University of Heidelberg

Department of Economics

Discussion Paper Series | No. 462

Application of the Generalized Method of Moments for Estimating Continuous-Time Models of U.S. Short-Term Interest Rates

Balázs Cserna

January 2008

Application of the Generalized Method of Moments for Estimating Continuous-Time Models of U.S. Short-Term Interest Rates

Balázs Cserna*

Abstract

We show by Monte Carlo simulations that the jackknife estimation of QUENOUILLE (1956) provides substantial bias reduction for the estimation of short-term interest rate models applied in CHAN ET AL. (1992) – hereafter CKLS (1992). We find that an alternative estimation based on NOWMAN (1997) does not sufficiently solve the problem of time aggregation. We provide empirical distributions for parameter tests depending on the elasticity of conditional variance. Using three-month U.S. Treasury bill yields and the Federal fund rates, we demonstrate that the estimation results can depend on both the sampling frequency and the proxy that is used for interest rates.

Keywords: Elasticity of conditional variance, generalized method of moments, jackknife estimation, stochastic differential equations, short-term interest rate.

JEL: C16, C52.

* Balázs Cserna, Alfred-Weber-Institute, University of Heidelberg, Chair of Statistics, Hauptstrasse 126, D-69117 Heidelberg (Germany). E-mail: balazs.cserna@awi.uni-hedielberg.de. Comments and suggestions are welcome.

1 Introduction

Many asset pricing models use one-factor stochastic differential equations – hereafter SDE – to capture the dynamics of the short-term interest rate. The properties of such a SDE are determined by its drift and diffusion functions. Since different functions generate significantly different prices for interest rate sensitive assets, the choice of the corresponding specification is of great importance. Models that permit closed-form solutions for contingent claims provide greater analytical insights, whereas the lack of empirical evidence may result in model risk.

In this work, we focus our attention on problems involved with the estimation procedure applied by CKLS (1992). In CKLS (1992) the empirical validity of several continuous-time models is analyzed by means of the generalized method of moments – hereafter GMM – estimation of HANSEN (1982). We show by Monte Carlo simulations that the estimation procedure suffers from significant estimation bias that arises from the estimation of autoregressive models. This bias is likely to have considerable effects on pricing derivatives. We prove by Monte Carlo simulations that the so-called jackknife estimation of QUENOUILLE (1956) achieves substantial bias reduction under the assumption that the dynamics of the short-term interest rate can be approximated by means of the discrete-time process used in CKLS (1992).

Moreover, the estimates cannot be assumed to follow the normal distribution in general. For some empirical investigations, this deviation from the normal distribution can question the importance of mean-reversion for the underlying short-term interest rate process. Using Bartlett weights suggested by NEWEY AND WEST (1987a) for variance-covariance matrix estimation, we determine empirical distributions of the associated *t*-statistics under the null hypothesis that the drift function is zero. Our findings indicate that the distributions depend on the elasticity of the conditional variance of changes in the short-term interest rate. In addition, we consider the empirical distribution of the corresponding likelihood ratio – hereafter LR – statistics under the null models that are examined in this work. We find that the distributions do not strictly and exclusively depend on the number of restrictions imposed by the underlying null models but also on the given model that is considered.

An additional problem involved with the estimation of CKLS (1992) is that discrete-time approximations introduce discretization bias, since – as a result of time aggregation – they neglect internal dynamics between sampling points. As shown in MELINO (1994), this feature can result in inconsistent estimators. To deal with the problem caused by discretization bias, we consider an alternative GMM technique based on the discrete-time version of the general SDE proposed by NOWMAN (1997). However, our Monte Carlo simulations demonstrate that the alternative estimation procedure does not sufficiently solve the problem of time aggregation.

We apply the estimation of CKLS (1992) – using our the empirical distributions of the associated test-statistics – for daily, weekly and monthly observations of the three-month U.S. Treasury bill yield and the Federal fund rate respectively, both from 04.01.1954 through 02.03.2006. Furthermore, we apply the unit root tests of SAID AND DICKEY (1984). We demonstrate that the estimation results can depend on both the sampling frequency and the proxy that is used for the short-term interest rate.

As indicated by the results, the three-month U.S. Treasury bill yields seem to be non-stationary, such that the role of mean-reversion appears to be negligible from the empirical point of view. As in CKLS (1992), we find that the conditional variance of changes in the three-month U.S. Treasury bill yield is highly sensitive to the yield level, whereas an exact specification of the elasticity is more important for daily observations, since the null models are rejected more often. Therefore, the sensitivity of the conditional variance on the yield level can be lower then assumed by CKLS (1992). On the other hand, the corresponding jackknife estimation applied for the three-month U.S. Treasury bill yield results in higher values for the elasticity.

In contrary to that result, daily observations of the Federal fund rate exhibit significant mean-reversion, whereas weekly and monthly observations seem to appear non-stationary. Another empirical result for the Federal fund rate is that the elasticity seems to be lower for daily observations then for the three-month U.S. Treasury bill yield. Although weekly and monthly observations of the Federal Funds rate exhibit higher elasticities, the corresponding null model that is not rejected for daily observations at the 5% level, cannot be rejected for weekly and monthly observations either.

The remainder of this work is organized as follows. Section 2 discusses the stochastic properties of several continuous-time models with regard to the short-term interest rate. Section 3 presents a theoretical analysis of the estimation procedure applied by CKLS (1992) and an alternative estimation. In Section 4 the results of the Monte Carlo experiments are presented. Section 5 reports the corresponding empirical results. Section 6 concludes.

2 Continuous-Time Models of the Short-Term Interest Rate

In CKLS (1992) the short-term interest rate follows a continuous-time stochastic process $\{R_t \mid t \geq 0\}$ which solves a time-homogenous, one-factor, diffusion-type SDE, namely

$$dR_t = (\alpha + \beta \cdot R_t)\,dt + \sigma \cdot R_t^{\gamma}\,dW_t\,, \tag{1}$$

where $\{W_t \mid t \geq 0\}$ is a standard Brownian motion on the filtered probability space $(\Omega, F, \{F_t \mid t \geq 0\}, P)$ and the parameter vector is $\theta = (\alpha, \beta, \sigma, \gamma)'$. The initial value R_0 is assumed to be fixed and positive. This specification allows both the conditional mean and variance of changes in the short-term interest rate to depend on the short-term interest rate level, whereas the drift function is linear in R_t. It follows from (1) that the conditional variance of changes in the short-term interest rate increases with the level of the interest rate if $\gamma > 0$. Using the properties of stochastic integrals, it can be shown that

$$\lim_{t \to \infty} E(R_t) = -\alpha / \beta \text{ for } \beta < 0\,,$$

where $E(..)$ denotes the associated expectation operator to P, such that $-\alpha / \beta$ is regarded as the long-run mean of R_t. This illustrates that in case of $\beta < 0$ the process is mean-reverting which means that there is an adjustment to the unconditional long-run mean measured by the level of β. We can obtain various well-known models of the short-term interest rate using corresponding parameter restrictions. The specifications that are examined in this work are summarized in Table 1.

Table 1 Summary of Alternative Models of the Short-Term Interest Rate

Model	SDE	α	β	γ
Unrestricted	$dR_t = (\alpha + \beta \cdot R_t)\,dt + \sigma \cdot R_t^{\gamma}\,dW_t$	-	-	-
MERTON (1973)	$dR_t = \alpha\,dt + \sigma\,dW_t$	-	0	0
VASICEK (1977)	$dR_t = (\alpha + \beta \cdot R_t)\,dt + \sigma\,dW_t$	-	-	0
DOTHAN (1978)	$dR_t = \sigma \cdot R_t\,dW_t$	0	0	1
CIR (1980)	$dR_t = \sigma \cdot R_t^{1.5}\,dW_t$	0	0	1.5
BS (1980)	$dR_t = (\alpha + \beta \cdot R_t)\,dt + \sigma \cdot R_t\,dW_t$	-	-	1
CIR (1985)	$dR_t = (\alpha + \beta \cdot R_t)\,dt + \sigma \cdot R_t^{0.5}\,dW_t$	-	-	0.5
GBM	$dR_t = \beta \cdot R_t\,dt + \sigma \cdot R_t\,dW_t$	0	-	1

Explanation: Table 1 summarizes the specifications of alternative continuous-time models of the short-term interest rate with their corresponding parameter restrictions that are imposed on the parameter vector of the unrestricted model.

