CEM

Center for Empirical Macroeconomics

Working Paper No. 39

Modeling Short Term Interest Rates

by

Chih-ying Hsiao and Willi Semmler

University of Bielefeld
Department of Economics
Center for Empirical Macroeconomics
P.O. Box 100 131

33501 Bielefeld, Germany

Modeling Short Term Interest Rates

Chih-ying Hsiao and Willi Semmler

Abstract

The objective of this paper is to model short term interest rate data. Three problems are studied in this paper. First, if we assume that short term interest rates are discrete-time observations of a diffusion process, which discretization methods are preferable for parameter estimation and prediction? Second, can we accept the assumption that the short term interest rates are from a discretely observed diffusion process? Third, if they are not, how can we improve the modeling of the short term interest rates? We commence by taking the diffusion process model suggested by Chan et al (1992) as the data generating process. We employ three discretization methods: the Euler method, the Milstein method and the new local linearization method to obtain discrete-time approximate models. In our numerical experiment three approximate models can be accepted as correctly specified and the Euler model, in contrast to some other results in the literature, is not inferior to the other two models. Then we apply these discrete-time approximate models to the short term interest rate data of Germany, the United Kingdom and the U.S. In contrast to the numerical results, all discrete-time models fail to pass the specification test. Compared to the numerical results this indicates that the model suggested by Chan et al.(1992) is very unlikely to be the data generating process for the short term interest rate. Therefore, we search for suitable models for the short rate. We do not find an appropriate model of diffusion processes which can reproduce the stylized facts we are concerned with in this paper. Therefore, we turn to a discrete-time framework in this search. We employ an ARMA-ARCH model with level-dependent volatility for the short term interest rates. The new model can provide better level and volatility forecasts.

JEL classification: C5; E47
Keywords: short term interest rates, diffusion processes, discrete-time approximations, ARCH, level-dependent volatility

[1] Dept. of Economics, University of Bielefeld, 33615 Bielefeld, Germany, Email:chsiao@wiwi.uni-bielefeld.de
[2] Center for Empirical Macroeconomics, University of Bielefeld, 33615 Bielefeld, Germany, Email:wsemmler@wiwi.uni-bielefeld.de.

1 Introduction

A *diffusion process* is the solution of a *stochastic differential equation* (SDE)

$$dX_t = b(X_t, \theta)dt + a(X_t, \theta)dW_t,$$

where $(W_t)_{t \geq 0}$ is a *Brownian motion.* [3] In modern finance theory, diffusion processes are often used to model financial time series data, for example the *short term interest rate.* The short term interest rate is important in characterizing the *term structure* of interest rates, which means the structure of interest rates with different maturities, and in pricing *interest rate contingent-claims.* There is some pioneering work, for example by Vasicek (1997) and Cox, Ingersoll and Ross (1985). A survey of recent work is given in Chan, Karolyi, Longstaff and Sanders (1992). Chan et al. (1992) show that a wide variety of well-known one-factor models for short rates can be nested within the following SDE

$$dX_t = (c - \beta X_t)dt + \sigma X_t^{\gamma} dW_t. \tag{1}$$

The feature of this equation is that it has a mean-reverting drift coefficient [4] and a level-dependent diffusion coefficient.

Such *continuous-time* framework can provide elegant expressions in theory, but it entails some difficulty in empirical research. The first problem is how to estimate the parameters of this continuous-time models. Many methods are developed to implement the estimations, for example, the indirect inference method by Gouriéroux, Monford and Renault (1993), the approximate likelihood method by Perdersen (1995), the general method of moment with respect to diffusion generators by Hansen and Scheinkman (1995) and Duffie and Glynn (2001), the efficient method of moment by Gallant and Tauchen (1996), the nonparametric method by Ait-Sahalia (1996) and Ait-Sahalia (1997), the density-approximation method by Dacunha-Castelle and Florens-Zmirou (1986) and Ait-Sahalia (1999), the Milstein method by Elerian (1998) and in this paper [5], the new local linearization (NLL) method developed by Shoji and Ozaki (1997) and (1998) .

The second problem, coming from the continuous-time modeling — which is more basic and important for the empirical research — is to judge the specification of the employed model with respect to the empirical data. Thompson

[3] The stochastic integration with respect to dW_t is the *Itô integration*, see Karatzas and Shreve (1991).

[4] If the process deviates from $\frac{c}{\beta}$ (the mean), for example, $X_t > \frac{c}{\beta}$, then the process is drifting down and it is pulled up when $X_t < \frac{c}{\beta}$.

[5] The application of the Milstein method for approximating diffusion processes is independently developed by the authors. In the appendix of this paper we present our application and show that it is equivalent to that of Elerian (1998).

(2002) provides specification tests for diffusion processes.

In this paper we employ three discretization methods so that we can solve the two problems mentioned above at the same time. The three discretization methods are *the Euler method, the NLL (new local linearization) method and the Milstein method*. These three methods deliver discrete-time approximate models for discrete-time observed data of a diffusion process. We can implement the *maximum likelihood estimation* (ML estimation) and prediction quite easily by using these approximate models. For testing the model specification of the three discrete-time models we pursue the following strategy. By using the discrete-time approximations, we can easily transform the economical time series into a *white noise process* which is independently and identically (i.i.d) normally distributed. [6] So we test whether the estimated white noise in each discrete-time approximate models is i.i.d. normally distributed. The intuition thus is that, if the discrete-time approximation can represent the *data generating process* correctly, then we can remove all deterministic structure correctly. It is important for data prediction: the more we know about the deterministic structure, the better we can predict data.

We will compare the performance of the three discrete-time approximations. The Euler approximation is the easiest and the most used as discrete-time approximation. Its disadvantage is well-known: the Euler estimator [7] is not consistent [8]. The Milstein and NLL approximations are shown to improve the Euler approximation, see Elerian (1998:11,Table 1) and Shoji and Ozaki (1997:494-501). The improvement in their papers is represented by smaller errors of the *parameter estimations* in the numerical experiments.

Our paper will discuss the inconsistency of the Euler estimator and investigate those improvements. For evaluating discrete-time approximate models, besides considering the accuracy of parameter estimation, we still consider the accuracy of prediction. For the SDE (1) where the drift coefficient is linear, we find that the Euler and the NLL approximations are equivalent under reparametrization. Therefore they have the same predictor. [9] Moreover, we can derive a functional relation between the estimate of the Euler approximation and the estimate of the NLL approximation. Using this relation we can explain why the NLL approximation performs better than the Euler approximation in Shoji and Ozaki (1997). [10] Thus, in the numerical experiment, we need not to consider the NLL method.

We compare the Euler and the Milstein approximations in our numerical ex-

---

[6] The white noise in the discrete-time models is represented by Brownian increments ΔW_t

[7] It means the ML estimator by using the Euler method.

[8] See Lo (1988).

[9] See Section 3.

[10] We presume this is why Shoji and Ozaki applied the NLL method for the nonlinear drift case in Shoji and Ozaki (1998).

periment using Monte-Carlo-simulations. Our results, however, in contrast to Elerian (1998), does not verify the superiority of the Milstein approximation over the Euler approximation. The parameter estimations and the one-step ahead predictions of the two models are very similar. The reason is the small size of our drift parameters. The small size of parameters has a similar effect as small discretization steps, because the observed variable evolves less for smaller parameters or during shorter evolution intervals. We know if the discretization steps are small, then the effect of the discretization is also small. The reason why we employ such small parameters is because they are suggested by our empirical results of the short rate data[11].

By considering the model specification we diagnose the estimated white noise of the Euler and the Milstein approximate models. The estimated white noise of the two approximations pass our specification test for most simulations. This means that the Euler and Milstein approximate models can recognize the deterministic structure of the real data correctly. We also find that the Milstein approximation can reduce the continuous-time effect better than the Euler approximation with respect to the distribution of the estimated white noise. We observe that for a large γ the rejection frequency of the distribution test of the Milstein method is smaller than that of the Euler method.

Beside the numerical experiment we also apply the Euler and Milstein approximate models to the short term interest rate data of Germany, United Kingdom and the U.S. We take data after the oil crisis, for 1983.01 - 2000.06, because many researchers have found evidence of regime changes for the crisis period 1979-1982 . As in the numerical experiment, we implement the ML estimation, the one-step ahead prediction and test the model specification. Two approximate models perform quite similarly. The results here indicate a significant difference between the simulated and real data: none of the short rate data can pass our specification test. The estimated white noise of all the three countries has high autocorrelation and thick tails. It is not the case for the simulated data. Therefore, we conclude that the real short rate data are very unlikely to be generated by the diffusion process of the equation (1).

The next step is to find new models which can explain the autocorrelation and the thick tails of the noise. In the continuous-time framework there is some work pointing out the shortcomings of the one-factor diffusion process of the equation (1), see for example Ait-Sahalia (1996) and Andersen and Lund (1997). However, the data simulated by those continuous-time models still can not explain the high autocorrelation of the estimated white noise either.

Since we can not find a suitable model in the continuous-time framework we turn to the discrete-time framework. We employ the autoregressive-moving-average (ARMA) model to fit the high autocorrelations of the estimated white noise. We will see, in Section 7, that we can model the autocorrelation of the

[11] See Section 6.

estimated noise by taking more lags in the models. To model the thick tails in the estimated white noise we follow the work of Brenner et al. (1996) and Koediji et al. (1997). They employ the autoregressive conditional heteroscedastic (ARCH) model suggested by Engle (1982) and Bollerslev (1986) to model the thick tail. In addition, they keep the dependence of the conditional variance on the leverage of short term rates as in the diffusion coefficient of SDE model (1). Summarizing the two modeling strategies above we employ the model with *ARMA-ARCH* and *level-dependent volatility*. Our model generalizes the model of Brenner et al. (1996) by using the ARMA-structure

The remainder of the paper is organized as follows. Section 2 introduces the three discretization methods. In Section 3 we introduce the prediction briefly and show the following two properties. We show an example that the Euler approximate model provides a consistent predictor although it's parameter estimator is inconsistent. We show also that the Euler and NLL approximate models for the SDE (1) have the same predictor, because the two models are equivalent under reparametrization. In Section 4 we introduce our specification tests. In Section 5 we carry out a numerical experiment with Monte-Carlo-simulations. In Section 6 the Euler and the Miltein approximations will be applied to the real short rate data. There we can find the evidence of the model misspecification. In Section 7 we observe at first the misspecification of the two further continuous-time models. Then we employ the ARMA-ARCH model with level-dependent volatility to model the short term interest rates. Section 8 concludes the paper.

2 Discrete-Time Approximation

The difficulty of the maximum likelihood (ML) estimation based on discrete-time observation is well-known in the literature, see Lo (1988). In this paper we employ discrete-time approximate models so that the ML estimation, prediction and the model specification test are feasible. Here we introduce briefly the three methods of discrete-time approximation: the Euler, the Milstein and the new local linearization (NLL) method.

2.1 Euler Method

The idea of the Euler method is to replace dt in the equation (1) by a time interval δt and we have a discrete-time approximation for the diffusion process X

$$X_{t_{i+1}} - X_{t_i} = b(X_{t_i}, \theta)\Delta t_i + a(X_{t_i}, \theta)\Delta W_{t_i} \ . \tag{2}$$

2.2 Milstein Method

The Milstein method approximates the SDE by the following scheme:

$$
\begin{aligned}
X_{t_{i+1}} - X_{t_i} =& b(X_{t_i}, \theta)\Delta t_i + a(X_{t_i}, \theta)\Delta W_{t_i} \\
&+ \frac{1}{2} a(X_{t_i}) a'(X_{t_i})((\Delta W_{t_i})^2 - \Delta t_i)
\end{aligned}
\tag{3}
$$

where $\Delta t_i = (t_{i+1} - t_i)$ and $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$.[12] It is Itô-Taylor expansion of convergence order 1.0. It has one more term then the Euler method of the equation (2) which is the Itô-Taylor expansion of convergence order 0.5. [13] The application of the Milstein method as a stochastic model can be found in Elerian(1998) and in the appendix of this paper.

