zero.

The updating equations are given by

$$
\begin{aligned}
\mathbf{A}_{t_i} &= \mathbf{H}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}) \mathbf{V}_{t_i|t_{i-1}} \mathbf{H}^T(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}) - E_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}] E_{i-1}^T[\tilde{\mathbf{B}}_{t_i|t_{i-1}}] \\
&\quad + \boldsymbol{\Sigma}_{t_i} && (34) \\
\mathbf{K}_{t_i} &= \mathbf{V}_{t_i|t_{i-1}} \mathbf{H}^T(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}) \mathbf{A}_{t_i}^{-1} && (35) \\
\hat{\mathbf{X}}_{t_i|t_i} &= \hat{\mathbf{X}}_{t_i|t_{i-1}} + \mathbf{K}_{t_i} \left\{ \mathbf{Y}_{t_i} - \mathbf{h}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}) - E_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}] \right\} && (36) \\
\mathbf{V}_{t_i|t_i} &= \mathbf{V}_{t_i|t_{i-1}} - \mathbf{K}_{t_i} \mathbf{H}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}) \mathbf{V}_{t_i|t_{i-1}} && (37)
\end{aligned}
$$

where $\mathbf{H}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi})$ is defined as the $m \times d$ matrix

$$
\mathbf{H}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}) = \left. \frac{\partial \mathbf{h}(t_i, \mathbf{x}; \boldsymbol{\psi})}{\partial \mathbf{x}} \right|_{\mathbf{x} = \hat{\mathbf{X}}_{t_i|t_{i-1}}} \tag{38}
$$

and the bias-correction term $E_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}]$ is a $m \times 1$-vector with the kth component given by

$$
E_{i-1}^k[\tilde{\mathbf{B}}_{t_i|t_{i-1}}] = \frac{1}{2} \mathrm{tr} \left\{ \frac{\partial^2 \mathbf{h}^k(\mathbf{x}; \boldsymbol{\psi})}{\partial \mathbf{x}^2} \mathbf{V}_{t_i|t_{i-1}} \right\} \Bigg|_{\mathbf{x} = \hat{\mathbf{X}}_{t_i|t_{i-1}}} \tag{39}
$$

Higher order filters can be obtained by including higher order terms from the Taylor series expansions of $\mathbf{f}$ and $\mathbf{g}$. However, the severe computational disadvantages make such filters infeasible, and it is generally recommended to use the first or second order filters on better models. The numerical work is considerably more demanding for the multivariate case, i.e. it involves the numerical solution of $d + \frac{d}{2}(d+1) = \frac{d}{2}(d+3)$ ODEs for the conditional first and second order central moments given by (27)–(28) between each sampling instant.

Frey and Runggaldier (1999) proposes a methodology that may be viewed as a nonlinear filtering method for discretely observed stochastic differential equations (in particular, stochastic volatility models) without observation noise, where the sampling instants t_i are modelled as a marked point process (Björk, Kabanov and Runggaldier, 1996; Björk, Masi, Kabanov and Runggaldier, 1997).

4 Quasi maximum likelihood method

In this section a QML method for estimation of the parameters in the continuous-discrete time state space model (1) and (22) is presented. It is assumed that the nonlinear filter based on the first two conditional moments from Section 3 is used to generate the one-step ahead prediction errors

$$
\boldsymbol{\varepsilon}_{t_i}(\boldsymbol{\psi}) \equiv \mathbf{Y}_{t_i} - \mathbf{h}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}) \tag{40}
$$

Assuming that the prediction errors are Gaussian, the Quasi log-likelihood function is given by

$$
Q_N(\boldsymbol{\psi}; \mathbf{Y}_{t_N}) = \sum_{i=1}^{N} l_i(\boldsymbol{\psi}) \tag{41}
$$

where

$$
\begin{aligned}
l_i(\boldsymbol{\psi}) &= -\frac{1}{2} \log \left| \mathbf{H}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi} \mathbf{V}_{t_i|t_{i-1}} \mathbf{H}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}^T + \Sigma_{t_i} \right| \\
&\quad -\frac{1}{2} \left(\boldsymbol{\varepsilon}_{t_i}^T(\boldsymbol{\psi}) \left[\mathbf{H}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi} \mathbf{V}_{t_i|t_{i-1}} \mathbf{H}(t_i, \hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\psi}^T + \Sigma_{t_i} \right]^{-1} \boldsymbol{\varepsilon}_{t_i}(\boldsymbol{\psi}) \right). && (42)
\end{aligned}
$$

