5-9-2015

Two Essays on Estimation and Inference of Affine Term Structure Models

Qian Wang

Two essays on estimation and inference of affine term structure models

By

Qian Wang

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Finance
in the Department of Finance and Economics

Mississippi State, Mississippi

May 2015

Two essays on estimation and inference of affine term structure models

By

Qian Wang

Approved:

Kenneth D. Roskelley
(Director of Dissertation/Graduate Coordinator)

Michael J. Highfield
(Committee Member)

Brandon N. Cline
(Committee Member)

Randall C. Campbell
(Committee Member)

Joe H. Sullivan
(Committee Member)

Sharon L. Oswald
Dean
College of Business

Name: Qian Wang

Date of Degree: May 9, 2015

Institution: Mississippi State University

Major Field: Finance

Major Professor: Kenneth D. Roskelley

Title of Study: Two essays on estimation and inference of affine term structure models

Pages in Study: 147

Candidate for Degree of Doctor of Philosophy

Affine term structure models (ATSMs) are one set of popular models for yield curve modeling. Given that the models forecast yields based on the speed of mean reversion, under what circumstances can we distinguish one ATSM from another? The objective of my dissertation is to quantify the benefit of knowing the "true" model as well as the cost of being wrong when choosing between ATSMs. In particular, I detail the power of out-of-sample forecasts to statistically distinguish one ATSM from another given that we only know the data are generated from an ATSM and are observed without errors. My study analyzes the power and size of affine term structure models (ATSMs) by evaluating their relative out-of-sample performance.

Essay one focuses on the study of the one-factor ATSMs. I find that the model's predictive ability is closely related to the bias of mean reversion estimates no matter what the true model is. The smaller the bias of the estimate of the mean reversion speed, the better the out-of-sample forecasts. In addition, my finding shows that the models' forecasting accuracy can be improved, in contrast, the power to distinguish between

different ATSMs will be reduced if the data are simulated from a high mean reversion process with a large sample size and with a high sampling frequency.

In the second essay, I extend the question of interest to the multi-factor ATSMs. My finding shows that adding more factors in the ATSMs does not improve models' predictive ability. But it increases the models' power to distinguish between each other. The multi-factor ATSMs with larger sample size and longer time span will have more predictive ability and stronger power to differentiate between models.

DEDICATION

This dissertation endeavor has been dedicated to my parents, my husband, my brother, and my SCCC members.

ACKNOWLEDGEMENTS

I acknowledge the support, the encouragement, and the insightful guidance for this dissertation by my dissertation chair, Dr. Kenneth D. Roskelley. I have been extremely lucky to have an advisor who cared so much about my work and my career. I would also like to thank Dr. Joe H. Sullivan for the help and comments. I am also indebted to Dr. Michael J. Highfield, Dr. Brandon N. Cline, and Dr. Randall C. Campbell, who supported my research and gave their best suggestions. With their help and support, I have been able to complete this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

LITERATURE REVIEW

The term structure of interest rates, or called yield curve, describes the relationship between the yields-to-maturity (YTM) of a set of zero-coupon bonds and their time-to-maturity. How to model the yield curve and forecast their future movements are widely discussed in the finance literature. Affine term structure models (ATSMs) are one set of popular models for yield curve modeling.

ATSMs have been used since the early 1980s. One big advantage of ATSMs is its tractability because: a) ATSMs hypothesize that the interest rates are a linear function of a set of observed/latent factors; b) ATSMs have less computational burden, comparing with Monte Carlo methods or solution methods for partial differential equations.

The rest of this chapter will give the background of ATSMs. Starting with the statement of some basic definitions and notations, I will first introduce one-factor ATSMs, followed by the Multifactor ATSMs.

1.1 Definitions and notations

Consider a zero coupon bond sold at time t and due to mature at time $T > t$. Denote $\tau = (T - t)$ as the time to maturity. Suppose the bond has the market value $P(t, T)$ at time t and has a unit payoff at maturity, i.e., $P(T, T) = 1$. At time t, the yield to maturity

$R(t,T)$ is the continuously compounded rate of return on a zero-coupon bond with unit value at time *T*.

$$P(t,T) = e^{-R(t,T)(T-t)} \tag{1.1}$$

$$\Updownarrow$$

$$R(\mathrm{t},\mathrm{T}) = -\frac{\ln P(t,T)}{T-t} = -\frac{\ln P(t,t+\tau)}{\tau} \tag{1.2}$$

Restricted to a continuous time process, the formulation of the instantaneous short rate is

$$r(t) = \lim_{T\to t} R(t,T) = \lim_{\tau\to 0} -\frac{\ln P(t,t+\tau)}{\tau} \tag{1.3}$$

If the dynamics of the instantaneous short rate are known, the price of the zero-coupon bond, $P(t,\ T)$, may be written as a function of $r(t)$,

$$P\big(t,T,r(t)\big) = E_t[\text{Present Value}] = E_t\left[e^{-\int_t^T r(u)du}\right] \tag{1.4}$$

1.2 Single-factor ATSMs

Before considering the complex multi-factor ASTMs, it is necessary to introduce relatively simple models: one-factor affine term structure models. Just like the name implies, there is only one factor deciding the dynamics of the entire term structure. That factor is the instantaneous short rate $r(t)$, written as r_t for simplicity. One-factor ATSMs assume that the instantaneous short rate follows a continuous Markov process. This means that, given all the information up to and including time *s*, the conditional probability distribution of r_t at time *s* is completely determined by the value of r_s.

Or

$$p_r(\, r_t \mid r_s,\, r_u, 0 \le u < s < t) = p_r(\, r_t \mid r_s \,) \tag{1.5}$$

where $p_r(\cdot \mid \cdot\,)$ is the conditional probability of the process $r(t)$.

A continuous Markov process is called the diffusion process. It can be described by a stochastic differential equation (SDE). So the instantaneous short rate can be expressed as the form

$$dr_t = f(r_t, t)dt + \rho(r_t, t)dW_t \tag{1.6}$$

where W_t is a Wiener process with incremental variance dt. So $dW_t \sim N(0, dt)$. $f(r_t, t)$ and $\rho(r_t, t)$ are the drift and diffusion terms of r_t, respectively.

Application of Ito's theorem yields that the bond price *P* satisfies a SDE

$$dP = P\mu(t, T, r_t)dt + P\sigma(t, T, r_t)dW_t \tag{1.7}$$

where parameters $\mu(t, T, r_t)$ and $\sigma(t, T, r_t)$ are given by

$$\mu(t, T, r_t) = \frac{1}{P}\left[\frac{\partial P}{\partial t} + f\frac{\partial P}{\partial r} + \frac{1}{2}\rho^2\frac{\partial^2 P}{\partial r^2}\right] \tag{1.8}$$

$$\sigma(t, T, r_t) = \frac{1}{P}\,\rho\frac{\partial P}{\partial r} \tag{1.9}$$

To formulate the fair values of zero coupon bonds, I assume an arbitrage-free market and construct a risk-free portfolio. The idea of this approach is to continuously adjust the portfolio weights to replicate the unexpected movements in the bond. This eliminates all the risk from the portfolio.

Consider two zero coupon bonds, 1 and 2, with maturity time T_1 and T_2, respectively. Suppose an investor issues an amount of B_1 of bond 1 and simultaneously buys an amount of B_2 of bond 2 at time *t*. Then the value of the portfolio is $B = B_2 - B_1$

and the increment of the portfolio is $dB = dB_2 - dB_1$. Assume the instantaneous short rate satisfies equation (1.6). According to Ito's differentiation rule,

$$dB_1 = B_1\mu_1(r_t,t)dt + B_1\sigma_1(r_t,t)\,dW_t \tag{1.10}$$

$$dB_2 = B_2\mu_2(r_t,t)dt + B_2\sigma_2(r_t,t)\,dW_t \tag{1.11}$$

then

$$dB = (B_2\mu_2(r_t,t) - B_1\mu_1(r_t,t))dt + (B_2\sigma_2(r_t,t) - B_1\sigma_1(r_t,t))\,dW_t \tag{1.12}$$

To simplify the writing, let $\mu_1 \equiv \mu_1(r_t,t), \mu_2 \equiv \mu_2(r_t,t), \sigma_1 \equiv \sigma_1(r_t,t), \sigma_2 \equiv \sigma_2(r_t,t)$, the above equation can be written as

$$dB = (B_2\mu_2 - B_1\mu_1)dt + (B_2\sigma_2 - B_1\sigma_1)\,dW_t \tag{1.13}$$

In order to eliminate the risk, choose B_1 and B_2 as

$$B_1 = (B_2 - B_1)\frac{\sigma_2}{\sigma_1-\sigma_2} \tag{1.14}$$

$$B_2 = (B_2 - B_1)\frac{\sigma_1}{\sigma_1-\sigma_2} \tag{1.15}$$

Substituting Eqs. (1.14) and (1.15) into (1.13) obtains

$$dB = (B_2 - B_1)\frac{\sigma_1\mu_2-\sigma_2\mu_1}{\sigma_1-\sigma_2}dt \tag{1.16}$$

Therefore, the diffusion term in Eq. (1.13) is eliminated, $B_2\sigma_2 - B_1\sigma_1 = 0$. This means the unexpected component of the portfolio has been removed and the portfolio is completely predictable all the time. So the portfolio *B* is risk-free. Since the market is assumed to be no arbitrage, the portfolio *B* has to appreciate at the risk free rate r_t over a time interval, *dt*. So the increment of *B*, *dB*, over *dt* is expressed as

$$dB = Br_tdt = (B_2 - B_1)r_tdt \tag{1.17}$$

Combination of Eq. (1.16) and (1.17) yields

$$(B_2 - B_1)\frac{\sigma_1\mu_2 - \sigma_2\mu_1}{\sigma_1 - \sigma_2}dt = (B_2 - B_1)r_t dt \tag{1.18}$$

Rearranging the above equation gets,

$$\frac{\sigma_1\mu_2 - \sigma_2\mu_1}{\sigma_1 - \sigma_2} = r_t \tag{1.19}$$

or equivalently,

$$\frac{\mu_1(r_t,t) - r_t}{\sigma_1(r_t,t)} = \frac{\mu_2(r_t,t) - r_t}{\sigma_2(r_t,t)} \triangleq \Lambda \tag{1.20}$$

Λ is named as the market price of risk and it represents the standardized excess return for holding a zero-coupon bond. We know a zero-coupon bond pays its face value with certainty at maturity. However, the instantaneous short rate always changes as the bond approaches maturity. So Λ is introduced to account for the uncertainty of the change of the instantaneous short rate. The detail description on the market price of risk will be given in Section 1.4.

If I substitute Eqs. (1.8) and (1.9) into Eq. (1.20), the price of a zero-coupon bond has a partial differential equation with the form

$$\frac{\partial P}{\partial t} + (f + \rho\Lambda)\frac{\partial P}{\partial r} + \frac{\rho^2}{2}\frac{\partial^2 P}{\partial r^2} - rP = 0 \tag{1.21}$$

and is subject to the boundary condition $P(T, T) = 1$. Once the processes of r_t and Λ are specified, the bond prices can be obtained by solving Eq. (1.21).

There are two well-defined one-factor affine term structure models, which are the Vasicek model and Cox, Ingersoll and Ross model. The following two sections will describe these two models in detail.

1.2.1 Vasicek model (1977)

Assuming the instantaneous short rate follows a mean-reverting Gaussian process, Vasicek (1977) obtains the explicit forms of $R(t,T)$ and $P(t, T)$ of a zero-coupon bond. Specifically, Vasicek assumes r_t follows an Ornstein-Uhlenbeck process, with the choice $f(r_t,t) = \kappa(\theta - r_t)$ and $\rho(r_t,t) = \sigma$ in equation (1.6). So the Vasicek model has the expression

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t \tag{1.22}$$

where θ is the long-run mean, κ is the speed of mean reversion, and W_t is a Wiener process.

Based on the process of equation (1.22), the transition density function of r_t can be derived explicitly, which is a conditional normal distribution. Given the value of the instantaneous short rate at time s, r_s, the transition density function of r_t ($t > s > 0$) is

$$r_t|r_s \sim N(\theta\left(1 - e^{-\kappa(t-s)}\right) + e^{-\kappa(t-s)}r_s, \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right)) \tag{1.23}$$

Equations (1.22) and (1.23) imply that the Vasicek model allows the existence of negative interest rates, which is not feasible in the real world.

One advantage of the Ornstein-Uhlenbeck process is that the bond price can be expressed as a closed-form, i.e., given the current instantaneous rate, $r_t \equiv r$, the zero-coupon bond price $\{P(t,T,r) \equiv P(t,T),\ t \in [0,T]\}$ has the form:

$$P(t,T) = e^{A(t,T) - B(t,T)r} \tag{1.24}$$

where $A(t,T)$ and $B(t,T)$ are deterministic functions. So equation (1.24) is termed as *an affine term structure model*. If thinking T is a fixed parameter, $P(t,T)$, $A(t,T)$ and $B(t,T)$ are only functions of t. We can then denote $P(\tau) \equiv P(t,T),\ A(\tau) \equiv A(t,T)$ and

$B(\tau) \equiv B(t,T)$ for notational simplicity. Thus, I can rewrite equation (1.24) in the following form:

$$P(\tau) = e^{A(\tau)-B(\tau)r} \tag{1.25}$$

where

$$A(\tau) = \left(\theta - \frac{\sigma\lambda}{\kappa} - \frac{\sigma^2}{2\kappa^2}\right)(B(\tau) - \tau) - \frac{\sigma^2 B^2(\tau)}{4\kappa} \tag{1.26}$$

$$B(\tau) = \frac{1}{\kappa}(1 - e^{-\kappa\tau}) \tag{1.27}$$

The zero bond yields have the expression

$$R(t,T) \equiv R(\tau) = -\frac{A(\tau)-B(\tau)r}{\tau} \tag{1.28}$$

with the $A(\tau)$ and $B(\tau)$ the same definitions as those in Eq. (1.25).

In Eqs. (1.25) and (1.28), I introduce a new parameter λ, which is the market price of risk. λ is the specific definition to Λ and is a constant in the Vasicek model.

1.2.2 Cox, Ingersoll, and Ross model (1985)

In the Vasicek model, the conditional volatility of the instantaneous short rate is constant and independent of the level of the interest rate. Different from the Vasicek model, Cox, Ingersoll, and Ross (1985) (hereafter CIR) assume that the conditional volatility is proportional to the square root of the instantaneous short rate. So the CIR model is a mean-reverting square root process. The dynamics of the instantaneous short rate of the CIR model can be expressed as:

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t \tag{1.29}$$

where κ and θ have the same interpretations as those in the Vasicek model. Imposing the restriction $2\kappa\theta \geq \sigma^2$ makes sure that the instantaneous short rate cannot reach zero. The model structure implies that the instantaneous short rate process is always nonnegative.

In this model, r_t has a conditional non-central chi-square distribution. Conditional on the value at time *s*, the transition density function of r_t (*t* > *s* > *0*) is

$$\begin{aligned} f(r_t|r_s) &= ce^{-u-v}\left(\frac{v}{u}\right)^{\frac{q}{2}} I_q\left(2(uv)^{\frac{1}{2}}\right)) \\ &\sim Non-central\ \chi^2(2cr_t; 2q+2, 2u) \end{aligned} \tag{1.30}$$

where $I_q(\cdot)$ is the modified Bessel function of the first kind of order q, and

$$c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa(t-s)})} \tag{1.31}$$

$$u = cr_s e^{-\kappa(t-s)} \tag{1.32}$$

$$v = \ cr_t \tag{1.33}$$

$$q = \frac{2\kappa\theta}{\sigma^2} - 1 \tag{1.34}$$

The conditional first and second moment of r_t are:

$$E(r_t|r_s) = r_s e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right) \tag{1.35}$$

$$\text{VAR}(r_t|r_s) = r_s\left(\frac{\sigma^2}{\kappa}\right)(e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}) + \theta\left(\frac{\sigma^2}{2\kappa}\right)\left(1 - e^{-\kappa(t-s)}\right)^2 \tag{1.36}$$

As *t* and *T* (*t* < *T*) become large, the conditional distribution will approach a Gamma distribution. The transition density function becomes:

$$f(r_\infty|r_s) = \frac{\omega^\nu}{\Gamma(\nu)} r^{\nu-1} e^{-\omega r} \tag{1.37}$$

where $\omega = \frac{2\kappa}{\sigma^2}$ and $\nu = \frac{2\kappa\theta}{\sigma^2}$.