Here we need to make two remarks: (i) As mentioned we can apply the strong Itô-Taylor expansion of different convergence orders to obtain diverse discrete-time approximations for the diffusion process. Such models are usually used for simulation but not for estimation. If we employ such discrete-time models for the maximum likelihood estimation, their density functions are complicated and the maximization of the likelihood function is usually unstable. (ii) The Milstein method is a better simulation method for diffusion processes only when the size of the simulation step goes to zero. If the time steps are fixed by the observation times $\{t_0, t_1, \cdots, t_N\}$ as in our case, then we can not say anything about the superiority of the Milstein method.

2.3 New Local Linearization Method

The new local linearization (NLL) method is suggested by Shoji and Ozaki(1997), p.490-491. We introduce their idea briefly: the Euler method holds constant the drift and the diffusion coefficients for $s \in [t_i, t_{i+1})$, while the Shoji and Ozaki approximate the drift coefficient $b(X_s)$ up to the second order terms by using the Itô formula

$$
\begin{aligned}
dX_s =& \Big(b(X_{t_i}) + b'(X_{t_i})(X_s - X_{t_i}) + \frac{1}{2} b''(X_{t_i}) a^2(X_{t_i})(s - t_i)\Big) ds \\
&+ a(X_{t_i}) dW_s.
\end{aligned}
\tag{4}
$$

[12] See Kloeden and Platen, 1992:345.
[13] See Kloeden and Platen, 1992:Chap.10.

The diffusion coefficient is still kept as a constant. The equation (4) can be solved analytically and the solution at t_{i+1} is given by

$$\begin{aligned}
& X_{t_{i+1}} - X_{t_i} \\
= & \frac{b(X_{t_i})}{b'(X_{t_i})}(e^{b'(X_{t_i})(t_{i+1}-t_i)} - 1) \\
& + \frac{b''(X_{t_i})}{(b'(X_{t_i}))^2}\frac{a(X_{t_i})^2}{2}(\ e^{b'(X_{t_i})(t_{i+1}-t_i)} - 1 - b'(X_{t_i})(t_{i+1}-t_i)\ ) \\
& + a(X_{t_i})\int_{t_i}^{t_{i+1}} e^{b'(X_{t_i})(t_{i+1}-z)} dW_z.
\end{aligned} \tag{5}$$

The distribution of the last term can be represented by

$$a(X_{t_i})\int_{t_i}^{t_{i+1}} e^{b'(X_{t_i})(t_{i+1}-z)} dW_z \overset{\text{dis.}}{\sim} \mathcal{N}\Big(0, a(X_{t_i})^2 \int_{t_i}^{t_{i+1}} e^{2b'(X_{t_i})(t_{i+1}-z)} dz\Big). \tag{6}$$

3 Prediction and Related Discussions

We will briefly introduce the prediction procedure. Then we discuss two aspects related to the prediction. First, we give an example where the Euler predictor is consistent although the Euler estimator is inconsistent. Second, we show that the NLL predictor is exactly the same as the Euler predictor in our case of the SDE (1).

3.1 prediction

Let $\mathcal{F}_t$ represent the information set before t. Let

$$E[\cdot|\mathcal{F}_t]$$

denote the *conditional expectation* given $\mathcal{F}_t$. If we only have information up to the period t, the best possible approach to $X^{\theta}_{t+\Delta t}$ is the conditional expectation $E[\cdot|\mathcal{F}_t]$. The expression "best" is in sense of the mean square error criterion. [14] It is to remark that we can achieve this optimum only when we already know the true parameter. Usually, we have to estimate it.

It is easy to obtain the conditional expectations for our three discrete-time models. Let $F_*(x_t, \theta, \Delta t)$ denote the conditional expectation of $X_{t+\Delta t}$ given $\mathcal{F}_t$

[14] See Hamilton(1994) chap.4, p.129.

by using the $*$-method:

$$
\begin{aligned}
F_{eu}(x_t,\theta,\Delta t) &= x_t + b(x_t,\theta)\Delta t, && (7)\\
F_{ms}(x_t,\theta,\Delta t) &= x_t + b(x_t,\theta)\Delta t, && (8)\\
F_{nll}(x_t,\theta,\Delta t) &= x_t + \frac{b(x_t)}{b'(x_t)}(e^{b'(x_t)\Delta t}-1) && (9)\\
&\quad + \frac{b''(x_t)}{b'(X_t)^2}\frac{a(X_t)^2}{2}\Big(e^{b'(x_t)\Delta t}-1-b'(x_t)\Delta t\Big). && (10)
\end{aligned}
$$

Let $\hat{\theta}_*((X_{t_i})_{i=0,\cdots,N})$ be the ML estimators based on the observations $(X_{t_i})_{i=0,\cdots,N}$ using the $*$-method. Then the one-step predictor is given by

$$\hat{X}_{*,t_{N+1}}|_{\mathcal{F}_{t_N}} = F_*\Big(X_{t_N},\hat{\theta}_*((X_{t_i})_{i=0,\cdots,N}),\Delta t_N\Big). \tag{11}$$

3.2 Decomposition of prediction errors

Now we consider the squared error of the one-step prediction. We decompose the expected prediction errors

$$
\begin{aligned}
&E\Big[(\hat{X}_{*,t_{N+1}}|_{\mathcal{F}_{t_N}} - X^{\theta}_{t_{N+1}})^2\Big]\\
&= E\Big[(\hat{X}_{*,t_{N+1}}|_{\mathcal{F}_{t_N}} - F(X^{\theta}_{t_N},\theta,\Delta t) + F(X^{\theta}_{t_N},\theta,\Delta t) - X^{\theta}_{t_{N+1}})^2\Big]\\
&=E[\Big(\hat{X}^{\theta}_{*,t_{N+1}}|_{\mathcal{F}_{t_N}} - F(X^{\theta}_{t_N},\theta,\Delta t)\Big)^2] + Var[X^{\theta}_{t_{N+1}}|\mathcal{F}_{t_N}]
\end{aligned}
$$

into two terms: the first term is the distance between the predictor of the discrete-time model and the best $\mathcal{F}_t$-approximator, the second term is the conditional variance of $X^{\theta}_{t_N}$. Only the first term is related to the approximation quality of discrete-time models. This error decomposition is based on that the expectation of the cross product

$$E[\Big(\hat{X}_{*,t_{N+1}}|_{\mathcal{F}_{t_N}} - F(X^{\theta}_{t_N},\theta,\Delta t)\Big)\Big(F(X^{\theta}_{t_N},\theta,\Delta t) - X^{\theta}_{t_{N+1}}\Big)]$$

is equal to zero. In the numerical experiment we will use the average to represent the expectation. However, the average cross term is not necessarily equal to zero. If the first term is small, then the average prediction error is disturbed by the average cross term and we can not judge the quality of the discretization correctly. Therefore we will take the average of the first term as the criterion to evaluate the prediction in the numerical experiment later.

We know already that the Euler estimator is inconsistent. [15] Here we give an example where, although the Euler estimator is inconsistent, the Euler predictor defined in (11) is consistent, i.e. the Euler predictor converges to the best prediction – the conditional expectation.

We take the SDE (1) with $\gamma = 0$ [16]. The observation times are equidistant $t_i = i\Delta t$, for $i = 0, \cdots, N$, for a fixed Δt and $N\Delta t = T$. We know the solution of this SDE is [17]

$$X_{(i+1)\Delta t} = \frac{c}{\beta}(1 - e^{-\beta\Delta t}) + e^{-\beta\Delta t} X_{i\Delta t} + \sigma \int_0^{\Delta t} e^{-\beta(\Delta t - s)} dW_{i\Delta t + s}, \qquad (12)$$

where the last term is i.i.d. $\mathcal{N}\Big(0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\Delta t})\Big)$-distributed. We rewrite this discrete-time process into the following equation

$$Z_i = a_0 + a_1 Z_{i-1} + u_i,$$

with $Z_i = X_{i\Delta t}$, $a_0 = \frac{c}{\beta}(1 - e^{-\beta\Delta t})$, $a_1 = e^{-\beta\Delta t}$ and u_i has the distribution specified above. Let $\hat{a}_{0,N}$, $\hat{a}_{1,N}$ be the ML estimators for a_0 and a_1 We know these estimators are consistent [18]

$$\operatorname*{plim}_{N\to\infty} \hat{a}_{0,N} = a_0 \quad \text{and} \quad \operatorname*{plim}_{N\to\infty} \hat{a}_{1,N} = a_1. \qquad (14)$$

Now we consider the Euler approximation

$$X_{(i+1)\Delta t} = X_{i\Delta t} + (c_{eu} - \beta_{eu} X_{i\Delta t})\Delta t + \sigma_{eu} \Delta W_{i\Delta t}.$$

We see that the Euler approximation and the discrete-time observed diffusion process can be linked with the reparametrization

$$\begin{aligned} c_{eu}\Delta t &= a_0 = \frac{c}{\beta}(1 - e^{-\beta\Delta t}) \\ 1 - \beta_{eu}\Delta t &= a_1 = e^{-\beta\Delta t} \\ \sigma_{eu}^2 \Delta t &= \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\Delta t}). \end{aligned} \qquad (16)$$

[15] See Lo (1988).
[16] It is called the Ornstein-Uhlenbeck process.
[17] see Kloeden and Platen(1992:118).
[18] See Fuller(1996). Note that the process is stationary because $\beta > 0$, then $a_1 \in (0, 1)$.

Let $\hat{c}_{eu,N}$, $\hat{\beta}_{eu,N}$ be ML estimators for c_{eu}, β_{eu}. Using the proposition 1 in the appendix the ML estimators have also the same relation

$$\begin{aligned} \hat{c}_{eu,N}\Delta t &= \hat{a}_{0,N} \\ 1 - \hat{\beta}_{eu,N}\Delta t &= \hat{a}_{1,N}. \end{aligned} \tag{17}$$

Because of the consistency (14) and the equity (17), we have the consistency of the Euler predictor

$$\begin{aligned} \hat{X}^{eu}_{T+\Delta t}|_{\mathcal{F}_T} &= X_T + (\hat{c}_{eu,N} - \hat{\beta}_{eu,N}X_T)\Delta t = \hat{a}_{0,N} + \hat{a}_{1,N}X_T \\ &\underset{\substack{\text{plim} \\ N\to\infty}}{\longrightarrow} a_0 + a_1 X_T = E[X_{T+\Delta t}|\mathcal{F}_T]. \end{aligned}$$

With exactly the same reasoning we can also see that the Euler estimators are inconsistent

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} \hat{c}_{eu,N} &= \frac{a_0}{\Delta t} = c\Big(\frac{1-e^{-\beta\Delta t}}{\beta\Delta t}\Big) \neq c \\ \operatorname*{plim}_{N\to\infty} \hat{\beta}_{eu,N} &= \frac{1-a_1}{\Delta t} = \frac{1-e^{-\beta\Delta t}}{\Delta t} \neq \beta. \end{aligned}$$

