A Statistical Comparison of the Short-Term Interest Rate Models for Japan, U.S., and Germany

ISAO SHOJI¹ and TOHRU OZAKI²
¹Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba-shi, Ibaraki-ken 305, Japan. e-mail: shoji@shako.sk.tsukuba.ac.jp
²The Institute of Statistical Mathematics, 4-6-7 Minami Azabu, Minato-ku, Tokyo 106, Japan. e-mail: ozaki@ism.ac.jp

Abstract. We propose a simple and practical model selection method for continuous time models. We apply the method to several continuous time short-term interest rate models using discrete time series data of Japan, U.S. and Germany. All the models can be easily estimated from discrete observations, and their performances can be evaluated in a uniform statistical framework. The models that allow dependence of volatility on the level of interest rates tend to perform well empirically. The degree of volatility dependence on the interest rate levels seems to be different across the countries. For the German data, we observe that a model with nonlinear drift performs better than the best linear drift model.

Key words: interest rate, stochastic differential equation, local linearization, AIC, model selection.

1. Introduction

Many financial models such as option pricing models, discount bond pricing models, and other types of asset pricing models have been developed in continuous time framework. For example, Merton (1973), Vasicek (1977), Brennan and Schwartz (1979), Dothan (1978), Cox, Ingersoll and Ross (1985), Longstaff (1989) and Longstaff and Schwartz (1992) use stochastic differential equations to model discounted bond pricing processes.

This issue as to how these models perform compared to each other, is very important for two reasons: (1) These models' implications for valuing contingent claims and hedging risks of financial assets are fundamentally different from each other. (2) These models are often used to price bonds and to forecast interest rate levels. Obviously, a model with a better descriptive power may be expected to perform well in forecasting exercises too. Despite its importance, not much statistical work seems to have been carried out to compare predictive powers of alternative continuous time financial asset models. In times series econometrics, such discrete time empirical models as the ARCH models, have been proposed so that it can capture the dynamics of various financial assets. These econometric time series models are useful and they are easy to estimate since discrete time data are used. Unfortunately, they are not based upon continuous time finance theories. Therefore,
we need a statistical method that can be applied to continuous time finance models so that we may easily obtain important information from past financial data, just like a discrete time econometric time series model enables us to do.

In a seminal paper, Chan et al. (1992a) have proposed an econometric method to compare the performances of continuous time interest rate models and clarified their stochastic natures and empirical implications. Notice that there are some short-term interest rate models that allow dependence of volatility on the level of interest rates, and some do not. Chan et al. (1992a, 1992b) and Fong and Vasicek (1991) conclude that interest rate volatility specification is crucial to models' performances. They empirically analyzed the U.S. short-term interest rate data, and concluded that the models allowing dependence of volatility on the level, seemed to capture the dynamics of interest rates best.

It is not clear, however, if volatility's dependence on the interest rate levels should be the same across different countries. We shall look into this using monthly short-term interest rates for three countries: Japan, U.S. and Germany. One of the key techniques used in Chan et al. (1992a)'s analysis is the likelihood ratio test of overidentifying restrictions for discretized models derived from stochastic differential equations. Although the likelihood-ratio testing procedure is valid for comparing the nested models to the general unrestricted model, it cannot be used to discriminate non-nested models. Furthermore, Chan et al. assumed that the general unrestricted model is the best model. It is not clear whether this assumption is reasonable or not. Their test does not give any insight into this matter.

To obviate these problems we use two statistical techniques: a new discretization method and an information criterion, the AIC. The AIC has been used for model selection purposes in various different areas, and it is generally accepted as an efficient and easy model selection technique (as Akaike (1977, 1983)). As to the discretization of continuous time models, we use a local linearization method developed by Shoji and Ozaki (1994). The local linearization method is far more efficient in discretization than the conventional discretization methods, e.g., the Euler method. The Euler method is known to cause some numerical problems. For instance, it is well known that it often results in an explosive model when the original continuous time model has a nonlinear drift term. By contrast, our local linearization method has been demonstrated to be computationally stable even when the drift function is nonlinear. In addition, our method has shown to be useful in such a nonlinear time series analysis as Ozaki (1992a, 1992b, 1993).