In the CIR model, the market price of risk, denoted as $\Lambda(r_t)$, is assumed to be an affine function of r_t,

$$\Lambda(r_t) = \lambda r_t \text{ , } \lambda \text{ is constant} \tag{1.38}$$

The price and yield of a zero-bond have the same forms as those of Vasicek model

$$P(\tau) = e^{A(\tau)-B(\tau)r} \text{ or } R(\tau) = -\frac{A(\tau)-B(\tau)r}{\tau} \tag{1.39}$$

but with different $A(\tau)$ and $B(\tau)$:

$$A(\tau) = \frac{2\kappa\theta}{\sigma^2} ln[\frac{2\gamma e^{\frac{\tau}{2}(\kappa+\lambda+\gamma)}}{(\kappa+\lambda+\gamma)(e^{\gamma\tau}-1)+2\gamma}] \tag{1.40)I}$$

$$B(\tau) = \frac{2(e^{\gamma\tau}-1)}{(\kappa+\lambda+\gamma)(e^{\gamma\tau}-1)+2\gamma} \tag{1.41}$$

$$\gamma = \sqrt{(\kappa+\lambda)^2 + 2\sigma^2} \tag{1.42}$$

1.3 Multifactor Affine Term Structure Models

One-factor ASTMs are the first step in modeling the dynamic term structure of interest rates. In such models, the instantaneous short rate is the only factor. However, they are too simple to describe the complete movement of the entire yield curve. Therefore, it is necessary to introduce multifactor models. The multifactor ATSMs consider bond yields as functions of a few factors, latent or observable. For example, the factors can be macroeconomic and/or financial variables. Same as the Vasicek and CIR models, the factors in the multifactor ATSMs are assumed to follow continuous Markov processes.

In the single-factor model, the no-arbitrage approach is used to price a zero coupon bond. An alternative approach to formulate the value of a zero coupon bond is to find an equivalent martingale measure (or risk-neutral measure) Q such that the risk can be removed under this measure. Although the first approach is pricing the bond assuming no arbitrage and the second one is applied under an equivalent risk-neutral measure, two approaches are equivalent.

The key idea of the second approach is to find a risk-neutral measure Q such that under this measure discounting a zero coupon bond by risk free rates gets a martingale[1,2], which is expressed in a mathematical form as

$$P\big(t,T,r(t)\big) = E_t^Q\left[e^{-\int_t^{t+\tau} r(u)du}\right] = E_t^P\left[e^{-\int_t^{t+\tau} r(u)du}\right] \tag{1.43}$$

Duffie and Kan (1996) use the risk-neutral measure and formulate multifactor term structure models. They assume that the future movement of the instantaneous short rate r_t is a function[3], $F(\boldsymbol{X_t})$, of a vector of factors $\boldsymbol{X_t} = (X_1(t), X_2(t), \ldots, X_N(t))^\intercal$. They prove that a multifactor term structure model has an affine representation (i.e., Eq. (1.24) in the one-factor ATSMs) if and only if two requirements are satisfied.

(1) $F(\boldsymbol{X_t})$ is assumed to be affine, which is defined by

$$r(t) = F(\boldsymbol{X_t}) = \delta_0 + \boldsymbol{\delta_x^\intercal X_t} \tag{1.44}$$

---

[1] A process $\{P_t, t > 0\}$ is a martingale if $E_t[P_t] = P_s$, for all $t > s$, given all the information up to and including time s. That is, the best forecast of unobserved future values is the last observation on P_t.

[2] To find a risk-neutral measure Q means to find a probability distribution Q such that $E_t^Q[e^{-rt}P_{u+t}] = P_u$, for all $t > 0$.

[3] In this paper, bold symbols are vectors or matrices. Others are scalars.

(2) X_t has an affine diffusion process under a risk-neutral probability measure Q. This means X_t is assumed to satisfy a stochastic differential equation on the given probability space Q,

$$\boldsymbol{dX_t} = \boldsymbol{\mu^Q}(\boldsymbol{X_t})\boldsymbol{dt} + \boldsymbol{\sigma^Q}(\boldsymbol{X_t})\boldsymbol{dW_t^Q} \tag{1.45}$$

where $\boldsymbol{W_t^Q}$ is a standard N-dimensional Brownian motion under Q, with both the drifts and the conditional variances of $\boldsymbol{X_t}$ are affine and have expressions:

$$\boldsymbol{\mu^Q}(\boldsymbol{X_t}) = \boldsymbol{\kappa^Q}(\boldsymbol{\theta^Q} - \boldsymbol{X_t}) \tag{1.46}$$

$$\boldsymbol{\sigma^Q}(\boldsymbol{X_t}) = \boldsymbol{\Sigma}\sqrt{\boldsymbol{S}(\boldsymbol{X_t})} \tag{1.47}$$

where $\boldsymbol{\theta^Q} \in \mathbb{R}^{N\times 1}$ and $\boldsymbol{\kappa^Q}, \boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$. $\boldsymbol{S}(\boldsymbol{X_t})$ is a diagonal $N \times N$ matrix with the i-th diagonal element given by

$$[\boldsymbol{S}(\boldsymbol{X_t})]_{ij} = \begin{cases} \alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{X_t}, & \forall\, i = j; \\ 0, & \forall\, i \neq j. \end{cases} \tag{1.48}$$

where $\alpha_i \in \mathbb{R}$ and $\boldsymbol{\beta_i} \in \mathbb{R}^{N\times 1}$.

In general, the dynamics of state variables of the multifactor ATSMs has the form

$$\boldsymbol{dX_t} = \boldsymbol{\kappa^Q}(\boldsymbol{\theta^Q} - \boldsymbol{X_t})\,\boldsymbol{dt} + \boldsymbol{\Sigma}\begin{pmatrix} \sqrt{S_1(\boldsymbol{X_t})} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{S_n(\boldsymbol{X_t})} \end{pmatrix}\boldsymbol{dW_t^Q} \tag{1.49}$$

with $S_i(\boldsymbol{X_t}) = \alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{X_t}$

To make sure the solution for Eq. (1.44) exists and is unique, the necessary condition to Eq. (1.49) is

For all i:

(1) *For all x such that* $v_i(\boldsymbol{X_t}) = 0, \boldsymbol{\beta}_i^\top(ax + b) > \boldsymbol{\beta}_i^\top \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \boldsymbol{\beta_i}/2$.

(2) *For all j, if* $(\boldsymbol{\beta}_i^{\intercal}\boldsymbol{\Sigma})_j \neq 0, then\ v_i = v_j$.

The conditions ensure that $v_i(\boldsymbol{X_t})$ is nonnegative for all *i* and *t*. A special case is that $\boldsymbol{\beta_i}$ is zero, which means the factors have constant volatility. This case is what the Vasicek model assumes. So there is no condition required for the existence of unique solutions if $\boldsymbol{\beta_i}$ is zero.

When Eqs. (1.44) and (1.49) are satisfied, Duffie and Kan (1996) derive that the zero-coupon bond price has an exponential affine form that can be expressed as

$$P(\boldsymbol{X_t}, \tau) = e^{A(\tau) - \boldsymbol{B}(\boldsymbol{\tau})^{\intercal}\boldsymbol{X_t}} \tag{1.50}$$

where

$$\frac{dA(\tau)}{d\tau} = -\delta_0 - (\boldsymbol{\theta^Q})^{\intercal}(\boldsymbol{\kappa^Q})^{\intercal}\boldsymbol{B}(\boldsymbol{\tau}) + \frac{1}{2}\sum_{i=1}^{n}[\boldsymbol{\Sigma}^{\intercal}\boldsymbol{B}(\boldsymbol{\tau})]_i^2\alpha_i \tag{1.51}$$

$$\frac{d\boldsymbol{B}(\boldsymbol{\tau})}{d\tau} = \boldsymbol{\delta_x} + (\boldsymbol{\kappa^Q})^{\intercal}\boldsymbol{B}(\boldsymbol{\tau}) - \frac{1}{2}\sum_{i=1}^{n}[\boldsymbol{\Sigma}^{\intercal}\boldsymbol{B}(\boldsymbol{\tau})]_i^2\boldsymbol{\beta}_i \tag{1.52}$$

with initial conditions $A(0) = 0$ and $B(\boldsymbol{0}) = \boldsymbol{0}_{N\times 1}$. These are known as Ricatti equations.

1.4 The market price of risk

As stated in the above section, the multifactor affine term structure models derived by Duffie and Kan are based on the risk-neutral probability measure *Q*. The *Q* measure is a constructed measure, such that the fair value of a zero coupon bond can be obtained through discounting the future value of the bond by the risk free rate under the *Q* measure, or

$$P(t, T) = E_t^Q\left[e^{-\int_t^{t+\tau} r(u)du}\right] \tag{1.53}$$

This means, the uncertainty of the future yield changes can be removed under the *Q* measure. Since arbitrage free pricing has been built on the *Q*, the measure *Q* is often referred as the risk-neutral measure. On the contrary, the probability measure in the real world is named as the physical measure *P*. Under the *P*, we cannot ignore the interest rate risk. So the question is how to find a bridge connecting the risk-neutral world and the real world. In the literature, this connection is called the price of uncertainty, often referred to as the market price of risk. In the multifactor models, the market price of risk is a function of $\boldsymbol{X_t}$, denoted as $\boldsymbol{\Lambda}(\boldsymbol{X_t})$.

With the market price of risk connecting *Q* and *P*, the dynamics of factors under two measures with respect to the $\boldsymbol{X_t}$ have the forms:

$$\begin{aligned} \boldsymbol{dX_t} &= \boldsymbol{\mu^Q}(\boldsymbol{X_t})\boldsymbol{dt} + \boldsymbol{\sigma}(\boldsymbol{X_t})\boldsymbol{dW_t^Q} \\ &= \boldsymbol{\mu^P}(\boldsymbol{X_t})\boldsymbol{dt} + \boldsymbol{\sigma}(\boldsymbol{X_t})\boldsymbol{dW_t^P} \end{aligned} \tag{1.54}$$

where $\boldsymbol{W_t^P}$ is a *N*-dimensional Wiener process under the *P* and $\boldsymbol{W_t^Q}$ is a *N*-dimensional Wiener process under the *Q*.

In terms of diffusion processes, Girsanov's theorem provides us the machinery to write

$$\boldsymbol{dW_t^Q} = \boldsymbol{dW_t^P} + \boldsymbol{\Lambda}(\boldsymbol{X_t})\boldsymbol{dt} \tag{1.55}$$

for any process $\boldsymbol{\Lambda}(\boldsymbol{X_t}) \in \mathbb{R}^{N\times N}$.

In general,

$$\boldsymbol{\mu^P}(\boldsymbol{X_t}) = \boldsymbol{\mu^Q}(\boldsymbol{X_t}) + \boldsymbol{\sigma}(\boldsymbol{X_t})\boldsymbol{\Lambda}(\boldsymbol{X_t}) \tag{1.56}$$

Given the affine forms of the drifts and the conditional variances of $\boldsymbol{X_t}$, which are

$$\boldsymbol{\mu}^{\boldsymbol{Q}}(\boldsymbol{X}_t) = \boldsymbol{\kappa}^{\boldsymbol{Q}}(\boldsymbol{\theta}^{\boldsymbol{Q}} - \boldsymbol{X}_t) \tag{1.57}$$

$$\boldsymbol{\sigma}(\boldsymbol{X}_t) = \boldsymbol{\Sigma}\sqrt{\boldsymbol{S}(\boldsymbol{X}_t)} \tag{1.58}$$

The dynamics of $\boldsymbol{X}_t$ under the risk-neutral measure Q is

$$\boldsymbol{dX}_t = \boldsymbol{\kappa}^{\boldsymbol{Q}}(\boldsymbol{\theta}^{\boldsymbol{Q}} - \boldsymbol{X}_t)\, \boldsymbol{dt} + \boldsymbol{\Sigma}\sqrt{\boldsymbol{S}(\boldsymbol{X}_t)}\boldsymbol{dW}_t^{\boldsymbol{Q}} \tag{1.59}$$

Once specifying a structure of the market price of risk $\boldsymbol{\Lambda}(\boldsymbol{X}_t)$**,** we can derive the form of $\boldsymbol{X}_t$ under P, which is

$$\boldsymbol{dX}_t = \boldsymbol{\kappa}^{\boldsymbol{Q}}(\boldsymbol{\theta}^{\boldsymbol{Q}} - \boldsymbol{X}_t)\boldsymbol{dt} + \boldsymbol{\Sigma}\sqrt{\boldsymbol{S}(\boldsymbol{X}_t)}\boldsymbol{\Lambda}(\boldsymbol{X}_t)\boldsymbol{dt} + \boldsymbol{\Sigma}\sqrt{\boldsymbol{S}(\boldsymbol{X}_t)}\boldsymbol{dW}_t^{\boldsymbol{P}} \tag{1.60}$$

Eqs. (1.59) and (1.60) are equivalent but under different probability measures.

There are different ways to model the market price of risk. In the recent literature there are three main classes: *completely affine models, essentially affine models*, and *semi-affine models*, defining three different structures of the market prices of risk. These classes are all affine under the risk-neutral measure.

1.4.1 Completely affine models

Dai and Singleton (2000) assume that $\boldsymbol{\Lambda}(\boldsymbol{X}_t)$ is proportional to the conditional standard deviation of $\boldsymbol{X}_t$, which is given by

$$\boldsymbol{\Lambda}(\boldsymbol{X}_t) = \sqrt{\boldsymbol{S}(\boldsymbol{X}_t)}\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_1 \in \mathbb{R}^{N\times 1} \tag{1.61}$$

Then $\sqrt{\boldsymbol{S}(\boldsymbol{X}_t)}\,\boldsymbol{\Lambda} = \boldsymbol{S}(\boldsymbol{X}_t)\boldsymbol{\lambda}_1$, and it is affine in $\boldsymbol{X}_t$.

In completely affine models, the drift and diffusion processes under both the P and Q measures are affine. So this class of models is widely used and nests many other models, e.g., multifactor versions of the Vasicek and CIR models.

The specification of $\boldsymbol{\Lambda}(\boldsymbol{X_t})$ imposes two limitations. First, the variation in $\boldsymbol{\Lambda}(\boldsymbol{X_t})$ is entirely determined by the variation in $\boldsymbol{S}(\boldsymbol{X_t})$. So the temporal variation in expected excess returns on bonds is driven by the volatility of the factors through $\boldsymbol{S}(\boldsymbol{X_t})$. Second, the sign of each element of $\boldsymbol{\Lambda}(\boldsymbol{X_t})$ is fixed and the same as that of the element of $\boldsymbol{\lambda_1}$.

1.4.2 Essentially affine models

Duffee (2002) describes a broad class of models - Essentially affine models, which allows that the bond yields vary independently of the volatility of $\boldsymbol{X_t}$. Thus, the sign of any element of the market price of risk vector can be changed. In particular, the $\boldsymbol{\Lambda}(\boldsymbol{X_t})$ is defined as

$$\boldsymbol{\Lambda}(\boldsymbol{X_t}) = \sqrt{\boldsymbol{S}(\boldsymbol{X_t})}\boldsymbol{\lambda_1} + \sqrt{\boldsymbol{S}^-(\boldsymbol{X_t})}\boldsymbol{\lambda_2}\boldsymbol{X_t} \tag{1.62}$$

Or

$$\boldsymbol{\Lambda}(\boldsymbol{X_t}) = \sqrt{\boldsymbol{S}(\boldsymbol{X_t})}(\boldsymbol{\lambda_1} + \boldsymbol{S}(\boldsymbol{X_t})^{-1}\boldsymbol{I}^-\boldsymbol{\lambda_2}\boldsymbol{X_t}) \tag{1.63}$$

where $\boldsymbol{\lambda_1} \in \mathbb{R}^{N\times 1}$, $\boldsymbol{\lambda_2} \in \mathbb{R}^{N\times N}$, and $\sqrt{\boldsymbol{S}^-(\boldsymbol{X_t})}$ is a diagonal $N \times N$ matrix with the i-th diagonal element given by

$$[\sqrt{\boldsymbol{S}^-(\boldsymbol{X_t})}]_{ii} = \begin{cases} (\alpha_i + \boldsymbol{\beta}_i^\intercal \boldsymbol{X_t})^{-1/2}, & \text{if } \inf(\alpha_i + \boldsymbol{\beta}_i^\intercal \boldsymbol{X_t}) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{1.64}$$

and the diagonal matrix $\boldsymbol{I}^-$ with the i-th element given by

$$\boldsymbol{I}_{ii}^- = \begin{cases} 1 & \text{if } \inf(\alpha_i + \boldsymbol{\beta}_i^\intercal \boldsymbol{X_t}) > 0; \\ 0 & \text{otherwise.} \end{cases} \tag{1.65}$$

In the essential affine models, $\sqrt{\boldsymbol{S}(\boldsymbol{X_t})}\boldsymbol{\Lambda}(\boldsymbol{X_t}) = \boldsymbol{S}(\boldsymbol{X_t})\boldsymbol{\lambda_1} + \boldsymbol{I}^-\boldsymbol{\lambda_2}\boldsymbol{X_t}$ is affine in $\boldsymbol{X_t}$. So the dynamics of $\boldsymbol{X_t}$ is affine both under the *Q* and the *P*. So the dynamics of $\boldsymbol{X_t}$ under *P* is

$$\boldsymbol{dX_t} = \boldsymbol{\kappa^Q(\theta^Q - X_t)dt + \Sigma\,(S(X_t)\lambda_1 + I^- \lambda_2 X_t)dt + \Sigma\sqrt{S(X_t)}dW_t^P} \tag{1.66}$$

1.4.3 Semi-affine models

Duarte (2004) proposes the semi-affine models. The specification of the market prices of risk has the form

$$\boldsymbol{\Lambda(X_t) = \Sigma^{-1}\lambda_0 + \sqrt{S(X_t)}\lambda_1 + \sqrt{S^-(X_t)}\lambda_2 X_t} \tag{1.67}$$

where $\boldsymbol{\lambda_0}$, $\boldsymbol{\lambda_1} \in \mathbb{R}^{N\times 1}$ and $\boldsymbol{\lambda_2} \in \mathbb{R}^{N\times N}$. When the vector $\boldsymbol{\lambda_0}$ and the $\boldsymbol{\lambda_2}$ matrix are null, $\boldsymbol{\Lambda(X_t)}$ has the form of completely affine models. Essentially affine models are affine when the vector $\boldsymbol{\lambda_0}$ is null.