3.4 Equivalence of the Euler and NLL predictors

Next we show the Euler and NLL predictors of the SDE (1) are equivalent. The reason is the linearity of the drift coefficient in the equation (1). We can easily see that the Euler approximation

$$X_{(i+1)\Delta t} - X_{i\Delta t} = (c - \beta X_{i\Delta t})\Delta t + \sigma X^{\gamma}_{i\Delta t}\Delta W_{i\Delta t} \tag{18}$$

and the NLL approximation

$$X_{(i+1)\Delta t} - X_{i\Delta t} = \frac{h_1(\beta)}{\beta}(c - \beta X_{i\Delta t}) + \sigma h_2(\beta) X^{\gamma}_{i\Delta t} U_{i+1} \tag{19}$$

are equivalent under the reparametrization

$$\begin{aligned} \beta_{eu}\Delta t &= h_1(\beta_{nll}) = 1 - e^{-\beta_{nll}\Delta t} \\ c_{eu}\Delta t &= \frac{c_{nll}}{\beta_{nll}} h_1(\beta_{nll}) \\ \gamma_{eu} &= \gamma_{nll} \\ \sigma_{eu} &= \sigma_{nll} h_2(\beta_{nll}) = \sqrt{\frac{1-e^{-2\beta_{nll}\Delta t}}{2\beta_{nll}\Delta t}}\ , \end{aligned}$$

where $U_i, i = 1, \cdots$ are i.i.d $\mathcal{N}(0, \Delta t)$-distributed. We rewrite the reparametrization more compactly. Let $\theta = \{\beta, c, \gamma, \sigma\}$ and

$$H(\beta, c, \gamma, \sigma) = \Big(\frac{1 - e^{-\beta \Delta t}}{\Delta t}, \frac{c h_1(\beta)}{\beta \Delta t}, \gamma, \sigma h_2(\beta)\Big).$$

Then it is clearly seen that

$$H(\theta_{nll}) = \theta_{eu}.$$

Now we will show the equivalence of the predictors. Let $p_{eu}(x, y, \theta, \Delta t)$ be the conditional density for the Euler approximation (18) with $X_{t_{(i+1)\Delta t}} = y$ and $X_{i\Delta t} = x$. Following the definition of the predictor (11) we have

$$\hat{X}^{eu}_{T+\Delta t}|_{\mathcal{F}_T} = E[X^{eu}_{T+\Delta t}|\mathcal{F}_T] = \int y p_{eu}(X_T, y, \hat{\theta}_{eu}, \Delta t) dy.$$

Let $p_{nll}(x, y, \theta, \Delta t)$ be the conditional density for the NLL approximation (19). Recall that θ_{eu} and θ_{nll} are so chosen that the two equations (18) and (19) are exactly the same. Thus, we have the equivalence of the conditional densities

$$p_{nll}(x, y, \theta_{nll}, \Delta t) = p_{eu}(x, y, \theta_{eu}, \Delta t) = p_{eu}(x, y, H(\theta_{nll}), \Delta t).$$

Let $\hat{\theta}_{nll}$ and $\hat{\theta}_{eu}$ be the ML estimators. Using the proposition 1 we obtain

$$H(\hat{\theta}_{nll}) = \hat{\theta}_{eu}.$$

Thus

$$p_{nll}(x, y, \hat{\theta}_{nll}, \Delta t) = p_{eu}(x, y, H(\hat{\theta}_{nll}), \Delta t) = p_{eu}(x, y, \hat{\theta}_{eu}, \Delta t).$$

Therefore, the equivalence of the predictors follows

$$\begin{aligned}\hat{X}^{eu}_{T+\Delta t}|_{\mathcal{F}_T} &= \int y p_{eu}(X_T, y, \hat{\theta}_{eu}, \Delta t) dy \\ &= \int y p_{nll}(X_T, y, \hat{\theta}_{nll}, \Delta t) dy = \hat{X}^{nll}_{T+\Delta t}|_{\mathcal{F}_T}.\end{aligned}$$

4 Specification Test

By modeling empirical data one must demonstrate the suitability of the chosen model with respect to the data. Once a certain model is chosen the model will impose constraint on the data. *A specification test* is a test whether the constraint can be accepted or must be rejected.

Our idea for test the model specification is to undertake *diagnostic checking for estimated white noise.* In our discrete-time models of the equations (2), (3) and

(5) it is very easy to transform the data into white noise which is represented by Brownian increments ΔW_t and therefore is i.i.d. normally distributed . In other words, this transformation should remove the deterministic structure specified by the (discrete-time) models.

We employ two simple tests to test whether the estimated white noise is i.i.d normally distributed. The first one is to test the null hypothesis $H_0^{(1)}$: the series of the estimated white noise does not have any autocorrelation, which is a prerequisite for the independence. The second one is to test $H_0^{(2)}$: the distribution of the estimated white noise is normal.

4.1 Autocorrelation Checking

Let $U_1, \cdots, U_N$ be identically distributed random variables. Assumed that $E[U_i] = 0$, $Var[U_i] = 1$ and $E|U_i|^s < \infty$. The problem is to test $H_0^{(1)}$: $(U_i)_{i=1,\cdots,N}$ is not autocorrelated.

Let $\hat{R}_k$ be the sample autovariance function [19] represented by

$$\hat{R}_k = \frac{1}{N-k} \sum_{i=k+1}^{N} U_i U_{i-k}.$$

Under the null we have $E[\hat{R}_k] = 0$ and

$$Var[\hat{R}_k] = \frac{1}{N-k},$$

for $k \geq 1$. We normalized $\hat{R}_k$

$$\hat{r}_k = \frac{\hat{R}_k - E[\hat{R}_k]}{\sqrt{Var[\hat{R}_k]}} = \sqrt{N-k}\,\hat{R}_k = \frac{1}{\sqrt{N-k}} \sum_{i=k+1}^{N} U_i U_{i-k}.$$

Consider the sequence $(U_i U_{i-k})_{i=k+1,\cdots,N}$ for a fixed k. It is *near epoch dependent* on $(U_i)_{i=1,\cdots,N}$ [20]. Using the central limit theorem for near epoch processes [21], $\hat{r}_k$ converges to $\mathcal{N}(0,1)$ in distribution as $N \to \infty$. Applying the test for our discrete-time approximations, we let $U_i = W_i - W_{i-1}$.

We remark here that $\hat{r}_k \sim \mathcal{N}(0,1)$ means $\hat{R}_k \sim \mathcal{N}(0, \frac{1}{N-k})$. It is similar with the result $Var[\hat{R}_k] \sim 1/N$ in Box et al.(1994:32) when N is large enough.

[19] $\hat{R}_k$ is also sample autocorrelation function because $Var[U_i] = 1$.

[20] See Gallant and White (1988) Def. 3.13, p.27 with $Z_{ni} = U_i U_{i-k}$. One can see $v_m = 0$ when $m \geq k$.

[21] See Gallant and White (1988), Theorem 5.3, p.76 . The conditions of the theorem are satisfied because under null (U_i) is independent and $v_n = n - k$.

We have to remark that there is a shortcoming in this test. We use a "theoretical noise" to derive the distribution of the sample autocorrelation, but in reality we have only "estimated noise" available. The work of Durbin(1970), Box and Pierce(1970) point out that the sample autocorrelations will be underestimated for small k. [22] Fortunately, this under-estimation does not affect our finding of misspecification later [23].

4.2 Testing normality

We employ here χ^2-test for histogram to test whether the distribution of samples is $\mathcal{N}(0,1)$-distribution. [24] The idea is to compare the relative frequency of samples on intervals I_m

$$\hat{p}_m = \frac{\text{number of}\{i; U_i \in I_m\}}{N}$$

and p_m the probability of $\mathcal{N}(0,1)$-distribution on the intervals I_m where $\{I_m, m = 1, \cdots, M\}$ are disjoint intervals of the real line.

The weighted distance

$$d = \sum_{m=1}^{M} \frac{N}{p_m(1-p_m)} (\hat{p}_m - p_m)^2 \tag{20}$$

measures the distance between the sample and the normal distributions. It converges to $\chi^2(M-1)$ in distribution as $N \to \infty$.

5 Numerical Experiment with Monte Carlo Simulation Method

Here we will compare the performance of the Euler and Milstein approximation in a numerical experiment using Monte-Carlo-simulations. The intention is to find out (i)which discrete-time approximations perform better, and (ii) whether the discrete-time approximations are correctly specified models for the discrete-time observations of the diffusion process (1).

The Monte-Carlo-simulation method means that we undertake repeatly simulations. One simulation in our experiment includes: (i) generating data of the equation (1) with a finer time interval to imitate the dt. Then we observe the

[22] The is why the "Q-statistc" is developed, see Box and Pierce(1970) and Ljung and Box(1978) .

[23] See Section 6.

[24] See Breiman(1973:189).

generated data with a time interval which is much greater than for the generated data, (ii) applying ML estimation using the discrete-time approximate models on the observed data, (iii) testing the specification of the models for the observed data, (iv) forecasting.

The parameter values for the data generating process are given by

$$\begin{aligned} c &= 0.06 \\ \beta &= 0.01 \\ \sigma &= 0.05 \\ \gamma &= 0.2, 0.5, 0.8, 1.2. \end{aligned}$$

The values of c, β, σ are chosen from the empirical results of the short term interest rate data of the U.S., see Table 7. We vary the value of γ to see whether the discretization effect would be stronger for greater γ. Recall that if $\gamma = 0$, the Euler approximation is exactly the correct model. [25]

The other parameters are:

the generation interval δt	$= 0.01$
the observation interval Δt	$= 1$
the whole observation time N	$= 200$
each simulation is repeated 1000 times.	

The numerical results are reported in Tables 1 – 4. [26]

The results of the Euler and the Milstein approximations are in the first and the second columns. In the last column labeled with "true model" are the results using *all* simulated data. Hence, the estimation model is exactly the same as the data generation model. The $\overline{E}$ denotes the arithmetic average over all simulations. We compare two kinds of prediction errors: one is the difference between the predictor and the data, the other one is the difference between the predictor and the conditional expectation, see Section 3.2. We argued there that the second one is better than the first one. In these tables we can also find the rejection frequency of the specification tests. $H_0^{(1)}$ is the autocorrelation test and $H_0^{(2)}$ is the normality test described in Section 4. We reject the $H_0^{(1)}$ if

$$\max_{k=1,\cdots,10} |\hat{r}_k| \geq 2.8.$$

We reject $H_0^{(2)}$ if the p-value of the χ^2-test is smaller than 0.05.

[25] It needs to be reparametrized. See Section 3.3.
[26] In Table 1 for $\gamma = 0.2$, 5 simulations can not converge by using Milstein approximation.

Table 1
Numeric results for $\gamma = 0.2$

$\gamma = 0.2$	Milstein	Euler	finer Euler
$\overline{E}[\hat{c}_* - c]$	0.1547	0.1550	0.1594
$\overline{E}[\hat{\beta}_* - \beta]$	0.0259	0.0260	0.0267
$\overline{E}[\hat{\gamma}_* - \gamma]$	0.0151	0.0067	0.0019
$\overline{E}[\hat{\sigma}_* - \sigma]$	0.7946	1.0050	0.0010
$\overline{E}[\hat{c}_* - c]^2$	0.0443	0.0444	0.0468
$\overline{E}[\hat{\beta}_* - \beta]^2$	0.0012	0.0012	0.0013
prediction errors $\overline{E}(\hat{Y}^*_{(N+1)\Delta t} - Y_{(N+1)\Delta t})^2$	$5.230e^{-3}$	$5.223e^{-3}$	$5.222e^{-3}$
estimation errors of cond. expectation $\overline{E}(\hat{Y}^*_{(N+1)\Delta t} - F(Y_{N\Delta t}, \Delta t))^2$	$7.730e^{-5}$	$7.848e^{-5}$	$7.420e^{-5}$
rejection of $H^{(1)}$	3.5%	3.6%	
rejection of $H^{(2)}$	8.2%	8.1%	

