Bayesian Estimation of Econometric Multi-Factor Cox-Ingersoll-Ross-Models of the Term Structure of Interest Rates Via MCMC Methods *

Sylvia Frühwirth-Schnatter
Department of Statistics

Alois L.J. Geyer†
Department of Operations Research

Vienna University of Economics and Business Administration

Abstract

We present a Markov-chain-Monte-Carlo technique for the *exact* estimation of multi-factor models of the term structure of interest rates. We apply this method to the Cox-Ingersoll-Ross-model which provides an interesting case because of the highly non-normal structure of the underlying state space model. Our technique is based on hybrid "Metropolis within Gibbs"-sampling and makes explicit use of the structure of the underlying economic model when constructing the proposal densities for the Metropolis-Hastings algorithm. In an empirical study using US interest rates we find that the difference between approximate quasi maximum likelihood estimation and exact estimation may be substantial.

Key words: Markov-chain-Monte-Carlo methods, term structure models
JEL classification: C15, E43

*We would like to thank Stefan Pichler and Walter Schachermayer for helpful contributions and discussions.

†corresponding author's address: Department of Operations Research, University of Economics and Business Administration Vienna, Augasse 2–6, A-1090 Wien; phone: +43.1.31336.4559; fax: +43.1.31336.708; e-mail: geyer@wu-wien.ac.at.

1 Introduction

The topic of this paper is the estimation of multi-factor models of the term structure of interest rates from a multivariate time series of yields observed at discrete points in time. The underlying unobservable stochastic process of the factors is assumed to follow the multi-factor Cox-Ingersoll-Ross-model (Cox *et al*, 1985), CIR-model hereafter. Such a model may be called an "econometric" term structure model because it introduces observation errors which are not present in theoretical term structure models. Estimation of econometric term structure models has become a central research objective during the last few years. The CIR-model is studied by Chen and Scott (1995) and Geyer and Pichler (1995), Lund (1994) estimates the Vasicek-model (Vasicek, 1977), and Duan and Simonato (1995) obtain estimates for the CIR- and the Vasicek model. The basic idea underlying these papers is to rewrite the econometric term structure model as a state space model and to utilize the framework of the Kalman filtering (reviewed e.g. in Harvey, 1989) to estimate the parameters of the term structure model as well as the unobservable factors.

Econometric term structure models based on the Vasicek-model lead to a Gaussian linear state space model and Kalman filtering provides an optimal, easily implementable estimation methodology. The simplicity of Kalman filtering, however, is lost for econometric term structure models based on the CIR-model since the transition density of the corresponding state space model is the product of non-central χ^2-distributions. By substituting this complicated transition density by an approximate normal density, Chen and Scott (1995) perform *approximate* Kalman filtering in combination with quasi maximum likelihood (QML) estimation of the parameters of the CIR-model. The merit of this approach is its simplicity, the approximation error introduced by the normal approximation, however, is unknown. Till recently *exact* estimation of non-normal state space models seemed infeasible. With the advent of new estimation methods such as Markov-chain-Monte-Carlo methods in the early 90's (Gelfand and Smith, 1990), exact estimation tools for non-normal and/or non-linear state space models became available. Among the Markov-chain-Monte-Carlo algorithms designed for specific state space models we would like to mention Carlin *et al.* (1992), Frühwirth-Schnatter (1994), Carter and Kohn (1994), Shephard (1994), De Jong and Shephard (1995) and Jacquier *et al.* (1995).

The main contribution of the present paper is to suggest *exact* estimation of econometric term structure models based on the CIR-model by means of a Markov-

chain-Monte-Carlo technique. Our Markov-chain-Monte-Carlo technique is based on hybrid "Metropolis within Gibbs"-sampling (Tierney, 1995) and makes explicit use of the structure of the CIR-model when constructing the proposal densities for the Metropolis-Hastings algorithm. It should be pointed out that the normal approximation suggested by Chen and Scott (1995) plays a central role for the derivation of the proposal densities. A detailed case study for US interest rates demonstrates that the difference between approximate QML estimation and exact estimation may be substantial.

The paper is organized as follows. In section 2 we review the state space formulation of the multi-factor CIR-model. In Section 3 we suggest Bayesian estimation of the parameters and factors of the CIR-model via Markov-chain-Monte-Carlo methods. Section 4 contains a detailed discussion of empirical results for US interest rates. Section 5 concludes the paper.

2 State Space Formulation of the Multi-Factor Cox-Ingersoll-Ross-Model

The Cox-Ingersoll-Ross-model (Cox *et al.*, 1985) frequently is presented as a one-factor-model, but already Cox *et al.* (1985) show how to incorporate multiple factors. The nominal instantaneous interest rate is assumed to be the sum of K state variables (factors) $r_{t,j}$:

$$i_t = \sum_{j=1}^{K} r_{t,j},$$

where the state variables $r_{t,j}$ are assumed to be independently generated by a square root process:

$$d\, r_{t,j} = \kappa_j(\beta_j - r_{t,j})dt + \sigma_j\sqrt{r_{t,j}} \cdot dz_{t,j}, \quad j = 1, \ldots, K, \tag{1}$$

where $z_{t,j}$ is a Wiener process. β_j is the long-term mean. $r_{t,j}$ is pulled towards β_j at a rate governed by the speed of adjustment coefficient κ_j.

Based on Cox *et al.* (1985), Chen and Scott (1995) derive the solution for the nominal price $P_t(T)$ at time t for a pure discount bond with face value 1 maturing at time $t + T$ as follows:

$$P_t(T) = A_1(T) \cdots A_K(T) \cdot \exp\left(-B_1(T)r_{t,1} - \ldots - B_K(T)r_{t,K}\right), \tag{2}$$

where

$$A_j(T) = C_j(T)^{\frac{2\kappa_j\beta_j}{\sigma_j^2}}, \quad C_j(T) = \frac{2\phi_j \cdot \exp\left(T/2 \cdot (\kappa_j + \lambda_j + \phi_j)\right)}{2\phi_j + (\kappa_j + \lambda_j + \phi_j)(\exp(\phi_j T) - 1)}, \tag{3}$$

$$B_j(T) = \frac{2(\exp(\phi_j T) - 1)}{2\phi_j + (\kappa_j + \lambda_j + \phi_j)(\exp\left(\phi_j T\right) - 1)}, \tag{4}$$

with $\phi_j = \sqrt{(\kappa_j + \lambda_j)^2 + 2\sigma_j^2}$. Each state variable is associated with a parameter λ_j which is negatively related to the risk premium. The yield to maturity at time t of a pure discount bond which matures at time $t + T$ is defined as:

$$Y_t(T) = -\frac{\ln P_t(T)}{T}, \tag{5}$$

which is a linear function of the state variables $r_{t,1}, \ldots, r_{t,K}$.

Let $y_t = (y_{t,1}, \ldots, y_{t,n_t})' = (Y(T_{t,1}), \ldots, Y(T_{t,n_t}))'$ be the n_t-dimensional vector of yields observed at time t, where n_t is the number of observed yields which need not be the same at each date. Let $T_{t,i}, i = 1, \ldots, n_t$, be the associated times to maturity. To estimate the unobservable state variables from yields observed at discrete time intervals, Chen and Scott (1995) and – independently – Geyer and Pichler (1995) suggest to use a state space formulation of the CIR-model. However, none of the above cited studies uses the exact state space formulation. The exact state space formulation with state variable $x_t = (r_{t,1}, \ldots, r_{t,K})'$ and observation vector y_t is given by the following model assumptions:

I $x_0, x_1, x_2, \ldots, x_t$ is a Markov process with $x_0 \sim p(x_0)$ and $x_t|x_{t-1} \sim p(x_t|x_{t-1})$. $p(x_0)$ is called the prior and $p(x_t|x_{t-1})$ is called the transition density.

II $y_1, y_2, y_3, \ldots, y_t$ are conditionally independent given $x_0, x_1, x_2, \ldots, x_t$ and y_t is independent of $x_s, s \neq t$ given x_t with $y_t|x_t \sim p(y_t|x_t)$. $p(y_t|x_t)$ is called the observation density.

For the CIR-model the exact transition densities are known to be the product of K non-central χ^2-densities (Cox *et al.*, 1985; Chen and Scott, 1993):

$$p(x_t|x_{t-1}) = \prod_{j=1}^{K} p(r_{t,j}|r_{t-1,j}), \tag{6}$$

$$p(r_{t,j}|r_{t-1,j}) = c_j e^{-c_j(r_{t,j} + e^{-\kappa_j \Delta t} r_{t-1,j})} \cdot \left(\frac{r_{t,j}}{e^{-\kappa_j \Delta t} r_{t-1,j}}\right)^{\frac{q_j}{2}} \cdot I_{q_j}(2c_j\sqrt{r_{t,j} e^{-\kappa_j \Delta t} r_{t-1,j}}), \tag{7}$$

$$c_j = \frac{2\kappa_j}{\sigma_j^2(1 - e^{-\kappa_j \Delta t})}, \qquad q_j = \frac{2\kappa_j\beta_j}{\sigma_j^2} - 1, \tag{8}$$

where $I_{q_j}(\cdot)$ is the modified Bessel function of the first kind of order q_j.