Comparing with the essentially affine models, the additional term, $\boldsymbol{\Sigma^{-1}\lambda_0}$, in semi-affine models offers additional sign changing flexibility without limiting the volatility dynamics. Unfortunately, the introduction of $\boldsymbol{\Sigma^{-1}\lambda_0}$ makes the drift process of the state variables under the *P* is not the affine function of $\boldsymbol{X_t}$. This is because

$$\begin{aligned}
\boldsymbol{\mu^P(X_t)} &= \boldsymbol{\kappa^Q(\theta^Q - X_t) + \Sigma\sqrt{S(X_t)}\Lambda(X_t)} \\
&= \boldsymbol{\kappa^Q(\theta^Q - X_t) + \Sigma\sqrt{S(X_t)}(\Sigma^{-1}\lambda_0 + \sqrt{S(X_t)}\lambda_1 + \sqrt{S^-(X_t)}\lambda_2 X_t)} \\
&= \boldsymbol{\kappa^Q(\theta^Q - X_t) + \Sigma\sqrt{S(X_t)}\Sigma^{-1}\lambda_0 + \Sigma\,(S(X_t)\lambda_1 + I^-\lambda_2 X_t)}
\end{aligned} \tag{1.68}$$

Obviously, the term $\boldsymbol{\Sigma\sqrt{S(X_t)}\Sigma^{-1}\lambda_0}$ is not affine in $\boldsymbol{X_t}$, and for this reason the model is called “semi-affine”.

1.5 Classification of ATSMs

Based on the discussion of Section 1.4, we can see the diffusion term is affine in both the risk-neutral measure and the physical measure if the models are either completely affine or essentially affine. Dai and Singleton (2000) analyze the classification of ATSMs assuming the models are essentially affine.

Dai and Singleton (2000) classify ATSMs based on the number of state factors and make sure that all models in each class are "admissible". The model is "admissible", meaning a specification of an affine model can well define bond yields. In particular, the model is admissible if the conditional variances of factors (denoted as $[\boldsymbol{S}(\boldsymbol{X_t})]_{ii}$) are strictly positive. In each "admissible" class, the paper proves that there exists a "maximal" model that can nest econometrically all other models within the same class. A maximal model is a model that is sufficiently flexible to describe the cross-section movements in short- and long-term bond yields.

Define $\boldsymbol{X_t}$ is a N-dimensional factor vector[4], i.e. $\boldsymbol{X_t} = (X_1, X_2, \dots, X_N)^{\intercal}$. A general specification of a multi-factor affine model under P is

$$\boldsymbol{dX_t} = \boldsymbol{\kappa^P}(\boldsymbol{\theta^P} - \boldsymbol{X_t})\,\boldsymbol{dt} + \boldsymbol{\Sigma}\sqrt{\boldsymbol{S}(\boldsymbol{X_t})}\boldsymbol{dW_t^P} \tag{1.69}$$

and thus, the instantaneous short rate is defined as

$$r_t = \delta_0 + \boldsymbol{\delta_x^{\intercal}X_t} \tag{1.70}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{\boldsymbol{N\times 1}}$and $\boldsymbol{\kappa}, \boldsymbol{\Sigma} \in \mathbb{R}^{\boldsymbol{N\times N}}$. $\boldsymbol{S}(\boldsymbol{X_t})$ is a diagonal $N \times N$ matrix with the element given by

$$[\boldsymbol{S}(\boldsymbol{X_t})]_{ij} = \begin{cases} \alpha_i + \boldsymbol{\beta_i^{\intercal}X_t}, & i = j \\ 0, & i \neq j \end{cases} \tag{1.71}$$

where $\alpha_i \in \mathbb{R}$ and $\boldsymbol{\beta_i} \in \mathbb{R}^{\boldsymbol{N\times 1}}$.

---

[4] To distinguish a scale variable and a vector variable, vector variables are in bold.

Expand equation (1.69) in a matrix form,

$$\begin{pmatrix} dX_1 \\ dX_2 \\ \vdots \\ dX_N \end{pmatrix} = \begin{pmatrix} \kappa_{11} & \cdots & \kappa_{1N} \\ \vdots & \ddots & \vdots \\ \kappa_{N1} & \cdots & \kappa_{NN} \end{pmatrix} \begin{pmatrix} (\theta_1 - X_1)dt \\ (\theta_2 - X_2)dt \\ \vdots \\ (\theta_N - X_N)dt \end{pmatrix}$$

$$+ \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{pmatrix} \begin{pmatrix} \sqrt{\alpha_1 + \sum_{i=1}^{N} \beta_{i1} X_i} dW_1 \\ \sqrt{\alpha_2 + \sum_{i=1}^{N} \beta_{i2} X_i} dW_2 \\ \vdots \\ \sqrt{\alpha_N + \sum_{i=1}^{N} \beta_{iN} X_i} dW_N \end{pmatrix} \tag{1.72}$$

The general form of Eq. (1.69) is complex and difficult to do a specific analysis. So Dai and Singleton (2000) formalize the family of the admissible *N*-factor ATSMs, denoted as $A_m(N)$. $m = rank(\mathcal{B})$, where $\mathcal{B} = (\beta_1, \beta_2, \dots, \beta_N)$, represents the degrees of the dependence of the conditional variances on the number of factors. Based on the value of *m*, the family of the admissible *N*-factor ATSMs can be classified into (*N*+1) subfamilies. Those (*N*+1) admissible ATSMs subfamilies have notations: $A_0(N), A_1(N), \dots, A_m(N), \dots, A_N(N)$.

In a matrix form, the *Canonical Representations* of $A_m(N)$ can be written as

$$
\begin{pmatrix} dX_1 \\ dX_2 \\ \vdots \\ dX_m \\ \hline dX_{m+1} \\ dX_{m+2} \\ \vdots \\ dX_N \end{pmatrix} = \left(\begin{array}{ccc|ccc} \kappa_{11} & \cdots & \kappa_{1m} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \kappa_{m1} & \cdots & \kappa_{mm} & 0 & \cdots & 0 \\ \hline \kappa_{m+1,1} & \cdots & \kappa_{m+1,m} & \kappa_{m+1,m+1} & \cdots & \kappa_{m+1,N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \kappa_{N1} & \cdots & \kappa_{Nm} & \kappa_{N,m+1} & \cdots & \kappa_{NN} \end{array}\right) \begin{pmatrix} (\theta_1 - X_1)dt \\ (\theta_2 - X_2)dt \\ \vdots \\ (\theta_m - X_m)dt \\ \hline -X_{m+1}dt \\ -X_{m+2}dt \\ \vdots \\ -X_N dt \end{pmatrix}
$$

$$
+ \begin{pmatrix} \sqrt{X_1}dW_1 \\ \sqrt{X_2}dW_2 \\ \vdots \\ \sqrt{X_m}dW_m \\ \hline \sqrt{1+\sum_{i=1}^m \beta_{i,m+1}X_i}dW_{m+1} \\ \sqrt{1+\sum_{i=1}^m \beta_{i,m+2}X_i}dW_{m+2} \\ \vdots \\ \sqrt{1+\sum_{i=1}^m \beta_{i,N}X_i}dW_N \end{pmatrix} \tag{1.73}
$$

with the restrictions:

(1) $\delta_i \geq 0, \qquad m + 1 \leq i \leq n$

(2) $\sum_{j=1}^m \kappa_{ij}\theta_j > 0, \qquad 1 \leq i \leq m$

(3) $\kappa_{ij} \leq 0, \qquad 1 \leq i \leq m, j \neq i$

(4) $\theta_i \geq 0, \qquad 1 \leq i \leq m$

(5) $\mathcal{B}_{ij} \geq 0, \qquad 1 \leq i \leq m, m + 1 \leq j \leq n$

The sign restriction of $\mathcal{B}_{ij}$ in (5) assures that the conditional variances of state factors are positive. In order to have $(X_1, X_2, \ldots, X_m)^{\intercal}$ strictly positive, restrictions (2), (3) and (4) are imposed. The form of Eq. (1.73) plus five restrictions (1 - 5) make sure that any ATSM satisfying the canonical representation of $A_m(N)$ is admissible. In addition, the canonical representation is also maximal.

If m = 0 or an $A_0(N)$ subfamily, there is no factor affecting the volatility of the state factors at all. In this case, the state factors are homoskedastic and follow Gaussian diffusion processes. A typical example of such a case is the Vasicek model when $N = 1$. The state factors in the $A_0(N)$ models are perfectly correlated either conditionally or unconditionally at the cost of constant conditional variances ($m = 0$). Another extreme case is when $m = N$ so that all state factors determine the volatility structures. The CIR model, $A_1(1)$, is a one-factor example in this subfamily. Since all state factors contribute to the conditional volatility structures in the $A_N(N)$ models, the dynamics of yield curves have maximal flexibility at the cost of the null conditional correlations of the state factors and the non-negative unconditional correlations. Between the $A_0(N)$ and $A_N(N)$ lie the (N-1) subfamilies of ATSMs with time-varying conditional volatilities of the state factors ($1 \leq m \leq N - 1$) and admissible non-zero conditional correlations of state factors. Therefore, there is an important trade-off within the family of ATSMs between the flexibility in specifying conditional volatilities of state factors and the flexibility in modeling conditional correlations of state factors.

Taking the 3-factor ATSMs as an example, there are four subfamilies: $A_0(3), A_1(3), A_2(3), A_3(3)$. The Vasicek, BDFS, Chen and CIR models are one example of those four subfamilies, respectively. Comparing four extant models with their maximal counterparts in their own subfamilies, Dai and Singleton (2000) find that except for the Vasicek model, the other three all impose overidentifying restrictions.

1.6 References

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

Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.

Duarte, Jefferson, 2004, Evaluating an alternative risk preference in affine term structure models, *Review of Financial Studies* 17, 379-404.

Duffee, G.R., 2002, Term permia and interest rate forecasts in affine models, *Journal of Finance* 57, 405-443.

Duffie, D. and R. Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379-406.

Vasicek, Oldrich A., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

CHAPTER II

THE VALUE OF KNOWING THE TRUTH AND THE COST OF BEING WRONG: AN APPLICATION TO ONE-FACTOR AFFINE TERM STRUCTURE MODELS

2.1 Introduction

Modeling the evolvement of yield curves over time is an important issue in the finance literature because the term structure of interest rates can help people determine the prices of interest rate contingent claims such as interest rate options, callable bonds, and interest rate futures. Affine term structure models (ATSMs) are one set of popular models for yield curve modeling. These models assume that the dynamics of the term structure of interest rates are a linear function of a set of observed and/or latent factors (Duffie and Kan, 1996). In this chapter, I focus on two well-defined one-factor ATSMs: Vasicek model (Vasicek, 1977) and Cox, Ingersoll and Ross (CIR hereafter) model (Cox, Ingersoll, and Ross, 1985).

To test and/or compare different models, scholars generally calibrate models using historical treasury yields and then check how far the predicted values are away from the real data. While much of the prior literature focuses on estimating and evaluating these cross-sectional pricing errors for different ATSMs, little is known about the data's power to distinguish one ATSM from another[5]. Given that the models forecast yields based on the speed of mean reversion, under what circumstances can we distinguish one ATSM from another? The objective of this chapter is to quantify the benefit of knowing the "true" model as well as the cost of being wrong when choosing between ATSMs. In particular, I detail the power of out-of-sample forecasts to statistically distinguish one ATSM from another given that we only know the data are generated from an ATSM and are observed without errors.

To explore the question of interest, I use a parsimonious Monte Carlo study to simulate six-month zero-coupon data from both the Vasicek and CIR models and then, in turn, estimate both the CIR and Vasicek models using that simulated data. In addition, I add a non-linear diffusion model (Aït-Sahalia, 1996), a first-order autoregressive model (AR(1)), and a Martingale model as three alternative estimation approaches. Then, I construct out-of-sample forecasts for all estimation models over 1-, 6-, and 12-month horizons. The model's forecasting accuracy is measured by the mean squared error (MSE). Last, I use the Giacomini and White (2006) test to examine the models' relative predictive ability as well as the model's power to statistically distinguish between different ATSMs. The test statistic is the standardized difference of two models' MSEs. If the test statistic is different from zero at the 5% significance level, we say that the two competing models can be statistically differentiated. The larger the absolute value of the

[5] In order for a K-factor model to fit the M bonds ($K > M$), an auxiliary error model is introduced. Most studies evaluate the ATSM based on the size of this error, though Duffee (2002) and Cheridito et al (2005) evaluate the models' (estimated with the auxiliary error) ability to forecast out-of-sample.

test statistics, the stronger the power to distinguish one model from another, and the higher the relative predictive ability of the model with the smaller MSE.

In this chapter, I use the maximum likelihood approach to estimate the ATSMs. Because of the high persistence of finite bond samples (Bauer, Rudebusch and Wu, 2012), bias in the parameter estimates, particularly the mean reversion estimate, can be exhibited by all estimation approaches, including the maximum likelihood estimation (Tang and Chen, 2009). When the mean reversion is near the unit root, the problem becomes even more serious. Duffee and Stanton (2012) investigate the maximum likelihood estimation under the circumstances that the ATSMs include complex forms of the price of risk. They find that the maximum likelihood estimation produces a strong bias in the mean reversion estimates when the price of risk has a flexible specification. That is, the maximum likelihood estimators based on the highly persistent bond samples differ largely from their true values, and then the corresponding bond yield forecasts have little precision. To increase the forecast precision, I need a technique to correct the estimator biases. The literature proposes several methods (Andrews, 1993; Smith, 1993; Phillips and Yu, 2009; Tang and Chen, 2009). In this chapter, I follow the bootstrap bias correction process proposed by Tang and Chen (2009). My study shows that the bootstrap process can largely reduce the parameter biases, particularly the bias of the mean reversion estimate.

Considering the historical bond data are highly persistent, I simulate the data from the ATSMs whose mean reversion parameters are near the unit root with an autoregressive coefficient ρ of 0.99. The autoregressive coefficient ρ is defined as $\rho = e^{-\kappa \Delta t}$, where κ is the mean reversion parameter and Δt is the time interval in years.

Comparing the affine term structure models' forecasting performance based on the estimates of the Vasicek model and CIR models, both before and after running the bootstrap bias correction processes, I find the model's relative predictive ability is closely related to the size of mean reversion bias no matter what the data-generating model is. The smaller the bias in the estimate of the mean reversion, the better out-of-sample forecasts the model produces. The probability that the true models are distinguished from the non-data-generating ATSMs, or the "wrong" models, is over 80% at the 1-month ahead horizon, over 65% at the 6-month ahead horizon, and over 55% at the 12-month ahead horizon. The reason that the power to differentiate between models is decreasing as the increase of the forecast horizons is that ATSMs are the mean-reversion processes. That is, the models will approach their long-run means as long as the time horizon is. In addition, the ATSMs generally have the higher predictive ability than three alternative models. The martingale model has the least forecasting accuracy over all models.

Since the ATSMs' predictive ability is related to the mean reversion parameters, a variety of experiments are conducted to explore how the predictive ability varies with the change of the ATSMs' mean reversion parameters. The first experiment (Setting A) is designed to push the mean reversion parameter far away from the unit root. In particular, I simulate the data with the new model whose half life of a shock to the interest rate is 1/6 of that of the old model whose mean reversion parameter is near the unit root. The half life of a shock is defined as the time horizon at which the effect of the shock is one-half. In the Vasicek and CIR models, the half life is calculated as $-\ln 0.5/\kappa$, which gives us important information about how quickly the instantaneous short rates converge to their long-run mean after a shock. The new model refers to a high speed mean reversion process and the autoregressive coefficient of the new model is 0.93. Then I repeat the same estimation, forecasting and testing processes as the above. My finding shows that the new models' predictive ability is generally higher than the old models'. However, the new model's power to distinguish between different ATSMs is largely reduced since the probabilities to differentiate the true models from the wrong models are below 30%, below 25% and below 20% at the 1-, 6- and 12-month ahead horizon, respectively. The possible explanation is that the faster mean reversion means that the model forecasts are simply equal to the long-run mean.