Table 2
Numeric results for $\gamma = 0.5$

$\gamma = 0.5$	Milstein	Euler	finer Euler
$\overline{E}[\hat{c}_* - c]$	0.1612	0.1610	0.1666
$\overline{E}[\hat{\beta}_* - \beta]$	0.0268	0.0268	0.0277
$\overline{E}[\hat{\gamma}_* - \gamma]$	-0.0243	-0.0293	0.0012
$\overline{E}[\hat{\sigma}_* - \sigma]$	0.0829	0.0705	0.0002
$\overline{E}[\hat{c}_* - c]^2$	0.0487	0.0486	0.0527
$\overline{E}[\hat{\beta}_* - \beta]^2$	0.0013	0.0013	0.0015
prediction errors $\overline{E}(\hat{Y}^*_{(N+1)\Delta t} - Y_{(N+1)\Delta t})^2$	0.0147	0.0147	0.0147
estimation errors of cond. expectation $\overline{E}(\hat{Y}^*_{(N+1)\Delta t} - F(Y_{N\Delta t}, \Delta t))^2$	$2.446e^{-4}$	$2.440e^{-4}$	$2.427e^{-4}$
rejection of $H^{(1)}$	4%	4%	
$H^{(2)}$	8.8%	9.5%	

Table 3
Numeric results for $\gamma = 0.8$

$\gamma = 0.8$	Milstein	Euler	finer Euler
$\overline{E}[\hat{c}_* - c]$	0.1516	0.1517	0.1553
$\overline{E}[\hat{\beta}_* - \beta]$	0.0256	0.0256	0.0262
$\overline{E}[\hat{\gamma}_* - \gamma]$	-0.0230	-0.0253	-0.0014
$\overline{E}[\hat{\sigma}_* - \sigma]$	0.0171	0.0154	0.0002
$\overline{E}[\hat{c}_* - c]^2$	0.0425	0.0425	0.0444
$\overline{E}[\hat{\beta}_* - \beta]^2$	0.0012	0.0012	0.0013
prediction errors $\overline{E}(\hat{Y}^*_{(N+1)\Delta t} - Y_{(N+1)\Delta t})^2$	0.0457	0.0457	0.0459
estimation errors of cond. expectation $\overline{E}(\hat{Y}^*_{(N+1)\Delta t} - F(Y_{N\Delta t}, \Delta t))^2$	$7.146e^{-4}$	$7.150e^{-4}$	$6.916e^{-4}$
rejection of $H^{(1)}$	3.8%	3.9%	
rejection of $H^{(2)}$	6.7%	8.4%	

Table 4
Numierc results for $\gamma = 1.2$

$\gamma = 1.2$	Milstein	Euler	finer Euler
$\overline{E}[\hat{c}_* - c]$	0.1330	0.1331	0.1345
$\overline{E}[\hat{\beta}_* - \beta]$	0.0253	0.0252	0.0256
$\overline{E}[\hat{\gamma}_* - \gamma]$	-0.0505	-0.0432	-0.0006
$\overline{E}[\hat{\sigma}_* - \sigma]$	0.0074	0.0062	0.0001
$\overline{E}[\hat{c}_* - c]^2$	0.0358	0.0359	0.0370
$\overline{E}[\hat{\beta}_* - \beta]^2$	0.0012	0.0012	0.0012
prediction errors $\overline{E}(\hat{Y}^*_{(N+1)\Delta t} - Y_{(N+1)\Delta t})^2$	0.2610	0.2614	0.2607
estimation errors of cond. expectation $\overline{E}(\hat{Y}^*_{(N+1)\Delta t} - F(Y_{N\Delta t}, \Delta t))^2$	0.0043	0.0043	0.0040
rejection of $H^{(1)}$	3.5%	3.4%	
rejection of $H^{(2)}$	7.4%	15.1%	

Comparing the performance of the Euler and Milstein methods, the results of drift parameter estimations c, β, are similar. Hence we have also similar results of the predictions because they are only based on drift parameters. For the diffusion parameters σ, γ none of the two approximations is clearly superior.

The rejection quotes of $H_0^{(1)}$ are about 4% for the both approximations and for all γ's. This means for about 96% simulation the maximal normalized autocorrelation coefficients are smaller than 2.8. The rejection frequency for $H_0^{(2)}$ would be around 5% if the noise were normally distributed. In the tables we see all the rejection frequencies of $H_0^{(2)}$ are greater than 5%. The rejection frequencies of the Euler approximation are greater than those of Milstein approximation for $\gamma = 0.5, 0.8, 1.2$. Especially for $\gamma = 1.2$, the rejection frequency of the Euler approximation is more than double as that of the Milstein approximation: 15.1% for the Euler and 7.4% for the Milstein approximations. This means the discretization effect in the estimated noise for a greater γ can be lowered more by the Milstein approximation than by the Euler approximation.

The estimated drift coefficients c, β show large estimation errors, for example for $\gamma = 0.5$, the error of β is 269% ($= 0.1611/0.06$) relative to the true β and 268% ($= 0.268/0.01$) relative to the true c. These errors are not caused by the discretization because the errors of the true model at the last column are even larger than those of the Euler and Milstein approximations. It does not coincide some results where the parameter estimations are better by using more frequent data, for example in Shoji nd Ozaki (1997). We remark here that the ML estimator is biased, because the correlation between the explanatory variables and the noises. [27]

However, in spite of the larger errors of parameter estimation, the true model can offer a better prediction. We can see that the true model has smaller errors of conditional expectation estimations.

6 Empirical Results on Modeling Short Term Interest Rates

We apply the Euler and the Miltein approximations on short rate data in this section. The short rate data we choose are interest rates with a one-day maturity: the overnight interbank rate of the United Kingdom, the federal funds rate of the U.S. and the call money rate of Germany. All data are monthly data. [28] We take the time period 1983.1 – 1997.12 (180 observations) for esti-

[27] See Frohn (1995).

[28] The source is “International Statistical YearBook”. See http://www.ub.uni-bielefeld.de/english/library/databases/, then choose International Statistical YearBook 2000, for "Datenbank" choosing "OECD" and "main economic indicators", for "Period" choose "monthly data", for "Search" choose "indicator-search", then

mation and 1998.1 – 2000.6 (30 observations) for prediction. The time series of the rates are plotted in Figures 1, 4 and 7.

In the Tables 5, 6 and 7 the empirical results are reported. In the first two columns are results of the Euler and Milstein approximations for the CKLS model (1). The notations of the parameters are changed because we will consider more general models below[29]. Their parameter estimations, their t-statistics for the estimates[30] and their predictions are very similar. The estimated white noise from the two approximations is also very similar. We plot their distributions in Figures 2, 5 and 8.

We also plot the normalized autocorrelations for the Euler approximation in Figures 3, 6 and 9. We see the first normalized autocorrelation are about 3.5 for Germany and the U.K. and about 5 for the U.S. It indicates strong autocorrelation between the time series of the estimated white noise. We remarked in Section 4.1 already that our autocorrelation test under-estimates the sample autocorrelation for short lags. This under-estimation does not affect the fact that the estimate noise have strong (even stronger) autocorrelation. For comparing this result of our numerical experiment, we take the result of $\gamma = 0.8$ as a benchmark for the result of the U.S. With 1000 simulations, 96% of the simulations have the first ten normalized autocorrelations smaller than 2.8. The maximal value of the autocorrelations in the numerical result is only 4.2. It indicates that the continuous-time CKLS model (1) can not reproduce the high autocorrelation of the noise as the empirical data exhibit. In other words, the CKLS model is misspecified.

We can also observe that the estimated white noise is highly concentrated around zero than standard normal distribution, which means they have thick tails.[31] This fact can be inferred from in Figures 2, 5 and 8 and the values of the χ^2-test and their p-values in Tables 5 - 7. Comparing this result to the numerical experiment, such large values do not occur. It means again the model (1) is misspecified.

We also see that the estimated drift coefficients do not significantly differ from zero. When they are zero, it means that we can not forecast tomorrow's data better than just using the data today. In order to see whether there is a reduction of the forecasting error by using the models we compare forecasting errors of the models relative to those of the "naive" forecast, — just using the data today. The results in the tables show that we do not need such a model.

"interest rates", then "immediate rates".

[29] In the parentheses are the old notations.

[30] in the parentheses

[31] Because the variance is normalized to 1. The concentration of the distribution around 0 let the variance smaller. In order to keep the variance as 1, there must be more weight in the tail.

Fig. 1. Call money rate, Germany

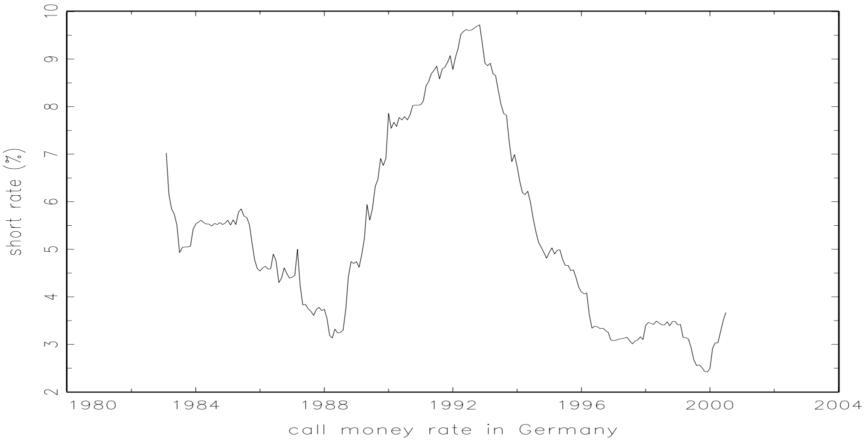


Fig. 2. Distribution of estimated white noise (I), Germany

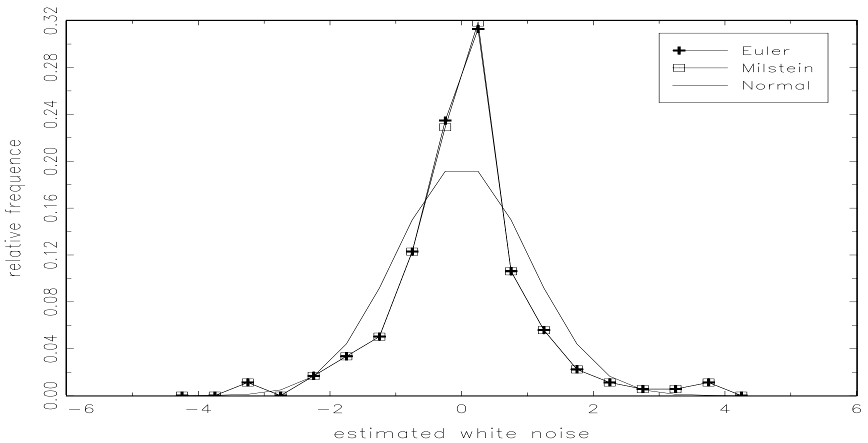


Fig. 3. Normalized autocorrelation of the estimated noise, Germany

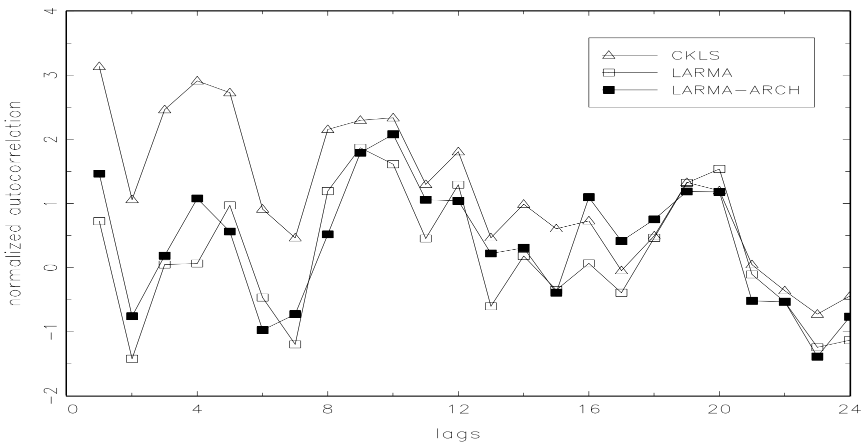


Fig. 4. Interbank rate, U.K.