With the local linearization method and the AIC, we can easily estimate complex stochastic differential equations, and evaluate their goodness of fit. In this paper, introducing stochastic differential equations with a nonlinear drift term, we check whether the linear assumption for the drift term is supported by the historical data. Indeed, the drift term of all conventional interest rate models, are either constant or linear. Hence, testing the drift term linearity and constancy of the existing interest rate models, has an important implication for interest rate model building.
Table I. The interest rate models.

<table>
<thead>
<tr>
<th>Model#</th>
<th>Model Name</th>
<th>Differential Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Merton</td>
<td>( dx_1 = \alpha , dt + \sigma , dB_t )</td>
</tr>
<tr>
<td>2</td>
<td>Dothan</td>
<td>( dx_1 = \sigma x_t , dB_t )</td>
</tr>
<tr>
<td>3</td>
<td>Geometric Brownian Motion</td>
<td>( dx_1 = x_1 \beta , dt + \sigma x_t , dB_t )</td>
</tr>
<tr>
<td>4</td>
<td>Vasicek</td>
<td>( dx_1 = (\alpha + \beta x_t) , dt + \sigma , dB_t )</td>
</tr>
<tr>
<td>5</td>
<td>CIR SR</td>
<td>( dx_1 = (\alpha + \beta x_t) , dt + \sigma \sqrt{x_t} , dB_t )</td>
</tr>
<tr>
<td>6</td>
<td>Brennan–Schwartz</td>
<td>( dx_1 = (\alpha + \beta x_t) , dt + \sigma x_t , dB_t )</td>
</tr>
<tr>
<td>7</td>
<td>CIR VR</td>
<td>( dx_1 = \sigma x_t^{1/2} , dB_t )</td>
</tr>
<tr>
<td>8</td>
<td>CEV</td>
<td>( dx_1 = \beta x_t , dt + \sigma x_t^\gamma , dB_t )</td>
</tr>
<tr>
<td>9</td>
<td>Unrestricted</td>
<td>( dx_1 = (\alpha + \beta x_t) , dt + \sigma x_t^\gamma , dB_t )</td>
</tr>
</tbody>
</table>

List of the sources of the above models:
1: Merton (1973),
2: Dothan (1978),
3: Marsh and Rosenfeld (1983),
4: Vasicek (1977),
5: CIR (1985),
6: Brennan and Schwartz (1980),
7: CIR (1980),
8: Cox and Ross (1976).

The organization of this paper is as follows. In Section 2, we present the stochastic differential equations of the short-term interest rate models that are examined in this paper. We describe the method of model selection in Section 3. In Section 4, we discuss the data used in this paper. Empirical results are given in Section 5. In the same section, we test nonlinearity of the drift term of a stochastic process. Concluding remarks are given in Section 6.

2. The Interest Rate Model

In Table I, we present nine short-term interest rate models. Notice that model 9 nests all the other models, i.e., models 1 through 8.

Models 1 through 8 are all well known stochastic processes that are used for such purposes as characterizing the term structure of interest rates, valuing bond options, futures, futures option and other types of contingent claims. Among the many properties of these models discussed in Chan et al. (1992a, 1992b), we note the dependence of volatility on changes in the level of an interest rate. It is often observed that the interest rate is more volatile when its level is relatively high. For instance, Chan et al. (1992a, 1992b) and Fong and Vasicek (1991) suggested empirical evidence in favor of the state dependent volatility interest rate model over the constant volatility models. Using our new approach to model identification, we shall present empirical evidences that support this view in Section 5.
3. Discretization and Model Selection

The first step in estimating parameters of a continuous time stochastic process on discrete time data is the discretization of the original continuous time process. Models 1, 2 and 3 are simple enough so that the crude Euler method may be used to obtain the exact discretized process. The Euler method is the simplest method to discretize ordinary differential equations and stochastic differential equations, see, e.g., Kloeden and Platen (1992). In the Euler method, we only need to assume that the the derivatives (or the coefficients of the differential) to be locally constant. When we deal with more complex continuous time stochastic processes, however, we need to be very cautious. For example, Ozaki (1992a, 1992b, 1993) demonstrated that the Euler method often results in computational explosion even if the original continuous time process is stationary. In this section, following Shoji and Ozaki (1994), we introduce an alternative discretization method, which we call ‘a local linearization method’. Our method is applicable not only to the more complex interest rate models like models 4 through 9, but also to a wide class of nonlinear stochastic differential equations.