Tang and Chen (2009) conclude that the bias of the mean reversion estimator is not a function of the number of observations but a decreasing function of the time span. This implies that the number of observations and the time span in an estimation window may influence the estimator of the mean reversion parameter. Therefore, I design two experiments to investigate the sensitivity of the model's predictive ability to the sample size and to the time span of the sample, respectively.

In my second experiment (Setting B), I change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span and other parameters the same as those of the initial model's. Accordingly, the sample size rises over a fixed time span. The latter corresponds to the high frequency data. The results show the increase of the sample size by the way of increasing the sampling frequency has little influence on the size of the mean reversion bias. However, the larger sample size improves the model's predictive ability. In contrast, the probability of the model's power to distinguish between different ATSMs drops by about 50%.

My third experiment (Setting C) studies the impact of the sample's time span on the model's predictive ability. I change the sampling frequency of the data from a monthly basis to a daily basis but keep the sample size and other parameters unchanged. The new model corresponds to a shorter time span but maintains the same sample size in comparison with the initial model. My findings are consistent with Tang and Chen. That is, the bias of the mean reversion estimate of the new model is larger than the corresponding value of the initial model. The shorter time span with the same sample size slightly increases the models' predictive ability. But the power to differentiate the true model from the wrong models is reduced by more than 10%. In addition, the comparison between the results from Setting B and Setting C shows that the larger sample size with the same sampling frequency improves the model's predictive ability but decreases the model's power to distinguish between different models.

In summary, I find that the a mean reversion process with a higher speed, the larger sample size, and the higher sampling frequency can improve ATSMs' predictive ability in contrast with the decrease of the models' power to differentiate between different ATSMs.

The rest of this chapter is organized as follows: Section 2.2 describes the parameter estimation methods and the Giacomini and White test. The detail data simulation procedure is described in Section 2.3. Section 2.4 examines the ATSMs' out-of-sample performance when the models are near the unit root. In addition, Section 2.4 discusses the role of the mean reversion parameter, the sample size, and the sample's time span on the models’ predictive ability. Section 2.5 concludes the paper and discusses the possible extensions to this research.

2.2 Model estimation and inference

This section introduces the estimation models for the affine term structure models. Maximum likelihood is my first choice because of its asymptotic efficiency. In addition, I can invert the pricing equations to infer the state factors and express the likelihood function as the equations of the bond prices. This is the approach used by Pearson and Sun (1994). I also outline three alternative forecast techniques which I compare to the ATSMs: Non-Linear diffusion model (ND), First-order autoregressive model (AR(1)), and Martingale model. After that, this section describes the bootstrap bias correction process of Tang and Chen (2009) because the maximum likelihood estimation exhibits a bias in the mean reversion estimator. Lastly, I introduce the Giacomini and White test to examine and compare the models' predictive ability.

2.2.1 Maximum likelihood estimation of Vasicek model (MLE-Vasicek)

This section states the process of zero yields estimation using the maximum likelihood function of the Vasicek model. The Vasicek model is a one-factor ATSM and derived by Vasicek in 1977. The model has the expression:

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t \tag{2.1}$$

where r_t is the instantaneous short rate, θ is the long-run mean, κ is the speed of mean reversion, and W_t is a Wiener process. One advantage of the Vasicek model is that its transition density is explicitly expressed, which is a Gaussian model. So we can form a log-likelihood function of the instantaneous short rate and apply the MLE method. Usually the instantaneous short rates are not observable in the market, while zero bond yields at different maturities are observed. Therefore, we have to find a way to derive a log-likelihood function of zero bond prices. In this section I show how to transform the instantaneous short rate's log-likelihood function to a log-likelihood function of observable zero-coupon bond prices.

The probability density function for r_t conditional on r_{t-1}, with a Δt time step between them, is given by

$$h_V(r_t|r_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t{}^2}} e^{-\frac{[r_t-(\theta+(r_{t-1}-\theta)e^{-\kappa\Delta t})]^2}{2\sigma_t{}^2}} \tag{2.2}$$

with
$$\sigma_t{}^2 = \sigma_t{}^2(r_t|r_{t-1}) = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa\Delta t}\right) \tag{2.3}$$

Thus, given the initial short rate r_1, the log-likelihood function of a set of (n - 1) observations on the short rate, $\{r_2, \cdots r_{n-1}, r_n\}$, is given by

$$\mathcal{L}(\kappa, \theta, \sigma_t) = \ln \prod_{i=2}^{n} h_V(r_i|r_{i-1}) = \sum_{i=2}^{n} \ln h_V(r_i|r_{i-1})$$

$$= -\frac{n-1}{2}\ln(2\pi) - (n-1)\ln(\sigma_t)$$

$$-\frac{1}{2\sigma_t{}^2}\sum_{i=2}^{n}[(r_i-\theta)-(r_{i-1}-\theta)e^{-\kappa\Delta t}]^2 \tag{2.4}$$

Define $\alpha = e^{-\kappa\Delta t}$, then Equation (2.4) can be written as

$$\mathcal{L}(\alpha, \theta, \sigma_t) = -\frac{n-1}{2}\ln(2\pi) - (n-1)\ln(\sigma_t)$$

$$-\frac{1}{2\sigma_t{}^2}\sum_{i=2}^{n}[(r_i-\theta)-(r_{i-1}-\theta)\alpha]^2 \tag{2.5}$$

Eq. (2.5) specifies the log-likelihood function of the short rate. So we can estimate three parameters, θ , κ, and σ, based on iterating log-likelihood maximization process, which are the numerical parameter estimates. Also, we can use the differential calculus to derive the analytical maximum likelihood estimates of parameters θ , κ, and σ. Brigo and Mercurio (1974, PP62) provide the calculations and equations:

$$\hat{\alpha} = \frac{(n-1)\sum_{i=2}^{n} r_i r_{i-1} - \sum_{i=2}^{n} r_i \sum_{i=2}^{n} r_{i-1}}{(n-1)\sum_{i=2}^{n} r_{i-1}^2 - \left(\sum_{i=2}^{n} r_{i-1}\right)^2} \tag{2.6}$$

$$\hat{\theta} = \frac{\sum_{i=2}^{n}[r_i - \hat{\alpha} r_{i-1}]}{(n-1)(1-\hat{\alpha})} \tag{2.7}$$

$$\widehat{\sigma_t^2} = \frac{1}{n-1}\sum_{i=2}^{n}[r_i - \hat{\alpha}\, r_{i-1} - \hat{\theta}(1-\hat{\alpha})]^2 \tag{2.8}$$

Rearranging the above equations gives

$$\hat{\theta} = \frac{\sum_{i=2}^{n} r_i \sum_{i=2}^{n} r_{i-1}^2 - \sum_{i=2}^{n} r_{i-1} \sum_{i=2}^{n} r_i r_{i-1}}{(n-1)\left(\sum_{i=2}^{n} r_{i-1}^2 - \sum_{i=2}^{n} r_i r_{i-1}\right) - \left(\left(\sum_{i=2}^{n} r_{i-1}\right)^2 - \sum_{i=2}^{n} r_i \sum_{i=2}^{n} r_{i-1}\right)} \tag{2.9}$$

$$\hat{\kappa} = -\frac{1}{\Delta t}\ln\left[\frac{\sum_{i=2}^{n} r_i r_{i-1} - \hat{\theta}\sum_{i=2}^{n} r_{i-1} - \hat{\theta}\sum_{i=2}^{n} r_i + (n-1)\hat{\theta}^2}{\sum_{i=2}^{n} r_{i-1}^2 - 2\hat{\theta}\sum_{i=2}^{n} r_{i-1} + (n-1)\hat{\theta}^2}\right] \tag{2.10}$$

$$\hat{\alpha} = e^{-\hat{\kappa}\Delta t} \tag{2.11}$$

$$\widehat{\sigma_t^2} = \frac{1}{n-1}\Big(\sum_{i=2}^{n} r_i^2 - 2\hat{\alpha}\sum_{i=2}^{n} r_i r_{i-1} + \hat{\alpha}^2 \sum_{i=2}^{n} r_{i-1}^2$$

$$-2\hat{\theta}(1-\hat{\alpha})\left(\sum_{i=2}^{n} r_i - \hat{\alpha}\sum_{i=2}^{n} r_{i-1}\right) + (n-1)\left[\hat{\theta}(1-\hat{\alpha})\right]^2\Big) \tag{2.12}$$

$$\widehat{\sigma^2} = \widehat{\sigma_t^2}\,\frac{2\hat{\kappa}}{1-\hat{\alpha}^2} \tag{2.13}$$

The above process is based on the transition density function of the instantaneous short rates. However, the usual case is that the instantaneous short rates are not observable in the market and the available data are bond prices. So it is necessary to find a way to derive the log-likelihood function of the zero bond prices. Pearson and Sun (2004) provide a process to derive the transition density function of the zero bond prices by re-writing the factors as a function of the bond prices. I follow the method of Pearson and Sun (2004) in this paper. Equation (2.14) enables us to infer the instantaneous short rates from the zero coupon prices.

$$P_t(\tau) = e^{A(\tau)-B(\tau)r_t} \tag{2.14}$$

where τ is the time to maturity[6] and:

$$A(\tau) = \left(\theta - \frac{\sigma^2}{2\kappa^2}\right)(B(\tau)-\tau) - \frac{\sigma^2 B^2(\tau)}{4\kappa} \tag{2.15}$$

$$B(\tau) = \frac{1}{\kappa}(1 - e^{-\kappa\tau}) \tag{2.16}$$

Fixing τ and simplifying $P_t(\tau)$ as P_t for convenience, the factor r_t can be written as:

$$r_t = \frac{A(\tau) - lnP_t}{B(\tau)} \tag{2.17}$$

Taking the derivatives of r_t with respect to P_t, we can get the Jacobian of the transformation, which is

$$J = \frac{dr_t}{dP_t} = -\frac{1}{P_t B(\tau)} \tag{2.18}$$

By dividing the time interval [0, *T*] into *n* subintervals of equal width and letting $t_0 = 0$ and $t_n = T$, we have *n* time knots $t_1, t_2, \cdots t_{n-1}, t_n$, and $\Delta t = t_i - t_{i-1},\ i = 1,2,\dots,n$. Assume the zero bond prices at each time knot are $P_1, P_2, \cdots, P_n$. At each time knot t_i, the *Jacobian* can be written as

$$J = \frac{dr_{t_i}}{dP_i} = -\frac{1}{P_i B(\tau)} \tag{2.19}$$

---

[6] The full definition is $A(\tau) = \left(\theta - \frac{\sigma\lambda}{\kappa} - \frac{\sigma^2}{2\kappa^2}\right)(B(\tau)-\tau) - \frac{\sigma^2 B^2(\tau)}{4\kappa}$, where λ is the market price of risk. λ is set as zero for convenience.

Therefore, the transition density function of bond prices is obtained by multiplying the transition density of the short rates with the *Jacobian*.

$$h_V(P_i|P_{i-1}) = |J|h_V\left(r_{t_i}\middle|r_{t_{i-1}}\right) = \left|-\frac{1}{P_iB(\tau)}\right|\frac{1}{\sqrt{2\pi{\sigma_t}^2}}e^{-\frac{\left(\frac{A(\tau)-lnP_i}{B(\tau)}-\mu\right)^2}{2\sigma_t^2}}$$

$$= \frac{1}{\sqrt{2\pi{\sigma_t}^2}P_i|B(\tau)|}e^{-\frac{(\frac{A(\tau)-lnP_i}{B(\tau)}-\mu)^2}{2\sigma_t^2}} \tag{2.20}$$

where $\mu = \theta + \left(\frac{A(\tau)-lnP_{i-1}}{B(\tau)} - \theta\right)e^{-\kappa\Delta t}$ and $\sigma_t^2 = \frac{\sigma^2}{2\kappa}\,(1 - e^{-2\kappa\Delta t})$.

Taking the natural logarithm of Eq. (2.20) gets

$$\ln h_V(P_i|P_{i-1}) = \ln(\frac{1}{\sqrt{2\pi{\sigma_t}^2}P_i|B(\tau)|}) - \frac{\left(\frac{A(\tau)-lnP_i}{B(\tau)}-\mu\right)^2}{2\sigma_t^2} \tag{2.21}$$

Replace μ in Eq. (2.21) with the full expression and get,

$$\ln h_V(P_i|P_{i-1}) = \ln(\frac{1}{\sqrt{2\pi\sigma_t^2}P_i|B(\tau)|})$$

$$-\frac{\left(\frac{A(\tau)-lnP_i}{B(\tau)}-\theta-\left(\frac{A(\tau)-lnP_{i-1}}{B(\tau)}-\theta\right)e^{-\kappa\Delta t}\right)^2}{2\sigma_t^2} \tag{2.22}$$

The log-likelihood function of zero coupon bond prices, $P_1, P_2, \cdots, P_n$, is thus

$$\mathcal{L}(\kappa,\theta,\sigma) = \ln h_V(P_n, P_{n-1}, \cdots, P_2|P_1)$$

$$= \ln\prod_{i=2}^{n} h_V(P_i|P_{i-1})$$

$$= \sum_{i=2}^{n} \ln h_V(P_i|P_{i-1}) \tag{2.23}$$

Substituting Eq. (2.22) into Eq. (2.23) gets

$$\mathcal{L}(\kappa,\theta,\sigma) = \sum_{i=2}^{n} \ln(\frac{1}{\sqrt{2\pi\sigma_t^2}P_i|B(\tau)|})$$

$$-\sum_{i=2}^{n} \frac{\left(\frac{A(\tau) - lnP_i}{B(\tau)} - \theta - \left(\frac{A(\tau) - lnP_{i-1}}{B(\tau)} - \theta\right)e^{-\kappa\Delta t}\right)^2}{2\sigma_t^2}$$

$$= (\mathrm{n}-1)\ln(\frac{1}{\sqrt{2\pi\sigma_t^2}|B(\tau)|}) + \sum_{i=2}^{n} \ln(\frac{1}{P_i})$$

$$-\sum_{i=2}^{n} \frac{\left(\frac{A(\tau) - lnP_i}{B(\tau)} - \theta - \left(\frac{A(\tau) - lnP_{i-1}}{B(\tau)} - \theta\right)e^{-\kappa\Delta t}\right)^2}{2\sigma_t^2}$$

$$= -(\mathrm{n}-1)\ln(\sqrt{2\pi\sigma_t^2}|B(\tau)|) - \sum_{i=2}^{n} \ln(P_i)$$

$$-\frac{1}{2\sigma_t^2} \sum_{i=2}^{n} \left(\frac{A(\tau)-lnP_i}{B(\tau)} - \theta - \left(\frac{A(\tau)-lnP_{i-1}}{B(\tau)} - \theta\right)e^{-\kappa\Delta t}\right)^2 \qquad (2.24)$$

Equation (2.24) is the final equation of the log-likelihood function based on prices for the zero-coupon bond maturing at time τ. Knowing Eq. (2.24), we can iterate the MLE process and estimate the parameters of the Vasicek model: κ, θ, σ.

Once we get the parameter estimators: $\hat{\kappa}, \hat{\theta}$, and $\hat{\sigma}$, we can forecast 1-step ahead or *h*-step ahead instantaneous short rates based on

$$E(r_{t+1}|r_t) = \hat{\theta}\left(1 - e^{-\hat{\kappa}\Delta t}\right) + e^{-\hat{\kappa}\Delta t}r_t \qquad (2.25)$$

$$E(r_{t+h\Delta t}|r_t) = \hat{\theta}\left(1 - e^{-\hat{\kappa}h\Delta t}\right) + e^{-\hat{\kappa}h\Delta t}r_t \qquad (2.26)$$

Thus, the corresponding forecasts of zero-coupon bond yields can be expressed as

$$\hat{y}_{t+1}(\tau) = -\frac{\hat{A}(\tau)-\hat{B}(\tau)\hat{r}_{t+1}}{\tau} \qquad (2.27)$$

$$\hat{y}_{t+h\Delta t}(\tau) = -\frac{\hat{A}(\tau) - \hat{B}(\tau)\hat{r}_{t+h\Delta t}}{\tau} \tag{2.28}$$

where

$$\hat{A}(\tau) = \left(\hat{\theta} - \frac{\hat{\sigma}^2}{2\hat{\kappa}^2}\right)\left(\hat{B}(\tau) - \tau\right) - \frac{\hat{\sigma}^2\hat{B}^2(\tau)}{4\hat{\kappa}} \tag{2.29}$$

$$\hat{B}(\tau) = \frac{1}{\hat{\kappa}}(1 - e^{-\hat{\kappa}\tau}) \tag{2.30}$$

Fixing τ and simplifying $\hat{y}_{t+1}(\tau)$ and $\hat{y}_{t+h\Delta t}(\tau)$ as $\hat{y}_{t+1}$ and $\hat{y}_{t+h\Delta t}$, respectively, for convenience. Note that the fact that the yields are linear in the factor means that our yield forecasts, which are based on the factor forecasts, will not suffer from Jensen's inequality.