The observation density $p(y_t|x_t)$ is based on the linear relationship (5) between yields and the state variable x_t. The distribution of observed yields given the state variable $y_t|x_t$ is derived from the following measurement equation:

$$y_t = a_t + B_t \cdot x_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, V_t), \quad V_t = \mathrm{Diag}(V_{t,11} \cdots V_{t,n_t n_t}), \tag{9}$$

where a_t is a n_t-dimensional vector and B_t is a $n_t \times K$ matrix. Both quantities are derived from (2) – (5):

$$a_{t,i} = -\sum_{j=1}^{K} \frac{\ln A_j(T_{t,i})}{T_{t,i}}, \ 1 \le i \le n_t, \quad B_{t,ij} = \frac{B_j(T_{t,i})}{T_{t,i}}, \ 1 \le i \le n_t, 1 \le j \le K. \tag{10}$$

ε_t is a n_t-dimensional random vector reflecting pricing errors caused by market imperfections. Geyer and Pichler (1995) assume that the errors of each maturity have the same variance σ_h^2, whereas Chen and Scott (1995), Duan and Simonato (1995) and Lund (1994) assume different variances $\sigma_{\varepsilon_i}^2$ for different maturities. The following choice combines individual variances for n maturities which are identical for each t, and a common variance for the remaining ones:

$$V_{t,ii} = \begin{cases} \sigma_{\varepsilon_i}^2, & 1 \le i \le n, \\ \sigma_h^2, & n+1 \le i \le n_t. \end{cases} \tag{11}$$

The observation density $p(y_t|x_t)$ is then the product of n_t normal densities

$$p(y_t|x_t) = \prod_{i=1}^{n_t} p_N(y_{t,i}; \hat{y}_{t,i}, V_{t,ii}), \qquad \hat{y}_{t,i} = \sum_{j=1}^{K} \left(-\frac{\ln A_j(T_{t,i})}{T_{t,i}} + \frac{B_j(T_{t,i})}{T_{t,i}} r_{t,j} \right). \tag{12}$$

To complete the state space formulation the prior $p(x_0)$ has to be chosen. We assume that $r_{0,1}, \ldots, r_{0,K}$ are independent apriori:

$$p(x_0) \quad = \quad \prod_{j=1}^{K} p(r_{0,j}). \tag{13}$$

For the distribution of the individual components $r_{0,j}$ one may either assume a vague normal prior $p_N(r_{0,j}; E(r_{0,j}), V(r_{0,j}))$ for each state variable $r_{0,j}$ with $V(r_{0,j})$ being large, or assume – like Chen and Scott (1995) – that each state variable $r_{0,j}$ is distributed according to the stationary gamma distribution (see Cox *et al.*, 1985):

$$p(r_{0,j}) = p_G(r_{0,j}; q_j+1, \frac{2\kappa_j}{\sigma_j^2}) = r_{0,j}^{q_j} e^{-\frac{2\kappa_j}{\sigma_j^2} r_{0,j}} \left(\frac{2\kappa_j}{\sigma_j^2}\right)^{q_j} \frac{1}{\Gamma(q_j+1)}, \quad (14)$$

$$E(r_{0,j}) = \beta_j, \qquad V(r_{0,j}) = \beta_j \frac{\sigma_j^2}{2\kappa_j}. \quad (15)$$

3 Bayesian Estimation of the CIR-Model via Markov-chain-Monte-Carlo Methods

3.1 Bayesian Estimation of state space models via Markov-chain-Monte-Carlo Methods

Estimation of the unobservable factors x_t within the exact state space model is not simple because of the non-normal transition density $p(x_t|x_{t-1})$. Furthermore, the multi-factor CIR-Model depends on the D-dimensional model parameter θ, $D = 4K + n + 1$, $\theta = (\kappa_1, \ldots, \kappa_K, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2, \lambda_1^2, \ldots, \lambda_K^2, \sigma_{\varepsilon_1}^2, \ldots, \sigma_{\varepsilon_n}^2, \sigma_h^2)^T$, which, too, has to be estimated from the observations. Estimation with approximate Kalman filtering in combination with QML estimation of the model parameters can be carried out by substituting the exact transition density by a normal density:

$$x_t|x_{t-1} \sim \mathrm{N}(F_t x_{t-1} + u_t, Q_t) \quad (16)$$

Chen and Scott (1995) determined F_t, u_t and Q_t in such a way that the first two moments of the approximate normal and the exact transition density are equal:

$$F_t = \mathrm{Diag}\left(\begin{array}{ccc} e^{-\kappa_1 \Delta t} & \cdots & e^{-\kappa_K \Delta t} \end{array} \right) \quad (17)$$

$$u_t = \left(\begin{array}{ccc} (1-e^{-\kappa_1 \Delta t})\beta_1 & \cdots & (1-e^{-\kappa_K \Delta t})\beta_K \end{array} \right)' \quad (18)$$

$$Q_t = \mathrm{Diag}\left(\begin{array}{ccc} Q_{t,11} & \cdots & Q_{t,KK} \end{array} \right),$$

$$Q_{t,jj} = \sigma_j^2 \frac{1-e^{-\kappa_j \Delta t}}{\kappa_j} \left((1-e^{-\kappa_j \Delta t}) \frac{\beta_j}{2} + e^{-\kappa_j \Delta t} r_{t-1,j} \right), \; j = 1, \ldots, K. \quad (19)$$

In this section we are going to discuss exact joint estimation of the whole state process $x^N = (x_0, \ldots, x_N)$ and the model parameters θ from the whole batch of data $y^N = (y_1, \ldots, y_N)$. The Bayesian solution to the estimation problem is to derive the joint posterior distribution $p(x^N, \theta|y^N)$ of x^N and θ given y^N. This "smoothing density" uses all observations to infer on x_t with $t \leq N$ and is to be distinguished

from the filtering density $p(x_t|y^t)$ which uses only observations $y^t = (y_1, \ldots, y_t)$. From this joint posterior the marginal $p(\theta|y^N)$ may be derived to infer on the model parameter θ. The marginal $p(x^N|y^N)$ is the solution to the problem of inference about the unobservable state process, given all the data. Note that this marginal also accounts for the uncertainty associated with θ. The joint posterior distribution $p(x^N, \theta|y^N)$ is given by Bayes' theorem:

$$\begin{aligned} p(x^N, \theta|y^N) &\propto \tilde{p}(x^N, \theta|y^N) \\ \tilde{p}(x^N, \theta|y^N) &= p(y^N|x^N, \theta)p(x^N|\theta)p(\theta), \end{aligned} \tag{20}$$

and is proportional to the following three densities:

- the "complete data likelihood" $p(y^N|x^N, \theta)$ which due to assumption II is equal to the product of the observation densities $p(y_t|x_t, \theta)$ given by (12):

$$p(y^N|x^N, \theta) = \prod_{t=1}^{N} p(y_t|x_t, \theta)$$

- the "prior density" $p(x^N|\theta)$ which due to assumption I is equal to the product of the transition densities $p(x_t|x_{t-1}, \theta)$ given by (6) times the prior $p(x_0|\theta)$ given by (13):

$$p(x^N|\theta) = \left(\prod_{t=1}^{N} p(x_t|x_{t-1}, \theta) \right) \cdot p(x_0|\theta)$$

- the marginal prior $p(\theta)$ of the model parameter θ.

Among all the densities envolved, the marginal prior $p(\theta)$ of the model parameter is the only one which is not defined by the state space model. To complete the specification of the CIR-model, we assume that the components θ_d of θ are independent a priori with the following un-informative priors:

$$\begin{aligned} &p(\beta_j) \propto c, \quad p(\kappa_j) \propto c, \quad p(\sigma_j^2) \propto 1/\sigma_j^2, \quad p(\lambda_j) \propto c, \quad j = 1, \ldots, K; \\ &p(\sigma_{\varepsilon_i}^2) \propto 1/\sigma_{\varepsilon_i}^2, \quad i = 1, \ldots, n; \quad p(\sigma_h^2) \propto 1/\sigma_h^2. \end{aligned} \tag{21}$$

Because of the non-normality of the transition densities $p(x_t|x_{t-1}, \theta)$, the joint posterior distribution $p(x^N, \theta|y^N)$ has no simple analytical form. During the last years, Bayesian analysis of models with complicated posterior distributions increasingly has been carried out by Markov-chain-Monte-Carlo (MCMC hereafter) methods (for a general review see e.g. Smith and Roberts, 1993; Tierney, 1995). These

techniques include the Gibbs sampler (Gelfand and Smith, 1990) and the Metropolis-Hastings algorithm (Hastings, 1970; Chib and Greenberg, 1994). Instead of trying to find an analytical approximation to a complicated posterior distribution such as $p(x^N, \theta|y^N)$, a (dependent) sample $(x^N, \theta)^{(m)}$, $m = 1, \ldots, M$ is drawn from $p(x^N, \theta|y^N)$ and quantities of interest such as parameter estimates or marginal densities are computed from this sample.

To start a MCMC technique an initial value $(x^N, \theta)^{(0)}$ has to be assigned to (x^N, θ). The next value $(x^N, \theta)^{(m)}$ is generated from $(x^N, \theta)^{(m-1)}$ via a Markov chain. The transition kernel of this Markov chain is chosen in such a way, that the stationary distribution of the Markov chain is equal to $p(x^N, \theta|y^N)$. If the kernel is ergodic, $(x^N, \theta)^{(m)}$ will be – usually after a burn-in phase to reach equilibrium – a (dependent) sample from $p(x^N, \theta|y^N)$ and the posterior expectation of any function $g(x^N, \theta)$ can be estimated as the ergodic average of $g((x^N, \theta)^{(m)})$.

For the Gibbs sampler the density of the transition kernel is chosen as the product of the full conditional densities $p(\theta_d|\theta_{-d}, x^N, y^N)$ and $p(r_{t,j}|r_{t,-j}, x^N_{-t}, \theta, y^N)$:

$$\begin{aligned} K((x^N, \theta)^{(m)}|(x^N, \theta)^{(m-1)}) &= \prod_{t=0}^{N} \prod_{j=1}^{K} p(r_{t,j}^{(m)}|r_{t,<j}^{(m)}, r_{t,>j}^{(m-1)}, (x^N_{<t})^{(m)}, (x^N_{>t})^{(m-1)}, \theta^{(m-1)}, y^N) \cdot \\ &\quad \cdot \prod_{d=1}^{D} p(\theta_d^{(m)}|\theta_{<d}^{(m)}, \theta_{>d}^{(m-1)}, (x^N)^{(m)}, y^N), \end{aligned} \tag{22}$$

where θ_{-d} denotes all components of θ except θ_d, $\theta_{<d}$ and $\theta_{>d}$ denotes all components $\theta_{d'}$ of θ with $d' < d$ and $d' > d$, respectively, with a similar meaning for $r_{t,\cdot}$; x^N_{-t}, $x^N_{<t}$ and $x^N_{>t}$ is the whole state process except, before and after t, respectively. (22) is – in principle – a simple procedure to determine $(x^N, \theta)^{(m)}$ from $(x^N, \theta)^{(m-1)}$: first sample the components $x_t^{(m)}$ of the state process starting from $t = 0$; for each t, sample the state variables $r_{t,j}^{(m)}$ starting from j=1. Then sample the components $\theta_d^{(m)}$ of θ starting from $d = 1$. Whenever a component has been updated from $m-1$ to m, we use this value in the conditioning argument of the full conditional densities instead of the value at $m-1$.

To implement the Gibbs sampler we need to sample from the full conditional densities, $p(\theta_d|\cdot)$ and $p(r_{t,j}|\cdot)$ for short, which are proportional to $p(x^N, \theta|y^N)$ which in turn is proportional to $\tilde{p}(x^N, \theta|y^N)$ (see (20)):

$$p(r_{t,j}|\cdot), p(\theta_d|\cdot) \propto p(y^N|x^N, \theta)p(x^N|\theta)p(\theta). \tag{23}$$

For the CIR-model only the full conditionals of the variances $\sigma^2_{\varepsilon_i}$ and σ^2_h of the observation density have a standard form and are straightforward to sample from.