Remark 4.1 *The assumption of Gaussianity may be tested using standard statistical tests for Gaussian white noise residuals.*

The consistent, asymptotically normally distributed and efficient estimators obtained using the ordinary Kalman filter and ML is lost for more general state space models with non-Gaussian transition densities as argued in e.g (Lund, 1997*a*). However, Bollerslev and Wooldridge (1992) shows that the nice properties of ML estimators are retained for QML estimators provided that the mean and variance are correctly specified. It is assumed that the approximate equations for the conditional mean and conditional covariance obtained by using a second order filter provide a better approximation to the true conditional moments than the ones used in the earlier cited work, so it is conjectured that the properties of the obtained estimators are most likely closer to those of (Bollerslev and Wooldridge, 1992) than those of (Lund, 1997*a*). A Monte Carlo study reported in the next Section supports this conjecture.

Remark 4.2 *The one-step ahead prediction errors defined by (40) are structurally in accordance with the innovations approach in the linear Kalman filter with a linear observation equation, i.e. with a linear observation equation* $\mathbf{H} \equiv \mathbf{0}$. *However, in the general nonlinear case, the expressions in the curly brackets in (36) contain the additional bias-correction terms given by (39). This is due to the approximative nature of a second order filter, and it suggests that the one-step ahead prediction errors (residuals) obtained from (40) may be confounded with some of the deficiencies of the filter in the general case.*[7]

5 Monte Carlo Analysis

In this Section a Monte Carlo study is performed to analyze the properties of the estimates provided by the methodology described in the previous two sections. In the study the SDE (1) is solved numerically using the Euler discretization scheme, see e.g. (Kloeden and Platen, 1995) for details.

Let $\delta = \Delta/K$ denote the length of the discretization time step, where $K > 1$ is the number of time steps in each interval $[t_{i-1}, t_i]$ for $i = 1, \ldots, N$ and $\Delta = t_i - t_{i-1}$ is the time between samples. Furthermore, introduce $\tau_{i-1,k} = t_{i-1} + k\delta$ for $k = 0, \ldots, K$, and let the stochastic process $\{\mathbf{Z}\}$ be a discrete-time approximation of $\{\mathbf{X}\}$. For the SDE (1) the ν'th component of the Euler discretization scheme is given by the stochastic difference equation

$$Z^{\nu}_{\tau_{i-1,k}} = Z^{\nu}_{\tau_{i-1,k-1}} + f^{\nu}(\tau_{i-1,k-1}, \mathbf{Z}_{\tau_{i-1,k-1}}; \boldsymbol{\psi})\delta + \sum_{j=1}^{d} g^{\nu j}(\tau_{i-1,k-1}, \mathbf{Z}_{\tau_{i-1,k-1}}; \boldsymbol{\psi})\delta W^{j}_{\tau_{i-1,k}} \quad (43)$$

for $\nu = 1, \ldots, d$ with the initial condition $\mathbf{Z}_{\tau_{i-1,0}} = \mathbf{X}_{t_{i-1}}$ and $\delta W^{j}_{\tau_{i-1,k}} = W^{j}_{\tau_{i-1,k}} - W^{j}_{\tau_{i-1,k-1}}$ is the $N(0, \delta)$ distributed increment of the jth component of the d-dimensional standard Wiener process $\mathbf{W}_t$.