2.2.2 Maximum likelihood estimation of CIR model (MLE-CIR)

In the Vasicek model, the conditional volatility of the instantaneous short rate is constant and independent of the level of the rate. Different from the Vasicek model, the CIR model assumes that the conditional volatility of the instantaneous short rates is proportional to the square root of the rate itself. So the CIR model is a mean-reverting square root process. The dynamics of the instantaneous short rate can be expressed as:

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t \tag{2.31}$$

Conditional on the value at time *s*, the transition density function of r_t ($t > s > 0$) is

$$h_{CIR}(r_t|r_s) = ce^{-u-v}\left(\frac{v}{u}\right)^{\frac{q}{2}} I_q\left(2(uv)^{\frac{1}{2}}\right)$$

$$= 2c \cdot Noncentral\ \chi^2(2cr_t; 2q+2, 2u)$$

$$= ce^{-c(r_t+r_s e^{-\kappa(t-s)})}\left(\frac{r_t}{r_s e^{-\kappa(t-s)}}\right)^{\frac{q}{2}} I_q\left(2c\sqrt{r_t r_s e^{-\kappa(t-s)}}\right) \quad (2.32)$$

where

$$c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa(t-s)})} \quad (2.33)$$

$$u = cr_s e^{-\kappa(t-s)} \quad (2.34)$$

$$v = cr_t \quad (2.35)$$

$$q = \frac{2\kappa\theta}{\sigma^2} - 1 \quad (2.36)$$

and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q.

In this model, the transition density for r_t is a noncentral chi-square distribution with $(2q+2)$ degrees of freedom and noncentrality parameter $2u$.

Because of the same reason as that stated in the Vasicek model, I use Pearson and Sun (2004) method to derive the log-likelihood function of zero-coupon bond prices based on the transition density function of the CIR model. The following states the details.

If dividing the time interval [0, *T*] into *n* subintervals of equal width and letting $t_0 = 0$ and $t_n = T$, we have *n* time knots $t_1, t_2, \cdots t_{n-1}, t_n$, and $\Delta t = t_i - t_{i-1},\ i = 1,2,\ldots,n$. Conditional on $r_{t_{i-1}}$, the transition density distribution of r_{t_i} can be written as:

$$h_{CIR}(r_{t_i}|r_{t_{i-1}}) = 2c \cdot Noncentral\ \chi^2(2cr_{t_i}; 2q+2, 2u)$$

$$= 2c \cdot Noncentral\ \chi^2\left(2cr_{t_i}; \frac{4\kappa\theta}{\sigma^2}, 2cr_{t_{i-1}}e^{-\kappa\Delta t}\right)$$

$$= ce^{-c(r_{t_i}+r_{t_{i-1}}e^{-\kappa\Delta t})}(\frac{r_{t_i}}{r_{t_{i-1}}e^{-\kappa\Delta t}})^{\frac{q}{2}} I_q\left(2c\sqrt{r_{t_i}r_{t_{i-1}}e^{-\kappa\Delta t}}\right) \quad (2.37)$$

where

$$c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa\Delta t})} \quad (2.38)$$

$$u = cr_{t_{i-1}}e^{-\kappa\Delta t} \quad (2.39)$$

$$v = cr_{t_i} \quad (2.40)$$

$$q = \frac{2\kappa\theta}{\sigma^2} - 1 \quad (2.41)$$

and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q.

Same as with the Vasicek model, I set the market price of risk of the CIR model to zero to simplify the calculations. Denote the zero bond prices at t_1, t_2, $\cdots$ t_{n-1}, t_n, as $P_1, P_2, \cdots, P_n$. Then the price P_i of a zero bond with maturity τ at time t_i, $(i = 1,2, \dots, n)$, has the expression:

$$P_i = e^{A(\tau)-B(\tau)r_{t_i}} \quad (2.42)$$

with

$$A(\tau) = \frac{2\kappa\theta}{\sigma^2} ln[\frac{2\gamma e^{\frac{\tau}{2}(\kappa+\gamma)}}{(\kappa+\gamma)(e^{\gamma\tau}-1)+2\gamma}] \quad (2.43)$$

$$B(\tau) = \frac{2(e^{\gamma\tau}-1)}{(\kappa+\gamma)(e^{\gamma\tau}-1)+2\gamma} \quad (2.44)$$

$$\gamma = \sqrt{\kappa^2 + 2\sigma^2} \quad (2.45)$$

From Equation (2.42), we can obtain the short rate r_{t_i}:

$$r_{t_i} = \frac{A(\tau) - lnP_i}{B(\tau)} \tag{2.46}$$

Then the *Jacobian* of the transformation from the instantaneous short rate to the zero-coupon bond price is:

$$J = \frac{dr_{t_i}}{dP_i} = -\frac{1}{P_i B(\tau)} \tag{2.47}$$

Therefore, the transition density function of P_i conditional on P_{i-1} is

$$\begin{aligned} h_{CIR}(P_i|P_{i-1}) &= |J| h_{CIR}\left(r_{t_i}\middle|r_{t_{i-1}}\right) \\ &= \left|-\frac{1}{P_i B(\tau)}\right| \cdot 2c \cdot Noncentral\ \chi^2\left(2cr_{t_i}; \frac{4\kappa\theta}{\sigma^2}, 2cr_{t_{i-1}}e^{-\kappa\Delta t}\right) \\ &= \frac{2c}{P_i|B(\tau)|} \cdot Noncentral\ \chi^2\left(2cr_{t_i}; \frac{4\kappa\theta}{\sigma^2}, 2cr_{t_{i-1}}e^{-\kappa\Delta t}\right) \end{aligned} \tag{2.48}$$

The natural logarithm of Eq. (2.48) is expressed as:

$$\begin{aligned} \ln h_{CIR}(P_i|P_{i-1}) &= \ln(\frac{2c}{P_i|B(\tau)|}) + \ln\left(Noncentral\ \chi^2\left(2cr_{t_i}; \frac{4\kappa\theta}{\sigma^2}, 2cr_{t_{i-1}}e^{-\kappa\Delta t}\right)\right) \\ &= \ln(2c) - \ln(|B(\tau)|) - \ln(P_i) \\ &\quad + \ln\left(Noncentral\ \chi^2\left(2cr_{t_i}; \frac{4\kappa\theta}{\sigma^2}, 2cr_{t_{i-1}}e^{-\kappa\Delta t}\right)\right) \end{aligned} \tag{2.49}$$

So, the logarithm of the likelihood function of zero-coupon bond prices $P_1, P_2, \cdots, P_n$ is

$$\mathcal{L}(\kappa,\theta,\sigma) = \ln h_{CIR}(P_n, P_{n-1}, \cdots, P_2|P_1)$$

$$= (n-1)\ln(2c) - (n-1)\ln(|B(\tau)|) - \sum_{i=2}^{n} \ln(P_i)$$

$$+ \sum_{i=2}^{n} \ln\left(Noncentral\ \chi^2\left(2cr_{t_i}; \frac{4\kappa\theta}{\sigma^2}, 2cr_{t_{i-1}}e^{-\kappa\Delta t}\right)\right) \tag{2.50}$$

Given Eq. (2.50), we can apply the method of maximum likelihood and estimate the parameters of the CIR model: κ, θ, σ. In addition, we can get the 1- or h-step ahead short rate forecasts.

$$E(r_{t+1}|r_t) = \hat{\theta}\left(1 - e^{-\hat{\kappa}\Delta t}\right) + e^{-\hat{\kappa}\Delta t} r_t \tag{2.51}$$

$$E(r_{t+h\Delta t}|r_t) = \hat{\theta}\left(1 - e^{-\hat{\kappa}h\Delta t}\right) + e^{-\hat{\kappa}h\Delta t} r_t \tag{2.52}$$

Then the corresponding forecasts of the zero-coupon bond yields have the expressions

$$\hat{y}_{t+1}(\tau) = -\frac{\hat{A}(\tau) - \hat{B}(\tau)\hat{r}_{t+1}}{\tau} \tag{2.53}$$

$$\hat{y}_{t+h\Delta t}(\tau) = -\frac{\hat{A}(\tau) - \hat{B}(\tau)\hat{r}_{t+h\Delta t}}{\tau} \tag{2.54}$$

with

$$\hat{A}(\tau) = \frac{2\hat{\kappa}\hat{\theta}}{\hat{\sigma}^2} ln\left[\frac{2\hat{\gamma}e^{\frac{\tau}{2}(\hat{\kappa}+\hat{\gamma})}}{(\hat{\kappa}+\hat{\gamma})(\mathrm{e}^{\hat{\gamma}\tau}-1)+2\hat{\gamma}}\right] \tag{2.55}$$

$$\hat{B}(\tau) = \frac{2(\mathrm{e}^{\hat{\gamma}\tau}-1)}{(\hat{\kappa}+\hat{\gamma})(\mathrm{e}^{\hat{\gamma}\tau}-1)+2\hat{\gamma}} \tag{2.56}$$

$$\hat{\gamma} = \sqrt{\hat{\kappa}^2 + 2\hat{\sigma}^2} \tag{2.57}$$

Fixing τ and simplifying $\hat{y}_{t+1}(\tau)$ and $\hat{y}_{t+h\Delta t}(\tau)$ as $\hat{y}_{t+1}$ and $\hat{y}_{t+h\Delta t}$, respectively, for convenience.

2.2.3 Non-linear diffusion model (ND)

Ait-Sahalia (1996) proposes a non-linear diffusion estimation process for the zero bond yields. The model has the form:

$$\Delta y_t = \alpha_0 + \alpha_1 y_t + \alpha_2 y_t^2 + \alpha_3 \frac{1}{y_t} \tag{2.58}$$

where y_t is the level of the yield at time *t* and Δy_t is the expected yield change from *t* to *t* + 1.

We can use the historical yield data to estimate parameters $\alpha_0, \alpha_1, \alpha_2$ and α_3, and then use those estimates to forecast the change in the bond yield, $\Delta\hat{y}_t$, as well as the expected bond yield, $\hat{y}_{t+1}$. The formulas for $\Delta\hat{y}_t$ and $\hat{y}_{t+1}$ are:

$$\Delta\hat{y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 y_t + \hat{\alpha}_2 y_t^2 + \hat{\alpha}_3 \frac{1}{y_t} \tag{2.59}$$

$$\hat{y}_{t+1} = y_t + \Delta\hat{y}_t \tag{2.60}$$

2.2.4 Autoregressive process of order one (AR(1))

The autoregressive process specifies that a time-series variable linearly depends on its previous values. The simplest form of an autoregressive process is a first-order autoregressive process, AR(1), which has the form:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t \tag{2.61}$$

where y_t and y_{t-1} are the outputs at time *t* and *t*-1, respectively. The AR(1) model shows that the future output can be described by its own one-lag behind value. In this paper y_t and y_{t-1} are the level of the zero bond yield at time *t* and *t*-1, respectively. And $\varepsilon_t \sim N(0, \sigma^2)$ is a white noise process. Assumption $|\beta| < 1$ makes the process AR(1) stationary.

In this paper I regress y_t on the yield at the previous time point, y_{t-1}, and get estimates of α and β. Then I use those estimates to forecast the future bond yields, which means $\hat{y}_{t+1} = \hat{\alpha} + \hat{\beta} y_t$.

2.2.5 Martingale model

A martingale process specifies that the expected value of the next output is equal to the current observation, given all prior observed values. The mathematical expression is

$$E(y_{t+1} | y_t, y_{t-1}, \dots, y_1) = y_t \tag{2.62}$$

In this paper, y_t, y_{t-1},..., y_1 are the zero bond yield at time *t*, *t*-1,...,1, respectively.

2.2.6 Bootstrap bias correction process

The prior literature shows that the empirical estimation of the mean reversion parameter κ can incur large bias (Yu and Phillips (2001)). The bias can be encountered by all the estimation approaches, including the maximum likelihood estimation (Tang and Chen (2009)). When κ is small and approaching zero, the problem becomes more serious. To reduce the bias in κ, current literature proposes three main methods: the median unbiased estimator (Andrews(1993)), the indirect inference method (Smith (1993), Phillips and Yu (2009)), and the bootstrap method (Tang and Chen (2009)). Specifically, Tang and Chen (2009) derive a simple bias correction formula for the Vasicek and CIR processes, which is $\hat{\kappa}_A = \hat{\kappa}_B - \frac{4}{T}$, where $\hat{\kappa}_A$ and $\hat{\kappa}_B$ are the estimator <u>a</u>fter and <u>b</u>efore the bias correction processes, respectively, and *T* is the data span in years. In this paper, the estimates calculated before the bootstrap bias correction are denoted as estimates with the subscript "B". Correspondingly, the estimates corrected after the bias correction process are denoted as values after the bias correction and signed as a subscript "A".

The following describes the bootstrap bias correction process of Tang and Chen (2009) to reduce the bias in estimatingκ. Let $\hat{\alpha}_B = [\hat{\kappa}_B, \hat{\theta}_B, \hat{\sigma}_B]$ be the parameter estimates before the bias correction process. First, generate a bootstrap sample path with the parameter estimates $\hat{\alpha}_B$ and the same sampling interval from the ATSM. Second, utilize the same estimation approach (in this case, MLE) and obtain new intermediate estimators $\hat{\alpha}^* = [\hat{\kappa}^*, \hat{\theta}^*, \hat{\sigma}^*]$. Third, repeat the first two steps *m* number of times and obtain *m* sets of intermediate parameter estimates $\hat{\alpha}_1^*, \hat{\alpha}_2^*, \hat{\alpha}_3^*, \ldots, \hat{\alpha}_m^*$. Let

$$\overline{\hat{\alpha}^*} = \frac{1}{N_S} \sum_{m=1}^{N_S} \hat{\alpha}_m^* \tag{2.63}$$

Then the estimated bias is $(\overline{\hat{\alpha}^*} - \hat{\alpha}_B)$ and so the bootstrap bias correction estimator is

$$\hat{\alpha}_A = \hat{\alpha}_B - (\overline{\hat{\alpha}^*} - \hat{\alpha}_B) \tag{2.64}$$

and the bootstrap estimator for the variance of $\hat{\alpha}_B$ is

$$\widehat{Var}(\hat{\alpha}_B) = \frac{1}{N_S} \sum_{m=1}^{N_S} (\hat{\alpha}_m^* - \overline{\hat{\alpha}^*})(\hat{\alpha}_m^* - \overline{\hat{\alpha}^*})' \tag{2.65}$$

To fully understand the bootstrap bias correction process, I replicate the results of Tang and Chen (2009). In my simulations, the true parameter values for the Vasicek model are κ=0.140, θ=0.0891, σ=0.0173, and the true values for the CIR model are κ=0.148, θ=0.09, σ=0.0707. Since my paper tests the performance of parameter estimation in the situation of near unit root, the three parameter values are chosen to make sure the autoregressive correlation, ρ, of the model is 0.99. The autoregressive correlation ρ is equal to $e^{-\kappa \Delta t}$, where κ is the mean reversion parameter and Δt is the time interval in years. All other conditions are the same for two models: the time increment Δt is 1/12 corresponding to monthly observations; the size of one sample path n is 300; the sample paths repeats N = 5,000 times and the number of the bootstrap resample m is 1,000.

The comparison between my replication results and Tang and Chen's is in Table 2.1. Checking the Vasicek model, we can see that the result from Tang and Chen shows that the bias of κ of is 0.190 and the relative bias is 133.046% before running the bias correction process. After the bias correction process, the bias of κ is reduced to 0.010 and the relative bias is reduced to 6.713%. My replication results show that the bias of κ is 0.183 and the relative bias is 130.714% before the bias correction. After running the bootstrap bias correction, the bias of κ is reduced to 0.012 and the relative bias becomes 8.571%. This shows that my replication results are close to Tang and Chen's. It is also true for the CIR model. The bias of κ of Tang and Chen's and mine for the CIR model are 0.192 and 0.203, respectively, before running the bootstrap process. After the bias correction, the corresponding biases are reduced to 0.007 and -0.0006, respectively. So my process to replicate the boot strap bias correction process is reliable.

2.2.7 Giacomini and White test

One way to compare term structure models is to examine the models' forecasting power. I use the Giacomini and White (2006) test to examine the models' predictive ability. The loss functions I examine are the mean squared error (MSE) and the absolute error (MAE), though other loss functions could be used. Suppose we want to compare the predictive ability of two models, Model *#1* and Model *#2*. Let n be the sample size of the estimation window and let p be the number of out-of-sample forecasts. If we do h-step-ahead forecasts, model #1 and model #2 will have predicted values $\hat{y}^1_{n+h+1}, \hat{y}^1_{n+h+2}, \dots, \hat{y}^1_{n+h+p-1}, \hat{y}^1_{n+h+p}$, and $\hat{y}^2_{n+h+1}, \hat{y}^2_{n+h+2}, \dots, \hat{y}^2_{n+h+p-1}, \hat{y}^2_{n+h+p}$, respectively. If the true values are $y_{n+h+1}, y_{n+h+2}, \dots, y_{n+h+p-1}, y_{n+h+p}$, we compute the squared errors at $i = n + h + 1, \dots, n + h + p$ as

$$L^1_i = L(\hat{y}^1_i | y_1, y_2, \dots, y_n) = (\hat{y}^1_i - y_i)^2 \tag{2.66}$$

$$L^2_i = L(\hat{y}^2_i | y_1, y_2, \dots, y_n) = (\hat{y}^2_i - y_i)^2 \tag{2.67}$$

The difference between L^1_i and L^2_i is the test function denoted as

$$Z_i = L^1_i - L^2_i \tag{2.68}$$

Define

$$\bar{Z} = \frac{1}{p} \sum_{i=n+h+1}^{n+h+p} Z_i \tag{2.69}$$

Then the test statistic of the Giacomini and White test is

$$T = \frac{\bar{Z}}{\hat{\Omega}/\sqrt{p}} \tag{2.70}$$

where $\hat{\Omega}$ is a covariance matrix constructed by the Newey-West method.