The short-term interest rate in MERTON (1973) evolves a so-called arithmetic Brownian motion. This specification implies that both the variance and the absolute value of the unconditional mean of the short-term interest rate increase by time, such that the resulting process is non-stationary. The model of VASICEK (1977) supposes that the short-term interest rate follows a so-called Ornstein-Uhlenbeck process. This process is asymptotic stationary if and only if $\beta < 0$. In both models, the conditional variance of changes in the short-term interest rate is constant. Since R_t is Gaussian for both models, it holds $P(R_t < 0) > 0$ which is not a desirable feature from the practical point of view.

The solution of the process suggested by DOTHAN (1978) follows a log-normal distribution. Since the variance increases as $t \to \infty$, the process is non-stationary. Assuming $R_0 = -\alpha / \beta$ for simplification, it can be shown that asymptotic stationary of the short-term interest rate process suggested by BRENNAN AND SCHWARTZ (1980) – hereafter BS (1980) – requires $\beta < 0$ and $2 \cdot \beta + \sigma^2 < 0$. Due to $\gamma = 1$, the models of DOTHAN (1978) and BS (1980) assume that the conditional volatility of changes in the short-term interest rate at time t is proportional to the rate level.

Assuming $R_0 = -\alpha / \beta$, it can be shown that $\beta < 0$ is necessary for asymptotic stationarity of the process proposed by CIR (1985), where the conditional distribution of the interest rate is non-central chi-square. However, a closed-form solution of the process is not known. Due to $\gamma = 0.5$, the model of CIR (1985) assumes that the conditional variance of changes in the short-term interest rate at time t is proportional to the rate level. If R_t follows a GBM, then asymptotic stationarity requires $\beta < 0$ and $2 \cdot \beta + \sigma^2 < 0$. The specification implies that in case of $\beta < 0$ and $2 \cdot \beta + \sigma^2 < 0$ both the mean and the variance of R_t converge to zero as $t \to \infty$.

3 Representation of the Estimation Techniques

Let $\{t_i \mid i = 0,1,..\}$ symbolize an equidistant discretization of time, where t_i denotes a point of time with $\Delta t := t_{i+1} - t_i > 0$ for all i and $t_0 := 0$.[1] For estimation, it is assumed in CKLS

[1] There is no loss of generality in assuming equidistant points of time, since there is no conclusive weekend effect in money market instruments; see e.g. AÏT-SAHALIA (1996).

(1992) that the solution of (1) can be approximated by a discrete-time stochastic process $\{r(t_i)\,|\,i=0,1,...\}$ that satisfies

$$r(t_{i+1})-r(t_i)=(\alpha_0+\beta_0\cdot r(t_i))\cdot\Delta t+\varepsilon_0(t_{i+1}), \qquad (2)$$

where $\theta_0=(\alpha_0,\beta_0,\sigma_0,\gamma_0)'$ denotes the parameter vector that is to be estimated.[2] The unobservable error term $\varepsilon_0(..)$ is allowed to be conditionally heteroskedastic, such that

$$E\big(\varepsilon_0(t_{i+1})\,|\,r(t_i)\big)=0 \text{ and } E\big(\varepsilon_0^2(t_{i+1})\,|\,r(t_i)\big)=\sigma_0^2\cdot r^{2\cdot\gamma_0}(t_i)\cdot\Delta t \text{ for all } i. \qquad (3)$$

It should be noted that the approximation (2), which is referred to as the Euler discretization of (1), neglects errors that are introduced as a result of time aggregation. To apply the GMM procedure of HANSEN (1982), the following vector function is suggested

$$f\big(r(t_i),\theta\big)=\left[\begin{pmatrix}\varepsilon(t_{i+1})\\ \varepsilon^2(t_{i+1})-\sigma^2\cdot r^{2\cdot\gamma}(t_i)\cdot\Delta t\end{pmatrix}\otimes\begin{pmatrix}1\\ r(t_i)\end{pmatrix}\right],$$

where $\otimes$ denotes the Kronecker product and $\varepsilon(t_{i+1})=r(t_{i+1})-r(t_i)-(\alpha+\beta\cdot r(t_i))\cdot\Delta t$. Using the law of iterated expectations, it can be shown that $E(f(r(t_i),\theta))=0$ for all i, such that $E(f(r(t_i),\theta))=0\Leftrightarrow\theta=\theta_0$ is assumed for all i. Given a finite observed sample of the short-term interest rate with $n+1$ observations, the sample moments of $f(r(t_i),\theta)$ satisfy

$$g(\theta)=\frac{1}{n}\cdot\sum_{i=0}^{n-1} f\big(r(t_i),\theta\big).$$

The GMM procedure consists of choosing an estimator $\hat\theta$ for θ_0, such that the criterion function $Q(\theta)$ given below is minimized with respect to θ, that is

$$\hat\theta=\arg\min_\theta\{Q(\theta)\}, \text{ where } Q(\theta):=g(\theta)'\cdot V\cdot g(\theta),$$

and V is a positive definite random weighting matrix, i.e. $V>0$, such that $Q(\theta)\geq 0$ for all θ and $Q(\theta)=0\Leftrightarrow g(\theta)=0$. It can be shown that minimizing $Q(\theta)$ is equivalent to solving $G(\theta)'\cdot V\cdot g(\theta)=0$, where $G(\theta)$ denotes the Jacobian of $g(\theta)$ with respect to θ.

Since θ_0 is exactly identified by the moment conditions in the unrestricted case, the estimator of the unrestricted parameter vector – denoted as $\hat\theta^{(0)}$ – can be obtained by solving the non-

[2] The statistical properties of the discrete-time approximation given in (2) are summarized by BROZE ET AL (1995). Note that the properties of the discrete-time approximation do not correspond with the continuous-time stochastic process as $\Delta t\to 0$; see e.g. BROZE ET AL (1995) and RODRIGUES AND RUBIA (2004) for details.

linear system $g(\theta)=0$, such that $\hat{\theta}^{(0)}$ does not depend on the choice of the weighting matrix. Since a closed-form solution is not attainable in general, $g(\theta)=0$ is solved numerically by the Newton-Raphson – hereafter NR – method. If restrictions on the parameter vector are imposed by an underlying short-term interest rate model, then the parameters are overidentified by the moment conditions. In this case, we cannot find a parameter vector θ that satisfies $g(\theta)=0$ and therefore, the corresponding estimator of the given restricted model – these estimators are denoted by $\hat{\theta}^{(s)}$ hereafter, where s is the number of restrictions – depends on the choice of the weighting matrix. The sample moments of the vector function are assumed to satisfy the central limit theorem, that is

$$\Sigma^{-0.5}\cdot\sqrt{n}\cdot g(\theta_0)\xrightarrow{d}N(0,I) \text{ as } n\to\infty,$$

where $\xrightarrow{d}$ denotes the convergence in distribution, $N(0,I)$ denotes a multivariate standard normal distribution with identity I, and Σ is a positive definite matrix that satisfies

$$\Sigma=\Gamma_0+\lim_{n\to\infty}\left\{\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)\cdot\left(\Gamma_j+\Gamma_j'\right)\right\}=\Gamma_0+\sum_{j=1}^{\infty}\left(\Gamma_j+\Gamma_j'\right),$$