In order to obtain a data set consisting of N observations, it is necessary to simulate $K \cdot N$ values of the state vector and pick out every K'th value of the state vector. To obtain a reasonable approximation to the continuous-time evolution of the state variables $K = 1000$ is chosen. The sampling time Δ is set to $\frac{1}{50}$ corresponding to weekly observations. Having simulated the evolution of the state variables the prices of the corresponding zero-coupon bonds are found by (14), where the functions $A(\tau), B(\tau), C(\tau)$ and $D(\tau)$ are fully determined by the model parameters and the market prices of risk. Gaussian white noise with variance $\sigma^2 = 1.0 * 10^{-6}$, which gives an approximate 97.5% fractile of the error of $\pm$ 42 basis point for the 6 month yield and $\pm$ 7 basis point for the 20 years yield, is added to the observations

[7] See (Tanizaki, 1996) for a discussion of this in discrete-time structural models.

In the simulation study presented in Table 1 zero-coupon bonds with the following maturities are used: 6 month, 1, 2, 3, 5, 7, 10, and 20 years. The market price of risks λ_r, λ_θ and λ_v are all equal to zero. The true parameter values are also provided in the table.

Parameter	True	Mean	Std.dev.	t-test
κ_1	0.4000	0.4011	0.0113	0.5052
κ_2	0.2000	0.2002	0.0052	0.1527
κ_3	0.1000	0.0936	0.0202	-1.5806
$\bar{\theta}$	0.1000	0.1000	0.0005	-0.2993
$\bar{v}$	0.0006	0.0006	0.0002	0.2806
ξ	0.1000	0.0998	0.0021	-0.5097
η	0.0100	0.0078	0.0023	-4.6948
r_0	0.1000	0.1001	0.0008	0.6724
θ_0	0.1000	0.0999	0.0012	-0.6167
v_0	0.0006	0.0006	0.0002	0.3461
$10^6\sigma^2$	1.0000	0.9992	0.0132	-0.3120

Table 1: QML estimates for the full three-factor model (simulation): Consider the special interest rate dynamics model $dr_t = \kappa_1(\theta_t - r_t)dt + \sqrt{v_t}dW_t^1$, $d\theta_t = \kappa_2(\bar{\theta} - \theta_t)dt + \xi\sqrt{\theta_t}dW_t^2$ and $dv_t = \kappa_3(\bar{v} - v_t)dt + \eta\sqrt{v_t}dW_t^3$, where r_t is the spot rate, θ_t is the central tendency, v_t is the stochastic volatility, $(W_t^1, W_t^2, W_t^3)^T$ is a three-dimensional Wiener process with uncorrelated elements, and κ_1, κ_2, κ_3 $\bar{\theta}$, ξ, $\bar{v}$ and η are constant parameters. The bond price satisfy the system of equations (16)-(19).

The results reported in Table 1 are based on 25 data set with 1000 observations. The mean and the standard deviation for the parameter estimates are presented as well as the t-test statistics under the null hypothesis that the estimated parameters are unbiased. It appears that unbiased estimates are obtained for all the model parameters except the η parameter which measures the volatility of the volatility process v_t. The reason for this parameter to be slightly downwards biased is the smoothing effect of the filter on the estimate of the state vector, which tends to reduce the volatility of the processes. Unbiased estimates of the initial values of the three state variables and the variance of the observation noise are also obtained. Other simulation studies (not reported here) show that similar results are obtained for different choices of parameter values.

Remark 5.1 *The smoothing effect of the nonlinear filter is also seen in simulation studies in (Nielsen et al., 2000), where it also has an unfortunate effect on the estimates of the drift parameters in the unobserved stochastic volatility process (13). This latter effect is not pronounced in the present study.*

6 An Empirical Study

The proposed econometric method is applied to a cross-section of daily observations of eight default-free Danish coupon-bearing bonds. The bonds considered have different time to maturity ranging from 2 to 10 years and the yearly coupon rate ranges from 6% to 9%. The period January 2, 1996 to December 31, 1997 is considered, which gives a data sample covering 499 days of observations. The selected bonds are some of the most traded bonds at the Danish bond market, so the bonds are fully liquid.