Let a continuous time stochastic process $x_t$ satisfy the following stochastic differential equation,

$$dx_t = f(x_t) \, dt + \sigma \, dB_t,$$  \hspace{1cm} (1)

where $f(\cdot)$ is a linear or nonlinear function of $x_t$, $\sigma$ is constant and $B_t$ is a standard Brownian motion. In the local linearization method, we approximate the drift function $f(x_t)$ in (1) locally using a linear function of $x_t$. Hence, we wish to investigate the local behavior of $f(x_t)$. Mathematically, a local behavior of a function may be expressed in terms of its differential, hence in the present case we need a differential of $f(x_t)$, which in turn can be characterized by the Ito formula:

$$df = \frac{\sigma^2}{2} f''(x_t) \, dt + f'(x_t) \, dx_t,$$ \hspace{1cm} (2)

where $f'$ and $f''$ are the first and second derivatives of $f$, respectively. In order to linearize $f$ with respect to $x_t$ and $t$, we assume that $f'(x_t)$ and $f''(x_t)$ are constant if $t$ belongs to a small time interval $[s, s + \Delta t]$, i.e., $t \in [s, s + \Delta t)$. We let,

$$df = \frac{\sigma^2}{2} f''(x_s) \, dt + f'(x_s) \, dx_t,$$

for $t \in [s, s + \Delta t)$ hereafter. Then, (2) can be solved as follows.

$$f(x_t) - f(x_s) = \frac{\sigma^2}{2} f''(x_s)(t - s) + f'(x_s)(x_t - x_s).$$ \hspace{1cm} (3)

Using this formula, we can obtain the following linear approximation of $f(x_t)$:

$$f(x_t) \approx L_s x_t + M_s t + N_s,$$ \hspace{1cm} (4)
where

\[ L_s = f'(x_s), \]
\[ M_s = \frac{\sigma^2}{2} f''(x_s), \]
\[ N_s = f(x_s) - f'(x_s)x_s - \frac{\sigma^2}{2} f''(x_s)s. \]

We have thus obtained the following linear stochastic differential equation

\[ dx_t = (L_s x_t + M_s t + N_s) \, dt + \sigma \, dB_t, \quad (5) \]

as an alternative to equation (1) supposing \( t \in [s, s + \Delta t] \). We now make a transformation of variable from \( x_t \) to \( y_t = \exp(-L_s t)x_t \). The stochastic differential equation in terms of \( y_t \) becomes,

\[
y_{s+\Delta t} = y_s + \int_s^{s+\Delta t} (M_s u + N_s) \exp(-L_s u) \, du + \sigma \int_s^{s+\Delta t} \exp(-L_s u) \, dB_u. \quad (6)
\]

Substituting \( y_t \) back into \( x_t \), we obtain a discretized process of \( x_t \):

\[
x_{s+\Delta t} = x_s + \frac{f(x_s)}{L_s} (\exp(L_s \Delta t) - 1) + \frac{M_s}{L_s^2} (\exp(L_s \Delta t) - 1 - L_s \Delta t) + \sigma \int_s^{s+\Delta t} \exp(L_s (s + \Delta t - u)) \, dB_u, \quad (7)
\]

where

\[ L_s = f'(x_s), \]
\[ M_s = \frac{\sigma^2}{2} f''(x_s). \]

After some calculations, we arrive at the following conditional variance expression:

\[ \text{Var}_s(x_{s+\Delta t}) = \left( \frac{\exp(2L_s \Delta t) - 1}{2L_s} \right) \sigma^2. \]

