Econometric Analysis of Continuous-Time Arbitrage-Free Models of the Term Structure of Interest Rates*

Jesper Lund†
Department of Finance
The Aarhus School of Business
Fuglesangs Alle 4
DK-8210 Aarhus V
Denmark

Phone: +45 8948-6362
Fax: +45 8615-1943
E-mail: jel@hha.dk

First draft: February 1994
Current draft: November 1997

*The most recent version of this paper is available at: `http://www.hha.dk/~jel`

†I thank Torben G. Andersen, Bent Jesper Christensen, Tom Engsted, Esben Høg, Svend Jakobsen, Nicholas Kiefer, and seminar participants at Aarhus University, Aarhus School of Business, the workshop on "Likelihood Ratio Analysis in Dynamic Programming Models" in Aarhus, August 1994, and the 1995 North American Winter Meetings of the Econometric Society in Washington, DC, for useful comments. Of course, I remain responsible for any errors. Financial support from the Danish Social Science Research Council is gratefully acknowledged.

Econometric Analysis of Continuous-Time Arbitrage-Free Models of the Term Structure of Interest Rates

Abstract

In this paper we discuss econometric analysis for exponential-affine term-structure models using a panel-data approach. We assume that all zero-coupon yields are observed with measurement error. Therefore, we take the state variables as unobserved, and use the Kalman filter to estimate the model parameters. For Gaussian models, this approach provides the exact likelihood function, but the main focus of the paper is on non-Gaussian, exponential-affine models. Since the exact filter and likelihood function are unknown in this case, the suggestion in the literature is to use a QML estimator obtained from the first and second conditional moments of the state variables. However, this QML estimator is not consistent.

We provide a detailed analysis of the problem, and, in that process, formulate general conditions for consistency of QML estimators for linear, non-Gaussian state space models. This analysis leads us to propose a modified QML estimator for exponential-affine term-structure models which ensures consistency. The basic idea is ignoring information about conditional heteroskedasticity, but this suggests that the finite sample properties may not improve. A small-scale Monte Carlo study lends some support to this conjecture, and also suggests that the asymptotic biases of the existing QML estimator are very small, and probably economically insignificant in most cases.

In the empirical part of the paper we apply the Kalman filter approach to estimate Gaussian and non-Gaussian central-tendency models on US swap market data, and we compare their implications for the yield curve.

JEL codes: C15, C51, G12, G13.

Keywords: Exponential-Affine Models, Kalman Filter, State Space Models, Quasi Maximum Likelihood (QML) Estimation.

1 Introduction

In this paper we consider estimation techniques for Markovian term-structure models, such as the models proposed by Vasicek (1977), Brennan and Schwartz (1979), Langetieg (1980), Cox et al. (CIR) (1985), Longstaff and Schwartz (1992), and Duffie and Kan (1996). These models have two separate, yet interrelated, implications for the yield curve. The state variables, and the stochastic process governing their dynamics, simultaneously determine the shape of the yield curve at a given time (cross-sectional implications) and the time-series movements of the yield curve (dynamic implications).

Early contributions to the literature tended to analyze only one of these implications. Among the most cited studies are Brennan and Schwartz (1979), who focus on the dynamic implications of their model, and Brown and Dybvig (1986) who estimate the CIR model from bond prices observed on a single day. The "panel data" approach, where the dynamic and cross-sectional implications are taken into account simultaneously, generally provides more efficient estimators of the model parameters. Not surprisingly, most recent studies have followed this line of research.[1]

In general, the number of zero-coupon yields on a given date exceeds the number of state variables, so the theoretical term-structure model cannot explain all variation in the data.[2] To overcome this problem, it is often assumed that the data are observed with some form of measurement error, originating from, e.g., non-synchronous trading, rounding of prices, and bid-ask spreads.[3] For example, Chen and Scott (1993) and Duffie and Singleton (1997) estimate the m-factor CIR model by assuming that yields at m maturities are measured without any error, whereas the remaining maturities in the data set are contaminated by measurement error. By inverting the bond-pricing formula, the first m yields are used as "instruments" for the state variables, and the exact likelihood function for the full data set is straightforward to construct. A similar approach is used in Daves and Ehrhardt (1993), and Pearson and Sun (1994).

Under an alternative assumption, all zero-coupon yields are measured with some error. In this case, it is convenient to cast the term-structure model in *state space* form, and use Kalman filter to estimate the model parameters. The basic idea underlying this approach was put forth by Pennacchi (1991), and it has subsequently been used in a series of papers dealing with the estimation of exponential-affine models, es-

[1] The main exception are papers that estimate Markovian term-structure models without a closed-form solution for bond prices, see, e.g., Aït-Sahalia (1996a, 1996b), Andersen and Lund (1997), Chan et al. (1992), Conley et al. (1997), Honoré (1997), Nowman (1997), Stanton (1997), Tauchen (1997), and Torous and Ball (1995). However, the rapidly increasing speed of computing equipment is likely to change this within the foreseeable future. See Honoré (1998) for further discussion, as well as an implementation of this approach.

[2] Throughout this paper, it is implicitly understood that the data consist of zero-coupon yields. Although such data may not be directly available to the econometrician, there is a variety of methods for estimating zero-coupon yields from prices of coupon bonds, see Bliss (1996) and Anderson et al. (1996) for recent surveys.

[3] There are a few exceptions, though. The "panel data" estimation technique proposed by Gibbons and Ramaswamy (1993) does not rely on assumptions about measurement errors.

pecially Gaussian and multi-factor CIR models. See Jegadeesh and Pennacchi (1996), Ball and Torous (1996), Duan and Simonato (1995), Chen and Scott (1995), Santa-Clara (1995), de Jong (1996), Zheng (1993), Gong and Remolona (1996a, 1996b), Geyer and Pichler (1996), and Babbs and Nowman (1997).

If the term-structure model is Gaussian, the exact likelihood function is obtained directly from the Kalman filter algorithm. This ensures that parameter estimators are consistent, asymptotically normally distributed, and efficient. Unfortunately, these desirable properties do not carry over to the case where non-Gaussian, exponential-affine models are estimated by the Kalman filter method. The exact likelihood function is not available in closed form, but a QML estimator can be constructed from the first and second conditional moments of the state variables. Clearly, this involves some efficiency loss relative to maximum likelihood, but the QML estimator also loses the consistency property, as pointed out by Duan and Simonato (1995).

In the present paper, we further analyze the problems associated with estimating non-Gaussian models by the Kalman filter method. Based on this analysis we propose a modified QML estimator which ensures consistency. Basically, the modification works by ignoring the conditional heteroskedasticity properties and non-negativity restrictions of the state variables. We compare the finite sample properties of the two QML estimators in a small Monte Carlo study.

The paper is composed as follows. Section 2 reviews the finance theory of Gaussian term-structure models, and in section 3 we discuss the Kalman filter method for Gaussian models. This discussion forms the basis for our extensive analysis of non-Gaussian (exponential-affine) models in section 4. Finally, section 5 contains an empirical illustration of the Kalman filter method, using US swap market data, and section 6 concludes the paper.

2 A review of Gaussian term-structure models

The basic assumption of Gaussian term-structure models is that prices of all bonds are functions of a $m \times 1$ vector of state variables, X_t, whose dynamics are governed by the stochastic differential equation

$$dX_t = \mathcal{K}\left(\Theta - X_t\right) dt + \sigma dW_t, \tag{1}$$

where Θ is a $m \times 1$ vector, while $\mathcal{K}$ and σ are $m \times m$ matrices. In this setup, the matrix of volatility coefficients, σ, is diagonal, and the m Brownian motions in W_t are correlated, with correlation matrix ρ. The short rate, or instantaneous interest rate, r_t, is given by a linear function of X_t:

$$r_t = r(X_t) = \sum_{i=1}^{m} w_i X_{it} = w' X_t \tag{2}$$

Generally, the vector w consists of either zeros or ones.[4]

[4]Identification conditions for general term-structure models, such as (1), are discussed by, *inter alia*, Dai and Singleton (1997), Pang and Hodges (1995), and Babbs and Nowman (1997).

A special case of this model is the well-known one-factor model of Vasicek (1977), where r_t is driven by the univariate stochastic differential equation

$$dr_t = \kappa(\mu - r_t)dt + \sigma dW_t \tag{3}$$

Here, the parameter κ controls the degree of mean reversion towards the unconditional mean μ. The parameter σ is the volatility, or diffusion coefficient, of r_t.