Some notation is required to cope with both the cross-section and time series information in the data sample. It is assumed that at most m bond prices are observed simultaneously and that the ν'th coupon-bearing bond carries J_ν coupons, $\nu = 1, \ldots, m$, where $C_{j\nu}$ denotes the value of the coupons for $j =$

$1, \ldots, J_\nu - 1$ and $C_{J_\nu \nu}$ denote the sum of the coupon and the face value. Thus, at time t_i, the value of the ν'th coupon-bearing bond is given by

$$P^c_{i\nu}(\boldsymbol{\psi}) + \text{accrued interest}(t) = \sum_{j=1}^{J_\nu} C_{j\nu} P(\tau_{ij\nu}; \boldsymbol{\psi}) I_{\{\tau_{ij\nu} \geq 0\}} \tag{44}$$

where $\tau_{ij\nu} = T_{j\nu} - t_i$ is the appropriate time-to-maturity, $I_{\{\cdot\}}$ is an indicator function (ensuring that the price of a particular bond that has matured does not contribute to the summation) and the price of a zero-coupon bond $P(\cdot, \cdot)$ is given by (14). Finally, the price vector $\mathbf{P}^c_i(\boldsymbol{\psi}) = (P^c_{i1}(\boldsymbol{\psi}), \ldots, P^c_{im}(\boldsymbol{\psi}))^T$ is substituted for $\mathbf{h}(t_i, \mathbf{X}_{t_i}; \boldsymbol{\psi})$ in (22), where the stochastic process $\{\mathbf{v}\}$ accounts for observation noise due to asynchronous trading, rounding off prices, bid-ask spreads, temporary deviations that are not arbitraged away and other market imperfections.

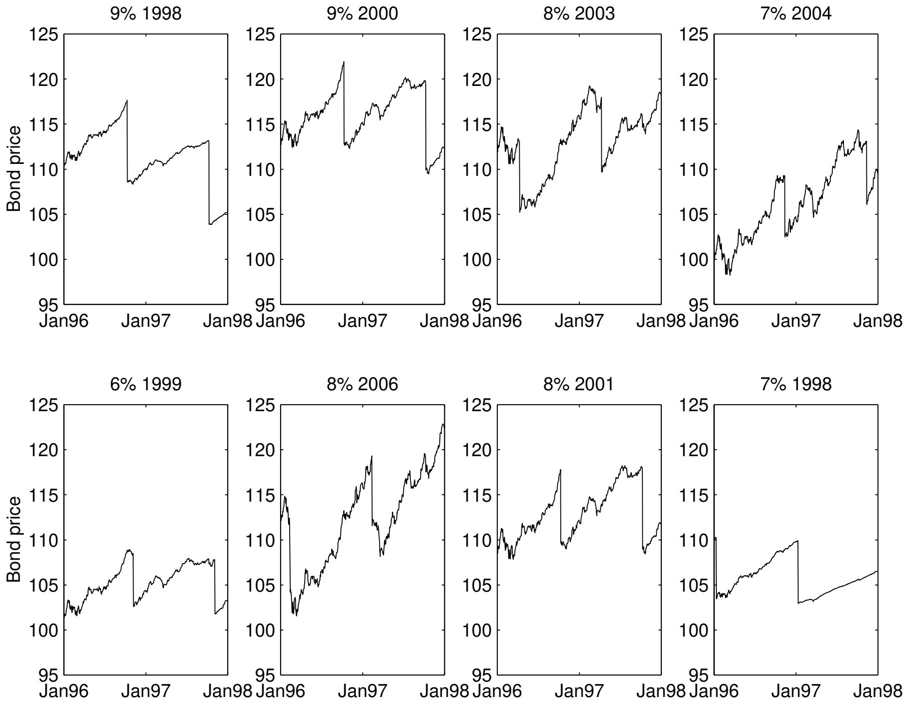

Figure 1: The sum of the price and the accrued interest rate for 8 Danish bonds in the period January 2. 1996 to December 31. 1997.

The market prices plus the accrued interest (including the observation noise) are plotted in Figure 1. The discontinuities in the time series are caused by the fact that the bond is *ex-coupon*, which means that the bond is sold without the right to receive the next interest payment.