For all other full conditionals the right hand side of (23) is a density with unknown normalizing constant and is not standard to sample from. For such cases it is quite common to use a step of the Metropolis-Hastings algorithm for sampling. Such a hybrid method is sometimes called "Metropolis within Gibbs" (Tierney, 1995).

Instead of sampling directly from $p(\theta_d|\cdot)$ and $p(r_{t,j}|\cdot)$, respectively, the Metropolis-Hastings algorithm samples candidate values $\theta_d^\star$ and $r_{t,j}^\star$ from a proposal density $q(\theta_d|\theta_d^{'},\cdot)$ and $q(r_{t,j}|r_{t,j}^{'},\cdot)$ with $\theta_d^{'} = \theta_d^{(m-1)}$ and $r_{t,j}^{'} = r_{t,j}^{(m-1)}$. The proposed values $\theta_d^\star$ and $r_{t,j}^\star$, however, are not accepted with probability one, but $\theta_d^\star$ only with probability $\alpha(\theta_d^\star|\theta_d^{(m-1)},\cdot)$ and $r_{t,j}^\star$ with probability $\alpha(r_{t,j}^\star|r_{t,j}^{(m-1)},\cdot)$:

$$\alpha(\theta_d^\star|\theta_d^{(m-1)},\cdot) = \min\left\{\frac{\tilde{p}(x^N,\theta_d^\star,\theta_{-d}|y^N)q(\theta_d^{(m-1)}|\theta_d^\star,\cdot)}{\tilde{p}(x^N,\theta_d^{(m-1)},\theta_{-d}|y^N)q(\theta_d^\star|\theta_d^{(m-1)},\cdot)},1\right\}, \tag{24}$$

$$\alpha(r_{t,j}^\star|r_{t,j}^{(m-1)},\cdot) = \min\left\{\frac{\tilde{p}(r_{t,j}^\star,r_{t,-j},x_{-t}^N,\theta|y^N)q(r_{t,j}^{(m-1)}|r_{t,j}^\star,\cdot)}{\tilde{p}(r_{t,j}^{(m-1)},r_{t,-j},x_{-t}^N,\theta|y^N)q(r_{t,j}^\star|r_{t,j}^{(m-1)},\cdot)},1\right\}. \tag{25}$$

If the proposed value $\theta_d^\star$ is accepted – e.g. by sampling u uniformly from $[0,1]$ and accepting if $u \leq \alpha(\theta_d^\star|\theta_d^{(m-1)},\cdot)$ – $\theta_d^{(m)}$ is defined to be equal to $\theta_d^\star$: $\theta_d^{(m)} = \theta_d^\star$. Otherwise the Markov chain remains in the old state: $\theta_d^{(m)} = \theta_d^{(m-1)}$. The same holds for $r_{t,j}^{(m)}$. A convenient property of this procedure is that the unknown normalizing constant of the posterior density $p(x^N,\theta|y^N)$ cancels in (24) and (25), and therefore need not to be known. In order to implement the Metropolis-Hastings algorithm, suitable proposal densities have to be specified. In Subsection 3.2 we will discuss in detail how to choose proposal densities for sampling from the full conditionals within the CIR-model.

3.2 Sampling from the Full Conditionals for the CIR-Model

For the CIR-model only the full conditionals of the variances $\sigma_{\varepsilon_i}^2$ and σ_h^2 of the observation density have a standard form. For all other full conditionals we use a step of the Metropolis-Hastings algorithm. In order to implement the Metropolis-Hastings algorithm, suitable proposal densities have to be specified. Considerable recent work has been devoted to the question of how these choices should be made (e.g. Tierney, 1995; Chib and Greenberg, 1994) and still there does not seem to be a clear answer. An efficient solution appears to be the construction of model specific proposal densities by modifying the full conditional densities (23) in such a way that simple densities are obtained. Such proposal densities will be suggested below. In what follows we will make extensive use of the following lemma:

Lemma 1. If a density $p(Z|\cdot)$ is proportional to the following product of "normal kernels":

$$p(Z|\cdot) \propto \prod_{s=1}^{S} \exp\left\{-\frac{1}{2c_s}(a_s Z - b_s)^2\right\}, \tag{26}$$

with a_s, b_s and c_s being independent of Z, then Z has a normal distribution with the following moments:

$$\begin{aligned} p(Z|\cdot) &= p_N(Z; \Sigma\mu, \Sigma), \\ \mu &= \sum_{s=1}^{S} \frac{a_s b_s}{c_s}, \quad \Sigma^{-1} = \sum_{s=1}^{S} \frac{a_s^2}{c_s}. \end{aligned} \tag{27}$$

A proof is obtained by completing squares.

Sampling of $\sigma^2_{\varepsilon_i}$ and σ^2_h. It is easy to verify that given the un-informative priors (21) for $\sigma^2_{\varepsilon_i}$ and σ^2_h, the full conditional posteriors $p(\sigma^2_{\varepsilon_i}|\cdot)$ and $p(\sigma^2_h|\cdot)$ of $\sigma^2_{\varepsilon_i}$ and σ^2_h are inverse gamma densities:

$$p(\sigma^2_{\varepsilon_i}|\cdot) = p_{IG}(\sigma^2_{\varepsilon_i}; N/2, 1/2\sum_{t=1}^{N}(y_{t,i} - \hat{y}_{t,i})^2), \tag{28}$$

$$p(\sigma^2_h|\cdot) = p_{IG}(\sigma^2_h; 1/2(\sum_{t=1}^{N} n_t - N \cdot n), 1/2\sum_{t=1}^{N}\sum_{i=n+1}^{n_t}(y_{t,i} - \hat{y}_{t,i})^2), \tag{29}$$

with $\hat{y}_{t,i}$ given by (12).

Proposal densities for λ_j. As $p(x^N|\theta)$ is independent of λ_j and the prior $p(\lambda_j)$ in (21) is un-informative, the full conditional $p(\lambda_j|\cdot)$ in (23) simplifies to:

$$p(\lambda_j|\cdot) \propto \prod_{t=1}^{N}\prod_{i=1}^{n_t} p_N(y_{t,i}; \hat{y}_{t,i}, V_{t,ii}), \tag{30}$$

where $\hat{y}_{t,i}$ depends on λ_j in a non-linear manner (see (12)). A normal proposal density $q(\lambda_j|\lambda_j', \cdot)$ is derived by linearizing $\hat{y}_{t,i}$ around the "old" value λ_j' and applying Lemma 1:

$$\begin{aligned} q(\lambda_j|\lambda_j', \cdot) &= p_N(\lambda_j; \hat{\lambda}_j^N, \Lambda_j^N), \\ \hat{\lambda}_j^N &= \lambda_j' + \Lambda_j^N \cdot \sum_{t=1}^{N}\sum_{i=1}^{n_t} \frac{y_{t,i} - \hat{y}_{t,i}(\lambda_j')}{V_{t,ii}} \cdot \frac{\partial \hat{y}_{t,i}}{\partial \lambda_j}(\lambda_j'), \\ (\Lambda_j^N)^{-1} &= \sum_{t=1}^{N}\sum_{i=1}^{n_t} \frac{(\frac{\partial \hat{y}_{t,i}}{\partial \lambda_j}(\lambda_j'))^2}{V_{t,ii}}. \end{aligned} \tag{31}$$

Proposal densities for $\beta_j, \kappa_j, \sigma_j$. Let γ_j be the following parameter vector: $\gamma_j = (\beta_j, \kappa_j, \tau_j)'$, $\tau_j^{-1} = \sigma_j^2$. If we exploit the transition density (6), we find that it factorizes into K products, each of which just depends on γ_j. Therefore the full conditional densities $p(\gamma_{j,k}|\cdot)$, $1 \leq k \leq 3$, in (23) simplify to:

$$p(\gamma_{j,k}|\cdot) \propto \prod_{t=1}^{N} \prod_{i=1}^{n_t} p_N(y_{t,i}; \hat{y}_{t,i}, V_{t,ii}) \cdot \prod_{t=1}^{N} p(r_{t,j}|r_{t-1,j}, \gamma_j) p(r_{0,j}|\gamma_j) \cdot p(\gamma_{j,k}). \quad (32)$$

$p(\gamma_{j,k})$ is the un-informative prior (21) of $\gamma_{j,k}$.

First, when constructing the proposal density for $\gamma_{j,k}$, we modify the observation densities in (32) by linearizing $\hat{y}_{t,i}$ around the "old" value $\gamma_{j,k}'$. Note, that $\hat{y}_{t,i}$ is already linear in β_j and no modifications are necessary when constructing the proposal density for β_j. Our second step is to substitute in (32) the exact transiton densities by the approximate normal transition densities (16) (Chen and Scott, 1995),

$$p(r_{t,j}|r_{t-1,j}, \gamma_j) \approx p_N(r_{t,j}; F_{t,jj} r_{t-1,j} + u_{t,j}, \sigma_j^2 \tilde{Q}_{t,jj}), \quad (33)$$

with F_t, u_t and Q_t given by (17) - (19) and $\tilde{Q}_{t,jj} = Q_{t,jj}/\sigma_j^2$ being independent of σ_j^2. Finally, in case we work with the stationary prior (14), the density $p(r_{0,j}|\gamma_j)$ is approximated by a normal density with the same first and second moment as the exact prior (13):

$$p(r_{0,j}|\gamma_j) \approx p_N(r_{0,j}; \beta_j, \beta_j \frac{\sigma_j^2}{2\kappa_j}). \quad (34)$$

In case we use a vague normal prior, $p(r_{0,j}|\gamma_j)$ is independent of γ_j and drops from (32).