It can be shown [Arnold (1974, p. 130)] that the solution of (1) is given by:

$$X_s = e^{-\mathcal{K}(s-t)}X_t + \int_t^s e^{-\mathcal{K}(s-v)}\mathcal{K}\Theta\, dv + \int_t^s e^{-\mathcal{K}(s-v)}\sigma dW_v, \tag{4}$$

where the matrix exponential function, $\exp(At)$, is formally defined as[5]

$$\exp(At) = \sum_{k=0}^{\infty} \frac{1}{k!}(At)^k\,.$$

If we condition on X_t, the only stochastic term in (4) is the last integral which can be shown to be normally distributed [Arnold (1974, p. 74)]. Hence, it follows that the conditional distribution of X_s is multivariate normal, with conditional mean

$$E[X_s \mid X_t] = e^{-\mathcal{K}(s-t)}X_t + \int_t^s e^{-\mathcal{K}(s-v)}\mathcal{K}\Theta\, dv$$

and conditional covariance matrix

$$\mathrm{Cov}[X_s \mid X_t] = \int_t^s e^{-\mathcal{K}(s-v)}\sigma\rho\sigma e^{-\mathcal{K}'(s-v)}\, dv.$$

We denote the current price of a zero-coupon bond, maturing at time T, by $P(t,T)$. A standard arbitrage argument, see e.g. Vasicek (1977), can be used to show that $P(t,T)$ satisfies the parabolic partial differential equation

$$\frac{1}{2}\mathrm{Tr}\left(\frac{\partial^2 P}{\partial X \partial X'}\sigma\rho\sigma\right) + \frac{\partial P}{\partial X'}\left[\mathcal{K}(\Theta - X) - \sigma\lambda\right] + \frac{\partial P}{\partial t} - r(X)P = 0, \tag{5}$$

subject to the boundary condition $P(T,T) = 1$. Note that the partial differential equation (5) is constructed from the parameters of (1), the stochastic process governing the state variables, and a vector λ containing the so-called market prices of risk, or risk premia.

It is straightforward to show that the bond price is given by:

$$P(t, t+\tau) = \exp\left[A(\tau) + B(\tau)'X_t\right], \tag{6}$$

where the functions $A(\tau)$ and $B(\tau)$ satisfy the joint system of ordinary differential equations (ODEs):

$$\frac{dB(\tau)}{d\tau} = -\mathcal{K}'B(\tau) - w \tag{7}$$

$$\begin{aligned}\frac{dA(\tau)}{d\tau} &= \frac{1}{2}\mathrm{Tr}\left[B(\tau)B(\tau)'\sigma\rho\sigma\right] + B(\tau)'\left[\mathcal{K}\Theta - \sigma\lambda\right] \\ &= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_i\sigma_j\rho_{ij}B_i(\tau)B_j(\tau) + \sum_{i=1}^{m}B_i(\tau)\left(\sum_{j=1}^{m}\mathcal{K}_{ij}\Theta_j - \lambda_i\sigma_i\right)\end{aligned} \tag{8}$$

[5]Techniques for computing the matrix exponential function, and integrals thereof, are discussed by Moler and Van Loan (1978), Van Loan (1978), and Golub and Van Loan (1989).

with boundary conditions $B(0) = 0$ and $A(0) = 0$. Langetieg (1980) derive the solution to (7) and (8) in the general case. Specifically, if $\mathcal{K}$ is non-singular, the solution to (7) is given by:

$$B(\tau) = \left(I_m - e^{-\mathcal{K}'\tau}\right)(\mathcal{K}')^{-1} w,$$

whereas $A(\tau)$ is given by a rather lengthy expression involving several integrals of the matrix exponential function.

The theoretical yield curve for a Gaussian model,

$$R(t, t+\tau) \;=\; \frac{-\log P(t, t+\tau)}{\tau} \;=\; -\left(\frac{A(\tau)}{\tau} + \frac{B(\tau)'}{\tau} X_t\right), \tag{9}$$

is an affine function of X_t, and by inverting this expression for m distinct maturities, $\tau_1, \ldots, \tau_m$, the bond prices in (6) can be expressed in terms of these yields. Such models are called yield-factor models by Duffie and Kan (1996). However, we argue in section 3 that the bond price equation with unobserved state variables (6) is the preferable starting point for the empirical analysis.

3 The empirical model for zero-coupon yields

For simplicity, we assume that the data set contains N zero-coupon yields with time-invariant maturities.[6] The observed yields at time t_k, for $k = 1, 2, \ldots, n$, are denoted by $R_k = (R_{1k}, \ldots, R_{Nk})$, where $R_{ik} = -\log P(t_k, t_k + \tau_i)/\tau_i$.

In general N is greater than m, the number of state variables, so if the observed zero-coupon yields correspond exactly to their theoretical counterparts, as defined by (9), the distribution of R_k is singular. Needless to say, this situation is unlikely to occur with real data. Moreover, due to exogenous factors such as non-synchronous trading, rounding of prices, and bid-ask spreads, we should expect (small) discrepancies from the theoretical term-structure model, and we commonly refer to these deviations as measurement errors.

The presence of measurement errors means that we have to distinguish between the theoretical term-structure model and the statistical model of the observed zero-coupon yields, henceforth called the empirical model. In our view, all N maturities are equally likely to be affected by measurement errors, and any selection of m "error-free" maturities is bound to be arbitrary. It is worth emphasizing, though, that the trade-off might be different for non-Gaussian models where the exact likelihood function cannot be computed with the linear Kalman filter.

3.1 The state space form of the empirical model

Under the usual assumption that measurement errors are additive and normally distributed, the so-called measurement equation of the state space model is given by:

$$R_k = d(\psi) + Z(\psi) X_k + \varepsilon_k \quad ; \; \varepsilon_k \sim N(0, H(\psi)), \tag{10}$$

[6] It is, of course, straightforward to generalize the Kalman filter method to cases where the number of yields and their maturities vary over time.

where the i'th row of the matrices d $(N \times 1)$ and Z $(N \times d)$ are given by $-A(\tau_i)/\tau_i$ and $-B(\tau_i)'/\tau_i$, respectively. The vector ψ contains all parameters of the state space model, and X_k is shorthand notation for X_{t_k}.

Since we are estimating a continuous-time model using discretely sampled observations, the transition equation is obtained from the exact discrete-time distribution of the state variables. In the Gaussian case, this distributions follows immediately from the SDE solution (4). Hence, as shown in section 2, the discrete-time distribution is a VAR(1) model with Gaussian innovations,

$$X_k = c_k(\psi) + \Phi_k(\psi) X_{k-1} + u_k, \tag{11}$$

where

$$\Phi_k(\psi) = e^{-\mathcal{K}(t_k - t_{k-1})} \tag{12}$$

$$c_k(\psi) = \int_{t_{k-1}}^{t_k} e^{-\mathcal{K}(t_k - v)} \mathcal{K}\Theta \, dv = \left[I - e^{-\mathcal{K}(t_k - t_{k-1})} \right] \Theta \tag{13}$$

$$V_k(\psi) \equiv \mathrm{Cov}(u_k) = \int_{t_{k-1}}^{t_k} e^{-\mathcal{K}(t_k - v)} \sigma\rho\sigma e^{-\mathcal{K}'(t_k - v)} \, dv \tag{14}$$

The second equality in (13) only holds if $\mathcal{K}$ is non-singular, corresponding to a strictly stationary stochastic process [Langetieg (1980)]. Finally, note that the VAR system matrices, c_k, Φ_k and V_k, are time-varying unless the observations are equally spaced in time.

3.2 The linear Kalman filter recursions

The state space model represented by (10)–(14) corresponds exactly to the standard linear, Gaussian state space model discussed in, e.g., Harvey (1989). Hence, the linear Kalman filter provides the exact likelihood function, and the estimators of the unobserved state variables correspond to conditional expectations given the observed data (zero-coupon yields), and so they are optimal in a MSE (mean square error) sense. To facilitate the discussion, let

$$\mathcal{F}_k = (R_1, R_2, \ldots, R_k) \tag{15}$$

denote the information set available at time t_k.

The Kalman filter algorithm consists of a sequence of prediction and update steps, and the likelihood function is obtained as a by-product of these recursions. First, the prediction step is given by

$$\hat{X}_{k|k-1} = E\left[X_k \,|\, \mathcal{F}_{k-1} \right] = c_k + \Phi_k \hat{X}_{k-1}, \tag{16}$$

with MSE matrix

$$\begin{aligned} \Sigma_{k|k-1} &= E\left[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})' \,|\, \mathcal{F}_{k-1} \right] \\ &= \Phi_k \Sigma_{k-1} \Phi_k' + V_k. \end{aligned} \tag{17}$$

Second, in the update step, the additional information provided by R_k is used to obtain a more precise estimator of X_k:

$$\hat{X}_k = E\left[X_k \mid \mathcal{F}_k\right] = \hat{X}_{k|k-1} + \Sigma_{k|k-1} Z' F_k^{-1} v_k, \tag{18}$$

with MSE matrix

$$\begin{aligned} \Sigma_k &= E\left[(X_k - \hat{X}_k)(X_k - \hat{X}_k)' \mid \mathcal{F}_k\right] \\ &= \Sigma_{k|k-1} - \Sigma_{k|k-1} Z' F_k^{-1} Z \Sigma_{k|k-1} \\ &= \left(\Sigma_{k|k-1}^{-1} + Z' H^{-1} Z\right)^{-1}, \end{aligned} \tag{19}$$

where

$$v_k = R_k - (d + Z\hat{X}_{k|k-1}) \tag{20}$$

$$F_k = Z\Sigma_{k|k-1} Z' + H \tag{21}$$

See Harvey (1989, ch. 3) for details of the derivations. The estimator of X_k in (18) is often called the *filtered* estimator.