Since the transition probability of the above discretized process follows the Normal distribution, the parameters can be easily estimated by the maximum likelihood technique. Suppose that discrete observations on \( \{x_t\}_{1 \leq t \leq N} \) are given,
where $t_i - t_{i-1} = \Delta t$. Using the Markov property of the discretized process, the log-likelihood function $\log(p(x_{t_1}, \ldots, x_{t_N}))$ becomes:

$$
\log(p(x_{t_1}, \ldots, x_{t_N})) = \log \left( p(x_{t_1}) \prod_{i=2}^{N} p(x_{t_i} | x_{t_{i-1}}) \right)
= -\frac{1}{2} \sum_{i=2}^{N} \left\{ \frac{(x_{t_i} - E_{t_{i-1}})^2}{V_{t_{i-1}}} + \log(2\pi V_{t_{i-1}}) \right\} + \log(p(x_{t_1}))
$$

(9)

where

$$
E_t = x_t + \frac{f(x_t)}{L_t} (\exp(L_t \Delta t) - 1) + \frac{M_t}{L_t^2} (\exp(L_t \Delta t) - 1 - L_t \Delta t),
$$

$$
V_t = \left( \frac{\exp(2L_t \Delta t) - 1}{2L_t} \right) \sigma^2,
$$

$$
L_t = f'(x_t),
$$

$$
M_t = \frac{\sigma^2}{2} f''(x_t).
$$

In some models of Section 2, the coefficient of the diffusion term is a function of $x_t$, not a constant. Such models do not fall into the class of stochastic differential Equation (1). Fortunately, such differential equations can be easily transformed into (1) by the Ito formula. Let $\phi(x)$ be a second continuously differentiable function of $x$ such that

$$
g(x) \phi'(x) = \sigma.
$$

Then the new process $y_t \equiv \phi(x_t)$ satisfies the following stochastic differential equation:

$$
dy_t = \left( \frac{g(x_t)^2}{2} \phi''(x_t) + f(x_t) \phi'(x_t) \right) dt + \sigma dB_t.
$$

(10)

Since the above stochastic differential equation has a constant coefficient diffusion term, the local linearization method can be applied and the discretized process with respect to $y_t$ can be easily obtained.

For estimation purposes, we need to obtain the log-likelihood of observations $(x_{t_1}, \ldots, x_{t_N})$. Let $p(x_{t_1}, \ldots, x_{t_N})$ and $p(y_{t_1}, \ldots, y_{t_N})$ be the joint density functions of $x_t$ and $y_t$, respectively. Then, by the transformation rule of a density function, we obtain the following:

$$
p(x_{t_1}, \ldots, x_{t_N}) = p(y_{t_1}, \ldots, y_{t_N}) \left| \frac{\partial (y_{t_1}, \ldots, y_{t_N})}{\partial (x_{t_1}, \ldots, x_{t_N})} \right|,
$$

(11)
Table II. The short-term interest rates of Japan, U.S., and Germany.

<table>
<thead>
<tr>
<th></th>
<th>Japan (1-month)</th>
<th>U.S. (1-month)</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>asset</td>
<td>CD rate</td>
<td>Treasury Bill rate</td>
<td>Call money rate*</td>
</tr>
<tr>
<td>sample period</td>
<td>Feb '77 – Dec '92</td>
<td>Jan '64 – Dec '92</td>
<td>Jan '60 – Dec '92</td>
</tr>
</tbody>
</table>

* Average of overnight, 1-month, and 3-month call money rates.

where

$$
\left| \det \left( \begin{array}{c}
\frac{\partial (y_{t_1}, \ldots, y_{t_N})}{\partial (x_{t_1}, \ldots, x_{t_N})}
\end{array} \right) \right|
$$

is the absolute value of the Jacobian and it is equivalent to

$$
\prod_{i=1}^{N} |\phi'(x_{t_i})|.
$$

Hence,

$$
\log(p(x_{t_1}, \ldots, x_{t_N})) = \log(p(y_{t_1}, \ldots, y_{t_N})) + \sum_{i=1}^{N} |\phi'(x_{t_i})|.
$$

Here, note that the formula of $\log(p(y_{t_1}, \ldots, y_{t_N}))$ is equivalent to (9) if $x_t$ is substituted for $y_t$ in (9).