It is easy to verify that the product of the approximate normal transition densities (33) times the approximate normal prior (34) times the un-informative prior (21) is a Gamma density in $\tau_j = \sigma_j^{-2}$. This suggests a gamma proposal density for τ_j. The information in the observation densities is exploited in the following way: as mentioned above, we linearize each $\hat{y}_{t,i}$ – considered as a function of τ_j – around the old value τ_j'. Applying Lemma 1 we obtain a normal approximation to the products of observation densities:

$$p(\tau_j|x^N, y^N) \approx p_N(\tau_j; \hat{\tau}_j^N, T_j^N), \quad (35)$$

$$\hat{\tau}_j^N = \tau_j' + T_j^N \cdot \sum_{t=1}^{N} \sum_{i=1}^{n_t} \frac{y_{t,i} - \hat{y}_{t,i}(\tau_j')}{V_{t,ii}} \cdot \frac{\partial \hat{y}_{t,i}}{\partial \tau_j}(\tau_j'),$$

$$(T_j^N)^{-1} = \sum_{t=1}^{N} \sum_{i=1}^{n_t} \frac{(\frac{\partial \hat{y}_{t,i}}{\partial \tau_j}(\tau_j'))^2}{V_{t,ii}}.$$

To this normal we fit a gamma density with the same first and second moment and combine it with the gamma density obtained from (33), (34) and (21). Finally, we end up with a gamma proposal for τ_j with the following parameters:

$$
\begin{aligned}
q(\tau_j|\tau_j^{'},\cdot) &= p_G(\tau_j;\nu_j^N,\Sigma_j^N), \qquad (36)\\
\nu_j^N &= \frac{(\hat{\tau}_j^N)^2}{T_j^N} + \frac{N}{2} + \nu_j^0\\
\Sigma_j^N &= \frac{\hat{\tau}_j^N}{T_j^N} + 1/2\sum_{t=1}^{N}\frac{(r_{t,j}-F_{t,jj}r_{t-1,j}-u_{t,j})^2}{\tilde{Q}_{t,jj}} + \Sigma_j^0,
\end{aligned}
$$

with $\nu_j^0 = 0.5$ and $\Sigma_j^0 = (r_{0,j}-\beta_j)^2\cdot\kappa_j/\beta_j$ in case we work with the stationary prior (14) and $\nu_j^0 = \Sigma_j^0 = 0$ in case we work with the a vague normal prior for the state variables. The proposal density for the original parameter σ_j^2 is then an inverse gamma density with the same parameters as the gamma density.

The approximate transitions densities (33) and the approximate normal prior (34) have to be modified further to obtain simple proposal densities for β_j and κ_j. It has been already mentioned, that the product of the observation densities is a normal kernel in β_j. This suggests a normal proposal density for β_j. The first moments of the approximate normal transition densities (33) and the approximate normal prior (34), too, are linear in β_j; the variances, however, are not independent of β_j. To obtain normal kernels in β_j, we evaluate these variances at the old value $\beta_j^{'}$. We then obtain from Lemma 1 the following normal proposal density $q(\beta_j|\beta_j^{'},\cdot)$:

$$
\begin{aligned}
q(\beta_j|\beta_j^{'},\cdot) &= p_N(\beta_j;B_j^N\hat{\beta}_j^N,B_j^N), \qquad (37)\\
\hat{\beta}_j^N &= -\frac{2\kappa_j}{\sigma_j^2}\left[\sum_{t=1}^{N}\sum_{i=1}^{n_t}\left(y_{t,i}-\sum_{k=1}^{K}\frac{B_k(T_{t,i})r_{t,k}}{T_{t,i}}+\sum_{k=1,k\neq j}^{K}\frac{\ln A_k(T_{t,i})}{T_{t,i}}\right)\frac{\ln C_j(T_{t,i})}{T_{t,i}V_{t,ii}}\right] +\\
&\quad +\sum_{t=1}^{N}\frac{(r_{t,j}-e^{-\kappa_j\Delta t}r_{t-1,j})(1-e^{-\kappa_j\Delta t})}{Q_{t,jj}(\beta_j^{'})} + \hat{\beta}_j^0\\
(B_j^N)^{-1} &= \frac{4\kappa_j^2}{\sigma_j^4}\left[\sum_{t=1}^{N}\sum_{i=1}^{n_t}\frac{(\ln C_j(T_{t,i}))^2}{T_{t,i}^2V_{t,ii}}\right] + \sum_{t=1}^{N}\frac{(1-e^{-\kappa_j\Delta t})^2}{Q_{t,jj}(\beta_j^{'})} + B_j^0,
\end{aligned}
$$

with $\hat{\beta}_j^0 = 2r_{0,j}\kappa_j/(\beta_j^{'}\sigma_j^2)$ and $B_j^0 = 2\kappa_j/\beta_j^{'}\sigma_j^2$ in case we work with the stationary prior (14) and $\hat{\beta}_j^0 = B_j^0 = 0$ in case we work with the a vague normal prior for the state variables.

Neither the observation density nor the transition density suggest any specific kernel for the proposal density of κ_j. We therefore construct a normal proposal density for κ_j. The first moments of the approximate normal transition densities

(33), which are non-linear functions of κ_j, are linearized around the old value $\kappa_j^{'}$ and the variances, which are not independent of κ_j, are evaluated at the old value $\kappa_j^{'}$ to obtain normal kernels in κ_j. The first moment of the approximate normal prior (34) is independent of κ_j; the variance, however, is not independent of κ_j. If we evaluate the variance at the old value $\kappa_j^{'}$, the approximate stationary prior is un-informative about κ_j. If we modify the observation densities by linearizing $\hat{y}_{t,i}$ around the "old" value $\kappa_j^{'}$, we obtain from Lemma 1 the following normal proposal density $q(\kappa_j|\kappa_j^{'},\cdot)$:

$$
\begin{aligned}
q(\kappa_j|\kappa_j^{'},\cdot) &= p_N(\kappa_j;\hat{\kappa}_j^N,K_j^N), \qquad (38)\\
\hat{\kappa}_j^N &= \kappa_j^{'} + K_j^N \cdot \left\{ \sum_{t=1}^{N}\sum_{i=1}^{n_t} \frac{y_{t,i}-\hat{y}_{t,i}(\kappa_j^{'})}{V_{t,ii}} \cdot \frac{\partial \hat{y}_{t,i}}{\partial \kappa_j}(\kappa_j^{'}) + \right.\\
&\qquad \left. + \sum_{t=1}^{N} \frac{\left[r_{t,j}-\beta_j-(r_{t-1,j}-\beta_j)e^{-\kappa_j^{'}\Delta t}\right](\beta_j-r_{t-1,j})\Delta t e^{-\kappa_j^{'}\Delta t}}{Q_{t,jj}(\kappa_j^{'})} \right\},\\
(K_j^N)^{-1} &= \sum_{t=1}^{N}\sum_{i=1}^{n_t} \frac{(\frac{\partial \hat{y}_{t,i}}{\partial \kappa_j}(\kappa_j^{'}))^2}{V_{t,ii}} + \sum_{t=1}^{N} \frac{(\beta_j-r_{t-1,j})^2(\Delta t)^2 e^{-2\kappa_j^{'}\Delta t}}{Q_{t,jj}(\kappa_j^{'})}.
\end{aligned}
$$

Proposal densities for $r_{t,j}$. The full conditional posterior $p(r_{t,j}|\cdot)$ in (23) simplifies to:

$$
p(r_{t,j}|\cdot) \propto \begin{cases} p(y_t|x_t,\theta)p(r_{t,j}|r_{t-1,j},\gamma_j)p(r_{t+1,j}|r_{t,j},\gamma_j), & 1 \le t \le N-1,\\ p(y_N|x_N,\theta)p(r_{N,j}|r_{N-1,j},\gamma_j), & t=N,\\ p(r_{0,j}|\gamma_j)p(r_{1,j}|r_{0,j},\gamma_j), & t=0. \end{cases}
$$

The observation density $p(y_t|x_t,\theta)$ is a normal kernel in $r_{t,j}$. Thus we suggest a normal proposal density for $r_{t,j}$. To obtain normal kernels from the transition densities we again use the approximate normal transition density (16). $p(r_{t,j}|r_{t-1,j},\gamma_j)$ approximated by (33) is already a normal kernel, $p(r_{t+1,j}|r_{t,j},\gamma_j)$ is not: although the mean is linear in $r_{t,j}$, the variance is not independent of $r_{t,j}$. To obtain a normal kernel we evaluate this variance at the old value $r_{t,j}^{'}$. If we work with the stationary prior (14) for $t=0$ we use the approximate normal prior (34), otherwise the prior $p(r_{0,j}|\gamma_j)$ is already normal.

We then obtain from Lemma 1 the following normal proposal densities $q(r_{t,j}|r_{t,j}^{'},\cdot)$:

$$
q(r_{t,j}|r_{t,j}^{'},\cdot) = p_N(r_{t,j};S_{t,j}m_{t,j},S_{t,j}), \qquad (39)
$$

$$m_{t,j} = \begin{cases} \sum_{i=1}^{n_t}(y_{t,i} - a_{t,i} - \sum_{k=1,k\neq j}^{K} B_{t,ik} r_{t,k})\frac{B_{t,ij}}{V_{t,ii}} + \frac{F_{t,jj} r_{t-1,j} + u_{t,j}}{Q_{t,jj}} + \frac{F_{t+1,jj}(r_{t+1,j} - u_{t+1,j})}{Q_{t+1,jj}(r'_{t,j})}, \\ \qquad 1 \leq t \leq N-1, \\ \sum_{i=1}^{n_N}(y_{N,i} - a_{N,i} - \sum_{k=1,k\neq j}^{K} B_{N,ik} r_{N,k})\frac{B_{N,ij}}{V_{N,ii}} + \frac{F_{N,jj} r_{N-1,j} + u_{N,j}}{Q_{N,jj}}, \quad t = N, \\ \frac{E(r_{0,j})}{V(r_{0,j})} + \frac{F_{1,jj}(r_{1,j} - u_{1,j})}{Q_{1,jj}(r'_{0,j})}, \quad t = 0; \end{cases}$$

$$S_{t,j}^{-1} = \begin{cases} \sum_{i=1}^{n_t} \frac{B_{t,ij}^2}{V_{t,ii}} + \frac{1}{Q_{t,jj}} + \frac{F_{t+1,jj}^2}{Q_{t+1,jj}(r'_{t,j})}, \quad 1 \leq t \leq N-1, \\ \sum_{i=1}^{n_N} \frac{B_{N,ij}^2}{V_{N,ii}} + \frac{1}{Q_{N,jj}}, \quad t = N, \\ \frac{1}{V(r_{0,j})} + \frac{F_{1,jj}^2}{Q_{1,jj}(r'_{0,j})}, \quad t = 0. \end{cases}$$

For $a_{t,i}$ and $B_{t,ij}$ see (10). If the proposed value is negative, it is rejected anyway. This imposes the constraint $r_{t,j} \geq 0$.