3.3 The exact likelihood function

By the prediction error decomposition, see Harvey (1989), the exact log-likelihood function for a linear, Gaussian state space model is given by:

$$\log L(R_1, ..., R_n\,; \psi) = \sum_{k=1}^{n} -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|F_k| - \frac{1}{2} v_k' F_k^{-1} v_k \tag{22}$$

The vector of prediction errors,

$$v_k = R_k - E\left[R_k \mid \mathcal{F}_{k-1}\right],$$

and their covariance matrix,

$$F_k = \mathrm{Cov}\left[R_k \mid \mathcal{F}_{k-1}\right] = E\left[v_k v_k' \mid \mathcal{F}_{k-1}\right],$$

are calculated in the update step of the linear Kalman filter, cf. equations (20) and (21).

The most time-consuming part of evaluating the likelihood function is the calculation of the inverse and the determinant of the $N \times N$ matrix F_k. Given that N is often much greater than m, we can improve the computational efficiency by using the formulae

$$\begin{aligned} F_k^{-1} &= H^{-1} - H^{-1} Z \left(\Sigma_{k|k-1}^{-1} + Z' H^{-1} Z\right)^{-1} Z' H^{-1} \\ |F_k| &= |H| \cdot |\Sigma_{k|k-1}| \cdot |\Sigma_{k|k-1}^{-1} + Z' H^{-1} Z| \end{aligned}$$

[see Harvey (1989, p. 108)]. Note that by using these formulae, we obtain the MSE matrix of $\hat{X}_k$ as a by-product, cf. the third line in equation (19).

4 The Kalman filter and exponential-affine models

4.1 Review of general exponential-affine models

Duffie and Kan (1996) propose a general class of term-structure models that include Gaussian models as a special case. Under the original (true) probability measure, the state variables, X_t, are governed the process

$$dX_t = \mathcal{K}\left(\Theta - X_t\right) dt + C\sigma(X_t) dW_t, \tag{23}$$

where $\sigma(X_t)$ is a $m \times m$ diagonal matrix whose i'th diagonal element given by

$$[\sigma(X_t)]_{ii} = \sqrt{\alpha_i + \beta_i' X_t}. \tag{24}$$

In equation (23), the m univariate Brownian motions are independent, and the dependence structure between the innovations to X_t is captured by the $m \times m$ matrix C. We refer to Duffie and Kan (1996) for a thorough discussion of conditions ensuring that (23) is a well-defined process.

The short rate is specified as in (2), i.e. $r_t = w' X_t$ for some vector w, usually consisting of either zeros or ones. Finally, to complete the model specification, we make the following assumptions about the market prices of risk:

$$\lambda(X_t) = \sigma(X_t)\lambda \tag{25}$$

With the model specification (23)–(25), Duffie and Kan (1996) show that the price of a zero-coupon bond, $P(t,T)$, is given by

$$P(t, t+\tau) = \exp\left[A(\tau) + B(\tau)' X_t\right],$$

where $A(\tau)$ and $B(\tau)$ are defined by the following ODE system:

$$\frac{dB(\tau)}{d\tau} = \frac{1}{2}\sum_{i=1}^{m}[C'B(\tau)]_i^2\,\beta_i - \mathcal{K}'B(\tau) - \sum_{i=1}^{m}\lambda_i[C'B(\tau)]_i\,\beta_i - w \tag{26}$$

$$\frac{dA(\tau)}{d\tau} = \frac{1}{2}\sum_{i=1}^{m}[C'B(\tau)]_i^2\,\alpha_i + B(\tau)'\mathcal{K}\Theta - \sum_{i=1}^{m}\lambda_i[C'B(\tau)]_i\,\alpha_i, \tag{27}$$

Here, $[C'B(\tau)]_i$ refers to the i'th element of the $m \times 1$ vector $C'B(\tau)$. Finding a general closed-form solution to this ODE system does not seem to be possible, but many special cases (models) can be solved in closed form. However, if a closed-form solution for (26) and (27) is not available, the equations can always be solved numerically with an ODE integrator, e.g. the Runge-Kutta method.[7]

To summarize, as in the Gaussian case there a linear (affine) relationship between the state variables and the yield curve. Furthermore, the process (23) may remedy

[7] In the appendix, we outline a method for calculating analytical derivatives of $B(\tau)$ and $A(\tau)$ with respect to the model parameters, ψ, which is applicable when the ODE system is solved numerically.

some of the weaknesses of Gaussian models, namely the possibility of negative realizations for the short rate, and the absence of conditional heteroskedasticity in the short-rate process. For an illustration, consider the one-factor CIR model,

$$dr_t = \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t,$$

which is similar to the Vasicek specification (3), except that the short rate is now restricted to the non-negative part of the real line. In addition, the CIR volatility specification introduces a form of conditional heteroskedasticity which, in the context, is often referred to as a "level effect", see Andersen and Lund (1997).

4.2 Discrete-time distribution of the state variables

Except for certain special cases, e.g. the CIR model, the exact discrete-time distribution for (23) is not available in closed form. However, even if the full conditional distribution could be derived, we are unable to exploit it when using the Kalman filter approach, cf. the discussion in the next section. Therefore, we focus on the first and second conditional moments of X_t for which closed-form expressions are straightforward to obtain, as we show in the following.[8]

First, it follows from Ito's lemma, as well as properties of the matrix exponential function, that

$$d\left(e^{\mathcal{K}t}X_t\right) = e^{\mathcal{K}t}\mathcal{K}X_t\,dt + e^{\mathcal{K}t}\,dX_t. \tag{28}$$

Thus, if we substitute dX_t from (23) into (28), we get

$$d\left(e^{\mathcal{K}t}X_t\right) = e^{\mathcal{K}t}\mathcal{K}\Theta\,dt + e^{\mathcal{K}t}C\sigma(X_t)dW_t, \tag{29}$$

After integrating (29) from t to s, and premultiplying by $\exp(-\mathcal{K}s)$ on both sides of the equation, we obtain the following representation for X_s:

$$X_s = e^{-\mathcal{K}(s-t)}X_t + \int_t^s e^{-\mathcal{K}(s-v)}\mathcal{K}\Theta\,dv + \eta(t,s), \tag{30}$$

where

$$\eta(t,s) = \int_t^s e^{-\mathcal{K}(s-v)}C\sigma(X_v)dW_v. \tag{31}$$

This representation is quite similar to (4), but since X_v still appears in the stochastic integral (31), it is not the formal solution to the SDE (23). However, since stochastic (Ito) integrals are martingales, the conditional expectation of $\eta(t,s)$, given X_t, is zero.

Hence, the conditional mean of X_s follows directly from (30). Assuming further that $\mathcal{K}$ is non-singular,

$$E\left[X_s \mid X_t\right] = e^{-\mathcal{K}(s-t)}X_t + \left(I_m - e^{-\mathcal{K}(s-t)}\right)\Theta\,. \tag{32}$$

[8]The expressions below for the first and second conditional moments are also derived in Duan and Simonato (1995), de Jong (1996), and Fisher and Gilles (1996). Their approaches are quite similar to ours, although they differ in various respects.

The conditional covariance matrix is given by:

$$\begin{aligned}
\mathrm{Cov}\left[X_s \,|\, X_t\right] &= E\left[\eta(t,s)\eta(t,s)' \mid X_t\right] \\
&= E\left[\left(\int_t^s e^{-\mathcal{K}(s-v)} C\sigma(X_v)dW_v\right)\left(\int_t^s e^{-\mathcal{K}(s-u)} C\sigma(X_u)dW_u\right)' \,\middle|\, X_t\right] \\
&= E\left[\int_t^s e^{-\mathcal{K}(s-v)} C\sigma^2(X_v)C' e^{-\mathcal{K}'(s-v)}\,dv \,\middle|\, X_t\right] \\
&= \int_t^s e^{-\mathcal{K}(s-v)} C E\left[\sigma^2(X_v) \mid X_t\right] C' e^{-\mathcal{K}'(s-v)}\,dv, \qquad (33)
\end{aligned}$$

where

$$E\left[\sigma^2(X_v) \mid X_t\right]_{ii} = \alpha_i + \beta_i' E\left[X_v \mid X_t\right]$$

and off-diagonal elements are zero. The equality in third line in (33) is also a consequence of the martingale property of stochastic integrals, see Arnold (1974, p. 65–67) for a proof. Even though the final expression in (33) still needs to be worked out in closed form, an inspection of the result shows that the conditional covariance matrix for X_s is an affine function of X_t.