Given the state space model (11)–(13) and the observation equation derived from the bond prices (as described above) the nonlinear filter (Section 3) and the QML method (Section 4) may be applied. As the observations are only available on trading days at the Copenhagen Stock Exchange, i.e. approximately 250 days per year excluding weekends and public holidays, the updating of the state variables are only done on trading days. The propagation equations (27)–(28) are used to predict the expected values of the state vector (and the associated covariance) on between trading days. Finally, the price of each particular zero-coupon bond is obtained by solving (16)-(19) using a Runge-Kutta method and the price of the coupon-bearing bond is obtained from (44).

For reasons of identifiability, it is assumed that the variance of the observation noise $\{\mathbf{v}\}$ is given by $\sigma^2 \mathbf{I}^{m\times m}$, where $\mathbf{I}^{m\times m}$ is a $m \times m$ identity matrix.

6.1 Empirical results

Using an implementation of the proposed methodology it is possible to estimate all the parameters in the general model (11)-(13) and test nested models thereof. Figure 2 provides an overview of the seven models considered in this study.

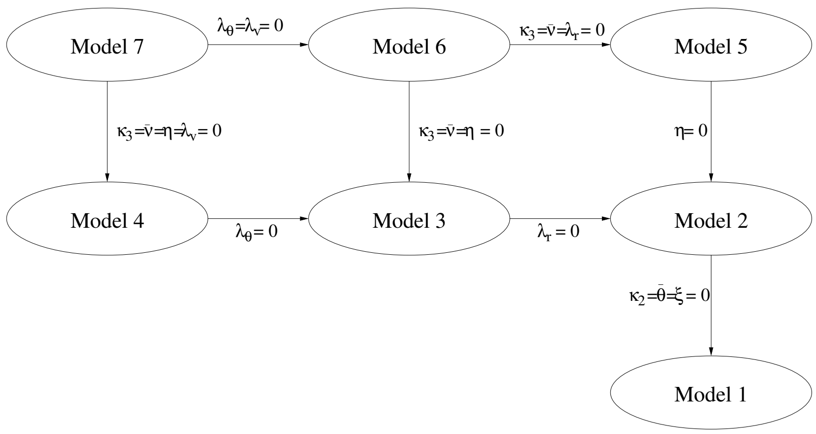

Figure 2: A diagram of the interrelations between the various special cases of the special interest rate dynamics model.

In the remainder of the paper model $\mathcal{M}_1$ will refer to the one-factor Vasicek model for which the empirical results are provided in Table 2. Models $\mathcal{M}_2$ through $\mathcal{M}_4$ refer to three two-factor models with constant volatility and stochastic level of mean reversion (central tendency), see Table 3, and models $\mathcal{M}_5$ through $\mathcal{M}_7$ are special cases of the general three-factor model (11)–(13), see Table 4.

Parameter	Model 1
κ_1	0.04115 (0.00034)
θ_0	0.21232 (0.00424)
v_0	0.000058 (0.000012)
r_0	0.04526 (0.00036)
$10^6\sigma^2$	37.53657 (0.86349)
$\ln L$	-18106.29

Table 2: QML estimates for the Vasicek model: Consider $dr_t = \kappa_1(\theta_0 - r_t)dt + \sqrt{v_0}dW_t^1$, where κ_1 is the speed-of-adjustment of the interest rate to the level of the mean θ_0 and v_0 is the volatility. The standard deviations are given in parenthesis.

Let $\mathcal{L}_i$ denote minus the optimal value of the log-likelihood function for model $\mathcal{M}_i$ for $i = 1, \ldots, 7$. A likelihood ratio test for the hypothesis that model $\mathcal{M}_i$ offers a significantly better description of the data than the model $\mathcal{M}_j$ is given by

$$\text{LR}_{ij} = 2(\mathcal{L}_i - \mathcal{L}_j) \sim \chi^2_{1-\alpha}(d); \quad i > j, \tag{45}$$

where $\chi^2_{1-\alpha}(d)$ denote the (1-α)–percentile of a χ^2-distribution with d degrees-of-freedom, d being the number of parameter restrictions imposed on $\mathcal{M}_i$ to obtain $\mathcal{M}_j$.