where $\Gamma_{i-j}:=E(f(r(t_i),\theta_0)\cdot f(r(t_j),\theta_0)')$.[3] Using the Taylor expansion, it follows that

$$\Lambda^{-0.5}\cdot\sqrt{n}\cdot(\hat{\theta}-\theta_0)\xrightarrow{d}N(0,I), \text{ where}$$

$$\Lambda^{0.5}=\left(G(\theta_0)'V\cdot G(\theta_0)\right)^{-1}\cdot\left(G(\theta_0)'V\cdot\Sigma\cdot V'\cdot G(\theta_0)\right)^{0.5}, \qquad (4)$$

such that $\Lambda=(G(\theta_0)'\Sigma^{-1}\cdot G(\theta_0))^{-1}$ if $V=\Sigma^{-1}$. Since $\Lambda-(G(\theta_0)'\Sigma^{-1}\cdot G(\theta_0))^{-1}\geq 0$ for all V, the optimal choice for the weighting matrix would be $V=\hat{\Sigma}^{-1}$, where $\hat{\Sigma}$ represents an estimator of Σ. Unfortunately, we cannot construct $\hat{\Sigma}$ by replacing the autocovariances Γ_j with their sample analogues $\hat{\Gamma}_j$ since the number of estimated autocovariances grows at the same rate as the sample size and $\hat{\Sigma}$ may be indefinite in finite samples. The solution is to construct an estimator in which the contribution of the sample autocovariances are weighted to reduce their role sufficiently for positive definiteness and have weights tend to one as $n\to\infty$ to ensure consistency, that is

$$\hat{\Sigma}=\hat{\Gamma}_0+\sum_{j=1}^{b(n)}\kappa(j,b(n))\cdot\left(\hat{\Gamma}_j+\hat{\Gamma}_j'\right),$$

[3] See e.g. HALL (2005).

where $\kappa(..)$ is known as the kernel and $b(n)$ as the bandwidth which depends on the sample size and must be positive. The bandwidth ensures that autocovariances corresponding to lags greater than $b(n)$ are given zero weight. The so-called Bartlett kernel introduced by NEWEY AND WEST (1987a) is given by

$$\kappa(j, b(n)) = 1 - j/(b(n)+1).$$

ANDREWS (1991) shows that the asymptotic mean square error of $\hat{\Sigma}$ is minimized by setting $b(n)$ to $O(n^{1/3})$ for the Bartlett kernel. This suggests that a form $b(n) = \rho \cdot n^{1/3}$ may be appropriate. However, this provides little practical guidance since ρ is not known. Unfortunately, statistical inference is often very sensitive to the choice of the bandwidth. Some methods of selecting the bandwidth require some prior knowledge or additional restrictions on the underlying process.[4] For estimation, we use $b(n) = \text{int}\{n^{1/3}\}$, where $\text{int}\{..\}$ is the integer part of the corresponding argument. The estimation of the restricted models requires a two-step GMM estimation; estimating the optimal weighting matrix in the first step using the unrestricted estimator, followed by estimating the parameters of the given restricted model.

To test the validity of the restrictions on the parameter vector given by the short-term interest rate models, we apply the methodology proposed by NEWEY AND WEST (1987b) which can be viewed as an extension to the GMM framework of the classical parameter tests from Maximum Likelihood – hereafter ML – theory. CKLS (1992) assume that the LR-statistic $n \cdot (Q(\hat{\theta}^{(s)}) - Q(\hat{\theta}^{(0)}))$ is asymptotically distributed chi-square under the given short-rate model with s degrees of freedom. As in CKLS (1992), the weighting matrix from the unrestricted model is used to calculate both $Q(\hat{\theta}^{(s)})$ and $Q(\hat{\theta}^{(0)})$.

Now we consider an alternative GMM estimation using the assumption of NOWMAN (1997) that the conditional variance of the short-term interest rate change remains unaffected over each unit observation period $[t_i, t_{i+1})$, such that (1) is simplified to

$$dr(\tau) = (\alpha + \beta \cdot r(\tau))\,d\tau + \sigma \cdot r^{\gamma}(t_i)\,dW_\tau \text{ for all } \tau \in [t_i, t_{i+1}). \qquad (5)$$

[4] ANDREWS (1991) and NEWEY AND WEST (1994) provide methods for bandwidth selection. The corresponding methods remove some of the judgmental aspects of selecting the bandwidth but do not completely solve the problem since the exact choice of several parameters is not specified; see e.g. HALL (2005) or HARRIS AND MÁTYÁS (1999). However, we are not interested in consistent estimation of the variance-covariance matrix but in the asymptotic behaviour of the test statistics.

This assumption facilitates the construction of a discrete time version of the model. It follows that the exact discrete-time model corresponding to (5) is given by

$$r(t_{i+1}) = \exp\{\Delta t \cdot \beta_0\} \cdot r(t_i) + \frac{\alpha_0}{\beta_0}\left(\exp\{\Delta t \cdot \beta_0\} - 1\right) + \varepsilon_0(t_{i+1}), \qquad (6)$$

where the error term $\varepsilon_0(..)$ satisfies

$$E\left(\varepsilon_0(t_{i+1}) \mid r(t_i)\right) = 0 \text{ and } E\left(\varepsilon_0^2(t_{i+1}) \mid r(t_i)\right) = \frac{\sigma_0^2}{2 \cdot \beta_0} \cdot \left(\exp\{\Delta t \cdot 2 \cdot \beta_0\} - 1\right) \cdot r^{2\cdot\gamma_0}(t_i). \quad (7)$$

Using (6) and (7), the subsequent GMM estimation – to which we refer as the alternative estimation – is then constructed analogously to the procedure applied by CKLS (1992). Since only the diffusion function is approximated in (6), this alternative estimation method is supposed to reduce aggregation bias relative to full discretization.

It is commonly known that standard procedures ML can lead to biased coefficient estimators while estimating autoregressive models. Since both discrete-time approximations given in (2) and (6) can be regarded as autoregressive models of order one with conditional heteroskedastic errors, the question of a bias reduction method arises. The so-called jackknife technique of QUENOUILLE (1956) to bias reduction is suggested by YU AND PHILLIPS (2005) within the framework of the ML procedure for estimating continuous-time models. Under the assumption that the bias of the estimates can be expanded asymptotically in a series of increasing powers of n^{-1}, it can be easily shown that the bias of the jackknife estimation is of order $O(n^{-2})$.

We apply the jackknife estimation within the GMM estimation framework of CKLS (1992) as follows. An observed sample with n observations is decomposed into $\lambda \geq 2$ consecutive sub-samples, each with τ observations, such that $n = \lambda \cdot \tau$. Then the jackknife estimator $\hat{\theta}_{jack}$ of the parameter vector θ is given by

$$\hat{\theta}_{jack} = \frac{\lambda}{\lambda - 1} \cdot \hat{\theta} - \frac{1}{\lambda \cdot (\lambda - 1)} \cdot \sum_{j=1}^{\lambda} \hat{\theta}_j \text{ for } \lambda \geq 2,$$

where $\hat{\theta}_j$ symbolizes a corresponding GMM estimate of θ_0 obtained from the j-th sub-sample with $j = 1,...,\lambda$. For estimation, the simple choice of two sub-samples is often suitable.