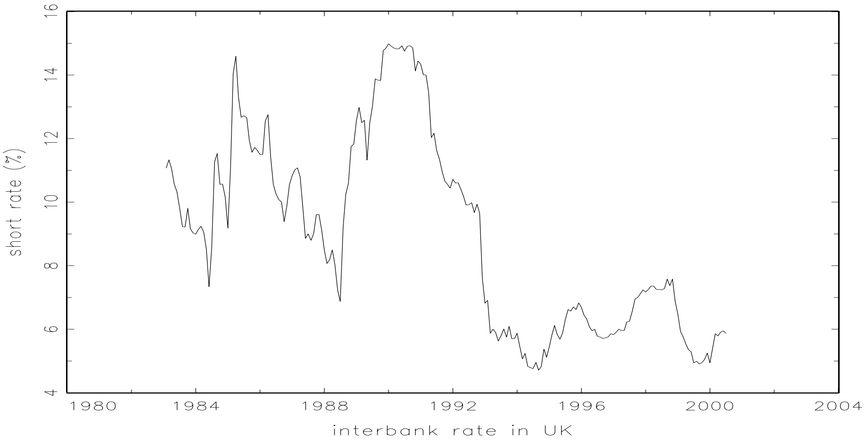


Fig. 5. Distribution of estimated white noise (I), U.K.

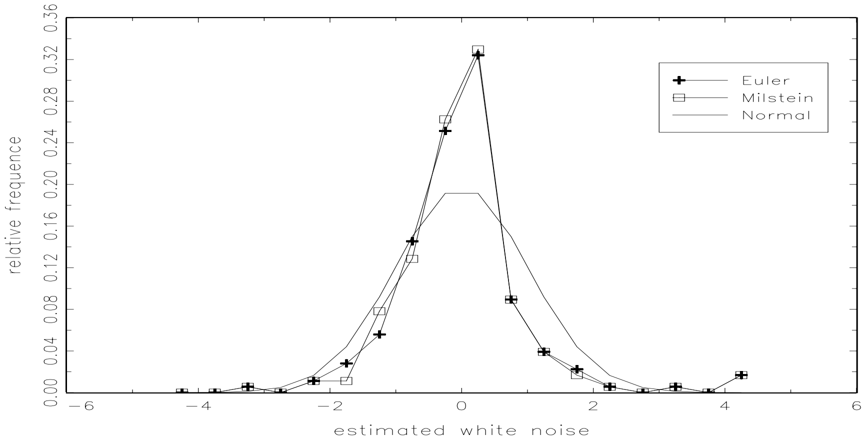


Fig. 6. Normalized autocorrelation of the estimated noise, U.K.

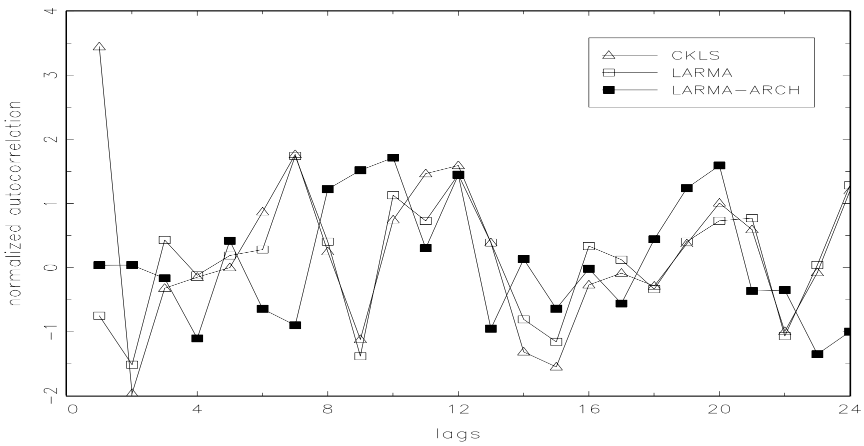


Fig. 7. Federal funds rate of the U.S.

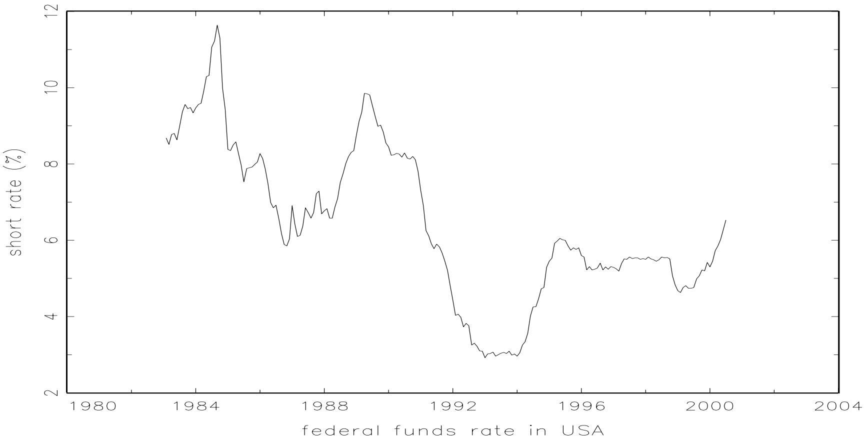


Fig. 8. Distribution of estimated white noise (I), U.S.

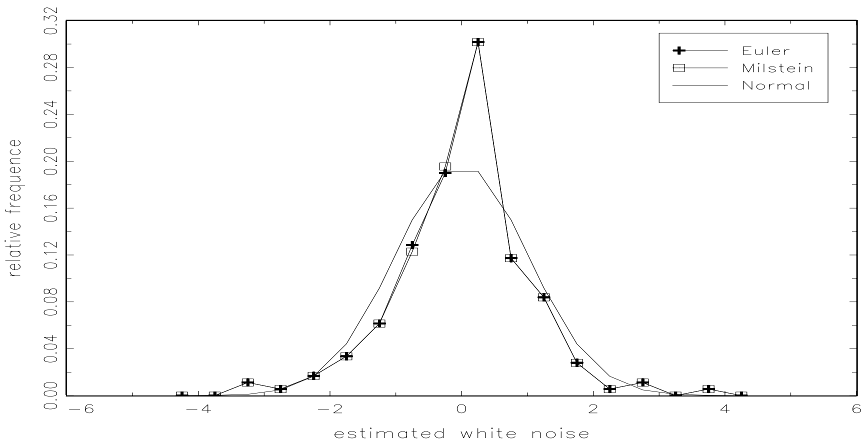
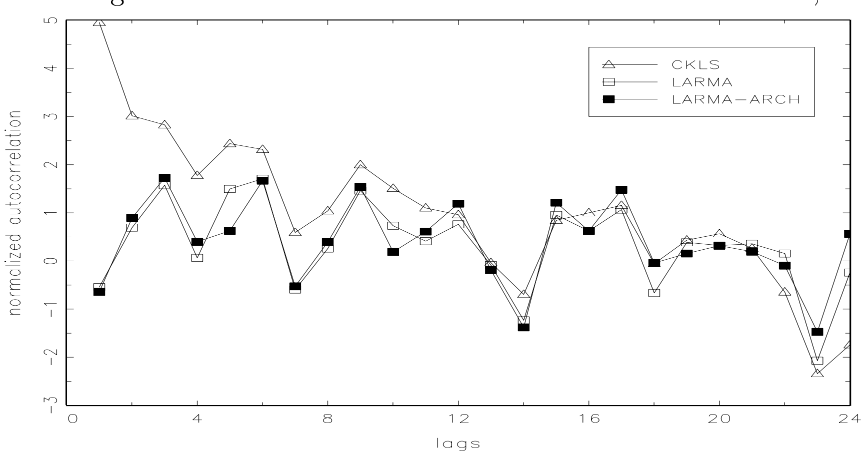


Fig. 9. Normalized autocorrelation of the estimated noise, the U.S.



7 Searching for New Models

Because of the misspecification of the model (1) shown in the last section we search for new models. They must be able to model the high autocorrelations and the thick tails in the estimated noise.

7.1 Improvement in the continuous-time framework

In literature, there are further works to improve the model (1) for modeling the short term rate in the framework of continuous-time models, for example Ait-Sahalia (1996) suggests an non-linear drift coefficient and Andersen and Lund (1997) suggests a stochastic volatility model.

We simulate data using the models specified in Ait-Sahalia (1996) and Andersen and Lund (1997).[32] We plot them in Figures 10 and 11. The model of Ait-Sahalia can not reproduce a similar time series of the real data. It stays always in a narrow band around the steady state. The normalized autocorrelation functions from these two models are plotted in Figure 12. We observe that there is no extreme autocorrelation in the estimated noise.

7.2 Modeling autocorrelations in the estimated noise

We employ the autoregressive-moving-average (ARMA) process[33] to model the autocorelation of the estimated noise

$$\Delta W_t = \sum_{i=1}^{p} \phi_i \Delta W_{t-i} + \sum_{j=0}^{q} \psi_i \epsilon_{t-j}. \tag{21}$$

We will transform the ARMA-structure of the noise into the ARMA-structure of the variable. We illustrate the transformation with an example, where the noise ΔW_t is an autoregressive process of order 1

$$\Delta W_t = \phi \Delta W_{t-1} + \epsilon_t.$$

We replace ΔW_t using (18), then we obtain

$$\frac{\Delta X_t - (c - \beta X_{t-1})}{\sigma X_{t-1}^{\gamma}} = \phi \, \frac{\Delta X_{t-1} - (c - \beta X_{t-2})}{\sigma X_{t-2}^{\gamma}} + \epsilon_t \, .$$

[32] We undertake simulation with an interval 0.01 and then pick up the simulated series with an interval 1.

[33] see Box, Jenkins and Reinsel (1994)

Fig. 10. Simulated data from Ait-Sahalia's model

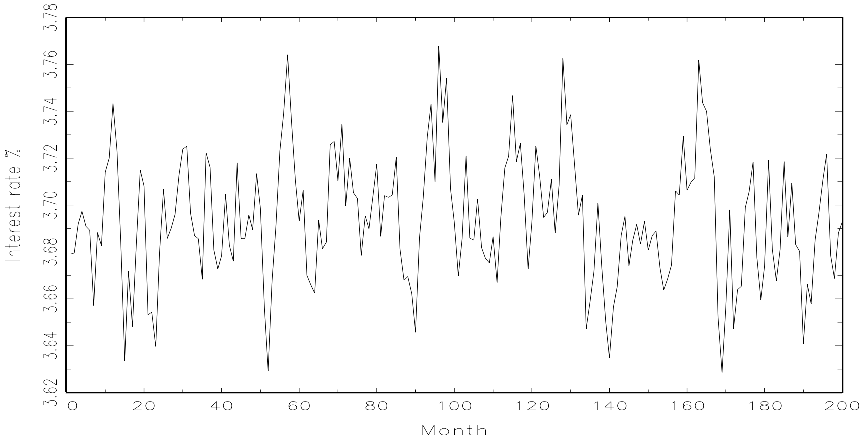


Fig. 11. Simulated data from Andersen-Lund's model

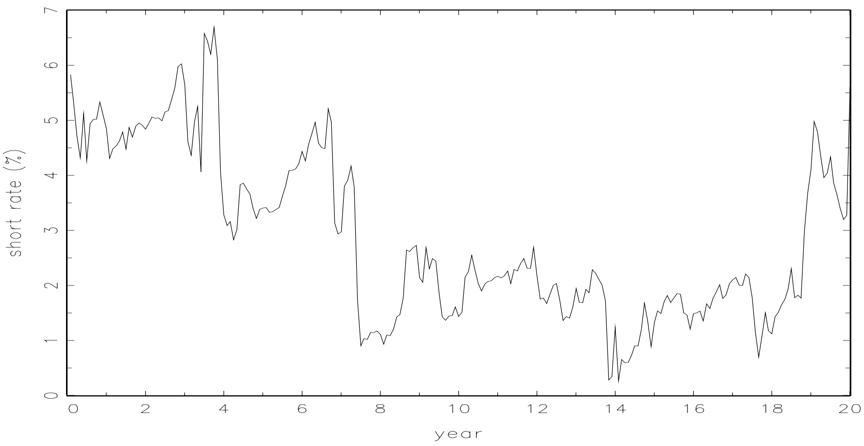
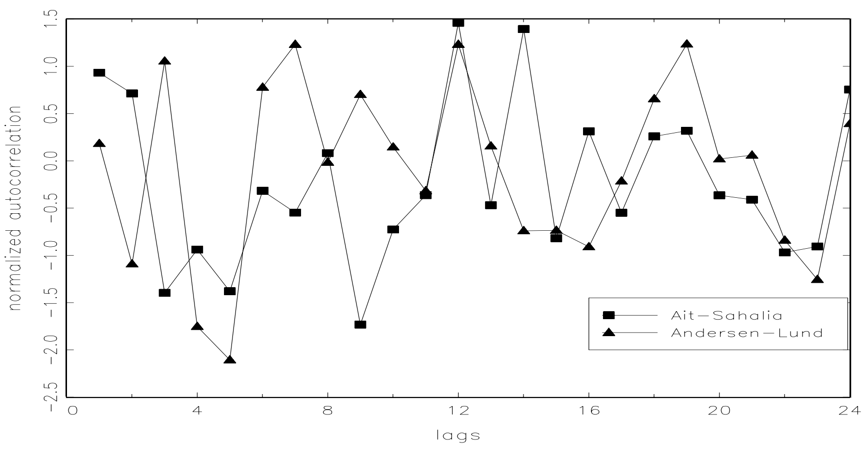