4 Empirical Analysis

4.1 Data description

To illustrate the estimation method suggested in this paper, we use a data set of 16 US interest rates observed monthly in the period January 1964 to December 1993 (360 observations). The data are collected from the CRSP Government Bond File making use of three different sources: (i) The rates for maturities from one to six months were collected from the Fama T-Bill Yield File based on average (mean of the bid and the ask quote) prices of 6-month T-Bills. (ii) The rates for maturities from seven to eleven months were collected from the Fama T-Bill Yield File based on average prices of 12-month T-Bills. (iii) The rates for maturities from one to five years were collected from the Fama-Bliss Discount Bond File extracted from prices of coupon-bearing Treasury issues.

We are going to fit a one-, a two-, and a three-factor CIR-model to this data. We make the simplifying assumption that each measurement error has the same variance $V_{t,ii} = \sigma_h^2$, $1 \leq i \leq n_t$ which may be relaxed in future investigations.

The prior of the factors $r_{0,j}$ is chosen to be a vague normal prior. The choice is based on results obtained by Chen and Scott (1995) and Geyer and Pichler (1995) who use a similar and the same data set, respectively. For two- and three-factor-

models they find the first factor to correlate highly with long maturity yields. The second factor is interpreted as a yield spread: that is the difference between yields with the shortest and longest maturity. These findings and equation (1) suggest the following choice: In a one-factor-model set $E(r_{0,1})$ equal to the yield with the shortest maturity observed in the first month. In two- and three-factor-models set $E(r_{0,1})$ equal to the yield with the longest maturity (five years) and $E(r_{0,2})$ to the yield spread. Since no specific interpretation exists about the third factor we simply set $E(r_{0,3}) = E(r_{0,2})$ in the three-factor-model. This choice takes into account available prior information about the factors and appears preferable to using the stationary distribution.

4.2 Starting Values, Assessing Convergence and Constructing the Final MCMC sample

In our empirical analysis we found that convergence of the Markov Chain designed in section 3 to the steady state might be slow. Thus some care should be given both to the choice of the starting value $(x^N, \theta)^{(0)}$ and to assessing convergence.

There has been an extensive discussion on the question whether to use a long single chain or many short multiple chains when implementing MCMC methods (see e.g. the comments following the papers of Gelman and Rubin, 1992, and Geyer, 1992). Finding it rather difficult to obtain convergence from a single chain, we used two seperate chains during the burn-in phase and monitored the algorithm simply by plotting the sampled values of the model parameters $\theta_d^{(m)}, m = 0, \ldots, M_0$ as a function of m for both chains. If the chains were not in equilibrium, we increased M_0 and continued the burn-in phase for each of the chains. We switched to a single chain when convergence was obvious from the plots. To obtain the final MCMC sample of length M we continued with this single chain for further M steps.

We found that the speed of convergence of the Markov chain towards the steady state heavily depends on the number of factors to be included in the model. Whereas the burn-in phase is rather short for the one-factor-model ($M_0 = 500$), it increases to $M_0 = 10000$ for the two-factor-model and is as long as $M_0 = 20000$ for the three-factor-model. For illustration purposes we include Figures 1 and 2 showing the parameters of the second factor both for the two- and the three-factor-model.

Figure 1 about here

Figure 2 about here

For the first starting value we use approximate Kalman filtering and QML estimation. The filtered estimates of the state variable $\hat{x}_{t|t}, t = 0, \ldots, N$, and the QML estimator $\hat{\theta}_{ML}$ served as starting value for one chain. The second starting value has been obtained by random choice. We sampled $200 \cdot K$ different values for θ from the cartesian product of univariate intervals, where these intervals were chosen large enough to cover a reasonable range of parameter values and computed the quasi likelihood function from the approximate normal model. The parameter with the largest functional value of the quasi likelihood function then served as starting value $\theta^{(0)}$ for the second chain. Starting values $(x^N)^{(0)}$ for the state process were obtained by approximate Kalman filtering conditional on $\theta^{(0)}$. The filtered estimates of the state variable $\hat{x}_{t|t}, t = 0, \ldots, N$, served as a starting value for the state process.

The QML estimator proved in general to be a good starting value in the sense that the corresponding chain reached the steady state quicker than the chain starting at a randomly chosen value (see Figures 1 and 2), except for the three-factor-model. For this model the chain starting at the QML estimator showed hardly any convergence to the steady state for the model parameters $(\kappa_3, \sigma_3, \beta_3, \lambda_3)$ of the third factor. However, after replacing the QML starting value by the randomly chosen starting value for the third factor only, the steady state was finally reached after a considerable burn-in phase.[1]

To construct the final MCMC sample we continued with the chain starting from the QML estimator for the one- and the two-factor-model, and starting from the QML estimator with the starting value for the model parameters of the third factor substituted by the randomly chosen starting value for the three-factor-model. The length of the final MCMC sample is chosen to be equal to $M = 9500$ for the one-factor-model and $M = 10000$ for the two- and the three-factor-model. We used Dickey-Fuller tests to test for stationarity of the final MCMC samples. The unit-root hypothesis was rejected for all parameters and factors at the 5% level, except for the MCMC samples of β_3 and λ_3.

We conclude this subsection with a discussion of the acceptance rates of the Metropolis-Hastings algorithm. In Section 3 we suggested to implement MCMC methods for the CIR-model by means of a Metropolis-Hastings step within Gibbs sampling, where sampling takes place from an approximate proposal density and a rejection step is incorporated in such a way that finally we sample from the correct

[1] Note that for the three-factor-model the chain denoted by QML in Figure 2 is the chain starting from this mixed value.

density. The larger the acceptance rates $\alpha(\theta_d^\star|\theta_d^{(m-1)},\cdot)$ for θ_d and $\alpha(r_{t,j}^\star|r_{t,j}^{(m-1)},\cdot)$ for $r_{t,j}^\star$ – see formula (24) and (25) – the better are the chosen proposal densities.

The empirical average acceptance rate $\hat{\alpha}_d = 1/(M_0+M)\sum_{m=1}^{M+M_0}\alpha(\theta_d^\star|\theta_d^{(m-1)},\cdot)$ for each model parameter θ_d, $1 \leq d \leq D$, for the chain starting at the QML estimator is reported in Table 1. The normal proposal density chosen for λ_j is extremely good, leading to an average acceptance rate of 99.7 – 99.9%. The normal proposal density chosen for β_j is also fine, leading to high average acceptance rates of 89.4 – 98.9%. The proposal densities chosen for κ_j and σ_j which are a normal and an inverse gamma density, respectively, are "worse" in the sense that the average acceptance rates range from 79.0 to 97.8% for κ_j and from 76.2 to 98.6% for σ_j. These rates are smaller than for λ_j and β_j but still in an acceptable range.

Table 1 about here

The acceptance rates $\hat{\alpha}_{t,j} = 1/(M_0+M)\sum_{m=1}^{M+M_0}\alpha(r_{t,j}^\star|r_{t,j}^{(m-1)},\cdot)$ for the state variable $r_{t,j}$ are analyzed in Table 2. For all factors and for all models the median of $\hat{\alpha}_{t,j}$ – where the median is taken over t, $0 \leq t \leq N$ – ranges from 96.8% to 99.6% and proves to be quite high. The 0.05-quantile of $\hat{\alpha}_{t,j}$ ranges from 66.6% to 98.5% which means that for 95% of the time points t, $0 \leq t \leq N$, the average acceptance rate is larger than the given number. Furthermore, for 92.5% – 100% of the time points t, $0 \leq t \leq N$, the average acceptance rate is larger than 0.9. Finally, for the one- and the two-factor-model the minimum of all average acceptance rates $\hat{\alpha}_{t,j}$ over t, $0 \leq t \leq N$ is larger than 62.7%. To sum up, in most cases the normal proposal density derived from the normal approximation of Chen and Scott (1995) is a perfect proposal density for the state variable $r_{t,j}$. For the three-factor-model, however, there are a few time points where this is not true. For the first and the second factor we find one and for the third factor even ten time points out of $N = 361$, where $\hat{\alpha}_{t,j}$ is smaller than 0.3.

Table 2 about here

4.3 Exploring the MCMC Sample

The final MCMC sample of length M may be explored in various ways. We focus on the difference between exact Bayesian and approximate QML estimation, and the economic relevance of the differences found.

We start by comparing the estimator $\hat{\theta}_{ML}$ obtained from the approximate quasi likelihood function with an exact Bayesian estimator $\hat{\theta}_B$ of the unknown model parameters θ. A Bayesian estimator of θ such as the posterior mean is obtained from the MCMC output simply by averaging over the sampled values $\theta^{(m)}, m = M_0 + 1, M_0 + M$. The estimation results are summarized in Table 3. (Q)ML-estimators are usually reported together with standard errors s_{ML} derived from the second derivative of the (quasi) likelihood function. We therefore include in this table both s_{ML} and the Bayesian analogon s_B which is simply estimated from the MCMC-sample as the standard deviation of the sampled parameter values from their mean. Furthermore we report empirical 95%-HPD-regions which are approximated by the shortest interval containing 95% of the sampled values.

Table 3 about here

Figures 3 and 4 contain exact marginal posterior densities of the model parameters – estimated by smoothed marginal histograms of the MCMC sample – and the approximate QML estimator.

Figure 3 about here
Figure 4 about here

Hardly any difference exists between exact Bayesian and approximate QML estimation for the one-factor-model. The difference between the estimators of β_1 is small compared to the standard deviation. For the two-factor-model the same is true for the first factor; however, the difference is considerable for the second factor, where the QML estimators of all parameters except σ_2 lie outside the 95%-H.P.D.-regions derived from Bayesian estimation. The differences are even stronger for the three-factor-model with the most striking deviation occuring for the third factor.

From the posterior densities and the 95%-H.P.D.-regions we conclude in an explorative manner that for the three-factor-model all model parameters $\kappa_j, \sigma_j, \beta_j$, $1 \leq j \leq 3$, differ from 0 and that at least three factors can be extracted for the US-data. The estimators obtained for κ_j, $1 \leq j \leq 3$, for the three-factor-model allow the following interpretation of the factors: The first factor with a small κ_1 exhibits long-term memory – the autocorrelation function of the square root process (1) over lag t equals $\exp(-\kappa_j t)$ – and behaves similar to a unit root. However, from the posterior densities in Figures 3 and 4 we find that $\kappa_1 > 0$ with extremely high

posterior probability and we conclude in an explorative manner that no unit root exists. The second factor exhibits medium-term memory with the mean-reverting behavior one would expect for a yield spread. The third factor with a large κ_3 is very close to a factor with no autocorrelation at all.