4 Monte Carlo Results

Within all Monte Carlo experiments, the realisations of standard normal pseudo-random variables are generated by the so-called ziggurat algorithm of MARSAGLIA AND TSANG (1984). Furthermore, the bias of the corresponding estimation is tested respectively using

$$\sqrt{m}\cdot\frac{\overline{\alpha}-\alpha_0}{\hat{\sigma}_{\hat{\alpha}}}\xrightarrow{d}N(0,1) \text{ as } m\to\infty\text{, under } H_0: E\left(\hat{\theta}\right)=\theta_0\text{,}$$

where m denotes the number of replications, $\overline{\alpha}$ symbolizes the sample mean of the estimates of α, $\hat{\sigma}^2_{\hat{\alpha}}$ represents an estimate for the variance, i.e. $\hat{\sigma}^2_{\hat{\alpha}}=\frac{1}{m-1}\sum_{j=1}^{m}(\hat{\alpha}_j-\overline{\alpha})^2$, and $\hat{\alpha}_j$ denotes an estimate of α computed at replication j with $j=1,...,m$. For the remaining parameters, the bias is tested analogously. Note that the normality of the estimators is not necessary for the normality of the estimation bias.

We construct simulated sample paths of (2) using $\alpha_0=0.04$, $\beta_0=-1$, $\sigma_0=0.2$, $\gamma_0=1.5$, $\Delta t=0.01$, and $r(t_0)=0.035$ from which the parameters are estimated by the method applied in CKLS (1992) respectively. We also compute the corresponding jackknife estimates using $\lambda=2$. The simulated sample paths consist of 12 500, 25 000, 50 000, and 100 000 observations and the Monte Carlo experiment is based on $m=5\,000$ replications. The results are presented in Table 2.

Table 2 The Bias of the CKLS (1992) Estimation – Monte Carlo Results

n	λ	α	β	σ	γ
12 500	1	0.04118** [17.1341]	-1.0297** [-17.1222]	0.2480** [18.1850]	1.4933* [-2.1878]
	2	0.039960 [-0.5565]	-0.9990 [0.5429]	0.2188** [6.7649]	1.4940 [-1.9358]
25 000	1	0.04060** [11.8145]	-1.0149** [-11.8146]	0.2237** [14.2299]	1.4964 [-1.6474]
	2	0.039981 [-0.3676]	-0.9995 [0.3525]	0.2041** [2.7486]	1.4966 [-1.5425]
50 000	1	0.04031** [8.6630]	-1.0077** [-8.6517]	0.2134** [11.9744]	1.4998 [-0.1332]
	2	0.039997 [-0.0815]	-0.9999 [0.0887]	0.2026** [2.5105]	1.5002 [0.1124]
100 000	1	0.04017** [6.9928]	-1.0044** [-6.9835]	0.2070** [9.1840]	1.5005 [0.4050]
	2	0.040009 [0.3722]	-1.0002 [-0.3659]	0.2006 [0.7846]	1.5007 [0.6355]

Explanation: Table 2 summarizes the Monte Carlo results for the estimation applied in CKLS (1992) obtained from simulated samples, where *n* denotes the sample size. The number of the sub-samples that is used for the jackknife estimation is denoted by λ, where $\lambda=1$ represents the non-jackknife estimates. The table reports the mean of the estimates of all parameters, where the test-statistics of the estimation bias are in brackets. The marking * (**) means that the null hypothesis that the bias is zero can be rejected at the 5% (1%) significance level respectively.

The null hypothesis is rejected at the 1% level for the estimates of α, β, and σ in all simulated paths. Only the bias of the estimates of γ does not significantly differ from zero at the 5% level – except for the estimates obtained from samples with $12\,500$ observations. The bias seems to be larger for less observations but it does not disappear even when very large samples – i.e. samples with $100\,000$ observations – are used. All in all, the estimator seems to converge in probability as $n \to \infty$ but not towards the true parameter vector.[5]

The biases of the jackknife estimators do not differ significantly from zero for α and β at the 5% level within all simulated sample paths. The same result holds for the estimates of γ. On the other hand, the null hypothesis is rejected at the 1% level for the jackknife estimates of σ that are obtained from simulated samples with $12\,500$, $25\,000$, and $50\,000$ observations. However, the null is not rejected at the 5% level for estimates of σ obtained from samples with $100\,000$ observations. Consequently, we find that the jackknife estimation provides a substantial improvement referring to the biasness of the estimation applied in CKLS (1992).

We now examine the properties of the CKLS (1992) estimation in case that the assumption that the dynamics of the short-term interest rate cannot be approximated by the discrete-time process given in (2). If the restriction $\gamma = 1$ on (1) is imposed, then the corresponding solution is given by

$$R_t = Z_t \cdot \left[R_0 + \int_0^t \frac{\alpha}{Z_s} ds \right], \text{ where } Z_t := \exp\left\{ \left(\beta - \tfrac{1}{2} \cdot \sigma^2 \right) \cdot t + \sigma \cdot W_t \right\}. \tag{8}$$

We confirm that $P(R_t > 0) = 1$ for all β and σ, if $\alpha > 0$ and $R_0 > 0$. Using the discretization of (8), namely

$$r(t_i) = z(t_i) \cdot \left[r(t_0) + \sum_{k=1}^{i} \frac{\alpha}{z(t_k)} \Delta t \right], \text{ where}$$

$$z(t_i) := \exp\left\{ \left(\beta - \tfrac{1}{2} \cdot \sigma^2 \right) \cdot i \cdot \Delta t + \sum_{k=1}^{i} \sigma \cdot \sqrt{\Delta t} \cdot \xi(t_k) \right\}, \tag{9}$$

[5] Note that sampling frequency has no consequential effects on the variance of the estimates of the drift function; see e.g. GOURIEROUX AND JASIAK (2001).

and where $\xi(..)$ is a standard normal white noise, we can generate simulated paths of (1) for $\gamma = 1$ without approximations. Hence, we refer to (9) as the exact discretization of (1) for $\gamma = 1$.

We generate simulated sample paths of (9) using $\alpha_0 = 0.04$, $\beta_0 = -1$, $\sigma_0 = 0.2$, $n = 20\,000$, $\Delta t = 0.01$, and $r(t_0) = 0.035$ from which the unrestricted model is estimated by the method applied in CKLS (1992) respectively. We also compute the corresponding jackknife estimates using $\lambda = 2$. The Monte Carlo experiment is based on $m = 5\,000$ replications. If we the bias of the estimates is found significant at the 5% level, then we refer to this feature as the aggregation bias of the corresponding estimation. For comparison, this experiment is repeated using the Euler approximation of (1) for $\gamma = 1$ to generate simulated sample paths. The results are reported in Table 3.

Table 3 Aggregation Bias of the CKLS (1992) Estimation – Monte Carlo Results

Discretization	λ	α	β	σ	γ
Exact	1	0.0410** [16.3680]	-1.0207** [-13.1079]	0.1991* [-2.3052]	0.9990 [-1.6815]
	2	0.0401* [2.1088]	-0.9986 [0.8700]	0.1980** [-5.0871]	1.0001 [0.4301]
Euler	1	0.04092** [14.9446]	-1.0232** [-14.7451]	0.2010** [2.6359]	0.9989 [-1.8672]
	2	0.04003 [0.3966]	-1.0005 [-0.3060]	0.1996 [-0.9960]	0.9995 [-0.7958]

Explanation: Table 3 summarizes the Monte Carlo results for the estimation applied in CKLS (1992) obtained from simulated samples generated by means of the exact and the Euler discretization of (1). The number of the sub-samples that is used for the jackknife estimation is denoted by λ, where $\lambda = 1$ represents the non-jackknife estimates. The table reports the mean of the estimates of all parameters, where the test-statistics of the estimation bias are in brackets. The marking * (**) means that the null hypothesis that the bias is zero can be rejected at the 5% (1%) significance level respectively.

If the CKLS (1992) estimation is applied for samples generated by the exact discretization, then only the bias of the estimates of γ and the jackknife estimates of β are not significant at the 5% level. If samples are generated by means of the Euler discretization, then the null hypothesis that the bias is zero cannot be rejected at the 5% level for the estimates of γ and for the jackknife estimates. These results demonstrate that, due to discretization bias, the parameter vector cannot be consistently estimated by CKLS (1992), if the assumption that short-term interest rate follows a discrete-time approximation is not satisfied. Note that the results probably depend on sampling frequency, since the discretization bias of the Euler discretization converges to zero as $\Delta t \to 0$, such that the aggregation bias of the estimation can be eliminated if high frequented data is used for estimation.