Fig. 12. Normalized autocorrelation of the estimated noise for the continuouts-time models



Rearranging it we obtain

$$\Delta X_t = (c - \beta X_{t-1}) + \phi\Big(\frac{X_{t-1}}{X_{t-2}}\Big)^{\gamma}\Big(X_{t-1} - (c - \beta X_{t-2})\Big) + \sigma X_{t-1}^{\gamma}\epsilon_t. \qquad (22)$$

This means, in order to eliminate the first autocorrelation of the noise ΔW_t, we must introduce the second lag as an explanatory variable.

The equation (22) give us a starting-point. We assume $\left(\frac{X_{t-1}}{X_{t-2}}\right)^{\gamma} \sim 1$. Rewriting (22) then we obtain a model with two lags in the drift term

$$\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \sigma X_{t-1}^{\gamma}\epsilon_t.$$

So, for modeling the noise ΔW_t in a general structure we employ the following equation

$$\Delta X_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i X_{t-i} + X_{t-1}^{\gamma}\Big(\sum_{i=0}^{q}\beta_i\varepsilon_{t-i}\Big) \qquad (23)$$

7.3 Modeling thick tails in the estimated noise

For modeling thick tails of the noise we employ the idea of Brenner et al.(1996) and Koedijk et al.(1997) . The common feature of their constructions is that they apply the autoregressive conditional heteroscadasticity (ARCH)[34] to model the thick tail.[35] Moreover, the conditional variance (the volatility) of X_t is level-dependent. Brenner et al. (1996) argue that both level- and ARCH-effects are significant for short-term rates.

We build the ARCH-structure in the model (23)

$$\begin{aligned} \varepsilon_t &\sim N(0, h_t), \qquad (24)\\ h_t &= c_0^2 + \sum_{i=1}^{k} c_i\varepsilon_{t-i}^2. \end{aligned}$$

We employ (23) and (24) as our model class to model short rates. For the unique specification of the parameter we normalize $\beta_0 = 1$. We make two remarks. First, our model generalizes the model of Brenner et al. (1996) by considering the ARMA-structure (23). We saw already that the ARMA-structure is used to model the autocorrelation of the noise which is found in the empirical results. Even in their results we can also find the evidence of autocorrelations of the residuals .[36] Second, we employ the ARCH-structure instead

[34] See Engle (1982).

[35] There is a thick tail effect if the kurtosis, defined as $\frac{E(\epsilon^4)}{(E(\epsilon^2))^2}$, is greater than 3 - the kurtosis of normal distribution.

[36] See Brenner et al. (1996) p.95 " *The Ljung-Box $Q(\epsilon_t/\sigma_t)$ statistics indicate that both models have significant serial correlation in the residuals.*"

of GARCH-structure in Brenner et al. (1996) . The GARCH model is a technical improvement over the ARCH-struture[37] when the lags of ε_t^2 are long. According to the results of our model identification we do not need to employ the GARCH-structure.

7.4 Model identification

By model identification we mean the determination of the orders p, q and k according to the data. We follow the Box-Jenkins-methodology in Box, Jenkins and Reinsel (1994). The first step is to choose a tentative model according to the *autocorrelation function* (ACF) and *partial autocorrelation function* (PACF). The second step is to check the tentative model. Then, according the diagnostic check we decide whether we accept the tentative model or we need go back to the first step to choose another model.

In our case, the choice of p and q is suggested by the ACF and PACF of ε_t. After that p and q are determined, the order k is chosen by the ACF and PACF of ε_t^2. By model diagnosic check we have to check three points: (i) overfitting: whether the estimated parameter differs from zero significantly. (ii) noise diagnostic checking: whether ϵ_t, ϵ_t^2 have extremely high autocorrelations. (iii) whether the chosen model can reproduce a similar time series as the empirical data. In summary, we choose a most parsimonious model in which the estimated noise does not have significant autocorrelations and which can reproduce a similar time series as the empirical data represent.

7.5 Results

In the Tables 5, 6 and 7 we report the empirical results for the short rate of Germany, the United Kingdom and the U.S. The first and second columns are already discussed in Section 6. In the third and fourth columns are results of the chosen ARMA model and the ARMA-ARCH model. The abbreviation "LARMA" denotes "level + ARMA" – ARMA with level effect. In the lowest box of the three tables we can find the forecast errors of the models, where "in" and "out" represent "in sample" and "out of sample". The predictor of X_{t+1} [38] in the LARMA and the LARMA-ARCH model is given by

$$\hat{X}_{t+1} = E_t[X_{t+1}](\hat{\theta}) = X_t + \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i X_{t-i+1} + X_t^{\hat{\gamma}} \Big(\hat{\beta}_1 \varepsilon_t + \cdots + \hat{\beta}_q \varepsilon_{t-q+1} \Big).$$

[37] see Bollerslev (1986).

[38] Recall the definition(11) in this paper.

Thus the forecast error of level is the difference

$$X_{t+1} - \hat{X}_{t+1} = X_t^{\hat{\gamma}} \varepsilon_{t+1}$$

and the forecast error of volatility is given by

$$(X_{t+1} - \hat{X}_{t+1})^2 - E_t\Big[(X_{t+1} - \hat{X}_{t+1})^2\Big] = (X_t^{\hat{\gamma}} \varepsilon_{t+1})^2 - X_t^{2\hat{\gamma}} \hat{h}_t.$$

The percentage rates in the "% to naive" record the (squared) forecast errors in proportion to the naive forecast. Recall that the naive forecast means forecast just using the data in the last period.

We observe the following (i) the drift parameters become more significantly different from zero (except the drift parameter in the U.K.) by introducing the new structures. In contrast to the CKLS model where the t-statistics are quite small, our new models have a better explanatory power than the naive forecast, (ii) the forecasts both of level and volatility are improved, (except the level forecast for U.K.) The improvement is between 10% and 43%. We can see that the major improvement of the forecast is due to the introduction of the ARMA-structure. This is because the ARCH-structure is not considered for improving level forecast. And for volatility forecast, if we can improve level forecast, then the squared errors become smaller, hence the squared errors of volatility also become smaller. For the data of Germany and the U.S., the volatility forecast is further improved by introduction of the ARCH-structure, (iii) the parameter γ is significantly different from zero. This corresponds to the existence of the level-effect in Brenner et al. (1996). For the data of Germany and the U.K., the parameter γ is not significantly different from 0.5. [39]

The normalized autocorrelations with respect to lags are plotted in Figures 3, 6, 9. The normalized autocorrelations for the chosen LARMA and LARMA-ARCH models are controlled within $[-2, +2]$. The distributions of the noise can be found in Figures 13, 14, 15 and the χ^2-statistics for the normality test are reported in Tables 5 – 7. Although we already have reduced concentration of the distributions a lot by introducing the ARCH-structure, all of them are still significant different from the normal distribution at the level 5%. The distance is greatest for the short rate in the United Kingdom.

Now we reconstruct time series for the short rates using the specified models and the estimated parameters. The simulated data are plotted in Figures 16, 17 and 18.

Comparing all three countries we can observe that the modeling for the short rate of the U.K. is less successful. The t-statistics of the estimated parameters are not significantly different from zero and the distance between the distri-

[39] This value has been proposed by the CIR model, see Cox, Ingersoll and Ross (1985).

bution of the estimated noise and the normal distribution is still sizeable even after the introduction of the ARCH-structure.

Fig. 13. Distribution of estimated white noise (II), Germany

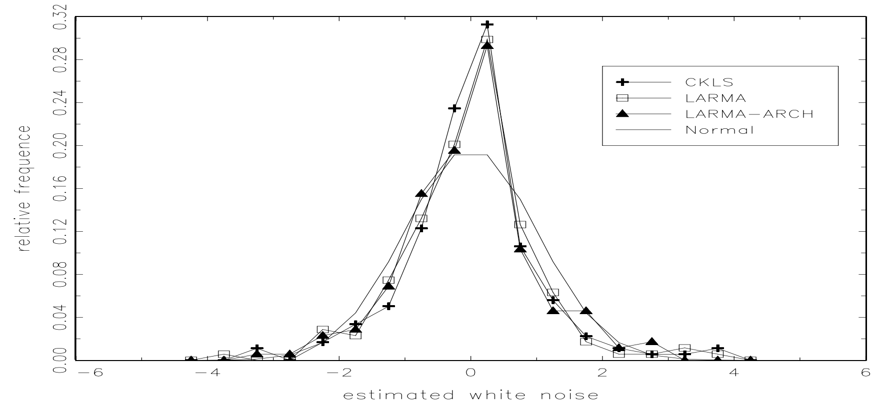


Fig. 14. Distribution of estimated white noise (II), U.K.

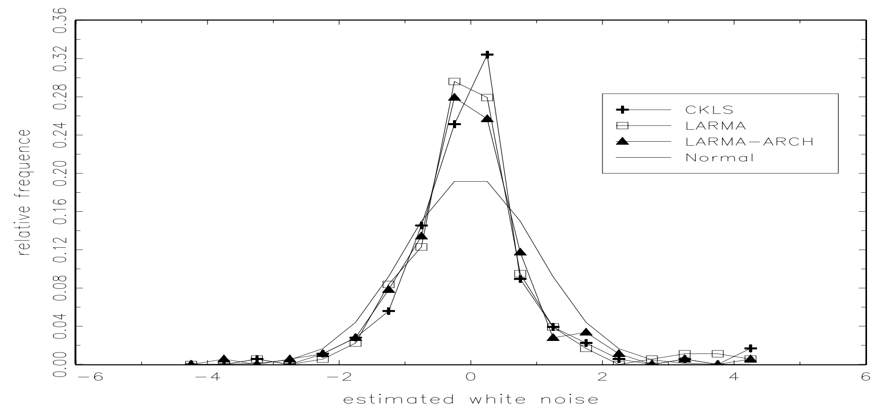


Fig. 15. Distribution of estimated white noise (II), U.S.