4.3 QML estimation of exponential-affine models: Part I

Generalizing the Kalman filter estimation technique to non-Gaussian, exponential-affine models turns out to be somewhat problematic. The structure of the measurement equation is identical to the Gaussian case,

$$R_k = d(\psi) + Z(\psi)X_k + \varepsilon_k, \qquad (34)$$

but the transition dynamics can no longer be represented by a Gaussian VAR(1) model. With a non-Gaussian distribution for the state variables, or the measurement errors for that matter, the linear Kalman filter is no longer optimal, and we do not obtain exact likelihood function. However, implementing the exact filtering recursions [see, e.g., Harvey (1989, pp. 162–165)] requires numerical integration which, except for one-factor models, is unlikely to be computationally feasible.[9]

Instead, Chen and Scott (1995) and Duan and Simonato (1995) propose a Gaussian QML approach based on the linear Kalman filter. The transition equation is obtained from the first and second conditional moments of the state variables:

$$X_k = c_k(\psi) + \Phi_k(\psi)X_{k-1} + u_k, \qquad (35)$$

where

$$\mathrm{Cov}\left[u_k \mid X_{k-1}\right] = V_k(X_{k-1}, \psi) \qquad (36)$$

[9] In a recent paper, Frühwirth-Schnatter and Geyer (1996) take a Bayesian approach to estimating the multi-factor CIR model, and coupled with modern Markov Chain Monte Carlo (MCMC) methods, they obtain optimal filtering for the state variables and likelihood inference for ψ. The MCMC approach is quite complex, though, and the authors do not report any computing times.

is obtained from (33). Note that there are two differences between this transition equation and the Gaussian counterpart (11). First, u_k is not normally distributed in (35), not even conditionally. Second, in (35) the conditional covariance matrix of u_k is an affine function of the lagged state vector, X_{k-1}, whereas it is constant for a Gaussian term-structure model.

The linear Kalman filter recursions are modified in two respects. First, since V_k is now state dependent, we evaluate V_k at $\hat{X}_{k-1}$ when computing $\Sigma_{k|k-1}$ in the prediction step (17). Second, in a non-Gaussian exponential-affine model, the state variables, or at least a subset thereof, are restricted to be non-negative. The linear update step (18) does not take such restrictions into account, and negative estimates of X_k could occur. If V_{k+1} is evaluated at a negative $\hat{X}_k$, the result may not be positive definite. Chen and Scott (1995) propose a solution to the problem that involves replacing negative estimates of $X_{k,j}$ with zero (assuming that the support of the j'th state variable is restricted).

Besides these modifications, we use exactly the same formulae for the Kalman filter recursions as in section 3.2. Moreover, we still obtain a sequence of prediction errors, v_k, which enter the QML estimation criterion,

$$Q_n(\psi) \;=\; \frac{1}{n}\sum_{k=1}^{n} l_k(\psi) \tag{37}$$

where

$$l_k(\psi) \;=\; -\log|F_k(\psi)| \;-\; v_k(\psi)'F_k^{-1}(\psi)v_k(\psi).$$

Apart from a constant and the scaling by $2/n$, $Q_n(\psi)$ is equivalent to the Gaussian log-likelihood function (22). The QML estimator, $\hat{\psi}_n$, is found by maximizing $Q_n(\psi)$ over the parameter space Ψ.

The asymptotic properties of QML for conditionally heteroskedastic models are discussed by Bollerslev and Wooldridge (1992). A sufficient condition for consistency of the QML estimator is given by:

$$\begin{aligned} E\left(v_k \mid \mathcal{F}_{k-1}\right) &= 0 \\ E\left(v_k v_k' \mid \mathcal{F}_{k-1}\right) &= F_k, \end{aligned}$$

where $\mathcal{F}_k$ is defined in (15). However, as Duan and Simonato (1995) point out, these conditions are not satisfied for the state space model represented by (34)–(36). If we condition on X_{k-1}, the mean and covariance matrix of R_k are both correctly specified, but the information set $\mathcal{F}_{k-1}$ only contains $\hat{X}_{k-1}$, not the unobserved X_{k-1}. Of course, QML would still be consistent if

$$\hat{X}_k = E[X_k \mid \mathcal{F}_k], \tag{38}$$

but (38) is a property of the exact filter, and, in general, it does not hold when applying the linear Kalman filter to a non-Gaussian state space model.[10] There is

[10] Censoring negative estimates of the state variables at zero is unlikely to improve the properties of the QML estimator, but some form of correction is, of course, necessary to ensure a positive definite covariance matrix. We suspect that the censoring leads to an upward bias in the estimated state variables which adversely affects the parameter estimates.

a subtle irony in the last argument. Since an exact filtering algorithm is deemed computationally infeasible, we rely instead on the linear Kalman filter and Gaussian QML — only to find that the QML estimator is inconsistent unless $\hat{X}_k$ is obtained from the exact filter.

Admittedly, the arguments for abandoning the Kalman filter (and QML) for non-Gaussian models are compelling. Besides being inconsistent, the QML estimator is, of course, inefficient relative to maximum likelihood. On the other hand, the Bayesian MCMC approach of Frühwirth-Schnatter and Geyer (1996), as well as the simulation-based EMM procedure used by Buraschi (1996) and Dai and Singleton (1997), provide consistent estimators which also attain full asymptotic efficiency. Alternatively, if m maturities are known to be observed without measurement error, the exact likelihood function is easy to compute, see Chen and Scott (1993) and Duffie and Singleton (1997).

However, these estimation procedures are either computationally involved (EMM, MCMC), or rely on more restrictive assumptions about the presence of measurement errors than the Kalman filter method. Furthermore, with a minor modification of the state space model (34) and (35), we actually obtain a consistent QML estimator. Altogether, this suggests that there is still some role for the Kalman filter method, the concerns about efficiency notwithstanding.

4.4 Properties of QML for non-Gaussian state space models

In this section, we consider a non-Gaussian state space model where the combination of the linear Kalman filter and QML does provide a consistent estimator. Apart from being interesting in its own right, this analysis serves two purposes.[11] First, it singles out the particular aspects of the non-Gaussian state space model (34) and (35) which cause QML to be inconsistent. Second, it motivates the subsequent modifications of the state space model that are necessary to ensure consistency.

The non-Gaussian state space model has the following general form:

$$y_k = d + ZX_k + \varepsilon_k \; ; \quad \varepsilon_k \sim D(0, H) \tag{39}$$

$$X_k = c + \Phi X_{k-1} + u_k \; ; \quad u_k \sim D(0, V), \tag{40}$$

where $D(0, \Sigma)$ denotes an arbitrary unconditional distribution with zero mean and covariance matrix Σ. The two error terms, ε_k and u_k are assumed to be cross-sectionally and serially uncorrelated,

$$\begin{aligned} E(\varepsilon_k u_s) &= 0 \quad \text{for all } k \text{ and } s, \\ E(\varepsilon_k \varepsilon_{k-j}) &= 0 \quad \text{for all } k, \text{ and } j > 0, \\ E(u_k u_{k-j}) &= 0 \quad \text{for all } k, \text{ and } j > 0. \end{aligned}$$

These restrictions do not rule out higher order dependencies, including conditional heteroskedasticity.

[11] Related studies that consider asymptotic properties of QML estimators for non-Gaussian state space models include Dunsmuir (1979), Watson (1989) and Ruiz (1994). The issue is also discussed by Hamilton (1994a, 1994b).

Because the state space model is non-Gaussian, the prediction and update steps of the linear Kalman filter no longer compute conditional expectations of the unobserved state variable, X_k, given the observed data. However, the filter is still optimal among all *linear* estimators of X_k, a property referred to as MMSLE (minimum mean square linear estimator), see Duncan and Horn (1972) and Harvey (1993). For the update step, this means that $\hat{X}_k$, given by equation (18), is the linear projection of X_k on $\mathcal{F}_k$, that is $R_k, \ldots, R_1$. Moreover, $\hat{X}_k$ is an unconditionally unbiased estimator of the unobserved X_k,

$$E\left[\hat{X}_k - X_k\right] = 0,$$

and Σ_k [equation (19)] is the unconditional covariance matrix (MSE matrix) of the estimation error $\hat{X}_k - X_k$. The same basic properties apply to the prediction step where $\hat{X}_{k|k-1}$ is the linear projection of X_k on $\mathcal{F}_{k-1}$, and so forth.

Consequently, since R_k is linear in X_k,

$$\hat{R}_k = d + Z\hat{X}_{k|k-1} \tag{41}$$

can be interpreted as the linear projection of R_k on $\mathcal{F}_{k-1}$. The prediction error, defined as $v_k = R_k - \hat{R}_k$, satisfies the following properties:

$$E(v_k) = 0, \tag{42}$$

$$E(v_k v_k') = Z\Sigma_{k|k-1}Z' + H = F_k, \tag{43}$$

$$E(v_k R_{k-j}) = 0, \quad \text{for all } j > 0. \tag{44}$$

The last result, (44), follows from $\hat{R}_k$ being the optimal linear estimator of R_k, given the variables in $\mathcal{F}_{k-1}$.

Having described the properties of the filtering algorithm (for a given value of the parameter vector ψ), we turn to the properties of the QML estimator, $\hat{\psi}_n$, which is obtained by maximizing (37). The following analysis is based on the general QML framework developed by White (1982, 1994).