Since we have successfully formulated the log likelihood of the discrete observations, we may use Akaike’s Information Criterion (AIC) for model selection. One of the advantages of using the AIC is in the fact that we do not need to worry whether the models under consideration are nested or nonnested. Rather we may use one single criterion, i.e., the AIC. Using the AIC, we may arrive at the best model out of the nine alternative models of short-term interest rates in Section 2, somewhat mechanically.

4. The Data

We use monthly short-term interest rates data. The data sources of Japan, U.S. and Germany are, respectively, Nikkei Needs, Salomon Brothers Asia Limited and Datastream International (Japan) K.K. Further discussions of the data are given in Table II.

The summary statistics for the three interest rates are presented in Table III. The mean of the U.S. treasury bill rate (0.06201) is the highest of the three interest rates. The standard deviation of Japanese CD rate (0.01578) is much smaller than those
Table III. Summary Statistics of the short-term interest rates of U.S., Japan, and Germany.

<table>
<thead>
<tr>
<th>Interest rate</th>
<th>$N$</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>$\rho_3$</th>
<th>$\rho_4$</th>
<th>$\rho_5$</th>
<th>$\rho_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese CD</td>
<td>191</td>
<td>0.05731</td>
<td>0.01578</td>
<td>0.95</td>
<td>0.11</td>
<td>0.03</td>
<td>-0.31</td>
<td>-0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>U.S. T-Bull</td>
<td>348</td>
<td>0.06201</td>
<td>0.02356</td>
<td>0.95</td>
<td>0.04</td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>German Call</td>
<td>396</td>
<td>0.05470</td>
<td>0.02468</td>
<td>0.91</td>
<td>0.22</td>
<td>0.35</td>
<td>-0.10</td>
<td>-0.14</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

$N$ is the number of observations, $\mu$ is the mean, $\sigma$ is the standard deviation, and $\rho_i$ represent the partial autocorrelation of the $i$th lag.

of the U.S. treasury bill rates (0.02356) and Germany call money rates (0.02468). The low volatility of Japanese interest rates is also reported by Chan et al. (1992b).

5. Empirical Analysis

5.1. The Results of the Comparison

In Tables IV through VI we show results of estimated parameters of the three interest rates, the Japanese CD rate, the U.S. treasury bill rate and the German call money rate. Let us present discussions of each table in the following.

The Japanese CD rate results are given in Table IV. According to the table, the best model, i.e., the one that has the lowest AIC value, is the CIR VR model, while the worst model is the Merton model. Although the best model of our study is different from that of Chan et al. (1992b), both the model that they chose and the CIR VR model allow the dependence of volatility on the interest rate level. By contrast, the Merton and Vasicek models which do not allow volatility dependence on the level, turned out to have smaller explanatory powers. In addition, our likelihood ratio test statistic value (not presented here) for the null of the Vasicek model versus the alternative of the unrestricted model, is highly significant. Therefore, we interpret these results to imply that the volatility of the Japanese CD rate is sensitive to the level of the interest rate.