Although the difference between exact Bayesian and approximate QML estimation of the model parameters is considerable, especially for models with more than one factor, one might wonder, whether this difference is of any economic relevance. We attempt to obtain an indication for the economic consequences by investigating the difference between estimates of the parameters which are relevant for valuation. These are σ_j and the parameter combinations $(\kappa+\lambda)_j = \kappa_j + \lambda_j$ and $(\kappa\beta)_j = \kappa_j\beta_j$. Bayesian point estimates, standard errors, 95%-H.P.D.regions and marginal posterior densities of these parameter combinations are easily obtained from the transformed MCMC sample $(\kappa+\lambda)_j^{(m)} = \kappa_j^{(m)} + \lambda_j^{(m)}$ and $(\kappa\beta)_j^{(m)} = \kappa_j^{(m)}\beta_j^{(m)}$ in the same fashion as above. In Table 4 these estimation results are compared to the plug-in estimator, where the transformed parameter is estimated by substituting each component by the QML estimator.

Table 4 about here

As regards the difference between estimates of σ_j, Table 3 shows that only the third factor's estimate in the three-factor-model is substantially different ($\hat{\sigma}_{ML,3} = 12.3$ and $\hat{\sigma}_{B,3} = 3.373$). On the basis of Table 4 we may conclude, however, that parameter combinations for all factors of the three-factor-model differ considerably, both in sign and magnitude. It is interesting to note, that the plug-in estimators lie outside the 95%-H.P.D.-regions. The differences for the one- and two-factor-model are minor and the plug-in estimators lie inside the 95%-H.P.D.-regions.

These results agree with those obtained from inspecting the factor loadings $B_j(T)$, which can also be used to indicate the economic relevance of differences in estimation approaches. The loadings associated with each factor are non-linear functions of the unknown model parameters (see (4)). In Figure 5 we compare the approximate plug-in estimator of $B_j(T)$ – plotted as a function of the time to maturity T – with the Bayesian estimator for the two- and the three-factor-model. The Bayesian estimator of $B_j(T)$ is the mean of the transformed MCMC sample $B_j^{(m)}(T)$, $m = M_0+1, \ldots M_0+M$, where the model parameters appearing in (4) are substituted by the parameters from the (stationary) MCMC sample. The plug-in estimator is derived by substituting the unknown parameters by the QML estimator.

Figure 5 about here

Figure 5 confirms that the Bayesian estimates of the parameters relevant for asset pricing in a three-factor-model differ strongly from QML estimates. Although the general tendency and shape of the loadings agrees for all three factors, the absolute values differ substantially, in particular for the first factor. Thus, if a one- or a two-factor-model is employed for asset pricing the additional effort implied by Bayesian estimation may not be worth while. We conclude, however, that significant differences in bond and derivative prices can result when a three-factor-model is used.

Some readers might consider Bayesian estimation rather envolved numerically and would prefer the QML approach, if the difference between the methods was not too substantial. It would be helpful to obtain information from the QML estimation which indicates that differences may be important. For that purpose we found an interesting relationship between the shape of the stationary distribution of the state process $r_{t,j}$ and the discrepancy between QML and Bayesian estimates. The parameters of the stationary gamma distribution of $r_{t,j}$ – the degrees of freedom ν_j and the scale parameter ξ_j – depend on the unknown model parameters κ_j, β_j and σ_j (see (14) and (8)) and may be estimated from the MCMC sample simply by

$$\hat{\nu}_{B,j} = \frac{1}{M} \sum_{m=M_0+1}^{M_0+M} \frac{\kappa_j^{(m)} \beta_j^{(m)}}{(\sigma_j^{(m)})^2}, \qquad \hat{\xi}_{B,j} = \frac{1}{M} \sum_{m=M_0+1}^{M_0+M} \frac{2\kappa_j^{(m)}}{(\sigma_j^{(m)})^2}. \tag{40}$$

If the number of degrees of freedom ν_j is large, the stationary distribution is close to a normal distribution with the variance proportional to the mean, whereas a more or less small degree of freedom leads to a more or less skew stationary distribution. Table 5 summarizes Bayesian estimators as well as plug-in estimators of the parameters of the stationary distribution for one-, two-, and three-factor-models. The stationary distribution of the only factor of the one-factor-model ($\hat{\nu}_{B,1} = 12.182$) is close to a normal distribution, the stationary distribution of the first factor of the two-factor-model ($\hat{\nu}_{B,1} = 4.297$) is quite skew, and the stationary distributions of the second factor of the two-factor-model ($\hat{\nu}_{B,2} = 1.021$) and of all factors of the three-factor-model ($\hat{\nu}_{B,1} = 1.48$, $\hat{\nu}_{B,2} = 0.448$, $\hat{\nu}_{B,3} = 1.253$) are extremely skew. It interesting to note that the differences between Bayesian and QML estimation of the model parameters (see Table 3 and Figure 4) are substantial, especially for factors with extremely skew stationary distributions. To a large extent this is true also for the plug-in estimator of the degrees of freedom derived from QML estimation.

Thus, if only QML estimation is carried out, the estimated degrees of freedom $\hat{\nu}_{ML,j} = \hat{\kappa}_{ML,j}\hat{\beta}_{ML,j}/\hat{\sigma}^2_{ML,j}$ may serve as a useful diagnostic tool: if any of the $\hat{\nu}_{ML,j}$s is extremely small, the approximate results derived from QML estimation may be somewhat doubtful.

Table 5 about here

5 Concluding Remarks

The MCMC approach presented in this paper is based on single-move-sampling, i.e. sampling a single component θ_d of θ from the full conditional density $p(\theta_d|\theta_{\_d}, x^N, y^N)$. If the correlation among the various components of θ is high, MCMC methods based on single-move-sampling are known to converge quite slowly towards the steady state. An alternative would be multi-move-sampling where highly correlated components are sampled simultaneously from an appropriate multivariate proposal density. The construction of multivariate model specific proposal densities for the parameters of the CIR-model is far from trivial and would be an interesting topic of further research.

By chosing the multi-factor CIR-model we confined ourselves in the present paper to one specific model among a much larger group of possible term-structure models (see Rogers, 1995 for a general review of term-structure models). From the view point of estimation, econometric term structure models based on the CIR-model are far more challenging than are other models, e.g. econometric term structure models based on the Vasicek-model, because of the highly non-normal structure of the underlying state space model. Bayesian estimation of econometric term structure models via MCMC methods, however, is not restricted to the CIR-model and is easily extended to other term structure models which fulfill the following requirements:

- The transition density $p(x_t|x_{t-1})$ is known analytically and the functional values of this density can be computed numerically.
- An explicit solution for the nominal prices $P_t(T)$ is available and possesses the affine structure given in (2) with $A_j(T)$ and $B_j(T)$ being known functions of the model parameters.

References

Carlin, B.P., N.G. Polson and D.S. Stoffer, 1992, A Monte Carlo approach to nonnormal and nonlinear state-space modeling, Journal of the American Statistical Association 87, 493–500.

Carter, C.K. and R. Kohn, 1994, On Gibbs sampling for state space models, Biometrika 81, 541–53.

Chen, 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, December, 14–31.

Chen, R. and L. Scott, 1995, Multi-factor Cox-Ingersoll-Ross models of the term structure: estimates and tests from a Kalman filter model, Working paper, University of Georgia.

Chib, S. and E. Greenberg, 1994, Understanding the Metropolis-Hastings algorithm, Working paper, Washington University.

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

De Jong, P. and N. Shephard, 1995, The simulation smoother for time series models, Biometrika 82, 339–350.

Duan, J., and J. Simonato, 1995, Estimating and testing exponential-affine term structure models by Kalman filter, Working paper, University of Montreal.

Frühwirth-Schnatter, S., 1994, Data augmentation and dynamic linear models, Journal of Time Series Analysis 15, 183–202.

Gelfand, A.E. and A.F.M. Smith, 1990, Sampling-based approaches to calculating posterior densities, Journal of the American Statistical Association, 85, 398–409.

Gelman, A. and D. Rubin, 1992, Inference from iterative simulation using multiple sequences, Statistical Science 7, 457–511.

Geyer, C., 1992, Practical Markov Chain Monte Carlo, Statistical Science 7, 473–511.

Geyer, A. and S. Pichler, 1995, A state space approach to estimate the term structure of interest rates: some empirical evidence, Working paper, University of Economics Vienna.

Hastings, W.K., 1970, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57, 97–109.

Harvey, A.C., 1989, Forecasting, structural time series models and the Kalman filter (Cambridge University Press, Cambridge).

Jacquier, E., Polson N.G. and P.E. Rossi, 1995, Bayesian analysis of stochastic volatility, Journal of Business and Economic Statistics 12, 371–417.

Lund, J., 1994, Econometric Analysis of continous-time arbitrage-free models of the term-structure of interest rates, Working paper, The Aarhus School of Business, Denmark.

Rogers, L.C.G., 1995, Which model for term-structure of interest rates should one use? in: M.H.A. Davis, D.Duffie, W.H. Fleming and S.E. Shreve, eds., Mathematical Finance, The IMA Volumes in Mathematics and its Applications, Vol. 65 (Springer, New York), 93–115.

Shephard, N., 1994, Partial non-gaussian state space models, Biometrika 81, 115–31.

Smith, A.F.M. and G. Roberts, 1993, Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods, Journal of the Royal Statistical Society B 55, 3–23.

Tierney, L., 1995, Markov chain for exploring posterior distributions, Annals of Statistics, to appear.