4.4.1 Consistency of the QML estimator

We assume that the uniform law of large numbers (ULLN) holds for $Q_n(\psi)$, which means that

$$Q_n(\psi) \to \bar{Q}(\psi) \quad \text{a.s. and uniformly in } \Psi, \tag{45}$$

where

$$\bar{Q}(\psi) = \lim_{n\to\infty} E_0\left[Q_n(\psi)\right] = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} E_0\left[l_k(\psi)\right]. \tag{46}$$

Here, $E_0[\,\cdot\,]$ denotes the expectation under the probability measure corresponding to the true DGP. Let $\bar{\psi}$ denote the (global) maximizer of the non-stochastic function $\bar{Q}(\psi)$. As a direct consequence of (45), it follows that

$$\hat{\psi}_n \to \bar{\psi} \quad \text{a.s.},$$

see White (1994) for a proof. Thus, in order to prove consistency of our QML estimator, we must demonstrate that $\bar{\psi} = \psi_0$, the true value of the parameter vector, or, in other words, that ψ_0 maximizes $\bar{Q}(\psi)$.

The k'th term in (46) is given by:

$$\begin{aligned} E_0\left[l_k(\psi)\right] &= -E_0\left\{\log|F_k(\psi)| + v_k(\psi)'F_k^{-1}(\psi)v_k(\psi)\right\} \\ &= -\log|F_k(\psi)| - \mathrm{Tr}\left\{F_k^{-1}(\psi)\,E_0[v_k(\psi)v_k(\psi)']\right\} \end{aligned} \tag{47}$$

The simplification in second line of (47) follows because $F_k(\psi)$ is non-stochastic as it does not depend on the data. This is easily verified by inspection of the linear Kalman filter recursions. Next, we rewrite $v_k(\psi) = R_k - \hat{R}_k(\psi)$ as

$$v_k(\psi) = R_k - \hat{R}_k^0 + \left[\hat{R}_k^0 - \hat{R}_k(\psi)\right] = v_k^0 + \left[\hat{R}_k^0 - \hat{R}_k(\psi)\right], \tag{48}$$

where the superscript 0 refers to the "true" projection of R_k on $\mathcal{F}_{k-1}$, corresponding, hypothetically, to running the linear Kalman filter with $\psi = \psi_0$. The term in brackets in (48) is given by the difference of two linear predictors of R_k which means that it is a linear function of $R_{k-1}, \ldots, R_1$. By the property (44), v_k^0 is uncorrelated with R_{k-j} for all $j > 0$. Hence, when squaring the right hand side of (48), and taking expectations, the cross product vanishes,

$$\begin{aligned} E_0[v_k(\psi)v_k(\psi)'] &= E_0\left[v_k^0 v_k^{0\prime}\right] + E_0\left\{\left[\hat{R}_k^0 - \hat{R}_k(\psi)\right]\left[\hat{R}_k^0 - \hat{R}_k(\psi)\right]'\right\} \\ &\equiv F_k^0 + G_k(\psi). \end{aligned} \tag{49}$$

Substituting (49) into (47) yields

$$E_0\left[l_k(\psi)\right] = -\log|F_k(\psi)| - \mathrm{Tr}\left[F_k^{-1}(\psi)\,F_k^0\right] - \mathrm{Tr}\left[F_k^{-1}(\psi)\,G_k(\psi)\right] \tag{50}$$

The third term in (50) can also be written as

$$-\mathrm{Tr}\left[F_k^{-1}(\psi)\,G_k(\psi)\right] = -E_0\left\{\left[\hat{R}_k^0 - \hat{R}_k(\psi)\right]' F_k^{-1}(\psi)\left[\hat{R}_k^0 - \hat{R}_k(\psi)\right]\right\},$$

which — for any positive definite matrix $F_k(\psi)$ — is maximized when

$$\hat{R}_k(\psi) = \hat{R}_k^0. \tag{51}$$

Define the remaining part of (50) as

$$g\left(F_k(\psi)\right) = -\log|F_k(\psi)| - \mathrm{Tr}\left[F_k^{-1}(\psi)\,F_k^0\right].$$

It follows from a standard matrix inequality, see Johnson and Wichern (1992, p. 146), that the maximizer of the function $g(F_k(\psi))$ is given by

$$F_k(\psi) = F_k^0. \tag{52}$$

Conditions (51) and (52) can only hold for all k when $\psi = \psi_0$, and this proves that the QML estimator is consistent, that is $\hat{\psi}_n \to \psi_0$ a.s.

4.4.2 Asymptotic normality of the QML estimator

Under certain regularity conditions, the QML estimator, $\hat{\psi}_n$, is asymptotically normally distributed,

$$\sqrt{n}\left(\hat{\psi}_n - \psi_0\right) \Rightarrow N\left(0,\, A_0^{-1} B_0 A_0^{-1}\right),$$

where

$$A_0 = \lim_{n\to\infty} E_0\left[-\frac{1}{n}\sum_{k=1}^{n}\frac{\partial^2 l_k(\psi_0)}{\partial\psi\partial\psi'}\right] \tag{53}$$

$$B_0 = \lim_{n\to\infty} \mathrm{Cov}_0\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\partial l_k(\psi_0)}{\partial\psi}\right). \tag{54}$$

Since (53) and (54) are unknown, we need to find consistent estimators of these matrices in order to compute standard errors. As shown in, e.g., Harvey (1989), the first and second order derivatives of $l_k(\psi)$ are given by

$$\frac{\partial l_k(\psi)}{\partial\psi_i} = -\mathrm{Tr}\left\{\left[F_k^{-1}\frac{\partial F_k}{\partial\psi_i}\right]\left[I - F_k^{-1}v_k v_k'\right]\right\} - 2\frac{\partial v_k'}{\partial\psi_i}F_k^{-1}v_k, \tag{55}$$

and

$$\begin{aligned}\frac{\partial^2 l_k(\psi)}{\partial\psi_i\psi_j} &= \mathrm{Tr}\left\{\left[\frac{\partial}{\partial\psi_j}\left(F_k^{-1}\frac{\partial F_k}{\partial\psi_i}\right)\right]\left[I - F_k^{-1}v_k v_k'\right]\right\} - \\ &\mathrm{Tr}\left\{F_k^{-1}\frac{\partial F_k}{\partial\psi_i}F_k^{-1}\frac{\partial F_k}{\partial\psi_j}F_k^{-1}v_k v_k'\right\} + \mathrm{Tr}\left\{F_k^{-1}\frac{\partial F_k}{\partial\psi_i}F_k^{-1}\left[\frac{\partial v_k}{\partial\psi_j}v_k' + v_k\frac{\partial v_k'}{\partial\psi_j}\right]\right\} \\ &-2\frac{\partial^2 v_k}{\partial\psi_i\partial\psi_j}F_k^{-1}v_k - 2\frac{\partial v_k'}{\partial\psi_i}\frac{\partial F_k^{-1}}{\partial\psi_j}v_k - 2\frac{\partial v_k}{\partial\psi_i}F_k^{-1}\frac{\partial v_k}{\partial\psi_j},\end{aligned} \tag{56}$$

respectively. The first and second order derivatives of v_k are linear functions of R_{k-j}, for $j > 0$, and thus, by (44), uncorrelated with v_k. Since F_k is non-stochastic, this implies that the expectation of the third, fourth and fifth terms in (56) are zero. Moreover, since $F_k = E_0(v_k v_k')$, taking expectations over (56) yields

$$E_0\left[\frac{\partial^2 l_k(\psi)}{\partial\psi_i\partial\psi_j}\right] = -\mathrm{Tr}\left[F_k^{-1}\frac{\partial F_k}{\partial\psi_i}F_k^{-1}\frac{\partial F_k}{\partial\psi_j}\right] - 2E_0\left[\frac{\partial v_k}{\partial\psi_i}F_k^{-1}\frac{\partial v_k}{\partial\psi_j}\right].$$

Hence, a consistent estimator of A_0 is given by:

$$A_n(\hat{\psi}_n) = \frac{1}{n}\sum_{k=1}^{n}\left\{\mathrm{Tr}\left[F_k^{-1}\frac{\partial F_k}{\partial\psi_i}F_k^{-1}\frac{\partial F_k}{\partial\psi_j}\right] + 2\frac{\partial v_k}{\partial\psi_i}F_k^{-1}\frac{\partial v_k}{\partial\psi_j}\right\} \tag{57}$$

Note that (57) only requires first order derivatives of v_k and F_k which can be computed analytically by setting up special derivative recursions alongside the regular Kalman filter recursions, see Harvey (1989) for details.

If v_k is a martingale difference sequence with respect to $\mathcal{F}_{k-1}$, the covariance matrix B_0 simplifies to the well-known OPG (outer product of the gradient) formula.