Let us turn to the estimates of $\gamma$ for the Japanese data. Since $\gamma$ enters the volatility term as $\sigma x_t^\gamma$, a constant volatility model is implied by $\gamma = 0$, while the state dependent volatility model has $\gamma \neq 0$. Moreover, we should note that the value of $\gamma > 1$ implies a nonlinear state dependency. Our estimated results show that the CEV and the unrestricted models have highly significant estimates of $\gamma$. In addition, the constant volatility models such as the Merton and Vasicek models have the largest AIC's among the nine models. Our empirical evidence is clear; the state dependent volatility models outperform the constant volatility models. This result is reasonable in view of the many empirical studies that report the important role of the state dependent volatility, e.g., Chan et al. (1992b) and the references cited therein.
Table IV. Parameter estimates of alternative models of the one-month Japanese CD rate.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma^2$</th>
<th>$\gamma$</th>
<th>logl</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merton</td>
<td>-0.002122</td>
<td>0.0</td>
<td>0.000286</td>
<td>0.0</td>
<td>916.09</td>
<td>-1828.19</td>
</tr>
<tr>
<td></td>
<td>(-0.50)</td>
<td>(20.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dothan</td>
<td>0.0</td>
<td>0.0</td>
<td>0.064762</td>
<td>1.0</td>
<td>951.09</td>
<td>-1900.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(15.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBM</td>
<td>0.0</td>
<td>0.010070</td>
<td>0.064664</td>
<td>1.0</td>
<td>951.10</td>
<td>-1898.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(15.04)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasicek</td>
<td>0.028922</td>
<td>-0.543966</td>
<td>0.000289</td>
<td>0.0</td>
<td>918.05</td>
<td>-1830.09</td>
</tr>
<tr>
<td></td>
<td>(1.82)</td>
<td>(-2.67)</td>
<td>(20.28)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIR SR</td>
<td>0.026615</td>
<td>-0.501853</td>
<td>0.004198</td>
<td>0.5</td>
<td>939.70</td>
<td>-1873.39</td>
</tr>
<tr>
<td></td>
<td>(1.88)</td>
<td>(-2.33)</td>
<td>(16.44)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brennan–Schwartz</td>
<td>0.022144</td>
<td>0.423602</td>
<td>0.065740</td>
<td>1.0</td>
<td>952.49</td>
<td>-1898.98</td>
</tr>
<tr>
<td></td>
<td>(1.63)</td>
<td>(-1.75)</td>
<td>(14.71)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIR VR</td>
<td>0.0</td>
<td>0.0</td>
<td>1.108590</td>
<td>1.5</td>
<td>956.01</td>
<td>-1910.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(15.17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEV</td>
<td>0.0</td>
<td>0.016264</td>
<td>1.820580</td>
<td>1.584674</td>
<td>956.13</td>
<td>-1906.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.27)</td>
<td>(1.30)</td>
<td>(12.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>0.016549</td>
<td>0.316336</td>
<td>1.462458</td>
<td>1.544327</td>
<td>956.98</td>
<td>-1905.95</td>
</tr>
<tr>
<td></td>
<td>(1.18)</td>
<td>(-1.11)</td>
<td>(1.32)</td>
<td>(12.09)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The parameters of the following stochastic differential equation are estimated by the maximum likelihood estimation based on the stochastic process discretized by local linearization. The interest rate $x_t$ is expressed in the annualized form and sampled from February 1977 to December 1992. Asymptotic $t$-statistics are given in parenthesis, and in computing the log-likelihood logl, the first observation and the constants are neglected.

$$dx_t = (\alpha + \beta x_t) \, dt + \sigma x_t^\gamma \, dB_t,$$

The results for the U.S. treasury bill rate are slightly different from those for the Japanese CD rate. The best model is the unrestricted model, while the worst is the Merton model. Similar to the Japanese CD rate result, the state dependent models seem to be preferred over the constant volatility model. Given the value of $\gamma$, the degree of volatility dependence on the level is stronger for the U.S. data compared to that of the Japanese data, since the estimates of the $\sigma^2$ parameter is almost twice as large as the estimates of the Japanese CD rates.

The results above are similar to those obtained by Chan et al. (1992a). Firstly, Chan et al. (1992a) showed that the Merton and Vasicek models have little explanatory powers, and the present study confirms this. Secondly, Chan et al. (1992a) assume that the unrestricted model performs best. We have reached the same conclusion although estimated $\gamma$ values are slightly different in our analysis and Chan et al.'s.

The results for the German call money rate are quite different from those of the U.S. and Japanese rates. Surprisingly, the best model is the CIR SR model. The CIR SR model performed poorly for the U.S. and Japanese data. The least favored
model on the German data, has been the CIR VR model. It is interesting to observe that most of the variance estimates are two to three times as large as those of the U.S. treasury bill rate. In particular, CIR VR model's $\sigma$ has a large estimated value. The CEV model and the unrestricted model have highly significant $\gamma$.