We now examine whether the alternative GMM estimation technique is able to reduce aggregation bias relative to full discretization. For this purpose, we repeat the Monte Carlo experiment reported in Table 3 using the alternative estimation. The results are presented in Table 4.

Table 4 Aggregation Bias of the Alternative Estimation – Monte Carlo Results

Discretization	λ	α	β	σ	γ
Exact	1	0.0411** [18.0074]	-1.0281** [-17.7332]	0.2024** [6.3688]	0.9995 [-0.9553]
	2	0.0402** [3.0986]	-1.0049** [-2.9608]	0.2010** [2.6278]	1.0001 [0.1941]
Euler	1	0.0412** [19.7794]	-1.0260** [-16.4416]	0.2002 [0.4127]	0.9990 [-1.6446]
	2	0.0403** [4.8381]	-1.0029 [-1.7852]	0.1991** [-2.3499]	1.0001 [0.1355]

Explanation: Table 4 summarizes the Monte Carlo results for the alternative estimation obtained from simulated samples generated by means of the exact and the Euler discretization of (1). For further explanations, see Table 3.

These results illustrate that the alternative estimation does not result in consistent estimates of the parameter vector, if it is applied for sample paths of continuous-time stochastic processes satisfying (2). Although the drift function is not approximated by (6), its parameters cannot be consistently estimated with the alternative method even if the corresponding jackknife estimation is applied. If the short-term interest rate process satisfies the discrete-time approximation given in (2) then we cannot consistently estimate the parameter vector by means of the alternative estimation either. Due to these results, the alternative estimation is no longer considered in this work.

We examine the asymptotic normality result given in (4) by applying the test of JARQUE AND BERA (1987) for the estimates that are obtained from the Monte Carlo study reported in Table 2. Under the null hypothesis that the estimated parameters are normally distributed, it follows that the test-statistic JB given below converges to the chi-square distribution with 2 degrees of freedom, i.e.

$$JB = m \cdot \left(\hat{\eta}_3^{\,2} / 6 + \left(\hat{\eta}_4 - 3 \right)^2 / 24 \right) \xrightarrow{d} \mathrm{X}^2 \left(df = 2 \right) \text{ as } m \to \infty ,$$

where $\hat{\eta}_3$ and $\hat{\eta}_4$ denote the sample analogues of the skewness and kurtosis of the estimated parameters. The corresponding results are reported in Table 5. Since the null-hypothesis $H_0 : \sigma_0 = 0$ would imply that the short-term interest rate is deterministic, the distribution of the estimates of σ is not considered.

Table 5 Asymptotic Normality – Monte Carlo Results

n	λ	α	β	γ
12 500	1	88.4176**	89.8376**	5.7272
		(0.0000)	(0.0000)	(0.0571)
	2	57.4285**	59.0414**	9.3410**
		(0.0000)	(0.0000)	(0.0094)
25 000	1	32.9353**	32.9154**	1.3970
		(0.0000)	(0.0000)	(0.4973)
	2	21.4347**	21.4307**	1.0067
		(0.0000)	(0.0000)	(0.6045)
50 000	1	9.4927**	9.5641**	3.7509
		(0.0000)	(0.0000)	(0.1533)
	2	8.3839*	8.4952*	2.9738
		(0.0151)	(0.0143)	(0.2261)
100 000	1	17.6187**	18.6999**	0.8835
		(0.0000)	(0.0000)	(0.6429)
	2	16.1780**	17.0215**	1.0524
		(0.0003)	(0.0002)	(0.5908)

Explanation: Table 5 shows the results of the normality tests that are applied for the estimates corresponding to Table 2. The table reports the resulting Jarque-Bera test-statistics respectively, where the *p*-values are in parenthesis. The number of the sub-samples that is used for the jackknife estimation is denoted by λ, where $\lambda = 1$ represents the non-jackknife estimates. The marking * (**) means that the null hypothesis that the parameter is normally distributed can be rejected at the 5% (1%) significance level respectively.

The results show that only the estimates of γ can be assumed to satisfy the normal distribution. This suggests that the application of significance tests on basis of the normal distribution does no seem to be appropriate for the estimates of the drift function. Due to this result, we analyze the distribution of the *t*-statistics for the estimates of β under the null hypothesis $H_0 : \alpha_0 = \beta_0 = 0$.[6] The alternative is $H_1 : \alpha_0 > 0 \wedge \beta_0 < 0$.[7] The *t*-statistics of the estimates for β are calculated by

$$t_{\hat{\beta}} = \sqrt{n} \cdot \frac{\hat{\beta}}{\sqrt{\hat{\Lambda}^{(2,2)}}},$$

where $\hat{\beta}$ is an estimate of β and $\hat{\Lambda}^{(2,2)}$ denotes the second component of the main diagonal of $\hat{\Lambda} := (G(\hat{\theta})' \hat{\Sigma}^{-1} \cdot G(\hat{\theta}))^{-1}$. We generate simulated samples by means of the Euler discretization of (1), each with $n = 1000$, under H_0 using $\sigma_0 = 0.5$, $\Delta t = 0.01$, and four

[6] In case of $\alpha_0 = \beta_0 = 0$ the short-term interest rate process has a unit root. It is well known that the tabulated distributions for the unit root tests of DICKEY AND FULLER (1979) assume that the errors are white noise. However, as shown in RODRIGUES AND RUBIA (2004), the resulting *t*-statistics do not seem to depend on the level of γ if the errors are of the form given in (3).

[7] Note that $\alpha_0 \leq 0 \wedge \beta_0 < 0$ would imply that the short-term interest rate follows a stationary process with non-positive mean. For practical reasons, this case is not considered in this work.

different values for γ respectively. The Monte Carlo experiment is based on $m = 30\,000$ replications. This experiment is repeated for the jackknife estimates with $\lambda = 2$, whereas the variance-covariance matrices are recalculated using the jackknife estimates. The results are presented in Table 6.

Table 6 Empirical Percentiles of the t-Statistics Depending on γ – Monte Carlo Results

γ	λ	0.5%	1%	5%	10%	50%	90%	95%	99%
0.0	1	-5.4445	-5.0488	-4.0131	-3.5361	-2.1717	-0.8655	-0.4257	0.4201
	2	-7.0040	-6.8055	-6.2242	-5.9051	-4.7792	-3.4583	-2.8316	-1.1035
0.5	1	-5.0691	-4.6604	-3.5231	-3.0179	-1.5972	-0.6037	-0.0765	1.0833
	2	-6.3508	-6.1211	-5.4167	-5.0250	-3.5730	-2.2367	-1.6158	0.9939
1.0	1	-4.0611	-3.7818	-3.0130	-2.6424	-1.4912	-0.6857	-0.4740	0.4626
	2	-6.3984	-6.1136	-5.1873	-4.6648	-3.1080	-2.0637	-1.8379	-1.4249
1.5	1	-3.8603	-3.5938	-2.8894	-2.5481	-1.4568	-0.3016	0.0003	0.6735
	2	-6.7506	-6.4567	-5.6701	-5.1942	-3.4947	-2.1947	-1.9035	-1.3826

Explanation: Table 6 contains the empirical percentiles of the distributions of the *t*-statistics of β that result from the corresponding estimation applied in CKLS (1992) under $H_0: \alpha = \beta = 0$ obtained from simulated samples. The number of the sub-samples that is used for the jackknife estimation is denoted by λ, where $\lambda = 1$ represents the non-jackknife estimates.