However, this result is unlikely to hold in the present context. In general, we can only assume that v_k is serially uncorrelated, and this is not sufficient to ensure that the score, $s_k = \partial l_k / \partial \psi$, in (55) is serially uncorrelated.[12] Of course, we can still estimate B_0, for example by using the Newey-West (1987) estimator,

$$B_n(\hat{\psi_n}) = \frac{1}{n}\left\{ \sum_{k=1}^{n} s_k s_k' + \sum_{h=1}^{L} \sum_{k=h+1}^{n} \left(1 - \frac{h}{L+1}\right)\left(s_k s_{k-h}' + s_{k-h} s_k'\right) \right\},$$

or another covariance matrix estimator which is robust to autocorrelation, see Andrews (1991), Andrews and Monahan (1992), Gallant and White (1988), and Newey and West (1994).

4.5 QML estimation of exponential-affine models: Part II

By comparing the discussion in section 4.3 and 4.4, it becomes apparent that consistency follows under much weaker conditions in the latter section. Specifically, in section 4.3 the first and second conditional moments of R_k must be correctly specified, whereas consistency in section 4.4 only requires that the linear projection is correctly specified (loosely speaking). As further argued in section 4.3, the first condition is impossible to satisfy unless the filter is optimal, and this more or less rules out QML applications.

Thus, it is important to realize that the basic problem with the QML estimator in section 4.3 is not related to the non-Gaussian distribution for u_k in the transition equation (35), but the fact that the covariance matrix V_k in (36) depends on the lagged state vector, X_{k-1}. Of course, this dependence is a direct consequence of modeling the first and second conditional moments of X_k, instead of the full conditional distribution (which would make the filtering problem non-linear). However, for the non-Gaussian state space model (39) and (40), we only require that u_k is serially uncorrelated, not a martingale difference process, and that V_k is the *un*conditional covariance matrix of u_k which, by definition, is independent of X_{k-1}.

This leads us to propose two modification of the QML estimator, or rather the implementation of the linear Kalman filter, which are sufficient to ensure consistency for all exponential-affine models. First, the transition equation is modified such that V_k is the unconditional covariance matrix of u_k. Since the conditional mean of u_k (given X_{k-1}) is zero, the unconditional covariance matrix of u_k is given by:

$$\begin{aligned} V_k &= E\left\{\mathrm{Cov}[X_{t_k} \mid X_{t_{k-1}}]\right\} \\ &= \int_{t_{k-1}}^{t_k} e^{-\mathcal{K}(t_k - v)} C E\left[\sigma^2(X_v)\right] C' e^{-\mathcal{K}'(t_k - v)}\, dv, \end{aligned} \tag{58}$$

where

$$E\left[\sigma^2(X_v)\right]_{ii} = \alpha_i + \beta_i' E\left[X_v\right] = \alpha_i + \beta_i' \Theta$$

[12] If further restrictions are imposed on the error terms, ε_k and u_k, in the state space model (39) and (40), the OPG formula may still obtain, see Dunsmuir (1979), Watson (1989) and Ruiz (1994).

and off-diagonal elements are zero. Note that the expression for V_k in (58) is completely analogous to the Gaussian transition equation (11).

Second, we do not impose any non-negativity restrictions on the state variables as this would invalidate the MMSLE property of the linear Kalman filter which is instrumental in proving consistency of the QML estimator. Since V_k no longer depends on X_{k-1}, ignoring these restrictions does not lead to any problems with non-positive definite covariance matrices. Essentially, the joint effect of these modifications is to treat the term-structure models as a "Gaussian" model, albeit with a measurement equation that is derived from the general exponential-affine ODE system (26)–(27).

In summary, the modified QML estimator achieves consistency by ignoring certain aspects of the term-structure dynamics, notably the conditional heteroskedasticity properties. We emphasize, therefore, that the procedure is not a panacea for all our problems with QML and exponential-affine models. On the contrary, we should expect that the cost of eliminating the asymptotic bias (more precisely, ensuring consistency) is further loss of efficiency. With a finite sample, the trade-off between bias and efficiency might easily be in favor of the unmodified QML estimator from section 4.3. A complete investigation of this issue is outside the scope of the present paper, but in the next section we perform a small-scale Monte Carlo study where the two QML estimators are compared for a particular exponential-affine model.

4.6 Finite sample properties of the two QML estimators

For reasons of space, we only consider the following two-factor model:

$$\begin{aligned} dr_t &= \kappa_1(\mu_t - r_t)dt + \sigma_1\sqrt{r_t}dW_{1t} && (59)\\ d\mu_t &= \kappa_2(\theta - \mu_t)dt + \sigma_2\sqrt{\mu_t}dW_{2t}, && (60) \end{aligned}$$

where W_{1t} and W_{2t} are independent Brownian motions. The market prices of risk are specified as

$$\begin{aligned} \lambda_1(\cdot) &= (\lambda_1/\sigma_1)\sqrt{r_t}\\ \lambda_2(\cdot) &= (\lambda_2/\sigma_2)\sqrt{\mu_t}. \end{aligned}$$

In this model, the short rate reverts towards a stochastic mean, μ_t, which is often called a *central-tendency* factor. The specification (59)–(60) is inspired by similar models put forth in Jacobs and Jones (1986), Beaglehole and Tenney (1991), Jegadeesh and Pennacchi (1996) and Balduzzi et al. (1996).

It is readily verified that the model is exponential-affine, and the trivariate ODE system, defining $A(\tau)$, $B_1(\tau)$ and $B_2(\tau)$, is given by:

$$\begin{aligned} B_1'(\tau) &= -(\kappa_1 + \lambda_1)B_1(\tau) + \frac{1}{2}\sigma_1^2 B_1^2(\tau) - 1\\ B_2'(\tau) &= -(\kappa_2 + \lambda_2)B_2(\tau) + \kappa_1 B_1(\tau) + \frac{1}{2}\sigma_2^2 B_2^2(\tau)\\ A'(\tau) &= \kappa_2\theta B_2(\tau) \end{aligned}$$

By adapting the proof in Chen (1996, appendix B), a closed-form solution for this ODE system can be determined, but the formulae involve infinite-order series expansions of complicated functions. Therefore, we solve the ODE numerically, using the fourth-order Runge Kutta method.

Besides the specification of the term-structure model, the setup of the Monte Carlo study can be described as follows:

- The simulated data consist of 1000 time-series observations, at the weekly frequency, thus spanning about 20 years.
- On each date ($k = 1, \ldots, 1000$) we observe 10 different zero-coupon yields with maturities of 1 and 3 months, 1, 2, 3, 5, 7, 10, 15 and 30 years.
- The individual measurement errors, ε_{ik} ($i = 1, \ldots, 10$, $k = 1, \ldots, 1000$), are independently, normally distributed with a common standard deviation, σ_ε, of 20 basis points.
- To simulate the state variables, (r_{t_k}, μ_{t_k}), we use the Euler discretization of (59) and (60) with 50 subdivisions per week. See Kloeden and Platen (1992) for details of the Euler discretization.
- We perform 1000 replications, and apply both QML estimators (cf. sections 4.3 and 4.5) to the simulated data.

The results of the Monte Carlo study are displayed in Table 1. Overall, we get quite similar results for the two QML estimators, but there are some differences. The standard errors for the modified QML estimator (QML II) are uniformly greater than for QML I (section 4.3). Since the former estimator ignores the information about conditional heteroskedasticity in order to ensure consistency, this result is in accordance with our expectations. However, it is somewhat surprising that the finite sample biases are actually greater for QML II. The differences are small, but nonetheless significant in a paired t-test.

In summary, these results suggest that the asymptotic bias of QML I is quite small.[13] Moreover, even in quite large samples (20 years of weekly data, with 10 points on the yield curve), QML I dominates QML II both in terms of efficiency and bias. It is, of course, unclear whether this result carries over to other models, or sample sizes for that matter. In particular, note that the non-negativity restrictions on r_t and μ_t are unlikely to be binding for the present model. On the other hand, we should expect these restrictions to be binding at some observations for a multi-factor CIR model where the short rate is given by the sum of m non-negative processes. This issue, however, is left for future research.

[13] This is consistent with the Monte Carlo evidence reported in Duan and Simonato (1995), and Ball and Torous (1996) who investigate the properties of QML (QML I) for the one-factor CIR model.

5 An empirical application

The data for the empirical analysis are collected from the US swap market. The swap rates are converted to zero-coupon yields using the so-called "bootstrap" method with linear interpolation, see Dattatreya and Fabozzi (1995) for details of the computations.[14] Although the original data set contains daily data, our analysis is limited to the weekly frequency (Wednesdays) in order to avoid problems associated with missing data points, discreteness of swap-rate changes, and other microstructure issues, such as day-of-the-week effects and bid-ask bounces. This leaves a sample of 332 weekly observations, spanning 1/3/90 to 5/8/96. The data contain eight different maturities: 3 months, as well as 1–5, 7 and 10 years.

We estimate two term-structure models. The first model is the Gaussian "double decay" model suggested by Beaglehole and Tenney (1991),

$$\begin{aligned} dr_t &= \kappa_1(\mu_t - r_t)\,dt + \sigma_1 dW_{1t} \\ d\mu_t &= \kappa_2(\theta - \mu_t)\,dt + \sigma_2 dW_{2t}, \end{aligned}$$

where the Brownian motions are correlated, with correlation coefficient ρ, and the market prices of risk are constant, λ_1 and λ_2. The second model, which is given by equations (59) and (60), is similar to the "double decay" specification. Since the volatilities depend on r_t and μ_t, respectively, negative interest rates are ruled out, contrary to the "double decay" model. Furthermore, the short-rate dynamics of the second model are conditionally heteroskedastic via the level effect. On the other hand, the non-Gaussian central-tendency model restricts the Brownian motions W_{1t} and W_{2t} to be independent.