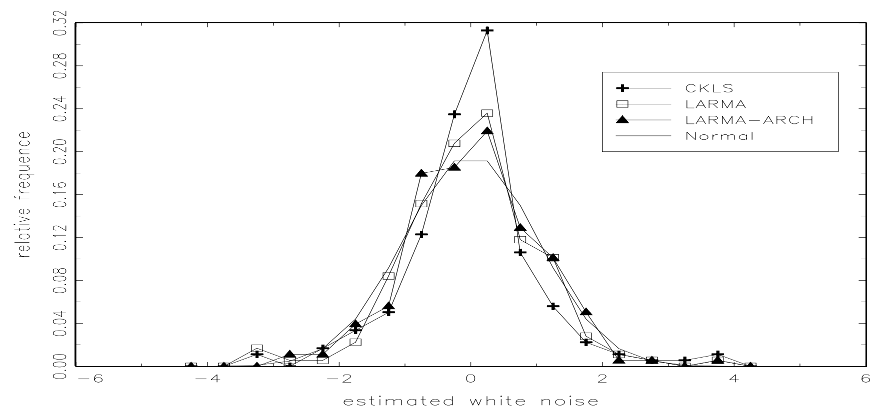


Fig. 16. Simulated rate of Germany

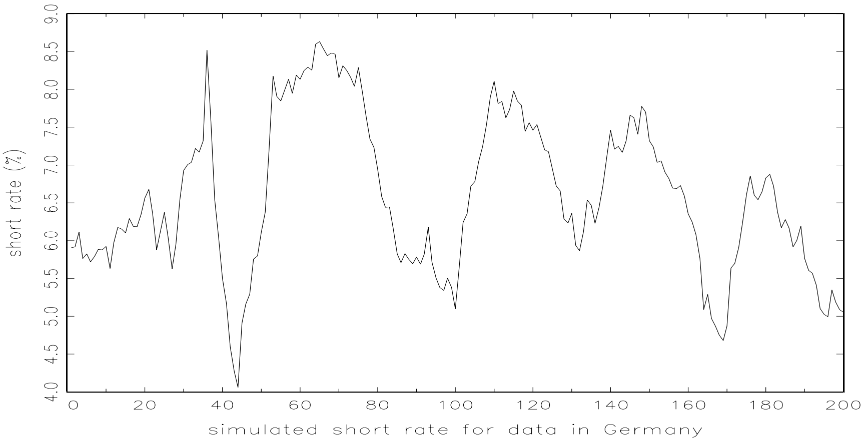


Fig. 17. Simulated rate of the U.K.

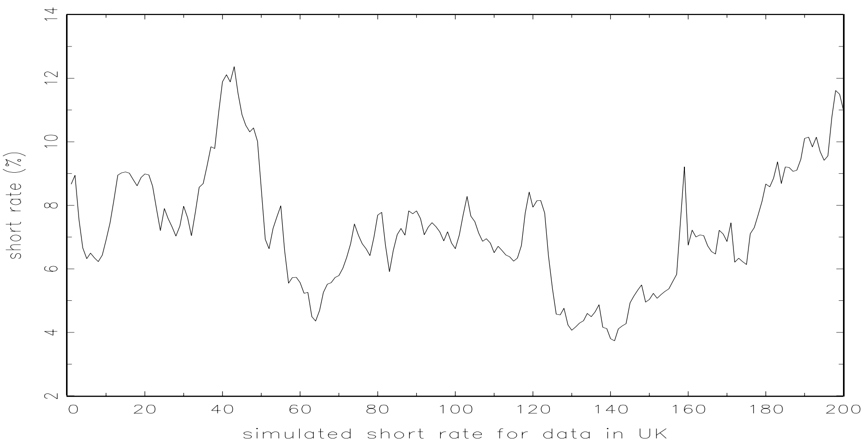
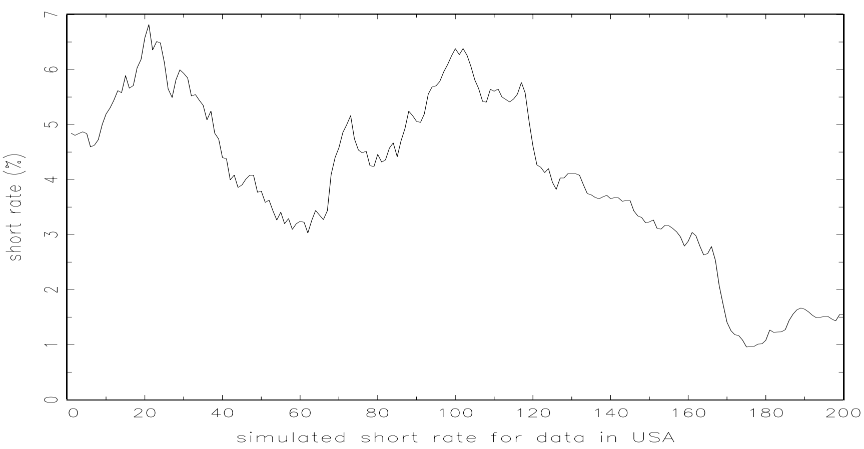


Fig. 18. Simulated rate of the U.S.



8 Conclusions

The objective of the paper is to empirically model short term interest rates. We begin with the contintuous-time CKLS model (1) and we apply the Euler, Milstein and NLL approximations. For evaluating the quality of the discrete-time approximations, we compare the errors of parameter estimations and the one step ahead predictions. Our results do not show an improvement of the NLL and Milstein approximations over the Euler approximation frequently found in the literature. The NLL approximation is equivalent to the Euler approximation due to the linearity of the drift coefficient. In our numerical experiment we do not find any superiority of the Milstein approximation over the Euler approximation.

We suggest two simple tests to test the model specification by checking whether there is still recognizable structure in the estimated white noise of models. We find that both approximations can be accepted as correctly specified. The superiority of the Milstein approximation is that the rejection frequency of the normality test of the Euler approximation is higher than that of the Milstein approximation. It means that the Milstein approximation can reduce the discretization effect on the noise distribution better than the Euler approximation.

We also apply the Euler and the Milstein approximations to the short term interest rates of Germany, the U.K. and the U.S.. In contrast to the numerical experiment we find strong evidence of model misspecification. The estimated white noise from the empirical data has high autocorrelation and thick tails. Since they do not occur in the numerical experiment where the data were simulated from the diffusion process (1), this indicates that the continuous-time model (1) of Chan et al. (1992) is not a suitable model for the short rate data.

We show that two further continuous-time models of Ait-Sahalia (1996) and Andersen and Lund(1997) can not model the autocorrelation of the estimated white noise either. Therefore, we decide to model short rates in a discrete-time framework. Our model is the ARMA-ARCH model class with level-dependent volatility. We choose the most parsimonious model without significant autocorrelation of the noise. The new model improvs the forecast from 10% to 43% for the short rate data of Germany and the U.S. The improvement of the forecast holds not only for the data in sample but also for the out of sample forecast. However, there is one exception, the out of sample forecast for U.K. is not improved.

This empirical result of the new model givse us an important message. It shows the necessity to check the estimated white noise. We can indeed improve the forecasts of the model by removing noticeable structure in the estimated noise.

Yet, there are still problems left. The modeling of the interest rate data of U.K. is less satisfactory and the normality test of the noise is rejected for all three countries. This might suggest to consider other distributions, like gamma distribution or stable distribution in the next step of the research.

A Appendix

A.1 The likelihood function of the Milstein approximation

Here we show the derivation of the likelihood function when using the Milstein method. Following (3), the dynamic of the SDE (1) is approximated by

$$X_{t_{i+1}} - X_{t_i} = (c - \beta X_{t_i})\Delta t_i + \sigma X_{t_i}^{\gamma} \Delta W_{t_i} + \frac{1}{2}\sigma^2 \gamma X_{t_i}^{2\gamma-1}(\Delta W_{t_i}^2 - \Delta t_i), \quad \text{(A.1)}$$

where $\Delta t_i = t_{i+1} - t_i$, $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$.

Let

$$Y_{t_{i+1}} = X_{t_{i+1}} - X_{t_i} - (c - \beta X_{t_i})\Delta t_i + \frac{1}{2}\sigma^2 \gamma X_{t_i}^{2\gamma-1}\Delta t_i.$$

Then (A.1) becomes

$$\frac{1}{2}\sigma^2 \gamma X_{t_i}^{2\gamma-1}(\Delta W_{t_i})^2 + \sigma X_{t_i}^{\gamma}\Delta W_{t_i} = Y_{t_{i+1}}. \quad \text{(A.2)}$$

Let $x_i \in \mathbb{R}$ still be the realizations of X_{t_i} and y_i be the realizations of Y_{t_i} for $i = 0, \cdots, N$ correspondingly. We solve the equation (A.2) to obtain the realizations of $\Delta W_{t_i} = u_{i+1}^{+}, u_{i+1}^{-}$, where

$$\begin{aligned} u_{i+1}^{+} &= \frac{-1 + \sqrt{1 + \frac{2\gamma y_{i+1}}{x_i}}}{\sigma \gamma x_i^{\gamma-1}} \\ u_{i+1}^{-} &= \frac{-1 - \sqrt{1 + \frac{2\gamma y_{i+1}}{x_i}}}{\sigma \gamma x_i^{\gamma-1}}. \end{aligned} \quad \text{(A.3)}$$

Then the conditional density is given by

$$\begin{aligned} p\Big(X_{t_{i+1}} = x_{i+1} \Big| X_{t_i} = x_i\Big) &= \frac{dP\Big(\{\Delta W_{t_i} = du_{i+1}^{+}\} \cup \{\Delta W_{t_i} = du_{i+1}^{-}\}\Big)}{dy_{i+1}} \\ =& \frac{dP\Big(\Delta W_{t_i} = du_{i+1}^{+}\Big)}{du_{i+1}^{+}} \left|\frac{du_{i+1}^{+}}{dy_{i+1}}\right| + \frac{dP\Big(\Delta W_{t_i} = du_{i+1}^{-}\Big)}{du_{i+1}^{-}} \left|\frac{du_{i+1}^{-}}{dy_{i+1}}\right| \\ =& \frac{1}{\sqrt{2\pi \Delta t_i}} \left(\exp\Big(-\frac{(u_{i+1}^{+})^2}{2\Delta t_i}\Big) + \exp\Big(-\frac{(u_{i+1}^{-})^2}{2\Delta t_i}\Big)\right) \left| \frac{1}{\sigma x_i^{\gamma} \sqrt{1 + \frac{2\gamma y_{i+1}}{x_i}}} \right|, \end{aligned} \quad \text{(A.4)}$$

Table 5
Results of estimation and forecast for Germany

Germany	Milstein CKLS	Euler CKLS	Euler LARMA	Euler LARMA-ARCH
Estimation [40]				
p		1	$\{1,\ 6\}$	$\{1,\ 6\}$
q		0	0	0
k		0	0	$\{1,\ 7\}$
$\alpha_0 \quad (c)$	0.020	0.017	0.068	0.065
(t-stat.)	(0.37)	(0.34)	(1.38)	(1.74)
$\alpha_1 \quad (-\beta)$	−0.007	−0.007	0.095	0.079
	(−0.73)	(−0.71)	(3.89)	(3.54)
α_6			−0.107	−0.091
			(−4.28)	(−4.01)
γ	0.417	0.378	0.186	0.485
	(2.21)	(2.01)	(1.00)	(2.14)
$c_0 \quad (\sigma)$	0.113	0.122	0.153	0.062
	(3.11)	(3.11)	(3.16)	(2.48)
c_1				0.297
				(2.43)
c_7				0.272
				(2.47)
log-lik	0.0054	0.0054	0.0061	0.0067
χ^2-test	161	160	95	31
(p-value)	($1.78e^{-25}$)	($2.91e^{-25}$)	($6.39e^{-13}$)	(0.02)
av. forecast error				
level (in)	0.0540	0.0541	0.0439	0.0441
% to naive	99%	99%	90%	90%
level (out)	0.0192	0.0193	0.0153	0.0156
% to naive	100%	100%	79%	81%
volatility (in)	0.0144	0.0144	0.0090	0.0082
volatility (out)	0.0017	0.0017	0.0015	0.0013