To examine whether model #1 and model #2 have equal predictive ability, the null hypothesis is a test of equal predictive ability of model #1 and model #2, $H_0: E[Z_i] = 0$ against the alternative hypothesis $H_{a1}: E[Z_i] \neq 0$. A level-α test can be conducted by rejecting the null hypothesis of equal predictive ability whenever $|T| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of a standard normal distribution. When α is 0.05, the critical value $z_{\alpha/2}$ is 1.96.

To examine whether model #1 has greater predictive ability than model #2, the null hypothesis is a test of $H_0: E[Z_i] = 0$ against $H_{a2}: E[Z_i] < 0$. A level-α test can be conducted by rejecting the null hypothesis whenever $T < z_\alpha$, where z_α is the α quantile of a standard normal distribution. When α is 0.025, the critical value z_α is -1.96.

2.3 Data simulation processes

To study the power and size of the Gaussian diffusion model of Vasicek and the square-root diffusion model of CIR, I simulate 10,000 paths of the instantaneous short rates based on the dynamics of the two one-factor ATSMs and then compute the zero bond prices and yields according to the corresponding ATSMs pricing equations. For each set of simulated data I then estimate yield forecasts using all models.

2.3.1 Vasicek data simulation process

The one-factor Vasicek model is tractable from a mathematical point of view. Given a spot rate r_s at time *s*, one can derive an explicit expression for the instantaneous short rate r_t at any time *t* (*t* > *s*) by Ito's Formula (Arnold (1974, PP130), Phillips (1972), Phillips and Yu (2009)),

$$r_t = e^{-\kappa(t-s)} r_s + \int_s^t \kappa\theta e^{-\kappa(t-u)} du + \int_s^t e^{-\kappa(t-u)} \sigma dW_t$$

$$= e^{-\kappa(t-s)} r_s + \theta\left(1 - e^{-\kappa(t-s)}\right) + \sigma \int_s^t e^{-\kappa(t-u)} dW_t \quad (2.71)$$

The last integral in Eq. (2.71) is the only stochastic element. Arnold (1974, PP77) shows that the integral is a normally distributed stochastic variable since it is independent of r_t . So r_t has a conditional normal distribution with the conditional mean and the variance (Vasicek (1977), Dixit and Pindyck (1994, PP76)):

$$\mu(r_t|r_s) = r_s e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right) \quad \forall\, s \le t \tag{2.72}$$

$$\sigma_t^2(r_t|r_s) \;=\; \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right), \forall\, s \le t \tag{2.73}$$

So the process of r_t can be written as an autoregressive form:

$$r_t = r_s e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right) + \varepsilon_t \;\;,\;\; t > s \tag{2.74}$$

where ε_t is normally distributed with mean zero and variance $\sigma_\varepsilon^2 = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right)$. The equivalent first-order autoregressive process, AR(1), is

$$r_t = r_{t-1} e^{-\kappa\Delta t} + \theta(1 - e^{-\kappa\Delta t}) + \varepsilon_t^1 \tag{2.75}$$

with $\varepsilon_t^1|\mathcal{F}_s \sim N(0, \frac{\sigma^2}{2\kappa}\,(1 - e^{-2\kappa\Delta t}))$. Specifically, Eq. (2.1) is the limiting case as $\Delta t \longrightarrow 0$ of the above AR(1) process (Eq. (2.75)). Thus, the sample path of r_t can be constructed by using the exact discrete-time expression:

$$r_t = r_{t-1} e^{-\kappa\Delta t} + \theta^Q(1 - e^{-\kappa\Delta t}) + \sigma\sqrt{\frac{1}{2\kappa}\,(1 - e^{-2\kappa\Delta t})}Z_t \;\;,\; Z_t \sim N(0,1) \tag{2.76}$$

In this chapter, I will use Eq. (2.76) to generate Vasicek-simulated short rates and then substitute those short rates into Eq. (2.14) to generate Vasicek-simulated zero bond yields. To simulate a series of short rates that are close to a unit root, I use the parameter values: $\theta = 0.0891$, $\kappa = 0.140$, and $\sigma = 0.0173$.[7] The autoregressive coefficient of the corresponding discrete time series data is 0.99. The size of a sample path *n* is 300 corresponding to monthly data series over a 25-year horizon (*T*=25 years). The time increment $\Delta t = \frac{1}{12}$. The initial value of the short rate is set equal to the long-run mean, or

---

[7] The parameter values come from Tang and Chen (2009).

$r_0 = 0.0891$. The data simulation path repeats N=10,000 times and the number of bootstrap resample is m = 500. Figure 2.1 presents the plot of one sample path of the Vasicek simulated short rate. The dashed line represents $E(r_t)$, the unconditional expected value, and the solid bold line represents $E(r_t) \pm 2\sigma$, two standard deviations above and below the mean.

2.3.2 CIR data simulation process

Given Eq. (2.37), the natural logarithm of the likelihood function of the short rates of the CIR model has the expression:

$$\ln h_{CIR}(r_t|r_{t-1}) = \ln(2c) + \ln\left(Noncentral\ \chi^2\left(2cr_t; \tfrac{4\kappa\theta}{\sigma^2}, 2cr_{t-1}e^{-\kappa\Delta t}\right)\right) \quad (2.77)$$

To simulate factors from the CIR model, we need to sample the random variables from the non-central chi-square distribution. Broadie and Kaya (2006) suggest a method of sampling directly from a non-central chi-square distribution.

Let $X'(x; d, \phi)$ denote a non-central chi-squared random variable with d (d > 1) degrees of freedom and non-centrality parameter ϕ. We can generate an ordinary chi-square distribution with (d-1) degrees of freedom, $X(x; d-1)$, and an independent standard normal random variable Z. For any d > 1, the following representation is valid:

$$X'(x; d, \phi) = X(x; d-1) + (Z + \sqrt{\phi})^2 \quad (2.78)$$

So if we can generate a series of standard normal random variables Z_1, Z_2, Z_3,…, Z_n, then we can simulate CIR short rates based on the density in Eq. (2.37),

$$X'\left(2cr_t; \tfrac{4\kappa\theta}{\sigma^2}, 2cr_{t-1}e^{-\kappa\Delta t}\right) = X\left(2cr_t; \tfrac{4\kappa\theta}{\sigma^2} - 1\right) + \left(Z_t + \sqrt{2cr_{t-1}e^{-\kappa\Delta t}}\right)^2 \quad (2.79)$$

I use Eq. (2.79) to generate CIR-simulated short rates and then substitute those short rates into Eq. (1.39) to generate CIR-simulated zero bond yields.

The parameter values used are $\theta = 0.09$, $\kappa = 0.148$, and $\sigma = 0.0707$[8]. The autoregressive coefficient of the corresponding discrete time model is 0.99. Such settings are for the examination of ATSMs when the parameters are near the unit root. A sample size *n* contains 300 months of zero bond yields (*T*=25 years) with maturity of 6 months and $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. The Monte Carlo simulation repeats *N* = 10,000 times. The initial value r_0 equals the long run mean of the process. The bootstrap bias correction process is based on *m* = 500 resamples. The plot of one sample path of the CIR-simulated short rates with a dashed line representing the expected value $E(r_t)$ and solid bond lines representing $E(r_t) \pm 2 \times s.d.$ presents in Figure 2.2.

2.3.3 Robustness checks

To determine the sensitivity of the power to distinguish different ATSMs to variations in the models' parameters, I conduct a variety of experiments.

Setting A: Change the model parameters: for the Vasicek model $(\kappa, \theta, \sigma) = (0.858, 0.0891, 0.0173)$, and for the CIR model $(\kappa, \theta, \sigma) = (0.892, 0.09, 0.0707)$. The mean reversion parameters of Setting A are 6 times the κ of the ones in Section 2.3.1 and 2.3.2. This implies that data series with Setting A have 1/6 of the half life of an interest rate shock of the data with the settings in the above two sections. In the Vasicek and CIR

[8] The parameter values come from Tang and Chen (2009).

models, the half life is calculated as $-\ln 0.5/\kappa$, which gives us important information about how quickly the instantaneous short rates would converge to their long-run mean after a shock. Setting A is designed to place the parameters far away from the unit root situation and the autoregressive coefficient of the new models is 0.93.

Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data.

Setting C: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$) while keeping the sample size n = 300. Correspondingly, the time span is reduced from $T = 25$ years to $T' = n \times \Delta t' = 300 \times \frac{1}{365} = 0.82$ years.

The aim of Setting B and C is to examine the sensitivity of the power to distinguish different ATSMs to the sample size and to the time span of the sample.

2.4 Results

To explore whether two ATSMs can be distinguished from each other, I perform a Monte Carlo study. The study investigates the power and size of ATSMs by evaluating their out-of-sample performance. I simulate data from different ATSMs and then use the data set to estimate each model, including those not used to generate the data. For ease of exposition, I consider a simple case when two one-factor ATSMs: Vasicek and CIR model, are investigated. Besides the two ATSMs, I add a nonparametric diffusion model (ND), an AR(1), and a martingale model as alternative forecasting models.

There are four subsections in this section. The first subsection states the results of the model estimation. Section 2.4.2 shows the models' out-of-sample forecast performance and Section 2.4.3 is the results on the evaluation of the models' forecasts given the simulated data are near the unit root. The last subsection discusses the power and size of the models' to distinguish between different ATSMs.

2.4.1 Model estimation results

This section presents model estimation and forecasting results. I use both Vasicek-simulated and CIR-simulated 6-month zero bond yields to estimate the ND, AR(1), Martingale, Vasicek, and CIR models. Since the parameter estimators of the Vasicek and CIR models have a bias problem, I apply the bootstrap bias correction process to those two models. This provides two sets of parameter estimates for the Vasicek and CIR models. Adding the other three forecasting models, in total, there are seven sets of parameter estimates. I then use those sets of estimates to forecast 1-, 6-, and 12-month ahead zero bond yields.

2.4.1.1 MLE of ATSMs based on simulated short rates

Table 2.2 presents the parameter estimates of the Vasicek and CIR model when the data are simulated instantaneous short rates. Panel A shows the Vasicek estimate statistics of analytical ML process before bias correction (Part 1), numeric ML before and after the bias correction (Parts 2 and 3, respectively). Since the bias correction process is based on the numeric iteration of MLE process, it is not possible to run the bias correction to the analytical ML process and then there is only before-bias-correction analytical ML estimates. Panel B presents the statistics of the CIR estimates of the numeric MLE before and after bias correction (Parts 1 and 2, respectively).

Checking the values in Part 1 of Panel A, we can see that the bias of the mean-reversion speed, $\hat{\kappa}_{B,AML}$, is 0.1283 or about 153% of the true value, while the biases of $\hat{\theta}_{B,AML}$and $\hat{\sigma}_{B,AML}$ are 0.0017 and 0.00098, respectively. This shows that the mean reversion estimate has more bias than the estimates of the long-run mean and volatility, which is consistent with the result of Tang and Chen (2009). The same results can be found in Part 1 of Panel B. In addition, the statistics of estimates in Part 2 of Panel A that are based on the numeric MLE are close to those in Part 1, which are based on the analytical MLE. This means that the numerical estimation is reliable. Part 3 of Panel A and Part 2 of Panel B show the estimates after applying the bias correction process. Comparing the mean reversion bias in Part 2 and Part 3 in Panel A, the mean reversion speed is reduced from 0.128 before the bias correction to 0.048 after the bias correction, while the biases of θ and σ before and after the bias correction are not bigger than those of the mean reversion parameter. Similar results can be found in Panel B for CIR. I also find that the mean reversion biases in Table 2.2 differ from the corresponding values in Table 2.1. Although the two tables use the same data simulating process, the different number of data simulating paths and the different number of bootstrap resamples can cause the difference in the final values.

To check the convergence of the parameters, I estimate both the Vasicek and CIR models using 3-month, 10-year, and 20-year zero coupon yields. The results are in Table 2.3. The parameter estimates in Table 2.3 are very close to the corresponding values in

Table 2.2. This means that the sample size used to estimate the models are large enough and the parameter estimates converge.

2.4.1.2 MLE of ATSMs based on simulated bond prices

Because the short rate is not observed in practice, I estimate the Vasicek and CIR models using the zero yields constructed from the simulated instantaneous short rates. Table 2.4 reports the statistics for the Vasicek and CIR models when the MLE is applied to the zero yields. Panel A represents results for the Vasicek simulated data and Panel B for CIR. When applying the MLE to the Vasicek data and without running the bias correction, the difference between the estimate and the true value of the mean reversion parameter is 0.1283, the relative bias of $\hat{\kappa}_{B,Vas}$ is 91.64%, and the root mean square error (RMSE) of $\hat{\kappa}_{B,Vas}$ is 0.2446. After running the bootstrap bias correction process, the bias of $\hat{\kappa}_{A,Vas}$ is reduced to 0.0483, the relative bias is 34.5%, and the RMSE of $\hat{\kappa}_{A,Vas}$ is also reduced to 0.1999. Checking the differences between the true values and estimated values of the long-run mean and the volatility, we can see that the bias of $\hat{\theta}_{B,Vas}$ and $\hat{\sigma}_{B,Vas}$ are 0.0017 and 0.0010, respectively, before bias correction, while the bias of $\hat{\theta}_{A,Vas}$ and $\hat{\sigma}_{A,Vas}$ is reduced to -0.0008 and 0.0002, respectively, after the correction. Obviously, the reduction of the biases of the long-run mean and the volatility is smaller than that of the mean reversion parameter. This result shows that the mean reversion parameter exhibits more substantial estimation bias than the other two parameters. This is consistent with Tang and Chen (2009) that the main objective of the bootstrap bias correction process is to reduce the bias of the mean-reversion parameter estimate. When the CIR MLE is applied to the Vasicek-simulated data, the RMSE of the estimates both before and after the bias correction are larger than the corresponding values based on the Vasicek MLE. In particular, the order of the bias of the mean reversion estimates is $\hat{\kappa}_{A,Vas} < \hat{\kappa}_{A,CIR} < \hat{\kappa}_{B,Vas} < \hat{\kappa}_{B,CIR}$.

Panel B of Table 2.4 presents the corresponding statistics based on the CIR-simulated data, both before and after running the bias correction process. Similar to the results in Panel A, the bootstrap bias correction decreases more bias of the mean reversion estimate than the biases of estimates of the long-run mean and the volatility. Since the data are simulated from the CIR model, the RMSEs of parameter estimates based on the Vasicek MLE are larger than the RMSEs based on the CIR MLE. The order of the bias of the mean reversion estimates is $\hat{\kappa}_{A,CIR} < \hat{\kappa}_{A,Vas} < \hat{\kappa}_{B,CIR} < \hat{\kappa}_{B,Vas}$.

2.4.1.3 Non-linear diffusion and AR(1) estimation

To compare the ATSMs to alternative models, I estimate the ND model and AR(1) using the Vasicek and CIR-simulated zero bonds, and then use the ND and AR(1) estimates to forecast the 1-, 6-, and 12-month ahead zero bond yields.

Table 2.5 reports the estimation results. The simulated data are Vasicek-simulated zero bond yields whose parameter values are given in the first row of Table 2.5. ND model has four parameters that capture the level, slope, and curvature of the zero bond yields, while AR(1) model has only two parameters, which capture the level and slope of the zero bond yields.

2.4.2 Models' out-of-sample forecasts

To compare the predictive ability of the models, I use the MLE parameter estimates summarized in Table 2.4 to compute the 1-month, 6-month, and 12-month out-of-sample zero yield forecasts using Eqs. (2.25), (2.26), (2.27) and Eq. (2.28) for the Vasicek model and Eqs. (2.51), (2.52), (2.53) and Eq. (2.54) for the CIR model. The purpose is to analyze the data's ability to distinguish one factor Vasicek from CIR models.