Vasicek, O.A., 1977, An equilibrium characterisation of the term structure, Journal of Financial Economics 5, 177–88.

model	factor	κ_j	σ_j	β_j	λ_j
one-factor	$j=1$	0.906	0.986	0.968	0.997
two-factor	$j=1$	0.978	0.983	0.989	0.999
	$j=2$	0.790	0.969	0.894	0.999
three-factor	$j=1$	0.949	0.952	0.967	0.999
	$j=2$	0.972	0.940	0.981	0.999
	$j=3$	0.882	0.762	0.977	0.998

Table 1: Empirical average acceptance rates of the Metropolis-Hastings algorithm for the model parameters θ_d

model	factor	$\tilde{\alpha}_j$	α_j^Q	$\frac{\#(\hat{\alpha}_{t,j} \geq 0.9)}{N+1}$	$\alpha_{\min,j}$	$\#(\hat{\alpha}_{t,j} \leq 0.3)$
one-factor	$j=1$	0.995	0.985	1.000	0.945	0
two-factor	$j=1$	0.996	0.968	1.000	0.779	0
	$j=2$	0.989	0.980	0.983	0.627	0
three-factor	$j=1$	0.995	0.859	0.942	0.239	1
	$j=2$	0.995	0.666	0.925	0.167	1
	$j=3$	0.968	0.864	0.942	0.000	10

$\tilde{\alpha}_j$ ... median of $\hat{\alpha}_{t,j}$, $0 \leq t \leq N$;

α_j^Q ... 0.05-quantile of $\hat{\alpha}_{t,j}$, $0 \leq t \leq N$;

$\alpha_{\min,j} = \min_{0 \leq t \leq N} \hat{\alpha}_{t,j}$;

Table 2: Empirical average acceptance rates of the Metropolis-Hastings algorithm for the state variables $r_{t,j}$

model	factor	θ_d	$\hat{\theta}_{ML,d}$	$\hat{\theta}_{B,d}$	$s_{ML,d}$	$s_{B,d}$	95%-H.P.D.-region	
one-factor	$j=1$	κ_1	0.144	0.141	0.007	0.008	0.127	0.156
		σ_1	0.277	0.277	0.006	0.007	0.263	0.291
		β_1	6.512	6.633	0.237	0.285	6.099	7.187
		λ_1	–0.185	–0.182	0.009	0.009	–0.199	–0.166
		h	0.465	0.465	0.004	0.005	0.456	0.474
two-factor	$j=1$	κ_1	0.089	0.078	0.007	0.008	0.062	0.094
		σ_1	0.261	0.256	0.007	0.006	0.245	0.269
		β_1	3.536	3.639	0.267	0.334	3.046	4.363
		λ_1	–0.291	–0.281	0.019	0.014	–0.311	–0.258
	$j=2$	κ_2	1.086	0.795	0.068	0.041	0.711	0.870
		σ_2	1.535	1.482	0.052	0.046	1.407	1.587
		β_2	1.945	2.817	0.157	0.143	2.521	3.090
		λ_2	–1.336	–0.985	0.105	0.057	–1.085	–0.858
		h	0.208	0.208	0.002	0.002	0.204	0.212
three-factor	$j=1$	κ_1	0.068	0.083	0.009	0.010	0.065	0.104
		σ_1	0.228	0.293	0.007	0.003	0.287	0.298
		β_1	3.111	1.539	0.313	0.200	1.194	1.922
		λ_1	–0.266	–0.632	0.018	0.015	–0.662	–0.606
	$j=2$	κ_2	0.523	0.178	0.038	0.030	0.124	0.236
		σ_2	0.939	1.069	0.029	0.024	1.023	1.111
		β_2	2.107	2.929	0.066	0.438	2.084	3.875
		λ_2	–0.302	–0.368	0.042	0.034	–0.427	–0.301
	$j=3$	κ_3	32.733	7.903	1.110	0.399	7.144	8.799
		σ_3	12.295	3.373	0.420	0.154	3.022	3.653
		β_3	0.307	1.794	0.019	0.067	1.690	1.943
		λ_3	–34.403	–3.114	1.229	0.255	–3.527	–2.611
		h	0.129	0.151	0.002	0.002	0.148	0.154

$\hat{\theta}_{ML,d}$ ... d-th component of the QML-estimator

$\hat{\theta}_{B,d}$ ... d-th component of the posterior mean estimated by the mean of the MCMC sample

$s_{ML,d}$ ... standard deviation of the QML estimator

$s_{B,d}$ ... standard deviation of the MCMC sample around $\hat{\theta}_{B,d}$

95%-H.P.D.-region ... estimated by the smallest interval containing 95% of the MCMC sample

Table 3: Comparing QML and Bayesian estimation for one-, two-, and three-factor-models

model	transformed parameter	plug-in	Bayes	95%-H.P.D.-region	
one-factor	$\kappa_1 + \lambda_1$	–0.041	–0.041	–0.052	–0.029
	$\kappa_1 \cdot \beta_1$	0.938	0.936	0.880	0.990
two-factor	$\kappa_1 + \lambda_1$	–0.202	–0.203	–0.226	–0.183
	$\kappa_1 \cdot \beta_1$	0.315	0.282	0.245	0.319
	$\kappa_2 + \lambda_2$	–0.250	–0.190	–0.273	–0.114
	$\kappa_2 \cdot \beta_2$	2.112	2.236	2.057	2.413
three-factor	$\kappa_1 + \lambda_1$	–0.198	–0.549	–0.574	–0.521
	$\kappa_1 \cdot \beta_1$	0.212	0.127	0.105	0.149
	$\kappa_2 + \lambda_2$	0.221	–0.189	–0.233	–0.136
	$\kappa_2 \cdot \beta_2$	1.102	0.513	0.397	0.666
	$\kappa_3 + \lambda_3$	–1.670	4.789	4.161	5.461
	$\kappa_3 \cdot \beta_3$	10.043	14.171	12.825	15.637

plug-in ... (approximate) estimator obtained from the QML estimator

Bayes ... (exact) posterior mean estimated by the the mean of the transformed MCMC sample

95%-H.P.D.-region ... estimated by the smallest interval containing 95% of the transformed MCMC sample

Table 4: Comparing QML and Bayesian estimation of transformed model parameters

model	factor	$\hat{\nu}_{B,j}$	$\hat{\nu}_{ML,j}$	$\hat{\xi}_{B,j}$
one-factor	$j = 1$	12.182	12.184	3.679
two-factor	$j = 1$	4.297	4.634	2.378
	$j = 2$	1.021	0.896	0.725
three-factor	$j = 1$	1.480	4.076	1.948
	$j = 2$	0.448	1.250	0.312
	$j = 3$	1.253	0.066	1.397

Table 5: Estimated parameters of the stationary distribution (degrees of freedom $\hat{\nu}$ and scale parameter $\hat{\xi}$) of factors for one-, two- and three-factor-models