Table 6
Results of estimation and forecast for United Kingdom

United Kingdom	Milstein CKLS	Euler CKLS	Euler LARMA	Euler LARMA-ARCH
Estimation				
p		$\{1\}$	$\{1\}$	$\{1\}$
q		0	$\{1\}$	$\{1\}$
k		0	0	$\{1\}$
α_0 (c)	0.153	0.155	0.289	0.210
(t-stat.)	(1.30)	(1.23)	(1.63)	(1.71)
α_1 $(-\beta)$	-0.018	-0.019	-0.034	-0.025
	(-1.25)	(-1.26)	(-1.67)	(-1.71)
β_1			0.431	0.313
			(5.38)	(2.18)
γ	0.974	0.742	0.574	0.527
	(4.97)	(3.45)	(2.91)	(2.21)
c_0 (σ)	0.067	0.115	0.157	0.136
	(2.31)	(2.11)	(2.29)	(1.86)
c_1				0.498
				(2.31)
log-lik	0.00038	0.00019	0.00052	0.00091
χ^2-test(p-value)	1639	1675	349	235
	(0.00)	(0.00)	$(9.59e^{-64})$	$(2.22e^{-40})$
av. forecast error				
level (in)	0.3668	0.3668	0.3155	0.3212
% to naive	99%	99%	85%	87%
level (out)	0.0701	0.0701	0.0777	0.0714
% to naive	105%	105%	116%	107%
volatility (in)	1.1705	1.1503	0.6469	0.8050
volatility (out)	0.0218	0.0306	0.0298	0.0277

Table 7
Results of estimation and forecast for United States

USA	Milstein CKLS	Euler CKLS	Euler LARMA	Euler LARMA-ARCH
Estimation				
p		$\{1\}$	$\{1,\ 2\}$	$\{1,\ 2\}$
q		0	0	0
k		0	0	$\{1,\ 6\}$
α_0 (c)	0.048	0.047	0.055	0.028
(t-stat.)	(1.03)	(1.01)	(1.23)	(0.64)
α_1 $(-\beta)$	−0.010	−0.010	0.361	0.456
	(−1.22)	(−1.20)	(5.219)	(6.11)
α_2			−0.371	−0.461
			(−5.39)	(−6.29)
γ	0.827	0.839	0.767	0.808
	(5.70)	(5.74)	(5.25)	(3.57)
c_0 (σ)	0.055	0.054	0.057	0.037
	(3.70)	(3.68)	(3.68)	(2.23)
c_1				0.225
				(1.26)
c_6				0.330
				(2.07)
log-lik	0.0050	0.0050	0.0054	0.0057
χ^2-test	65	63	76	36
(p-value)	$(1.32e^{-7})$	$(3.43e^{-7})$	$(2.33e^{-9})$	(0.0053)
av. forecast error				
level (in)	0.0732	0.0732	0.0614	0.0618
% to naive	99%	99%	82%	83%
level (out)	0.0252	0.0252	0.0190	0.0187
% to naive	102%	102%	77%	76%
volatility (in)	0.0256	0.0256	0.0183	0.0178
volatility (out)	0.0020	0.0020	0.0017	0.0014

as $1+\frac{2\gamma y_{i+1}}{x_i} > 0$. If $1+\frac{2\gamma y_{i+1}}{x_i} < 0$, then the density above is infinity. If $1+\frac{2\gamma y_{i+1}}{x_i} < 0$, which means there is no real solution of ΔW_{t_i} in (A.2) for such y_{i+1}, therefore the density is equal to zero

$$p\Big(X_{t_{i+1}} = dx_{i+1}\Big|X_{t_i} = x_i\Big) = 0.$$

Comparing the density function (A.4) and the density function in (2.5) p.7 in Elerian (1998), it is not difficult to show the identity of these two functions by some calculation.

By numerical operations of the ML estimations we must modify the density function, because when $1+\frac{2\gamma y_{i+1}}{x_i} = 0$, the value of the density function is infinity. Therefore we apply the following density function for the ML estimations:

$$\begin{aligned}
g_{mil}(x_i, x_{i+1}, \theta, \Delta t_i) &= \frac{dP\Big(X_{t_{i+1}} = dx_{i+1}\Big|X_{t_i} = x_i\Big)}{dx_{i+1}} \\
&= \frac{1}{\sqrt{2\pi\Delta t_i}}\Bigg(\exp\Big(-\frac{(u_{i+1}^{+})^2}{2\Delta t_i}\Big) + \exp\Big(-\frac{(u_{i+1}^{-})^2}{2\Delta t_i}\Big)\Bigg)\Bigg|\frac{1}{\sigma x_i^{\gamma}\sqrt{1+\frac{2\gamma y_{i+1}}{x_i}}}\Bigg|, \\
&\text{for} \quad 1+\frac{2\gamma y_{i+1}}{x_i} > 10^{-10} \\
&= 10^{-10}, \\
&\text{otherwise.}
\end{aligned}$$

A.2 ML estimators in equivalent models

Here we give a simple proof about equivalent models. Let $l_1(\theta_1, x)$, $l_2(\theta_2, x)$ be two log-likelihood functions of two models. We say these two model are equivalent under reparametrization, if there exists a bijective mapping H so that for every observations $x \in \mathbb{R}^n$ we have

$$l_1(\theta_1, x) = l_2(H(\theta_1), x)$$

and symmetrically

$$l_2(\theta_2, x) = l_1(H^{-1}(\theta_2), x).$$

Also we can represent this equivalence in a shorter form with respect to their parameters:

$$\theta_2 = H(\theta_1).$$

Proposition 1 *We have two equivalent models. Let $\hat{\theta}_1$, $\hat{\theta}_2$ be the ML estimators. If the two ML estimators exist uniquely, and if the differential $H'(\theta) \neq 0$,* *for all θ, then these ML estimators satisfy also the following equivalence*

$$\hat{\theta}_2 = H(\hat{\theta}_1)$$

Proof

Because $\hat{\theta}_1$ is the ML estimator, then we have

$$\begin{aligned} 0 = \frac{\partial}{\partial \theta_1} l_1(x, \theta_1)\Big|_{\hat{\theta}_1} &= \frac{\partial}{\partial \theta} l_2(x, H(\theta_1))\Big|_{\hat{\theta}_1} \\ &= \frac{\partial}{\partial \theta_2} l_2(x, \theta_2)\Big|_{H(\hat{\theta}_1)} H'(\hat{\theta}_1). \end{aligned}$$

Because $H'(\hat{\theta}_2) \neq 0$, it follows

$$\frac{\partial}{\partial \theta_2} l_2(x, \theta_2)\Big|_{H(\hat{\theta}_1)} = 0.$$

Therefore $H(\hat{\theta}_1)$ be the ML estimator for l_2.

Because the ML estimator exists uniquely, we have

$$H(\hat{\theta}_1) = \hat{\theta}_2.$$

References

[Ait-Sahalia(1996)] Ait-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. The review of financial studies 9 (2), 385–426.

[Ait-Sahalia(1997)] Ait-Sahalia, Y., 1997. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica .

[Ait-Sahalia(1999)] Ait-Sahalia, Y., August 1999. Transition densities for interest rate and other nonlinear diffusions. The journal of finance LIV (4), 1361–1395.

[Andersen and Lund(1997)] Andersen, T., Lund, J., 1997. Estimating continuous-time stochastic volatility models of the short-term interest rate. Journal of Econometrics 77, 343–377.

[Bollerslev(1986)] Bollerslev, T., 1986. Generalzied autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

[Box et al.(1994)Box, Jenkins, and Reinsel] Box, G., Jenkins, G., Reinsel, G., 1994. Time series analysis, 3rd Edition. Prentice Hall.

[Box and Pierce(1970)] Box, G., Pierce, D., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of Amer. Statist. Assoc. 65, 1509–1526.

[Breiman(1973)] Breiman, L., 1973. Statistics. Houghton Mifflin company.

[Brenner et al.(1996)Brenner, Harjes, and Kroner] Brenner, R., Harjes, R., Kroner, K., March 1996. Another look at models of the short-term interest rate. Journal of financial and quantitative analysis 31 (1), 85–107.

[Chan et al.(1992)Chan, Karolyi, Longstaff, and Sanders] Chan, K., Karolyi, G., Longstaff, F., Sanders, A., July 1992. An empirical comparison of alternative models of the short-term interest rate. The journal of finance XLVII (3), 1209–1227.

[Cox et al.(1985a)Cox, Ingersoll, and Ross] Cox, J., Ingersoll, J., Ross, S., March 1985a. An intertemporal general equilibrium model of asset prices. Econometrica 53 (2), 363–384.

[Cox et al.(1985b)Cox, Ingersoll, and Ross] Cox, J., Ingersoll, J., Ross, S., March 1985b. A theory of the term structure of interest rates. Econometrica 53 (2), 385–407.

[Dacunha-Castelle and Florens-Zmirou(1986)] Dacunha-Castelle, D., Florens-Zmirou, D., 1986. Estimation of the coefficients of a diffusion from discrete observations. Stochastics 19, 263–284.

[Duffie and Glynn(2001)] Duffie, D., Glynn, P., 2001. Estimation fo continuous time markov processes sampled at random time intervals (2001), via http://www.stanford.edu/ duffie/dg.pdf .

[Durbin(1970)] Durbin, J., 1970. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. Biometrika 38, 410–421.

[Elerian(1998)] Elerian, O., 1998. A note on the exisistence of a colsed form conditional transition density for the milstein scheme (1998), via http://www.nuff.ox.ac.uk/users/elerian/milpaper.pdf .

[Engle(1982)] Engle, R. F., July 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50 (4), 987–1007.

[Frohn(1995)] Frohn, J., 1995. Grundausbildung in Ökonometrie, 2nd Edition. Walter de Gruyter.

[Fuller(1996)] Fuller, W., 1996. Introduction to statistical time series, 2nd Edition. Wiley Series in probability and statistics.

[Gallant and Tauchen(1996)] Gallant, A., Tauchen, G., 1996. Which moments to match. Econometric Theory 12, 657–681.

[Gallant and White(1988)] Gallant, A., White, H., 1988. A unified theory of estimation and inference for nonlinear dynamic system. Basil Blackwell.

[Gouriéroux et al.(1993)Gouriéroux, Monfort, and Renault] Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of applied econometrics 8, S85–S118.

[Hansen and Scheinkman(1995)] Hansen, L., Scheinkman, J., 1995. Back to the future: generating moment implications for continuous-time markov processes. Econometrica 63 (4), 767–804.

[Karatzas and Shreve(1991)] Karatzas, I., Shreve, S., 1991. Brownian motion and stochastic calculus. Springer.

[Koedijk et al.(1997)Koedijk, Nissen, Schotman, and Wolff] Koedijk, K., Nissen, F., Schotman, P., Wolff, C., 1997. The dynamics of short-term interest rate volatility. Euroean finance review 1, 105–130.

[Ljung and Box(1978)] Ljung, G., Box, G., 1978. On measure of lack of fit in time series models. Biometrika 65, 297–303.

[Lo(1988)] Lo, A., 1988. Maximum likelihood estimation of generalized itô processes with dsicretely sampled data. Econometric Theory 4, 231–247.

[Perdersen(1995)] Perdersen, A., 1995. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scandinavian journal of statistics 22, 55–71.

[Shoji and Ozaki(1997)] Shoji, I., Ozaki, T., 1997. Comparative study of estimation methods for continuous time stochastic processes. Journal of time series analysis 18 (5), 485–506.

[Shoji and Ozaki(1998)] Shoji, I., Ozaki, T., 1998. Estimation for nonlinear stochastic differencial equations by local linearization method. stochastic analysis and applications 16 (4), 733–752.

[Thompson(2002)] Thompson, S., 2002. Evaluating the goodness of fit of conditional distributions, with an application to affine term structure models (2002), via http://post.economics.harvard.edu/faculty/thompson/papers/spectest.pdf .

[Vasicek(1977)] Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of financial economics 5, 177–188.