Parameter	Model 2	Model 3	Model 4
κ_1	0.24763 (0.00268)	0.24773 (0.00274)	0.24701 (0.00270)
κ_2	0.22265 (0.00362)	0.22325 (0.00370)	0.23924 (0.00559)
$\bar{\theta}$	0.010318 (0.00026)	0.10222 (0.00026)	0.09581 (0.00024)
λ_r	N/A	2.48145 (0.84473)	1.40550 (0.49692)
λ_θ	N/A	N/A	0.18851 (0.00729)
r_0	0.04319 (0.00028)	0.04319 (0.00028)	0.04318 (0.00028)
θ_0	0.08235 (0.00053)	0.08175 (0.00053)	0.08212 (0.00053)
v_0	0.00006 (0.00001)	0.00006 (0.00001)	0.00006 (0.00001)
ξ	0.08681 (0.00226)	0.08714 (0.0234)	0.08754 (0.00429)
$10^6\sigma^2$	5.68269 (0.13598)	5.68547 (0.13609)	5.68467 (0.13602)
$\ln L$	-21473.37	-21473.38	-21473.51

Table 3: QML estimates for the two-factor model: Consider $dr_t = \kappa_1(\theta_t - r_t)dt + \sqrt{v_0}dW_t^1$ and $d\theta_t = \kappa_2(\bar{\theta} - \theta_t)dt + \xi\sqrt{\theta_t}dW_t^2$, where r_t is the spot rate, θ_t is the central tendency, $(W_t^1, W_t^2)^T$ is a two-dimensional Wiener process with uncorrelated elements, and κ_1, κ_2, $\bar{\theta}$ and ξ are constant parameters.

Referring to Table 5, the test statistic $\text{LR}_{21} = 6734.2$ (with 3 degrees-of-freedom), that the Vasicek model is rejected compared to the two-factor models. This result was to be expected as the yields in a one-factor model is perfectly correlated and this property is, in general, not supported empirically. Note also that the variance of the observation noise decreases to about one seventh by increasing the number of factors from one to two. The test statistics $\text{LR}_{43} = 0.26$ and $\text{LR}_{32} = 0.02$, both with one degree-of-freedom, show that the restrictions $\lambda_r = 0$ and $\lambda_\theta = 0$, and $\lambda_\theta = 0$ cannot be rejected on a 5%-level.

Although the two-factor models are superior to the one-factor model, a more general model is needed to capture the dynamics of the Danish term structure of interest rates. Compared to the three-factor model with market price of risks, i.e. $\mathcal{M}_7$ in Table 4, the restrictions imposed to get the two-factor models in Table 3 is rejected at all levels. It is also rejected that the general three-factor model can be reduced to a three-factor model with the volatility process being a random walk, i.e. $\mathcal{M}_5$, because $\text{LR}_{65} = 524.82$ with 3 degrees-of-freedom. However, it cannot be rejected that the market price of risks λ_r and λ_θ are equal to zero, i.e. $\mathcal{M}_6$, because $\text{LR}_{76} = 1.96$ with 2 degrees-of-freedom.

7 Conclusion

In this paper we have presented a method to estimate the interest rate dynamics based on panel-data of prices of coupon bonds. We use a truncated second order filter to obtain estimates of the unobservable state variables in the three factor model presented, and the quasi maximum likelihood method is used to estimate the model parameters. The advantages of using a panel of bond prices instead of some points on the yield curve are that the bond prices can be observed directly at the market, which is not the case for the yield curve. The empirical analysis of the Danish bond market shows that at least a three-factor model is needed to describe the dynamics of the term structure of interest rates.