As the results show, the empirical distribution of the *t*-statistics, unlike the *t*-statistics that are obtained from unit root tests, seems to depend on the value of γ that is used to generate simulated sample paths. Using our results, however, we cannot identify any kinds of patterns for the percentiles concerning the dependence of the distribution on γ.

Using a Monte Carlo experiment based on $m = 30\,000$ replications, we also examine the distribution of the LR-statistics for all models. We generate simulated samples, each with $n = 1000$, by means of the Euler discretization of (1) using $\sigma_0 = 0.5$, $\Delta t = 0.01$, and the corresponding restrictions on the parameter vector imposed by the given null models respectively. The results are presented in Table 7.

Table 7 Empirical Percentiles of the LR-Statistics – Monte Carlo Results

Model	*s*	1%	5%	10%	50%	90%	95%	99%
MERTON (1973)	2	0.0224	0.1132	0.2264	1.4550	4.8592	6.3411	9.8118
VASICEK (1977)	1	0.0003	0.0072	0.0295	0.8293	4.8179	6.8982	12.2344
DOTHAN (1978)	3	0.6319	1.1431	1.5804	4.5747	10.6623	13.2419	19.6428

CIR (1980)	3	0.6684	1.1995	1.6488	4.6841	10.6841	13.4771	20.3675
BS (1980)	1	0.0004	0.0076	0.0302	0.8567	5.6433	8.4467	17.2100
CIR (1985)	1	0.0004	0.0103	0.0412	1.0029	4.9780	7.0076	12.5912
GBM	2	0.0243	0.1268	0.2671	1.8278	6.3591	8.3170	13.1324

Explanation: Table 7 contains the empirical percentiles of the distributions of the LR-statistics obtained from simulated samples under the given null models, where *s* denotes the number of restrictions.

As the results show, the distribution of the LR-statistics seems to depend on both the number of restrictions and the underlying null models. Only the empirical percentiles that result from the DOTHAN (1978) and from the CIR (1980) models – both use $s = 3$ restrictions – do not considerably differ from each other respectively. The resulting distributions of the LR-statistics do not tend to be well approximated by the chi-square distribution in general. Only the LR-statistics that are obtained from the MERTON (1973) model are assumed to follow the chi-square distribution under the corresponding null hypothesis.

5 Empirical Results

The concept of the short-term interest rate is not unambiguous from the practical point of view, such that the choice for an appropriate proxy has to be made. Since one-month Treasury bill yields are assumed to be affected by idiosyncratic variation – as shown in DUFFEE (1996) –, we use samples of daily, weekly and monthly observations of the three-month U.S. Treasury bill yield from 04.01.1954 through 02.03.2006 which are provided by the Board of Governors of the Federal Reserve System. We also use daily, weekly, and monthly observations of the Federal fund rate from 04.01.1954 through 02.03.2006.

We report the descriptive statistics for daily, weekly, and monthly observations of U.S. Treasury bill yields and Federal funds rates respectively. We test the normality of the observations using JARQUE AND BERA (1987). We also report the results from the application of the unit root test of SAID AND DICKEY (1984) using the corresponding critical values. For the test a constant is included. We use AKAIKE (1973) to determine the number of lagged differences. The results are presented in Table 8.

Table 8 Summary Statistics

Data	Frequency	*n*	Mean	Standard Deviation	*JB*	*ADF*	Lags
Treasury bill yield	*Daily*	13 028	5.1755%	2.8327%	4048**	-2.7421	221

	Weekly	2 721	5.1815%	2.8336%	831**	-2.5218	42
	Monthly	625	5.1816%	2.8326%	183**	-2.3816	12
Federal fund rate	*Daily*	18 873	5.7133%	3.4039%	7496**	-2.2553	211
	Weekly	2 695	5.7135%	3.3899%	1026**	-2.4260	52
	Monthly	619	5.7149%	3.3827%	222**	-2.4230	11

Explanation: contains descriptive statistics of daily, weekly, and monthly observations of three-month U.S. Treasury bill yields and Federal fund rates from 04.01.1954 through 02.03.2006. The number of observations is denoted by *n, JB* denotes the Jarque-Berra test-statistic, *ADF* denotes the *t*-statistic obtained from the test of SAID AND DICKEY (1984) with the corresponding number of lagged differences that are chosen using AKAIKE (1973). The maximum of the lagged differences that is considered is 250 for daily, 52 for weekly, and 12 for monthly observations. The critical values that we use are -2.8865 and -3.4752 for the 5% and the 1% level. The marking * (**) means that the corresponding null hypothesis is rejected at the 5% (1%) level respectively.

The null hypothesis of the Jarque-Bera test is rejected at the 1% level for both instruments and for all types of sampling frequencies respectively. The results of the unit root tests suggest that, from the empirical point of view, neither the Treasury bill yield nor the Federal fund rate can be assumed to follow a stationary process at the 5% level.

Since the parameters of the drift function are linear and exactly identified by the first two equations of the moment conditions that do not include the remaining parameters, only the estimation of σ and γ requires a numerical optimization for which the NR method is applied. The main problem with the NR method is that global optimization is not guaranteed since many local minima can be found in general depending on the initial value that is chosen. However, the solution of the optimization is only sensitive to the choice of the initial value of γ, while the solution seems to be very robust to the choice of σ. Since only one solution can be obtained from initial values of $0 \leq \gamma \leq 3$ respectively – this result holds for all types of observations of both three-month U.S. Treasury bill yields and Federal fund rate –, we assume that the local minima that are found are the solutions of the corresponding global optimizations.

Now, we estimate the unrestricted model by CKLS (1992) for daily, weekly and monthly observations of three-month Treasury bill yields and Federal fund rates, both from 04.01.1954 through 02.03.2006, by means of the NR method using the initial values $\sigma = 0.1$ and $\gamma = 0$ respectively. For estimation, we use $\Delta t = 1/250$ for daily, $\Delta t = 1/52$ for weekly, and $\Delta t = 1/12$ for monthly observations. We also compute the corresponding jackknife estimates using $\lambda = 2$. We test the null hypothesis $H_0 : \alpha_0 = \beta_0 = 0$ using the percentiles of the empirical distributions as critical values given in Table 6 depending on the estimation result of γ. Due to our Monte Carlo results, we assume that the estimates of γ are asymptotically normal under the null hypothesis $H_0 : \gamma_0 = 0$. The results are presented in Table 9.

Table 9 Estimation Results of the Unrestricted Model

Data	Frequency	λ	α	β	Critical Values	σ	γ	$-\alpha/\beta$
Treasury bill yield	*Daily*	1	0.0087	-0.1571	-2.89 (5%)	0.8782	1.4978	0.0556
				[-0.8098]	-3.59 (1%)		[19.1493]	
		2	0.0111	-0.2174	-5.67 (5%)	0.8882	1.5696**	0.0510
				[-1.1202]	-6.46 (1%)		[13.0189]	
	Weekly	1	0.0083	-0.1492	-2.89 (5%)	0.9441	1.5410**	0.0559
				[-0.8378]	-3.59 (1%)		[13.1749]	
		2	0.0110	-0.2177	-5.67 (5%)	0.9158	1.6474**	0.0506
				[-1.2212]	-6.46 (1%)		[6.8538]	
	Monthly	1	0.0095	-0.1728	-2.89 (5%)	1.3499	1.6595**	0.0552
				[-1.3017]	-3.59 (1%)		[7.5542]	
		2	0.0126	-0.2538	-5.67 (5%)	1.5718	2.0026**	0.0498
				[-1.9092]	-6.46 (1%)		[2.9000]	
Federal fund rate	*Daily*	1	0.0828	-1.4407**	-3.52 (5%)	0.2909	0.5713**	0.0574
				[-6.5962]	-4.66 (1%)		[7.9348]	
		2	0.0780	-1.2880*	-5.42 (5%)	0.0685	0.4128	0.0606
				[-5.9912]	-6.12 (1%)		[0.5109]	
	Weekly	1	0.0155	-0.2602	-3.01 (5%)	0.2321	0.8283**	0.0597
				[-1.2099]	-3.78 (1%)		[5.7759]	
		2	0.0200	-0.3646	-5.19 (5%)	0.1878	0.8758**	0.0548
				[-1.6857]	-6.11 (1%)		[2.7258]	
	Monthly	1	0.0110	-0.1808	-2.89 (5%)	1.6553	1.7666**	0.0609
				[-1.3229]	-3.59 (1%)		[5.4369]	
		2	0.0143	-0.2699	-5.67 (5%)	1.7200	1.9869**	0.0530
				[-1.9778]	-6.46 (1%)		[3.1031]	