The forecasting short rates and zero bond yields are in Table 2.6. Panel A reports the forecasts when the data are simulated from the Vasicek model and Panel B are the results when the data are simulated from the CIR model. "Vasicek-B" and "Vasicek- A" represent the sets of MLE-Vasicek estimates, with "B" indicating before the bias correction and "A" indicating after the bias correction. Similarly, "CIR-B" and "CIR-A" are the corresponding sets of estimates for the MLE-CIR. The results show that for any model, the RMSEs of the forecasting short rates and zero bond yields increase when the forecasting time horizon increases from 1 month to 6 months, and up to 12 months, while the standard deviations of the forecasting short rates and zero bond yields are decreasing with the increase of the forecasting time horizon. Comparing the four models in Panel A, we can see that Vasicek-A has the smallest forecast RMSE and CIR-B has the largest forecast RMSE, when fixing the forecasting time horizon. For example, the RMSE of 12-month ahead forecasting short rate of Vasicek-A is 0.0189 and the RMSE of the 12-month ahead short rate of CIR-B is 0.0319, while the RMSEs of the 12-month ahead short rates of Vasicek-B and CIR-A are 0.0294 and 0.0211, respectively. Similar results apply to the forecasting zero bond yields. If we rank the RMSEs of the forecasts from the smallest to the largest, we can see the order is the same as that of the mean reversion bias, although the true model is the Vasicek model.

If we check the forecasts in Panel B based on the data simulated from CIR, we find that the order of the RMSEs of the forecasts of the four models from the smallest to the largest are CIR-A < Vasicek-A < CIR-B < Vasicek-B. This order is the same as the order of the mean reversion bias. The results show that no matter what the true model is, the RMSE is closely related to the mean reversion bias. The smaller the mean reversion bias is, the smaller the RMSE of the forecasts is.

Table 2.7 reports the out-of-sample forecasting results of all seven models, including the two ATSMs before and after the bias correction: Vasicek-A, Vasicek-B, CIR-A, CIR-B, and three alternative models: ND, AR(1), Martingale model. In Table 2.7, I present the root mean square errors (RMSEs) and the mean absolute errors (MAEs) of the forecasting zero bond yields over 1-, 6-, and 12-month forecasting duration. Checking the 1-month ahead forecasts of zero bond yields in Panel A, we see that Vasicek A has the smallest RMSE and CIR B has the largest one among the four ATSMs. The order of the RMSEs of the forecasting zero bond yields of the four ATSMs is Vasicek-A < CIR-A < Vasicek-B < CIR-B, which is consistent with the order of the biases of the mean reversion estimates in Panel A of Table 2.4. Comparing the three alternative models in Panel A, the ND has the smallest RMSE, AR(1) is the second, and the Martingale model has the largest RMSE of forecasting zero bond yields.

If we compare the ATSMs with the alternative models in Panel A, the ATSMs have smaller forecast errors than the alternative models. In general, the order of the RMSEs of $\hat{y}_t$ from smallest to largest is Vasicek-A < CIR-A < Vasicek-B < CIR-B < ND < AR(1) < Martingale. The same result can be found on the 6- and 12-month ahead forecasts and can also be applied to the MAE of $\hat{y}_t$ over 1-, 6-, and 12-month ahead forecasts. Panel B shows the out-of-sample forecasting results based on the CIR simulated data. Similar to the results of Panel A, the four ATSMs have smaller RMSEs and MAEs of the forecasting zero bond yields than three alternative models. Among the four ATSM forecasts, the order of RMSEs or MAEs of $\hat{y}_t$ is consistent with the biases of mean reversion estimates in Panel B of Table 2.4. So the order of the RMSEs of $\hat{y}_t$ from smallest to largest in Panel B is CIR-A < Vasicek-A < CIR-B < Vasicek-B < ND < AR(1) < Martingale. Checking the RMSEs of $\hat{y}_t$ over the forecasting time horizons in both panels, we can see that the RMSEs of $\hat{y}_t$ increase with the increase of the forecasting time horizon. The same result can be found on the MAEs of $\hat{y}_t$ of any model in both Panel A and B.

2.4.3 Evaluation of out-of-sample forecasting performance

This section discusses how well two ATSMs can be statistically differentiated from each other. If the Vasicek model is the data generation process and the CIR model is the comparison model, the question becomes whether the Vasicek model's out-of-sample forecast are statistically superior to those from the CIR model.

To compare models' out-of-sample forecasts, I use the Giacomini and White (2006) test to assess the predictive ability of seven models. The results are reported in Table 2.8. The first two columns in Table 2.8 identify a baseline model, labeled as "*Model #1*", and a reference model, labeled "*Model #2*". I use the Giacomini and White (2006) test to examine the null hypothesis that Model #1 and Model #2 have equal predictive ability. The third column, titled "*% Paths #1 ≠ #2",* reports the proportion of sample replications where the null hypothesis has been rejected at the 5% significance level. The fourth column, "*% Paths #1 Beats #2*", reports the proportion of 10,000 sample replications where model #1 has greater predictive ability than the model #2 at the 5% significance level. Model #1 outperforms Model #2 in the out-of-sample forecasts if "*% Paths #1 Beats #2*" exceeds 50%. The larger the "*% Paths #1 Beats #2*" is, the greater predictive ability the Model #1 has.

Take Vasicek-A as Model #1 and CIR-A as Model #2 in Panel A where the data have been simulated from the Vasicek model. In the 1-month ahead forecasts, 84.71% of the 10,000 replications reject the hypothesis that the forecast errors of Vasicek-A and CIR-A have equal values at the 5% significance level. This means that the probability that Vasicek-A can be statistically differentiated from CIR-A is 84.71%. In addition, in 57.54% of the 10,000 replications Vasicek-A has significantly smaller MSE of forecasting zero bond yields than CIR-A, which means 5,754 sample paths out of 10,000 replications accept that Vasicek-A has greater out-of-sample predictive ability than CIR-A. Over a 6-month ahead forecast horizon, the percentage of the rejection of equal predictive ability is reduced to 65.24% and the possibility of Vasicek-A beating CIR-A is 48.75%. Once the forecasting time horizon increases to 12 months, the percentage of the unequal predictive ability of Vasicek-A and CIR-A is reduced to 57.13% and 41.61% paths accept that Vasicek-A outperforms CIR-A.

When checking "*% Paths #1 Beats #2*" over a fixed forecasting time horizon in Panel A, we see that Vasicek-A has the greatest predictive ability, then CIR-A, the third is Vasicek-B, and CIR-B has the lowest predictive ability. This order is consistent with the order of the MSE in the forecasts of zero bond yields in Panel A of Table 2.7. This is not surprising since the construction of Giacomini and White test statistic is based on the MSE of model forecasts. In addition, we see that the order of ATSMs predictive ability is also consistent with the order of the biases of mean reversion estimates in Panel A of Table 2.4. The smaller the bias of the mean reversion estimate, the stronger is the predictive ability of the model. If we compare the four ATSMs with the three alternative models, we see that ATSMs always beat the ND, AR(1) and Martingale models. The percentage of rejections of equal forecast ability between any ATSM and ND, AR(1) and Martingale is almost 100%, and the percentage that an ATSM beats ND, AR(1) and Martingale is over 99%. Following the same logic, we find that the Martingale model has the weakest predictive ability among seven models. In summary, we make an order of the relative predictive ability of seven models: Vasicek-A > CIR-A > Vasicek-B > CIR-B > ND > AR(1) > Martingale model. This order is consistent with the order of RMSE of ATSMs forecasted zero bond yields in Panel A of Table 2.7.

If we compare seven models based on Column "*% Paths #1≠ #2*" over a fixed time horizon in Panel A, we see that the percentage values are over 80% in 1-month forecasting, over 70% in 6-month forecasting, and greater than 60% in 12-month forecasting. This means that the probability that four ATSMs can be distinguished is above 80% when the forecasting horizon is 1 month. The power to distinguish between ATSMs is reduced to about 70% and 60% once the horizon increases up to 6 and 12 months, respectively. In addition, the probability of distinguishing four ATSMs from three alternative models is almost over 95% over 1-, 6-, and 12-month forecasting.

Panel B of Table 2.8 is the results based on the CIR simulated data. If fixing the forecasting time horizon, we can get the similar results to Panel B, which is, no matter what the true model is, the smaller the bias of the mean reversion estimate of the model is, the stronger is the predictive ability of the model. Also, ATSMs always beat three alternative models and the Martingale model is the worst model in forecasting zero bond yields. Therefore, the order of the relative predictive ability in Panel B is CIR-A > Vasicek-A > CIR-B > Vasicek-B > ND > AR(1) > Martingale model. Considering the power of distinguishing seven models, we see that the probability of differentiating four ATSMs is above 95% and the power of distinguishing ATSMs from alternative models is almost greater than 99%.

Another interesting result showed in Table 2.8 is that the power to differentiate models becomes weaker as the forecasting horizon becomes longer, because the proportion of rejections of equal predictive ability of two models decreases as the forecasting horizon increases from 1 month, 6 months to 12 months. The possible reason is that models will gradually lose predictive power as the time horizon becomes longer since the predicted values of all models will approach the long-run mean of the data.

The above results are based on the 6-month zero yields simulated from the Vasicek model. I rerun the Giacomini and White test using the simulated 20-year zero yields from the Vasicek model. The results are in Table 2.9. Comparing the numbers in Tables 2.8 and 2.9, we see that the choice of the zeros does not change the results of the Giacomini and White test.

2.4.4 Discussion of power and size of the ATSMs

So far I discuss the out-of-sample performance of the Vasicek and CIR models when the simulated data are near the unit root. To understand more about the impact factors on models' out-of-sample performance and their ability to differentiate different ATSMs, I study the power and size of ATSMs under different data simulating situations in this section.

2.4.4.1 The impact of size of κ on models out-of-sample forecasting performance

The Monte Carlo study in prior sections examines ATSMs' relative out-of-sample forecasting ability when the mean reversion parameter is close to unit root. The results show that the model's relative predictive ability is closely related to the bias size of mean reversion estimates no matter what the true model is. To understand more about of the impact of the size of the mean reversion parameter on ATSMs' out-of-sample forecasting performance, this section focuses on the situation that the mean reversion parameter rises away from the unit root case. Therefore, I repeat the same Monte Carlo simulation process in the prior section but use the parameter sets in Setting A: the Vasicek model $(\kappa, \theta, \sigma) = (0.858, 0.0891, 0.0173)$ and the CIR model $(\kappa, \theta, \sigma) = (0.892, 0.09, 0.0707)$.

Table 2.10 presents estimation results based on Setting A. Comparing Table 2.10 with Table 2.4, we can see the biases of mean reversion estimates in Table 2.10 are larger than the biases of κ of the corresponding models in Table 2.4. However, the order of the bias of mean reversion estimates in Table 2.10 is the same as the order in Table 2.4. If the data is simulated from the Vasicek model, the order is $\hat{\kappa}_{A,Vas} < \hat{\kappa}_{A,CIR} < \hat{\kappa}_{B,Vas} < \hat{\kappa}_{B,CIR}$, while the order of the bias of the mean reversion estimates is $\hat{\kappa}_{A,CIR} < \hat{\kappa}_{A,Vas} < \hat{\kappa}_{B,CIR} < \hat{\kappa}_{B,Vas}$ when the data is generated from the CIR model.

Table 2.11 reports the summary of models' forecasting zero yields. Examining Table 2.11 and Table 2.7, we see that the RMSEs of forecasting zero yields in Table 2.11 are smaller than the corresponding values in Table 2.7. This shows the forecast errors decrease but the models' predictive ability increases when the speed of the mean reversion is 6 times the κ of the near unit root case. In Table 2.11, the RMSEs of forecasting zero bond yields of four ATSMs are close to each other and smaller than the

RMSEs of ND, AR(1), and the Martingale model. In addition, the results show that the RMSE and MAE of $\hat{y}_t$ of the Martingale model are close in both tables and larger than the RMSEs and MAEs of $\hat{y}_t$ of the other six models. If ranking the RMSEs based on the Vasicek simulated data, the order of the RMSEs of seven models from smallest to the largest is Vasicek-A < CIR-A < Vasicek-B < CIR-B < ND < AR(1) < Martingale. The order of the RMSEs becomes CIR-A < Vasicek-A < CIR-B < Vasicek-B < ND < AR(1) < Martingale if the data are simulated from the CIR model. The same orders can be found on MAE of forecasted zero bond yields.

Since the RMSEs of models' forecasts are smaller in Table 2.11 than in Table 2.7, I can predict that the power to differentiate seven models will decrease once the half life of a shock to the interest rate is only 1/6 of the original value. This prediction is confirmed in Table 2.12. Comparing the values in Table 2.12 with those Table 2.8, we see that, except for the Martingale model, the values of "*%Paths #1 ≠ #2*" and "*%Paths #1 Beats #2*" are smaller in Table 2.12 than the corresponding values in Table 2.8. Take Vasicek-A and CIR-A as an example. 84.71% of the 10,000 replications reject the hypothesis that CIR-A and Vasicek-A have equal predictive ability in the 1-month ahead forecasts in Panel A of Table 2.8, while the rejection percentage is reduced to 19.90% in Panel A of Table 2.12 when the data are generated from the Vasicek model. The percentage that Vasicek-A beats CIR-A is 57.54% in Panel A of Table 2.8, while the corresponding value in Panel A of Table 2.12 is 9.35%. This shows that the probability to distinguish Vasicek-A from CIR-A decreases from 84.71% to 19.90% and the probability that Vasicek-A beats CIR-A is reduced from 57.54% to 9.5% when the half life of a shock to the interest rate is only 1/6 of the original value. However, the order of the models' relative predictive ability in Panel A of Table 2.12 is the same as the order in Table 2.8 and is consistent with the order of the RMSEs of the forecasts, no matter what the data generating process is. The comparison of results in Table 2.8 and in Table 2.12 reflects that the situation that the mean reversion parameter is close to the unit root lowers the models' predictive ability but increases the power of distinguishing between different ATSMs. The possible explanation is that the larger mean reversion speed will revert the forecasted values more quickly to model's unconditional mean.

Since the value of κ has a big impact on the model's predictive ability as well as model's ability to distinguish between each other, I estimate the models using the different κ values. I increase κ from the smallest 0.160, to 0.400, then 0.700, and up to 0.900. Based on those different setting of κ values, I have four settings of parameter estimates, which are in Table 2.13. From the results of MLE-Vasicek in Table 2.13, we can see that the relative bias of the estimate of κ has reduced as the κ increases from 0.160 to 0.900. The model's predictive ability tests are in Table 2.14. Based on the results of Table 2.14, we can see that the percentage of Model #1 ≠ #2 as well as the percentage of Model #1 Beats #2 are decreasing as κ increases. This means that the increase of the κ value reduces the model's power to distinguish between each other.

2.4.4.2 The impact of sample size n on models out-of-sample forecasting performance

To investigate the impact of the sample size on ATSMs' relative predictive ability, I increase the size of the estimation window by the way of increasing the sampling frequency while keeping the time span unchanged. That is I rerun the same Monte Carlo simulation as that in the prior sections but increase the frequency from the monthly basis

($\Delta t = \frac{1}{12}$) to the daily basis ($\Delta t' = \frac{1}{365}$). Correspondingly, the sample size n increases from 300 months to 9,125 days when the time span T is 25 years.

Table 2.15 presents the parameter estimates with the daily frequency. The comparison of Table 2.4 and Table 2.15 shows that increasing the sample size without changing the time span has no help to reduce the bias of the mean reversion estimates. For example, the mean of $\hat{\kappa}_{B,Vas}$ is 0.2683 and the RMSE is 0.2446 in Panel A of Table 2.4, while the corresponding values are 0.3415 and 0.2804, respectively, in Panel A of Table 2.15. In general, the biases of mean reversion estimates in Table 2.15 are greater than the biases of the corresponding estimates in Table 2.4. Tang and Chen (2009) concludes that the bias of the mean reversion estimator is not a function of the number of observations, which is consistent with my results. Therefore, increasing the sample size by the way of increasing the sampling frequency while keeping the time span unchanged, cannot reduce the bias of the mean reversion estimator. In addition, we see that the orders of the biases of the mean reversion estimates in Table 2.15 and Table 2.4 are the same.

Table 2.16 presents the summary of models out-of-sample performance forecasts with 1-, 6-, and 12-month horizon when the sampling frequency increases from the monthly basis to the daily basis. Comparing the numbers in Table 2.16 to the numbers in Table 2.7, we see that the RMSEs and MAEs of forecasting zero bond yields in Table 2.16 are smaller than those in Table 2.7. But the orders of the RMSEs and MAEs of the forecasts from smallest to largest in both tables are the same. This result shows that increasing the sample size without changing the time span increases the models' predictive ability but may reduce models' power to distinguish between models. This conclusion is proved in Table 2.17.