Our statistical analysis has uncovered an interesting aspect that differentiates various different interest rate models. Chan et al. (1992a) maintain that $\gamma$ is the most important parameter that differentiates the interest rate models. According to their analysis of the U.S. data, the models which allow $\gamma \geq 1$ show better performance than those which require $\gamma < 1$. One of our interests in the present study is to check this point with data of other countries. We have confirmed Chan et al.'s conclusion on our Japanese and the U.S. data. For the German data, however, estimated $\gamma$ values are quite different from that was obtained using the U.S. or Japanese data. Interestingly, the estimates of $\gamma$ for the three countries are all different from each other: around 1.5 for Japan, 1 for U.S., and 0.5 for Germany.

5.2. TESTING NONLINEARITY OF THE DRIFT TERM

All the models examined so far have constant or linear drift terms. In this section, in order to test for a nonlinearity of the drift term, we introduce a model that has polynomial function drift term in the interest rate.

$$dx_t = f(x_t) \, dt + \sigma x^\gamma \, dB_t$$ (13)
Table VI. Parameter estimates of alternative models of the German call money rate.

<table>
<thead>
<tr>
<th>Model</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \sigma^2 )</th>
<th>( \gamma )</th>
<th>logl</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merton</td>
<td>0.001448</td>
<td>0.0</td>
<td>0.001226</td>
<td>0.0</td>
<td>1617.33</td>
<td>-3230.65</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(49.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dothan</td>
<td>0.0</td>
<td>0.0</td>
<td>0.477815</td>
<td>1.0</td>
<td>1623.39</td>
<td>-3244.77</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td>(42.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBM</td>
<td>0.0</td>
<td>0.259180</td>
<td>0.472678</td>
<td>1.0</td>
<td>1625.77</td>
<td>-3247.53</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td>(40.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasicek</td>
<td>0.056540</td>
<td>-1.008587</td>
<td>0.001250</td>
<td>0.0</td>
<td>1625.49</td>
<td>-3244.98</td>
</tr>
<tr>
<td></td>
<td>(3.44)</td>
<td>(-4.83)</td>
<td>(46.61)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIR SR</td>
<td>0.055552</td>
<td>-0.990682</td>
<td>0.022264</td>
<td>0.5</td>
<td>1652.81</td>
<td>-3299.63</td>
</tr>
<tr>
<td></td>
<td>(4.32)</td>
<td>(-4.13)</td>
<td>(42.04)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brennan–Schwartz</td>
<td>0.047208</td>
<td>-0.778768</td>
<td>0.498510</td>
<td>1.0</td>
<td>1632.69</td>
<td>-3259.38</td>
</tr>
<tr>
<td></td>
<td>(3.84)</td>
<td>(-2.33)</td>
<td>(35.38)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIR VR</td>
<td>0.0</td>
<td>0.0</td>
<td>12.920058</td>
<td>1.5</td>
<td>1562.77</td>
<td>-3123.54</td>
</tr>
<tr>
<td></td>
<td>(3.84)</td>
<td>(41.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEV</td>
<td>0.0</td>
<td>0.039034</td>
<td>0.025778</td>
<td>0.5</td>
<td>1643.92</td>
<td>-3281.83</td>
</tr>
<tr>
<td></td>
<td>(3.32)</td>
<td>(4.38)</td>
<td>(13.63)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>0.055555</td>
<td>-0.988371</td>
<td>0.026401</td>
<td>0.5</td>
<td>1652.90</td>
<td>-3297.81</td>
</tr>
<tr>
<td></td>
<td>(4.30)</td>
<td>(-3.91)</td>
<td>(4.19)</td>
<td>(13.02)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table VII. Values of AIC for polynomial models