Figure 1: Assessing convergence for the second factor of the two-factor-model

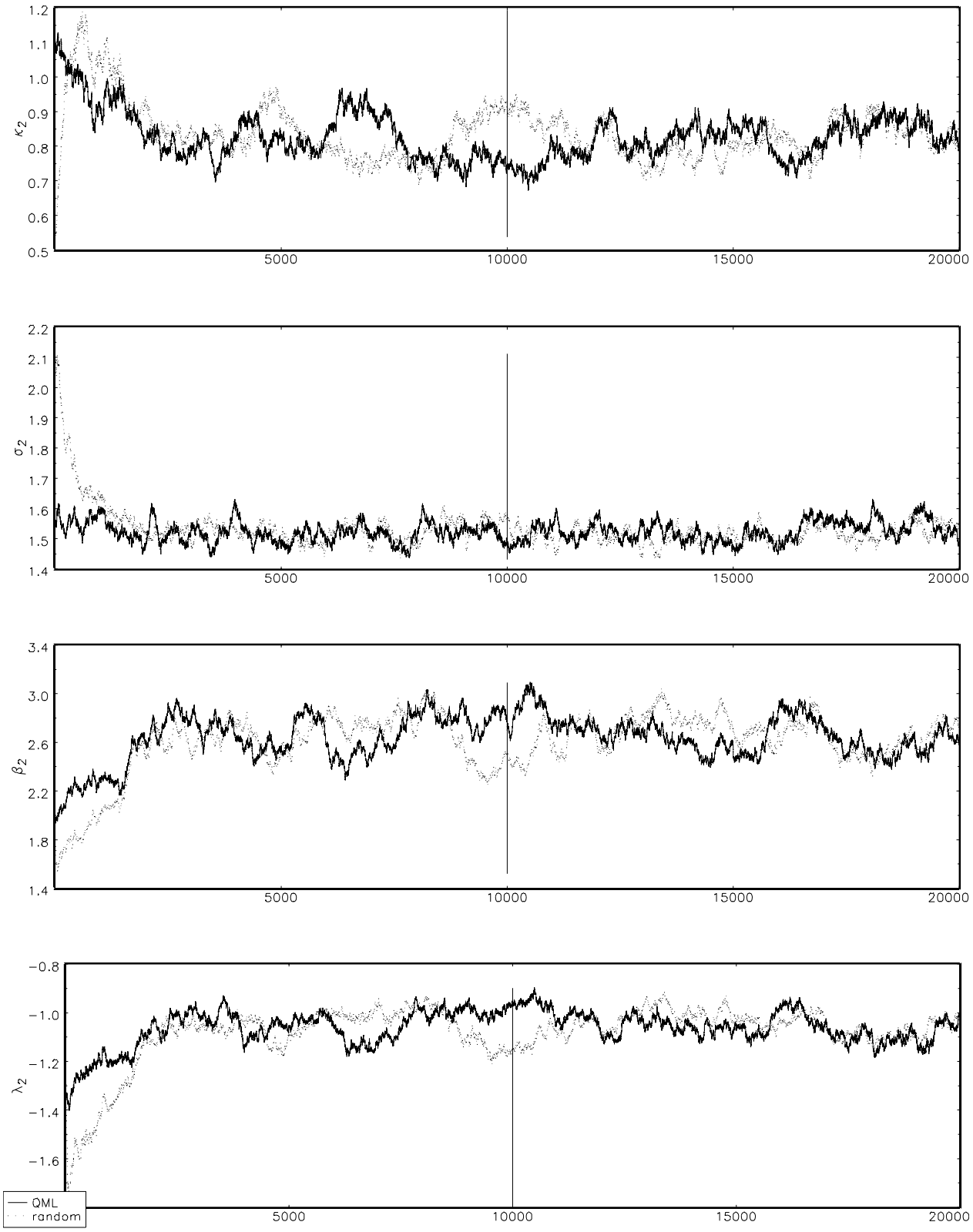

Figure 2: Assessing convergence for the second factor of the three-factor-model

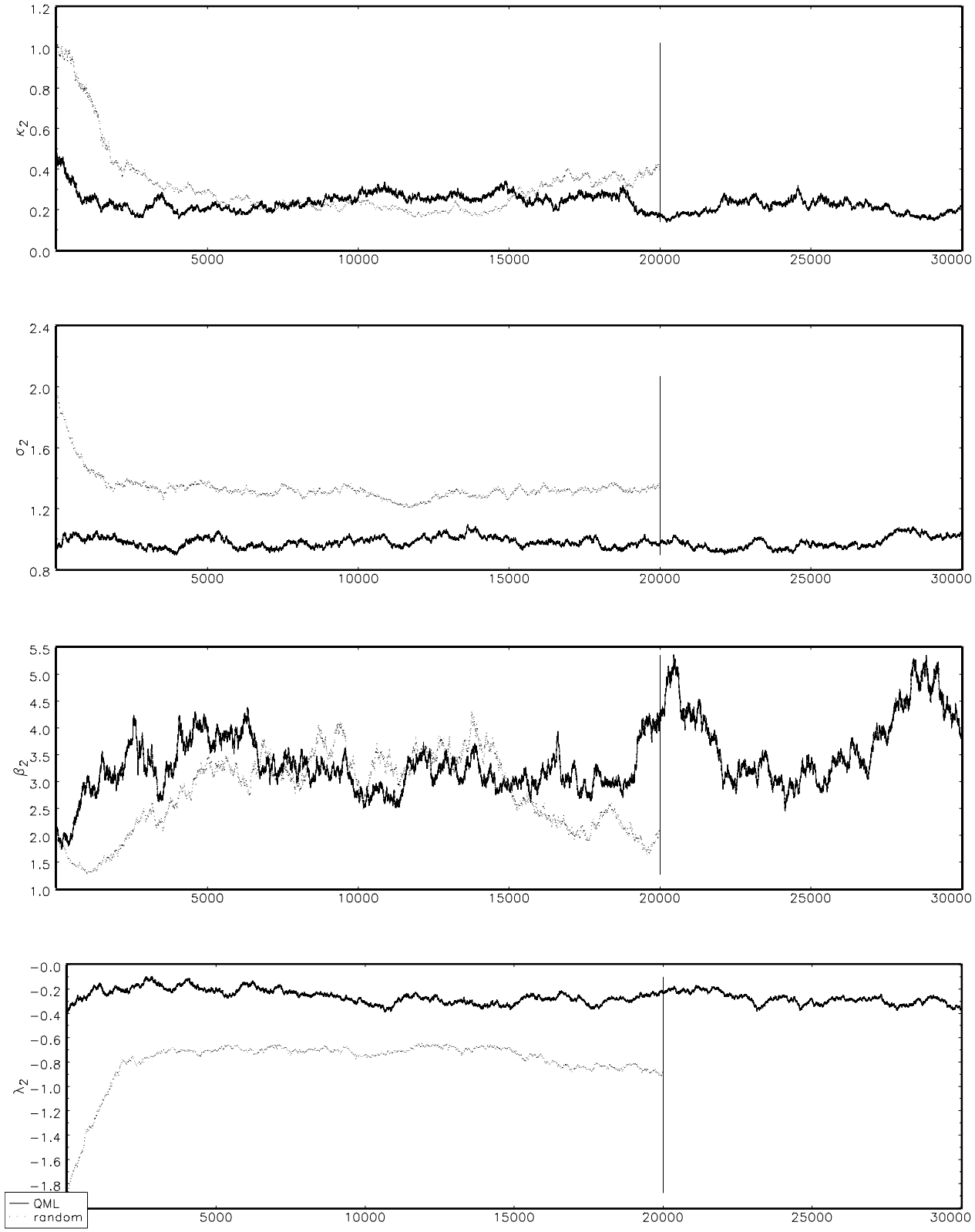

Figure 3: Marginal posterior densities and the QML estimator (vertical line)

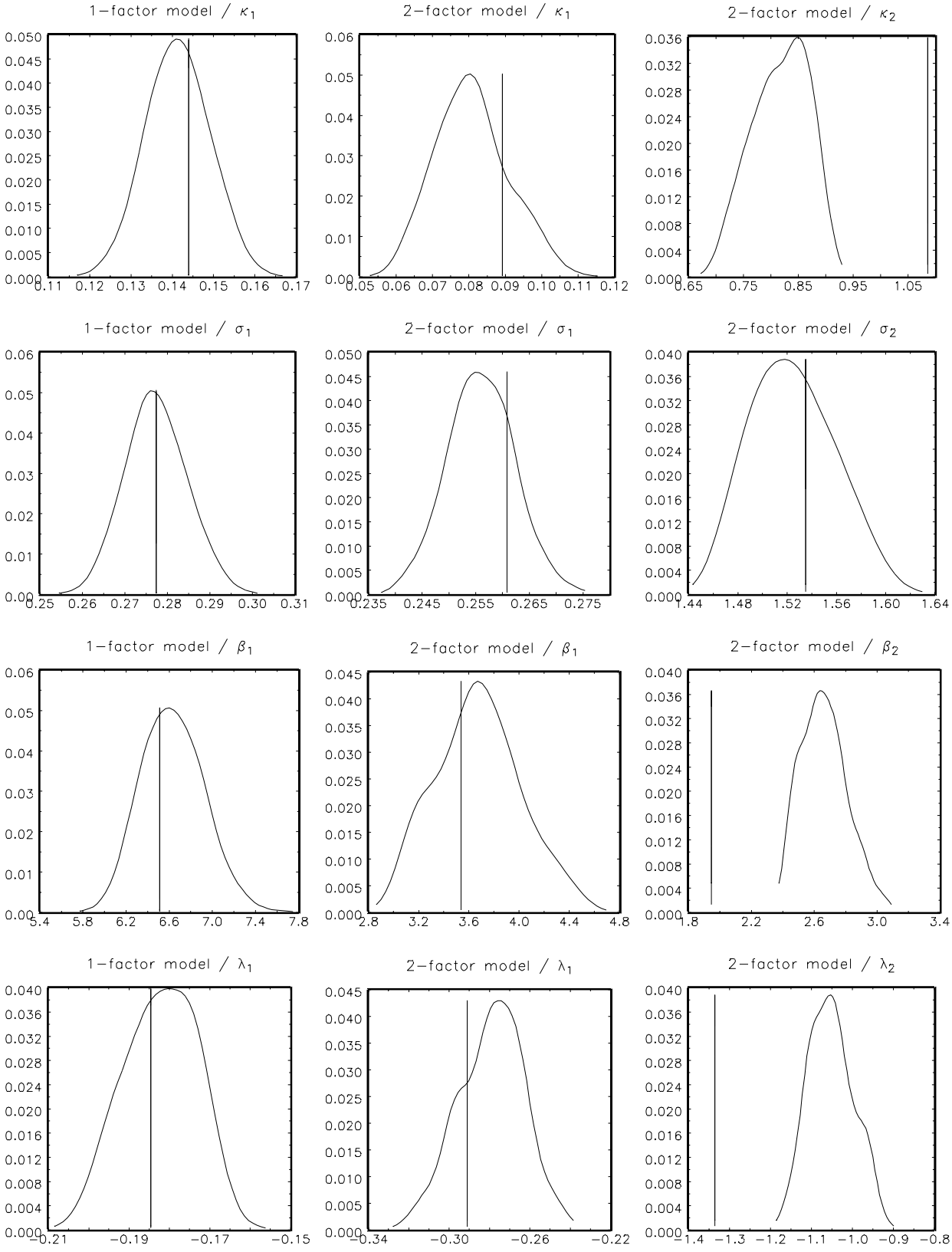

Figure 4: Marginal posterior densities and the QML estimator (vertical line)

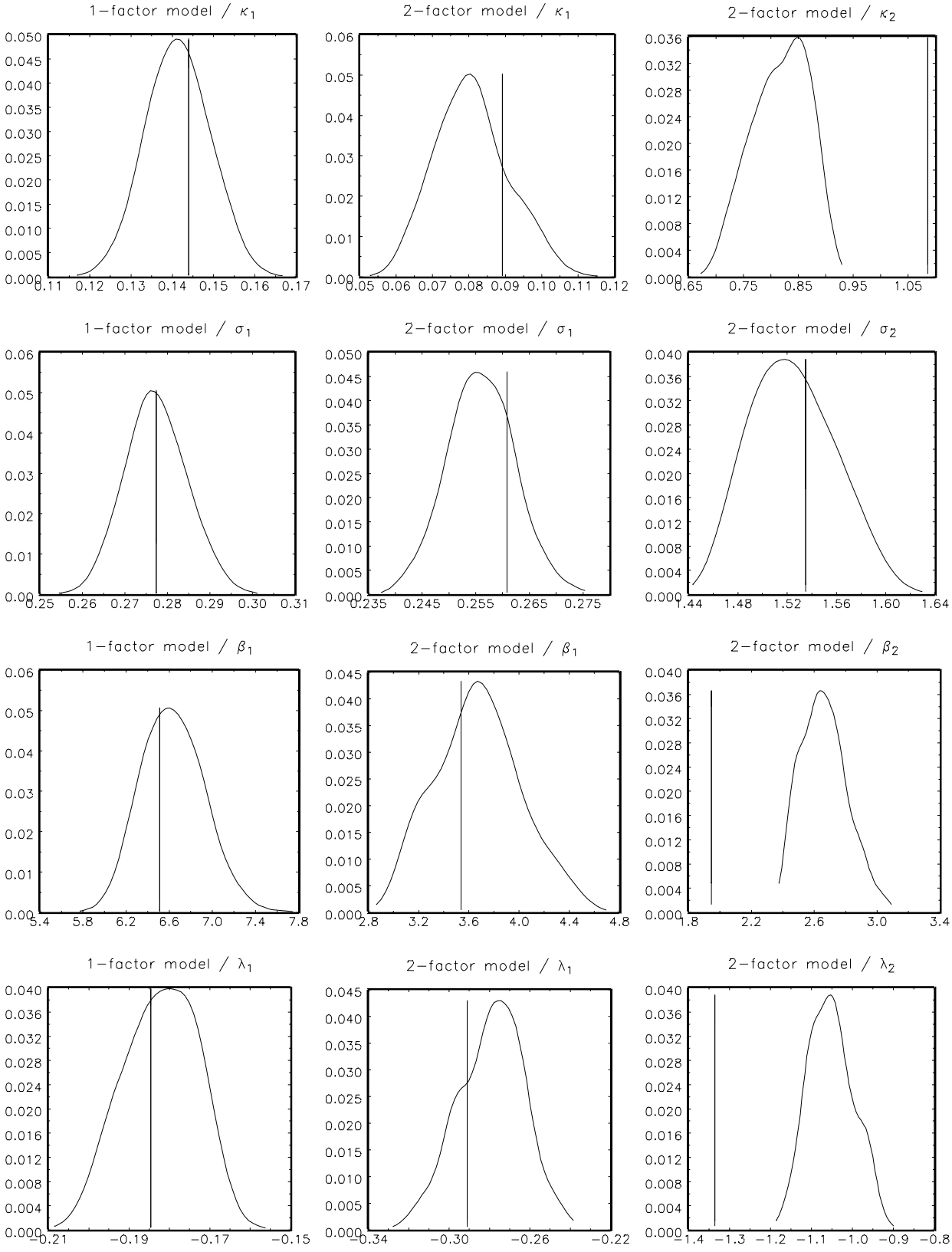

Figure 5: Comparing QML and Bayesian estimators of the factor loadings $B_j(T)$