Explanation: Table 9 contains the estimation results of the unrestricted model for daily, weekly, and monthly observations of three-month U.S. Treasury bill yields and Federal fund rates from 04.01.1954 through 02.03.2006 with the corresponding estimates. The number of the sub-samples that is used for the jackknife estimation is denoted by λ, where $\lambda = 1$ represents the non-jackknife estimates. The term $-\alpha/\beta$ represents the estimated long-run mean. The corresponding *t*-statistics are in brackets. The critical values are taken from Table 6. The marking * (**) means that the null hypothesis that the estimated parameter is zero is rejected at the 5% (1%) level.

As the results illustrate, the null hypothesis $H_0 : \alpha_0 = \beta_0 = 0$ cannot be rejected at the 5% level for all types of sampling frequencies of the three-month U.S. Treasury bill yields. From that it follows that there appears to be no empirical evidence for mean-reversion in the three-month U.S. Treasury bill yield. In contrast to that result, the null hypothesis is rejected at the 1% level for daily observations of the Federal fund rates, while, for the corresponding jackknife estimates, the null is rejected at the 5% level. Therefore we find that mean-reversion plays an important role for the specification of the Federal fund rate dynamics but only for daily observations.

Our results also indicate that the conditional variance of changes in the three-month U.S. Treasury bill yield is highly sensitive to the yield level. However, for the Federal fund rates, the estimated value of γ increases from daily to monthly observations. From the empirical point of view, the conditional variance of changes in daily observations of the Federal fund rate is approximately proportional to the rate level, while, for weekly and for monthly observations, the conditional variance seems to be more sensitive.

Now, we estimate the restricted models by means of the NR method using the estimates of the unrestricted case as initial values.[8] We calculate the corresponding LR-statistics. For critical values, we use the percentiles given in Table 7 – except for MERTON (1973) for which the chi-square distribution with 2 degrees of freedom is used. Since the empirical distribution of the LR-statistics under the given null model is known, we do not apply the jackknife estimation for the restricted models. The results are presented in Table 10.

Table 10 Estimation Results of the Restricted Models

Model	Data	Daily	Weekly	Monthly
MERTON (1973)	*Treasury bill yield*	109.8111**	42.9676**	10.33**
	Federal fund rate	59.8113**	22.0294**	7.5315*
VASICEK (1977)	*Treasury bill yield*	108.0119**	42.3872**	9.8946*
	Federal fund rate	34.9138**	20.3528**	5.9502
DOTHAN (1978)	*Treasury bill yield*	32.9966**	17.2741*	5.8133
	Federal fund rate	83.7127**	7.1675	5.1990
CIR (1980)	*Treasury bill yield*	2.6802	2.7342	3.2011
	Federal fund rate	117.4089**	26.1764**	3.6026
BS (1980)	*Treasury bill yield*	29.6006**	14.0948*	4.1562
	Federal fund rate	72.0333**	1.4664	2.5947
CIR (1985)	*Treasury bill yield*	74.5311**	30.9254**	7.5155*
	Federal fund rate	0.9317	4.5476	4.5119
GBM	*Treasury bill yield*	32.2474**	16.4096**	5.5239
	Federal fund rate	77.3434**	3.8977	4.9823

Explanation: Table 10 contains the LR-statistics of the restricted models for daily, weekly, and monthly observations of three-month U.S. Treasury bill yields and Federal fund rates from 04.01.1954 through 02.03.2006. Except for MERTON (1973), the critical values for the LR-statistics from Table 7 are used. The marking * (**) means that the corresponding null model is rejected at the 5% (1%) level.

To sum up, for daily and weekly observations of the three-month U.S. Treasury bill yield, all models except the model of CIR (1985) are rejected at the 5% level. This result indicates that the conditional variance of changes in daily and weekly observations of the three-month U.S.

[8] Note that the stopping rule of ANDREWS (1997), which is suggested for estimation of over-identified systems within the GMM framework, is not applicable, since the weighting matrix is calculated using of the estimates of the unrestricted case.

Treasury bill yield is highly sensitive to the yield level, while the mean reversion is negligible from the empirical point of view. Only the specifications of MERTON (1973), VASICEK (1977), and CIR (1985) can be rejected at the 5% level for monthly observations of the three-month U.S. Treasury bill yield, such that monthly observations allow a lower elasticity of the conditional variance of yield changes. The reason for this result is that the variance of the estimates of γ is lower for higher sampling frequencies. Since only the model of CIR (1980) is not rejected at the 5% level for all types of sampling frequencies, we find that the model of CIR (1980) best describes the dynamics of the three-month U.S. Treasury bill yield.

The results obtained from the Federal fund rates depend on the chosen sampling frequency. While, for daily observations, the mean- reversion appears to be more important, this feature is negligible for weekly and monthly observations. For weekly and monthly observations, the elasticity seems to be higher than for daily observations. As for the three-month U.S. Treasury bill yield, an exact specification of the elasticity of conditional variance is more important for daily observations. Only the specification of MERTON (1973) can be rejected at the 5% level for monthly observations of the Federal fund rate. Since only the model of CIR (1985) is not rejected at the 5% level for all types of sampling frequencies, we find that the model of CIR (1985) best describes the dynamics of the Federal fund rate.

Since several models are sub-nested within other models, the performance of a given model can also be measured relative to the model in which it is nested. The corresponding weighting matrix, which is used for both criterion functions, is obtained using the estimates of the alternative unrestricted model in each case. Following CKLS (1992), we assume that the resulting LR-statistics converge under the associated null hypothesis to the chi-square distribution with s degrees of freedom, where s denotes the number of restrictions imposed by the restricted nested model on the corresponding alternative unrestricted model. The results of the tests are summarized in Table 11.

Table 11 Pairwise Comparisons of Alternative Nested Models

Alternative Model	Restricted Nested Model	s	Data	Daily	Weekly	Monthly
VASICEK (1977)	MERTON (1973)	1	*Treasury bill yield*	1.1500 (0.2835)	0.3947 (0.5298)	0.4440 (0.5052)
			Federal fund rate	33.7398** (<0.0001)	1.6875 (0.1939)	1.3915 (0.2382)
GBM	DOTHAN (1978)	1	*Treasury bill yield*	0.9346 (0.3337)	0.9575 (0.3278)	0.3982 (0.5280)
			Federal fund rate	10.3885** (0.0013)	2.8434 (0.0917)	0.2871 (0.5921)

BS (1982)	DOTHAN (1978)	2	*Treasury bill yield*	2.8015 (0.2464)	2.4806 (2.4806)	1.4136 (0.4932)
			Federal fund rate	11.4605** (0.0032)	5.6384 (0.0597)	2.2511 (0.3245)
BS (1982)	GBM	1	*Treasury bill yield*	1.9217 (0.1657)	1.6608 (0.1975)	1.1549 (0.2825)
			Federal fund rate	2.9909 (0.0837)	2.4354 (0.1186)	2.0894 (0.1483)

Explanation: Table 11 contains the values of the LR-statistics that are obtained from pairwise comparisons of the alternative models for daily, weekly, and monthly observations of three-month U.S. Treasury bill yields and Federal fund rates from 04.01.1954 through 02.03.2006. The number of restrictions imposed by the restricted nested model relative to the alternative model is denoted by *s*. The corresponding *p*-values are in parenthesis. The marking * (**) means that the corresponding null hypothesis is rejected at the 5% (1%) level.