For both models, the covariance matrix of the measurement errors is parameterized with a single parameter σ_ε as in the Monte Carlo study, i.e. $H_k = \sigma_\varepsilon^2 I_N$. The non-Gaussian model, (59)–(60), is estimated using the QML estimator described in section 4.3 (QML I). We refrain from using the modified QML estimator, proposed in section 4.5, because of its rather disappointing performance in the Monte Carlo study.[15]

The estimation results can be found in Table 2. In general, the performance of the two models appears to be quite similar. First, the standard deviations of the measurement errors are 13 basis points in either case which clearly suggests that the two models provide the same level of fit to the data. Second, the shape of the yield curve is largely determined by the mean reversion coefficients under the risk-neutral measure. For the non-Gaussian model, the risk-neutral mean reversion coefficients are given by $\kappa_i + \lambda_i$, or 0.3222 and 0.1255, and these numbers are very close to the corresponding estimates of κ_1 and κ_2 in the "double decay" model. Third, the asymptotic levels of the yield curves, $\lim_{T\to\infty} R_i(t,T)$, differ by less than 4 basis points (7.56% vs. 7.60%), despite the fact that the maximum maturity used in the estimation is 10 years. Finally, in Figure 1 we show the theoretical yield curve implied

[14] I am grateful to Henrik Dahl and Alfred Berg Bank for supplying the US zero-coupon yields.

[15] The estimates obtained from QML II (the modified QML estimator) are very similar to those reported in Table 2, though.

by both models on the last day of the sample, 5/8/96. As can be seen from the figure, the two yield curves are virtually indistinguishable.[16]

6 Conclusion

In this paper we conduct a detailed investigation of the properties of QML for a general, non-Gaussian state space model, and we show that consistency basically follows under the same conditions which ensure that the Kalman filter is linearly optimal. This analysis, however, does not subsume state space model where the system matrices depend on the state vector. Consequently, if we estimate non-Gaussian term-structure models by matching the first and second conditional moments of the state variables in the transition equation, the resulting QML estimator is inconsistent. We propose a modified QML estimator which only uses first order properties of the state variable dynamics, thereby ignoring conditional heteroskedasticity properties. While this modification achieves consistency, the preliminary Monte Carlo evidence reported here suggests that the finite sample properties of the QML estimator may not improve.

[16]While the theoretical yield curve of Gaussian and non-Gaussian models may be very similar for realistic parameter values, prices of certain interest-rate derivative securities could still differ by a substantial amount, see Rogers (1996) for an illuminating discussion.

Appendix

In this appendix we provide a general method for computing analytical derivatives of the functions $A(\tau)$ and $B(\tau)$ which is applicable when the associated ODE system cannot be solved in closed form.

As shown in section 4.1, the $(m+1)$–dimensional ODE system is given by:

$$\frac{dB(\tau)}{d\tau} = \frac{1}{2}\sum_{i=1}^{m}\left[C'B(\tau)\right]_i^2\,\beta_i - \mathcal{K}'B(\tau) - \sum_{i=1}^{m}\lambda_i\left[C'B(\tau)\right]_i\,\beta_i - w \tag{61}$$

$$\frac{dA(\tau)}{d\tau} = \frac{1}{2}\sum_{i=1}^{m}\left[C'B(\tau)\right]_i^2\,\alpha_i + B(\tau)'\mathcal{K}\Theta - \sum_{i=1}^{m}\lambda_i\left[C'B(\tau)\right]_i\,\alpha_i, \tag{62}$$

with boundary conditions $B(0)=0$ and $A(0)=0$. The solution, $B(\tau)$ and $A(\tau)$, depends on ψ through the dependence of $\mathcal{K}$, Θ, C, α, β, λ and w on the parameter vector ψ. To keep the notation manageable, this dependence is implicitly understood in the following.

By construction, the ODE system (61) and (62) holds for all values of ψ. Hence, we may differentiate with respect to ψ_j on both sides of the equations, yielding

$$\begin{aligned}\frac{\partial^2 B(\tau)}{\partial\psi_j\partial\tau} &= \frac{1}{2}\sum_{i=1}^{m}\left[C'B(\tau)\right]_i^2\frac{\partial\beta_i}{\partial\psi_j} + \sum_{i=1}^{m}\left[C'B(\tau)\right]_i\left[\frac{\partial C'}{\partial\psi_j}B(\tau) + C'\frac{\partial B(\tau)}{\partial\psi_j}\right]_i\beta_i \\ &\quad -\frac{\partial\mathcal{K}'}{\partial\psi_j}B(\tau) - \mathcal{K}'\frac{\partial B(\tau)}{\partial\psi_j} - \sum_{i=1}^{m}\frac{\partial\lambda_i}{\partial\psi_j}\left[C'B(\tau)\right]_i\beta_i \\ &\quad -\sum_{i=1}^{m}\lambda_i\left[C'B(\tau)\right]_i\frac{\partial\beta_i}{\partial\psi_j} - \sum_{i=1}^{m}\lambda_i\left[\frac{\partial C'}{\partial\psi_j}B(\tau) + C'\frac{\partial B(\tau)}{\partial\psi_j}\right]_i\beta_i - \frac{\partial w}{\partial\psi_j}\end{aligned} \tag{63}$$

and

$$\begin{aligned}\frac{\partial^2 A(\tau)}{\partial\psi_j\partial\tau} &= \frac{1}{2}\sum_{i=1}^{m}\left[C'B(\tau)\right]_i^2\frac{\partial\alpha_i}{\partial\psi_j} + \sum_{i=1}^{m}\left[C'B(\tau)\right]_i\left[\frac{\partial C'}{\partial\psi_j}B(\tau) + C'\frac{\partial B(\tau)}{\partial\psi_j}\right]_i\alpha_i \\ &\quad +\frac{\partial B(\tau)'}{\partial\psi_j}\mathcal{K}\Theta + B(\tau)'\left(\frac{\partial\mathcal{K}'}{\partial\psi_j}\Theta + \mathcal{K}'\frac{\partial\Theta}{\partial\psi_j}\right) - \sum_{i=1}^{m}\frac{\partial\lambda_i}{\partial\psi_j}\left[C'B(\tau)\right]_i\alpha_i \\ &\quad -\sum_{i=1}^{m}\lambda_i\left[C'B(\tau)\right]_i\frac{\partial\alpha_i}{\partial\psi_j} - \sum_{i=1}^{m}\lambda_i\left[\frac{\partial C'}{\partial\psi_j}B(\tau) + C'\frac{\partial B(\tau)}{\partial\psi_j}\right]_i\alpha_i\,.\end{aligned} \tag{64}$$

Since $B(0)=0$ and $A(0)=0$ hold for all ψ, the boundary conditions in (63) and (64) are given as

$$\frac{\partial B(0)}{\partial\psi_j} = \frac{\partial A(0)}{\partial\psi_j} = 0.$$

By combining (63)–(64) with (61)–(62), we obtain a $2(m+1)$–dimensional ODE system which jointly defines $B(\tau)$ and $A(\tau)$, as well as $\partial B(\tau)/\partial\psi_j$ and $\partial A(\tau)/\partial\psi_j$. This system must be solved for each j, for example using the fourth-order Runge-Kutta method.

Alternatively, we may perturb the j'th element of ψ by a small number, say 10^{-6}, resolve the ODE system (61)–(62), and compute the derivatives of $B(\tau)$ and $A(\tau)$ by finite differences. Obviously, relying on numerical derivatives requires fewer algebraic manipulations, not to mention less computer programming, but, contrary to the approach outlined above, we have no direct way of controlling the accuracy of the resulting derivatives.

References

Aït-Sahalia, Y. (1996a), "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560.

Aït-Sahalia, Y. (1996b), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

Andersen, T.G. and J. Lund (1997), "Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77, 343–377.

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

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Andrews, D.W.K. and J.C. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–967.

Arnold, L. (1974), *Stochastic Differential Equations: Theory and Applications,* Wiley, New York.

Babbs, S. and K.B. Nowman (1997), "Kalman Filtering of Generalized Vasicek Term Structure Models," Manuscript, First National Bank of Chicago, London.

Balduzzi, P., S.R. Das and S. Foresi (1996), "The Central Tendency: A Second Factor in Bond Yields," Manuscript, Stern School of Business, New York University, forthcoming *Review of Economics and Statistics.*

Ball, C.A. and W.N. Torous (1996), "Unit Roots and the Estimation of Interest Rate Dynamics," *Journal of Empirical Finance*, 3, 215–238.

Beaglehole, D.R. and M.S. Tenney (1991), "General Solutions of Some Interest Rate-Contingent Claim Pricing Equations," *Journal of Fixed Income*, 1, Sept., 69–83.