<table>
<thead>
<tr>
<th>Country</th>
<th>degree</th>
<th>0.0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>1</td>
<td>-1830.09</td>
<td>-1873.39</td>
<td>-1898.98</td>
<td>-1907.89</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1829.57</td>
<td>-1871.73</td>
<td>-1896.98</td>
<td>-1906.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-1828.06</td>
<td>-1870.07</td>
<td>-1895.29</td>
<td>-1904.43</td>
</tr>
<tr>
<td>U.S.</td>
<td>1</td>
<td>-3069.88</td>
<td>-3172.75</td>
<td>-3218.25</td>
<td>-3207.57</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-3074.42</td>
<td>-3172.23</td>
<td>-3216.25</td>
<td>-3206.70</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-3072.77</td>
<td>-3170.24</td>
<td>-3214.25</td>
<td>-3204.77</td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
<td>-3244.98</td>
<td>-3299.59</td>
<td>-3259.38</td>
<td>-3144.22</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-3246.67</td>
<td>-3297.83</td>
<td>-3263.80</td>
<td>-3155.67</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-3255.00</td>
<td>-3304.95</td>
<td>-3265.41</td>
<td>-3154.00</td>
</tr>
</tbody>
</table>

We specify our nonlinear drift term model as follows:

\[
dx_t = f(x_t) \, dt + \sigma x_t^\gamma \, dB_t,
\]

where \( f(\cdot) \) is a polynomial function of \( x(t) \). The table below shows the value of AIC computed for the short-term interest rates of Japan, U.S. and Germany. In computing the AIC, the first observation and the constants are neglected.
where,

\[ f(x) = \sum_{i=0}^{m} \alpha_i x^i \]

and \( \sigma \) is constant. We assume \( 0 \leq m \leq 3 \) for practical reasons.

For fixed \( \gamma \), by varying the degree \( m \) from one to three, nonlinearity of the drift term can be tested. We may conclude that there is a nonlinearity in the drift term if the degree of the best polynomial model is greater than one. In the following tests, using the local linearization method, we estimate parameters of the interest rates models for Japan, U.S. and Germany. We then compute the AIC's of the estimated models for four cases; \( \gamma = 0, 0.5, 1, \) and \( 1.5 \).

Table VII shows the values of AIC for all polynomial models. For Japan and U.S., we do not observe nonlinearity in the drift term since the AIC is the smallest when \( m = 1 \), i.e., linear term case.

Unlike U.S. and Japan, for Germany, descriptive power of a polynomial model improves as the degree becomes higher. For \( \gamma = 0, 0.5 \) and \( 1 \), the model with \( m = 3 \) is the best, while for \( \gamma = 1.5 \) the model with \( m = 2 \) is the best. Recalling that the best linear model in the previous analysis is the CIR SR model, i.e., \( \gamma = 0.5 \), we may conclude that the polynomial model with \( \gamma = 0.5 \) and degree = 3 is the best among all models.

6. Conclusion

In this paper, using two statistical techniques, the local linearization method and the AIC, we proposed a method of model selection. We applied the method to the short-term interest rate models studied by Chan et al. (1992a, 1992b) using monthly short-term interest rates of the following three countries: one-month Japanese CD rate, one-month U.S. treasury bill rate, and Germany call money rate.

Our empirical comparisons show that different models perform best for the three countries: the CIR VR model performs best for the Japanese CD rate, the unrestricted model performs best for the U.S. treasury bill rate, and the CIR SR model performs best for the German call money rate. Chan et al. (1992a) supposed that unrestricted model is always the best model. We, however, found some evidence against such thesis. Therefore, we caution that such assumption should not be made, particularly for the Japanese and the German interest rate data. As to the drift term nonlinearity, we found that only the German call money rate has a nonlinear drift, but not for the U.S. treasury bill rate or the Japanese CD rate.

Our empirical investigations have shown that the model selection and the discretization method proposed in this paper, are not only quite robust for uncovering various different properties of financial data, but also simple and efficient to use.
Acknowledgment

The bulk of this research was carried out when the first author was a graduate student at the Graduate University for Advanced Studies. The authors would like to thank the editors and referees for their careful readings and helpful comments.

References