Parameter	Model 5	Model 6	Model 7
κ_1	0.29730 (0.00157)	0.56841 (0.04335)	0.57079 (0.04538)
κ_2	0.49427 (0.00425)	0.76515 (0.05251)	0.75886 (0.05432)
κ_3	N/A	0.12142 (0.01103)	0.04334 (0.01298)
$\bar{\theta}$	0.09974 (0.00008)	0.11230 (0.00564)	0.11357 (0.00598)
$\bar{v}$	N/A	0.00002 (0.00001)	0.00005 (0.00001)
λ_r	N/A	-378.95806 (69.98981)	-392.84243 (74.32463)
λ_θ	N/A	N/A	-0.07072 (0.08432)
λ_v	N/A	N/A	-34.40820 (19.6537)
r_0	0.04475 (0.00029)	0.04876 (0.00176)	0.04878 (0.00187)
θ_0	0.06864 (0.00068)	0.07908 (0.00665)	0.07948 (0.00698)
v_0	0.00076 (0.00006)	0.00005 (0.00001)	0.00005 (0.00001)
ξ	0.13671 (0.00216)	0.08622 (0.00915)	0.08585 (0.00964)
η	0.04571 (0.00205)	0.00233 (0.00034)	0.00227 (0.00038)
$10^6\sigma^2$	4.08327 (0.09861)	3.77699 (0.09243)	3.76348 (0.10042)
$\ln L$	-21872.22	-22134.63	-22135.61

Table 4: QML estimates for the full three-factor model: Consider the special interest rate dynamics model $dr_t = \kappa_1(\theta_t - r_t)dt + \sqrt{v_t}dW_t^1$, $d\theta_t = \kappa_2(\bar{\theta} - \theta_t)dt + \xi\sqrt{\theta_t}dW_t^2$ and $dv_t = \kappa_3(\bar{v} - v_t)dt + \eta\sqrt{v_t}dW_t^3$, where r_t is the spot rate, θ_t is the central tendency, v_t is the stochastic volatility, $(W_t^1, W_t^2, W_t^3)^T$ is a three-dimensional Wiener process with uncorrelated elements, and κ_1, κ_2, κ_3, $\bar{\theta}$, ξ, $\bar{v}$ and η are constant parameters. The bond price satisfy the system of equations (16)-(19).

References

Abken, P. A. (1993), Generalised method of moments tests of forward rate processes, Working paper 93-7, Federal Reserve Bank of Atlanta.

Anderson, N., Breedon, F., Deacon, M., Derry, A. and Murphy, G. (1996), *Estimating and Interpreting the Yield Curve*, Wiley, Chichester, UK.

Babbs, S. H. and Nowman, K. B. (1999), 'Kalman Filtering of Generalized Vasicek Term Structure Models', *Journal of Financial and Quantitative Analysis* **34**(1), ?

Balduzzi, P., Das, S. R. and Foresi, S. (1998), 'The Central Tendency: A Second Factor in Bond Yields', *Review of Economics and Statistics* **80**, 62–72.

Björk, T., Kabanov, Y. and Runggaldier, W. (1996), 'Bond market in the presence of marked point processes', *Mathematical Finance* **7**(2).

Björk, T., Masi, G. D., Kabanov, Y. and Runggaldier, W. (1997), 'Towards a General Theory of Bond Markets', *Finance and Stochastics* **1**(2), 141–174.

Bollerslev, T. and Wooldridge, J. M. (1992), 'Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometric Reviews* **11**(2), 143–172.

Buraschi, A. (1996), How Large Is the Inflation Risk Premium in the U.S. Nominal Term Structure, Technical report, Institute of Finance and Accounting, London Business School.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$	$\mathcal{M}_7$
$\mathcal{M}_1$	6734.2 (3) 0%					
$\mathcal{M}_2$		0.02 (1) 88.75%		797.7 (1) 0%		
$\mathcal{M}_3$			0.26 (1) 61.01%		1322.5 (3) 0%	
$\mathcal{M}_4$						1524.2 (4) 0%
$\mathcal{M}_5$					524.82 (3) 0%	
$\mathcal{M}_6$						1.96 (2) 37.53%

Table 5: For each of the nested tests shown in Figure 2, the likelihood ratio test statistic LR_{ij}, the degrees-of-freedom in parenthesis and the level of significance are given.

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**, 1209–1227.

Chen, L. (1996), *Interest Rate Dynamics, Derivatives Pricing, and Risk Management*, Vol. 435 of *Lecture Notes in Econom. and Math. Systems*, Springer, Berlin Heidelberg New York.

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

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.

Claessens, S. and Pennacchi, G. G. (1996), 'Estimating the Likelihood of Mexican Default from Market Prices of Brady Bonds', *Journal of Financial and Quantitative Analysis* **31**, 109–126.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–407.