For the three-month U.S. Treasury bill yield, as shown, none of the underlying nested models can be rejected against the corresponding alternatives at the 5% level. This result confirms that, for the three-month U.S. Treasury bill yield, none of the models with mean-reverting processes outperform the nested models with different specifications of the drift function respectively.

For weekly and monthly observations of the Federal fund rate, the corresponding null hypotheses cannot be rejected at the 5% either. On the other hand, for daily observations, the MERTON (1973) model is rejected at the 1% level against the VASICEK (1977) model, and the DOTHAN (1978) model is rejected at the 1% level against its both of its alternatives, namely the specification of BS (1980) and the GBM. That result also indicates that, for daily observations of the Federal Funds rate, there is strong evidence for mean-reversion from the empirical point of view. In contrast to that result, the GBM cannot be rejected at the 5% level against the alternative model of BS (1980).

6 Conclusion

This work illustrates that the GMM estimation applied in CKLS (1992) for estimating continuous-time models of the short-term interest rate suffers from significant estimation bias which is reduced by means of the jackknife estimation under the assumption that the dynamics of the short-term interest rate can be approximated by means of a discrete-time process.

We provide critical values for parameter tests, obtained from empirical distributions of the associated test-statistics. We show that the associated *t*-statistics of the drift parameters depend on the elasticity of the conditional variance of changes in the short-term interest rate under the null hypothesis that the drift function is zero, whereas the *t*-statistics obtained from unit root tests are robust to conditional heteroskedastic errors. We find that the distributions of the LR-statistics of the corresponding null models do not strictly and exclusively depend on the number of restrictions imposed by the underlying null models but also on the given model that is considered. Our Monte Carlo results also illustrate, that an alternative GMM estimation based on the discretization of NOWMAN (1997) does not sufficiently reduces the estimation bias for the drift parameters caused by neglecting internal dynamics between sampling points.

Using our estimation results obtained from daily, weekly, and monthly observations of the three-month U.S. Treasury bill yield and the Federal fund rate, we demonstrate that the models that are chosen can depend on both the sampling frequency and the proxy that is used for the short-term interest rate. While daily observations of the Federal fund rate seem to exhibit significant mean-reversion, the specification of the drift function seems to be of secondary importance for the dynamics of the three-month U.S. Treasury bill yield.

We also demonstrate that the jackknife estimation can result in higher values for the elasticity of conditional variance of changes in the yield level for the three-month U.S. Treasury bill yield. We find the conditional variance of changes in the yield to be highly sensitive to the yield level. This sensitivity appears to be lower for daily observations of the Federal fund rate. The results illustrate that an exact specification of the elasticity of conditional variance is more important for daily observations, since the null models are rejected more often. We find that – considering the alternative models that are examined – the model of CIR (1980) best describes the dynamics of the three-month U.S. Treasury bill yield, whereas, for the Federal funds rate, the corresponding model is CIR (1985).

References

AKAIKE, H. (1973): A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716-723.

AÏT-SAHALIA, Y. (1996): Testing Continuous-Time Models of the Spot Interest Rate. *The Review of Financial Studies*, Vol. 9, No. 2, pp. 385-426.

ANDREWS, D. W. K. (1991): Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, Vol. 59, No. 3, pp. 817-858.

ANDREWS, D. W. K. (1997): Stopping Rule for the Computation of Generalized Method of Moments Estimators. *Econometrica*, Vol. 65, No. 4, pp. 913-931.

BRENNAN, M. J., SCHWARTZ, E. S. (1980): Analyzing Convertible Bonds. *Journal of Financial and Quantitative Analysis*, Vol. 15, No. 4, pp. 907-929.

BROZE, L., SCAILLET, O., ZAKÖIAN, J. M. (1995): Testing for Continuous-Time Models of the Short-Term Interest Rate. *Journal of Empirical Finance*, Vol. 2, No. 3, pp. 199-223.

CHAN, K. C., KAROLYI, G. A., LONGSTAFF, A. F., SANDERS, A. B. (1992): An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *The Journal of Finance,* Vol. 68, No. 3, pp. 1209-1227.

COX, J. C., INGERSOLL, J. E., ROSS, S. A. (1980): An Analysis of Variable Rate Loan Contracts. *The Journal of Finance*, Vol. 35, No. 2, pp. 389-403.

COX, J. C., INGERSOLL, J. E., ROSS, S. A. (1985): A Theory of the Term Structure of Interest Rates. *Econometrica*, Vol. 53, No. 2, pp. 385-407.

DOTHAN, U. L., (1978): On the Term Structure of Interest Rates. *Journal of Financial Economics*, Vol. 6, No. 1, pp. 59-69.

DUFFEE, G. R. (1996): Idiosyncratic Variation of Treasury Bill Yields. *The Journal of Finance*, Vol. 51, No. 2, pp. 527-551.

GOURIEUX, C., JASIAK, J. (2001): Financial Econometrics, Problems, Models, and Methods. Princeton University Press, New Jersey.

JARQUE, C. M., BERA, A. K. (1987): A Test for Normality of Observations and Regression Residuals. *International Statistical Review*, Vol. 55, No. 2, pp. 163-172.

HALL, A. R. (2005): Generalized Method of Moments. Oxford University Press, New York.

HANSEN, L. P. (1982): Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, Vol. 50, No. 4, pp. 1029-1054.

HARRIS, D., MÁTYÁS, L. (1999): Introduction to the Generalized Method of Moments Estimation. Cambridge University Press, Cambridge.

MARSAGLIA, G., TSANG, W. (1984): A Fast, Easily Implemented Method for Sampling from Decreasing or Symmetric Unimodal Density Functions. *SIAM Journal on Scientific and Statistical Computing*, Vol. 5, No. 2, pp. 349-359.

MELINO, A. (1994): Estimation of Continuous-Time Models in Finance. Advances in Econometrics, Sixth World Congress, Vol. 2, Cambridge University Press, Cambridge.

MERTON, R. C. (1973): Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, Vol. 4, No. 1, pp. 141-183.

NEWEY, W. K., WEST, K. D. (1987a): A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, Vol. 55, No. 3, pp. 703-708.

NEWEY, W. K., WEST, K. D. (1987b): Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review*, Vol. 28, No. 3, pp. 777-787.

NEWEY, W. K., WEST, K. D. (1994): Automatic Lag Selection in Covariance Matrix Estimation. *The Review of Economic Studies*, Vol. 61, No. 209, pp. 631-653.

NOWMAN, K. B. (1997): Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates. *The Journal of Finance,* Vol. 52, No. 4, pp. 1695-1706.

QUENOUILLE, M. H. (1956): Notes on Bias in Estimation. *Biometrika*, Vol. 43, No. 3/4, pp. 353-360.

RODRIGUES, P. M. M., RUBIA, A. (2004): On the Small Sample Properties of Dickey Fuller and Maximum Likelihood Unit Root Tests on Discrete-Sampled Short-Term Interest Rates. Working Paper, Department of Financial Economics, University of Algarve.

SAID, S. E., DICKEY, D. A. (1984): Testing for Unit Root in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*, Vol. 71, No. 3, pp. 599-607.

VASICEK, O. A. (1977): An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, Vol. 5, No. 2, pp. 177-188.

YU, J., PHILLIPS, P. C. B. (2005): Jackknifing Bond Option Prices. *The Review of Financial Studies*, Vol. 18, No. 2, pp. 707-742.