Bliss, R.R. (1996), "Testing Term Structure Estimation Methods," Manuscript, Federal Reserve Bank of Atlanta, forthcoming in *Advances of Futures and Options Research.*

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

Brennan, M.J. and E.S. Schwartz (1979), "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance*, 3, 133–155.

Brown, S.J. and P.H. Dybvig (1986), "The Empirical Implications of the Cox, Ingersoll and Ross Theory of the Term Structure of Interest Rates," *Journal of Finance*, 41, 617–632.

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

Chan, K.C., G.A. Karolyi, F.A. Longstaff and A.B. Sanders (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*, Lecture Notes in Economics and Mathematical Systems, Vol. 435, Springer-Verlag, New York.

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

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

Conley, T., L.P. Hansen, E.J.G. Luttmer and J. Scheinkman (1997), "Short-Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies*, 10, 525–577.

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

Dai, Q. and K.J. Singleton (1997), "Specification Analysis of Affine Term Structure Models," Manuscript, Graduate School of Business, Stanford University.

Dattatreya, R.E. and F.J. Fabozzi (1995), "The Risk-Point Method for Measuring and Controlling Yield Curve Risk", *Financial Analysts Journal*, July-August, 45–54.

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

de Jong, F. (1996), "Time-Series and Cross-Section Information in Term Structure Models," Manuscript, Tilburg University.

Duan, J.C. and J.G. Simonato (1995), "Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter," Manuscript, McGill University.

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

Duffie, D. and K.J. Singleton (1997), "An Econometric Model of the Term Structure of Interest Rate Swap Yields," *Journal of Finance*, 52, 1287–1321.

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

Dunsmuir, W. (1979), "A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and Its Application to Models for a Signal Observed with Noise," *Annals of Statistics*, 7, 490–506.

Fisher, M. and C. Gilles (1996), "Estimating Exponential-Affine Models of the Term Structure," Manuscript, Board of Governors of the Federal Reserve System, Washington, DC.

Frühwirth-Schnatter, S. and A.L.J. Geyer (1996), "Bayesian Estimation of Econometric Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure of Interest Rates via MCMC Methods," Manuscript, University of Economics and Business Administration, Vienna.

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

Geyer, A.L.J. and S. Pichler (1996), "A State-Space Approach to Estimate the Term Structure of Interest Rates: Some Empirical Evidence," Manuscript, University of Economics and Business Administration, Vienna.

Gibbons, M.R. and K. Ramaswamy (1993), "A Test of the Cox, Ingersoll and Ross Model of the Term Structure," *Review of Financial Studies*, 6, 619–658.

Golub, G.H. and C.F. Van Loan (1989), *Matrix Computations, 2nd ed.*, John Hopkins University Press, Baltimore.

Gong, F.F. and E.M. Remolona (1996a), "Two Factors Along the Yield Curve," Manuscript, Capital Markets Function, Federal Reserve Bank of New York, forthcoming in *Manchester School*.

Gong, F.F. and E.M. Remolona (1996b), "A Three-Factor Econometric Model of the U.S. Term Structure," Manuscript, Capital Markets Function, Federal Reserve Bank of New York.

Hamilton, J.D. (1994a), *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hamilton, J.D. (1994b), "State-Space Models," chapter 50 in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics, Volume IV*, North-Holland, Amsterdam.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

Harvey, A.C. (1993), *Time Series Models, 2nd Edition*, Prentice-Hall, UK.

Honoré, P. (1997), "Maximum Likelihood Estimation of Non-Linear Continuous-Time Models," Manuscript, Department of Finance, Aarhus School of Business.

Honoré, P. (1998), "Panel-Data Estimation of Non-Linear Term-Structure Models," Manuscript in preparation, Department of Finance, Aarhus School of Business.

Jacobs, R.L. and R.A. Jones (1986), "A Two Factor Latent Variable Model of the Term Structure of Interest Rates," Manuscript, Simon Fraser University.

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

Johnson, R.A. and D.W. Wichern (1992), *Applied Multivariate Statistical Analysis, 3rd Edition*, Prentice-Hall, Englewood Cliffs, NJ.

Kloeden, P.E. and E. Platen (1992), *Numerical Solutions of Stochastic Differential Equations*, Springer-Verlag, New York.

Langetieg, T.C. (1980), "A Multivariate Model of the Term Structure," *Journal of Finance*, 35, 71–97.

Longstaff, F.A. and E.S. Schwartz (1992), "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, 47, 1259–1282.

Moler, C. and C.F. Van Loan (1978), "Nineteen Dubious Ways to Compute the Exponential of a Matrix," *SIAM Review*, 20, 801–836.

Newey, W. and K.D. West (1987), "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Newey, W. and K.D. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

Nowman, K.B. (1997), "Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates," *Journal of Finance*, 52, 1695–1706.

Pang, K. and S. Hodges (1995), "Non-Negative Affine Yield Models of the Term Structure," Manuscript, Financial Options Research Centre, University of Warwick.

Pearson, N.D. and T.S. Sun (1994), "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll and Ross Model," *Journal of Finance*, 49, 1279–1304.

Pennacchi, G.G. (1991), "Identifying the Dynamics of Real Interest Rates and Inflation: Evidence Using Survey Data," *Review of Financial Studies*, 4, 53–86.

Rogers, L.C.G. (1996), "Gaussian Errors," *Risk*, 9, January, 42–45.

Ruiz, E. (1994), "Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, 63, 289–306.

Santa-Clara, P. (1995), "Markovian Arbitrage-Free Models of the Term-Structure of Interest Rates," Manuscript, Anderson Graduate School of Management, University of California, Los Angeles.

Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Risk," Manuscript, Haas School of Business, University of California, Berkeley, forthcoming *Journal of Finance*, December 1997.

Tauchen, G.E. (1997), "New Minimum Chi-Square Methods in Empirical Finance," in D.M. Kreps and K.F. Wallis, eds., *Advances in Economic Theory and Econometrics, Seventh World Congress, Volume III*, Cambridge University Press, New York, NY.

Torous, W.N. and C.A. Ball (1995), "Regime Shifts in Short-Term Riskless Interest Rates", Manuscript, Owen Graduate School of Business, Vanderbilt University.

Van Loan, C.F. (1978), "Computing Integrals Involving the Matrix Exponential," *IEEE Transactions on Automatic Control*, 23, 395–404.

Vasicek, O.A. (1977), "An Equilibrium Characterization of the Term Structure of Interest Rates," *Journal of Financial Economics*, 5, 177-188.

Watson, M.W. (1989), "Recursive Solution Methods for Dynamic Linear Rational Expectations Models," *Journal of Econometrics*, 41, 65–89.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, New York, NY.

Zheng, C.C. (1993), "Testing Term Structure Models with Unspecified Dynamic Factors," Manuscript, First National Bank of Chicago, Chicago, IL.

Table 1:
Results of Monte Carlo study
Non-Gaussian Central-Tendency model

		QML I	QML II
Parameter	True value	Mean [Std.Err]	Mean [Std.Err]
σ_ε	0.2000	0.1999 [0.0016]	0.1999 [0.0016]
κ_1	1.2000	1.2041 [0.1027]	1.2120 [0.1061]
κ_2	0.2000	0.2033 [0.0215]	0.2038 [0.0218]
θ	0.0700	0.0700 [0.0092]	0.0694 [0.0094]
σ_1	0.1200	0.1199 [0.0033]	0.1208 [0.0047]
σ_2	0.0300	0.0299 [0.0019]	0.0299 [0.0021]
λ_1	-0.1000	-0.1038 [0.1003]	-0.1117 [0.1038]
λ_2	-0.0500	-0.0529 [0.0212]	-0.0536 [0.0216]

Notes: QML I and QML II refer to, respectively, the QML estimators described in sections 4.3 and 4.5. The latter is also called the "modified" QML estimator in the text. In each case, we report the sample mean and the standard error (in brackets) of the parameter estimates from 1000 Monte Carlo replications.

Table 2:
Parameter estimates for central-tendency models
Data: zero-coupon yields from US swap rates
Sample period: 1/1/90 – 5/8/96 (weekly)

	Beaglehole-Tenney Model (Gaussian)		Central-Tendency Model (non-Gaussian)	
Parameter	Estimate	Std.Err	Estimate	Std.Err
σ_ε	0.1279	(0.0046)	0.1284	(0.0047)
κ_1	0.3354	(0.0316)	0.5686	(0.0760)
κ_2	0.1286	(0.0162)	0.0966	(0.0124)
θ	0.0649	(0.0173)	0.0627	(0.0111)
σ_1	0.0083	(0.0006)	0.0397	(0.0025)
σ_2	0.0174	(0.0014)	0.0475	(0.0045)
ρ	0.4152	(0.0843)	—	—
λ_1	-1.6370	(0.3225)	-0.2464	(0.0731)
λ_2	0.1428	(0.1073)	0.0289	(0.0137)

Figure 1
Theoretical Yield Curves (May 8, 1996)