Cumby, R. E. and Evans, M. D. D. (1995), The Term Structure of Credit Risk: Estimates and Specification Tests. Manuscript, Georgetown University.

Dai, Q. and Singleton, K. J. (1997), Specification Analysis of Affine Term Structure Models, Technical report, Graduate School of Business, Stanford University. Forthcoming in *Journal of Finance*.

Daves, P. R. and Ehrhardt, M. C. (1993), 'Joint Cross-Section/Time-Series Maximum Likelihood Estimation for the parameters of the Cox-Ingersoll-Ross Bond Pricing Model', *The Financial Review* **28**(2), 203–237.

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.

Duffie, D. (1996), *Dynamic Asset Pricing Theory, Second Edition*, Princeton University Press, Princeton, New Jersey.

Duffie, D. and Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical Finance* **6**(4), 379–406.

Duffie, D. and Singleton, K. (1997), 'An Econometric Model of the Term Structure of Interest Rate Swap Yields', *Journal of Finance* **52**, 12587–1323.

Dybvig, P. H. (1989), *Bond and option pricing based on the current term structure*, Working paper, Washington University, St. Louis, Missouri.

Frey, R. and Runggaldier, W. J. (1999), A Nonlinear Filtering Approach to Volatility Estimation with a View Towards High Frequency Data, Technical report, University Zürich and University of Padua.

Frühwirth-Schnatter, S. (1994), 'Applied State Space Modelling of Non-Gaussian Time Series Using Integration-Based Kalman Filtering', *Statistics and Computing* **4**, 259–269.

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

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments', *Econometrica* **50**, 1029–1054.

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.

Heston, S. (1993), 'A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options', *Review of Financial Studies* **6**, 327–343.

Honoré, P. (1998), Five essays on financial econometrics in continuous-time models, PhD thesis, Department of Finance, The Aarhus School of Business.

Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Jegadeesh, N. and Pennacchi, G. G. (1996), 'The behavior of interest rates implied by the term structure of eurodollar futures', *Journal of Money, Credit and Banking* **28**, 426–446.

Karatzas, I. and Shreve, S. E. (1996), *Brownian Motion and Stochastic Calculus, Second Edition*, Springer-Verlag, New York.

Kloeden, P. E. and Platen, E. (1995), *Numerical Solutions of Stochastic Differential Equations, Second Edition*, Springer-Verlag, Heidelberg.

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

Lund, J. (1997*a*), Econometric analysis of continuous-time arbitrage-free models of the term structure of interest rates, Technical report, Department of Finance, The Aarhus School of Business.

Lund, J. (1997*b*), A model for studying the effect of emu on european yield curves, Technical report, Department of Finance, The Aarhus School of Business.

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

Nielsen, J. N. (1996), Nonlinear dynamics and time series analysis, Master's thesis, Department of Mathematical Modelling, Lyngby, Denmark.

Nielsen, J. N., Vestergaard, M. and Madsen, H. (2000), 'Estimation in continuous-time stochastic volatility models using nonlinear filters', *International Journal of Theoretical and Applied Finance* **3**(2), 1–30.

Øksendal, B. (1995), *Stochastic Differential Equations, 4th Edition*, Springer-Verlag, Heidelberg.

Pearson, N. J. 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* **XLIX**(4), 1279–1304.

Pennacchi, G. G. (1991), 'Identifying the dynamics of real interest rates and inflation: Evidence using survey data', *Review of Financial Studies* **4**(1), 53–86.

Protter, P. (1990), *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.

Steeley, J. (1991), 'Estimating the gilt-edged term structure: basis splines and confidence intervals', *Journal of Business Finance & Accounting* **18**(4), 513–29.

Stein, E. M. and Stein, J. C. (1991), 'Stock price distributions with stochastic volatility – an analytic approach', *Review of Financial Studies* **4**, 727–752.

Tanizaki, H. (1996), *Nonlinear Filters, Second Edition*, Springer Verlag, Heidelberg.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economics* **5**, 177–188.