Table 2.17 reports the evaluation results of models out-of-sample forecasting performance. The values of "*%Paths #1 ≠ #2*" in Table 2.17 are smaller than the corresponding values in Table 2.8. For example, the percentage of rejecting the hypothesis of equal predictive ability between Vasicek-A and CIR-A in the 1-month ahead forecasts is 84.71% in Panel A of Table 2.8, while the rejection percentage of the two models over the same forecasting time horizon is only 46.67% in Panel A of Table 2.17. This shows that the probability of the model differentiation is reduced when the sample size is increased by the way of increasing the sampling frequency without changing the time span. In addition, the power to differentiate the ATSMs from ND and AR(1) is also weaker in Table 2.17 than in Table 2.8. Take Vasicek-A and AR(1) in 1-month ahead forecasts as an example. When the data are from the Vasicek model, the rejection percentage of the two models is 99.96% in Panel A of Table 2.8, while the corresponding rejection percentage is reduced to 84.72% in Panel A of Table 2.17. If ordering the relative predictive ability of the models, we see that the order of the models' predictive ability in Table 2.17 is the same as that in Table 2.8, no matter what the data generating process is. Also, the order of the predictive ability of the four ATSMs is consistent with the order of the size of the bias of the mean reversion estimates in Table 2.15, no matter what the true model is. The smaller the bias of the mean reversion estimate is, the higher the model's predictive ability has.

2.4.4.3 The impact of time span *T* on models out-of-sample forecasting performance

Tang and Chen (2009) states that the bias of the mean reversion estimator is not a function of the number of observations but is a decreasing function of the time span. In details, the bias of the mean reversion estimator is a function of T^{-1}. The results in the last section prove the former part of their statement. This section will study the impact of the change in the time span T on ATSMs' relative predictive ability. Therefore, I use the same Monte Carlo study but changing the sampling frequency from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$) while keeping the sample size n = 300 unchanged. Correspondingly, the time span is reduced from $T = 25$ years to $T' = n \times \Delta t' = 300 \times \frac{1}{365} = 0.82$ years.

Table 2.18, Table 2.19, and Table 2.20 report the statistics of the models' estimates, the summary of the out-of-sample forecasting performance, and the results of the models' predictive ability tests, respectively. The results of Table 2.18 and Table 2.4 show that the bias of the mean reversion estimates substantially increases when the time span is reduced from 25 years to 0.82 years. This is consistent with Tang and Chen (2009). However, the order of the bias of the mean reversion estimates in both tables are the same.

Comparing the values in Table 2.19 to the corresponding values in Table 2.7 and Table 2.16, we see that the RMSEs and MAEs of forecasting zero bond yields in Table 2.19 are larger than the values in Table 2.16 but smaller than the values in Table 2.7. This shows that the models' predictive ability is increasing with the increasing of the sampling frequency while keeping the sample size unchanged. But when the sampling frequency is kept the same, the models' predictive ability is increasing with the increasing of the time span. In addition, the orders of RMSEs and MAEs of the forecasting zero bond yields from smallest to largest in Table 2.19 are the same as the orders in Table 2.7 and in Table 2.16. The order of the RMSEs of the forecasts of the ATSMs in Table 2.19 once again proves that the smaller the bias of the mean reversion estimate is, the smaller the RMSE and AME of the forecasted values are.

The comparison of Table 2.20 with Table 2.17 shows that the power to differentiate different ATSMs improves with the decreasing of the time span when the sampling frequency is the same, although the models' predictive ability is reduced. If comparing the values in Table 2.20 and the values in Table 2.8, we see that increasing the sampling frequency while keeping the sample size unchanged reduces the power to differentiate ATSMs.

2.5 Conclusion

A widely discussed issue in yield curve modeling is that ATSMs have performed poorly as forecasting tools. This paper discusses that the reason of ATSMs out-of-sample poor performance is there is no econometric model accurately describing the true data generating process. Therefore, the goal of the paper is to study how well the simulated data differentiate between the two affine models if there are no errors in the observed data. To explore the question of interest, I use a parsimonious Monte Carlo study to simulate six month zero-coupon data from both Vasicek and CIR models and then, in turn, estimate both the CIR and Vasicek models using that simulated data. The results show that the ATSMs' predictive ability is related to the mean reversion parameters. The smaller the bias of the estimate of the mean reversion is, the better out-of-sample forecasts the model has. Also, I find that a mean reversion process with a higher speed, the larger sample size, and the higher sampling frequency can improve ATSMs' predictive ability in contrast with the decrease of the models' power to differentiate between different ATSMs.

While we have gained understanding on the power and size of one-factor affine term structure models (ATSMs), there is a need to understand more on estimation of multifactor ATSMs. Another important issue is to understand how the existence of observational errors impact parameter estimates and out-of-sample forecasts since early empirical analysis of ATSMs focus primarily on estimation of the ATSMs with the auxiliary errors. Therefore, my future research will focus on exploring the ability of out-of-sample forecasts to statistically distinguish one ATSM from another a) when the data are simulated from multifactor ATSMs, and b) when the data is generated from an ATSM observed with errors.

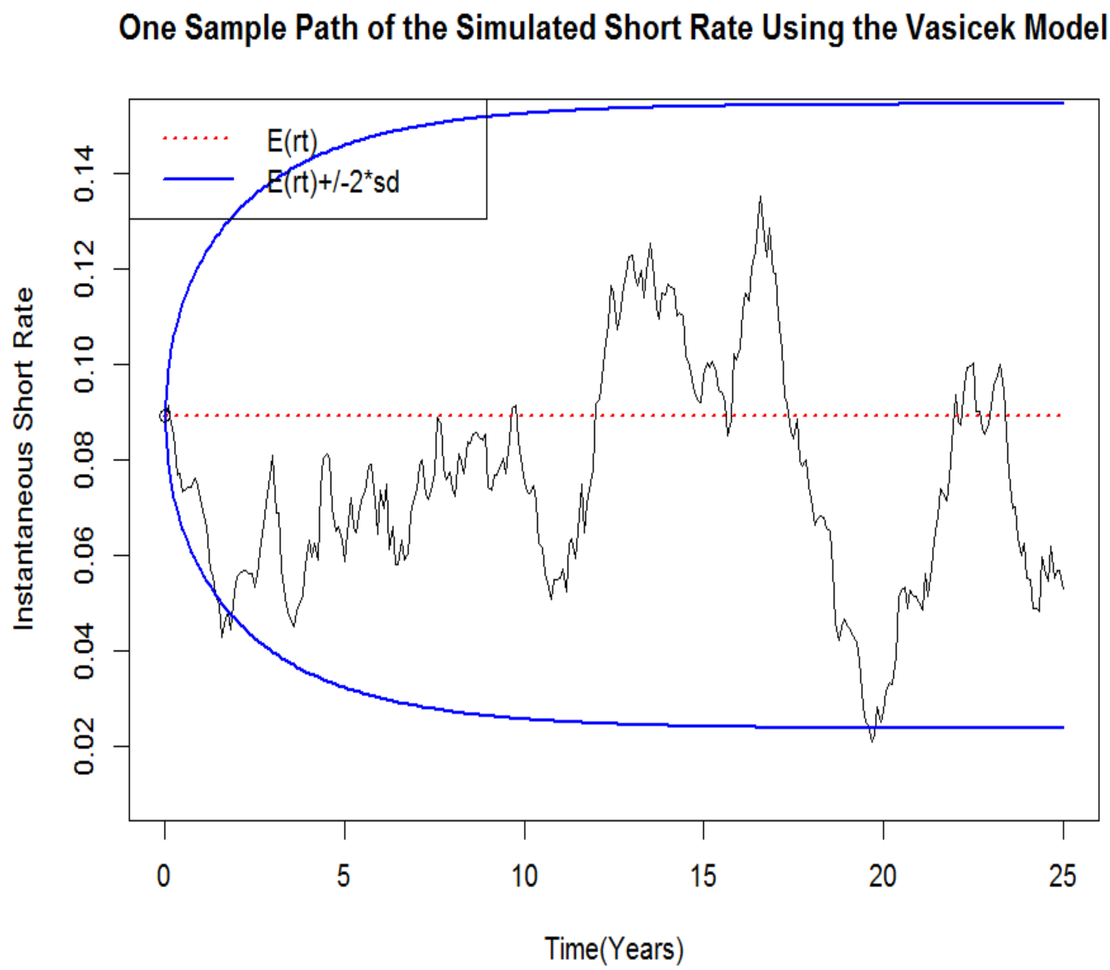

Figure 2.1 One sample path of the Vasicek simulated short rate

The Vasicek-simulated short rate is based on the parameter values of κ=0.140, θ=0.0891, and σ=0.0173. The size of a sample path n is 300 corresponding to monthly data series over a 25-year horizon (T=25 years). The time increment $\Delta t = \frac{1}{12}$. The initial value of the short rate is set equal to the long-run mean, or $r_0 = 0.0891$.

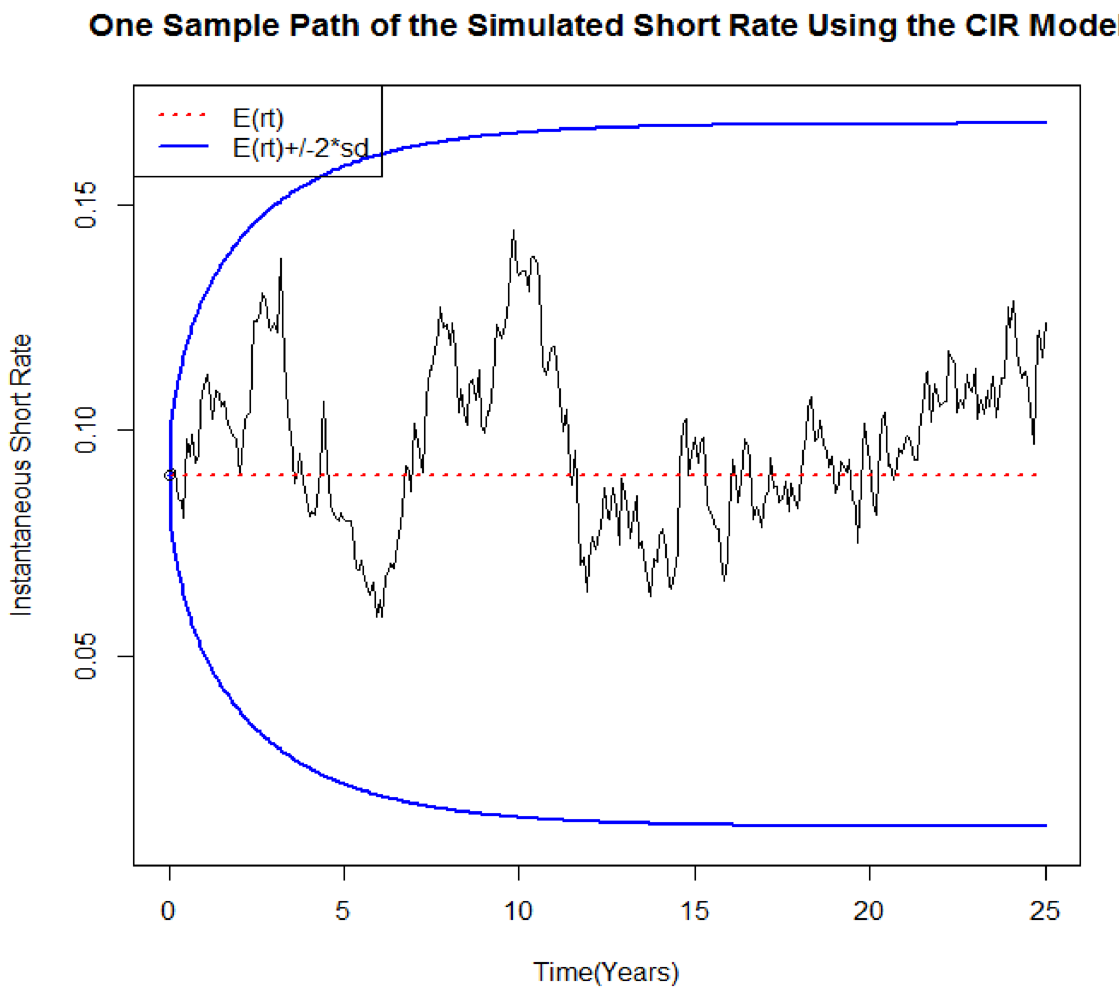

Figure 2.2 One sample path of the CIR simulated short rate

The CIR-simulated short rate is based on the parameter values of κ=0.148, θ=0.09, and σ=0.0707. The size of a sample path n is 300 corresponding to monthly data series over a 25-year horizon (T=25 years). The time increment $\Delta t = \frac{1}{12}$. The initial value of the short rate is set equal to the long-run mean, or $r_0 =$ 0.09.

Table 2.1 The results of replicating Table 3 in Tang and Chen (2009)

	Vasicek Model				CIR Model			
	True κ=0.140				True κ=0.148			
	Tang & Chen's		My Replication		Tang & Chen's		My Replication	
	$\hat{\kappa}_{B,TC}$	$\hat{\kappa}_{A,TC}$	$\hat{\kappa}_{B,MY}$	$\hat{\kappa}_{A,MY}$	$\hat{\kappa}_{B,TC}$	$\hat{\kappa}_{A,TC}$	$\hat{\kappa}_{B,MY}$	$\hat{\kappa}_{A,MY}$
Bias	0.190	0.010	0.183	0.012	0.192	0.007	0.203	-0.0006
R. Bias(%)	133.046	6.713	130.714	8.571	129.455	4.459	137.162	-0.405
S.D.	0.206	0.213	0.167	0.204	0.222	0.214	0.208	0.235

I replicate Table 3 in Tang and Chen (2009) and compare their results with mine. My simulation inputs are from Tang and Chen (2009). The true parameter values for the Vasicek model are κ=0.140, θ=0.0891, σ=0.0173,and for the CIR model are κ=0.148, θ=0.09, σ=0.0707. All other conditions are the same for two models: the time increment Δt is 1/12 corresponding to monthly observations; the size of one sample path n is 300; the sample paths repeats N = 5,000 times and the number of the bootstrap resample m is 1,000. $\hat{\kappa}_B$ and $\hat{\kappa}_A$ represent the κ estimators before and after the bootstrap bias correction process, respectively. The subscript "*TC*" denotes the results from Tang and Chan (2009) and "*MY*" means my replication. *Bias* is defined as the estimate minus true value. *R. Bias* is the relative bias. *S.D.* is the standard deviation.

Table 2.2 Parameter estimates based on Vasicek- and CIR-simulated instantaneous short rates

Panel A: Vasicek Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/12, n=300			
Part 1: Analytical MLE Before Bias Correction			
Estimates	$\hat{\kappa}_{B,AML}$	$\hat{\theta}_{B,AML}$	$\hat{\sigma}_{B,AML}$
Mean	0.2683	0.09080	0.01828
S.D.	0.4810	0.06950	0.01341
Median	0.2267	0.08901	0.01813
Skewness	1.1177	16. 123	0.2089
Kurtosis	2. 2290	263.00	1.4475
Part 2: Numeric MLE Before Bias Correction			
Estimates	$\hat{\kappa}_{B,NML}$	$\hat{\theta}_{B,NML}$	$\hat{\sigma}_{B,NML}$
Mean	0.2683	0.09080	0.01828
S.D.	0.2108	0.04695	0.00128
Median	0.2267	0.08901	0.01813
Skewness	1.1136	13.991	0.06502
Kurtosis	1.2122	265.11	0.56679
Part 3: Numeric MLE After Bias Correction			
Estimates	$\hat{\kappa}_{A,NML}$	$\hat{\theta}_{A,NML}$	$\hat{\sigma}_{A,NML}$
Mean	0.1883	0.08834	0.01732
S.D.	0.2043	0.04102	0.00071
Median	0.1341	0.08745	0.01731
Skewness	2.2247	16.461	0.08397
Kurtosis	5.6412	392.47	1.04774
Panel B: CIR Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/12, n=300			
Part 1: Numeric MLE Before Bias Correction			
Estimates	$\hat{\kappa}_{B,NML}$	$\hat{\theta}_{B,NML}$	$\hat{\sigma}_{B,NML}$
Mean	0.3475	0.09121	0.07488
S.D.	0.2035	0.04657	0.00492
Median	0.3001	0.08632	0.07386
Skewness	1.0485	11.133	0.64720
Kurtosis	0.9579	246.20	0.53838
Part 2: Numeric MLE After Bias Correction			
Estimates	$\hat{\kappa}_{A,NML}$	$\hat{\theta}_{A,NML}$	$\hat{\sigma}_{A,NML}$
Mean	0.2081	0.09163	0.07143
S.D.	0.2193	0.04681	0.00590
Median	0.1327	0.08638	0.07063
Skewness	1.7311	14.949	1.2565
Kurtosis	3.0457	346.72	4.1348

The numeric Vasicek MLE of Eq. (2.5) and the analytical Vasicek MLE of Eqs. (2.9 - 2.13) are applied to the Vasicek-simulated short rates. The results are showed in Panel A and the true parameter values are given in the first row. In Panel A, Part 1 presents statistics of analytical MLE before bias correction and Part 2 and Part 3 present the results of numeric MLE before and after bias correction process, respectively. The subscript "B" denotes before the bias correction and the "A" indicates after the bias correction. Since the bias correction process is based on the numeric iteration of ML process, it is no possible to find the after-bias-correction estimates to the analytical ML process and then there is only before-bias-correction analytical ML estimates. The numeric CIR MLE of Eq. (2.37) is applied to the CIR-simulated short rates whose true parameter values are given in the first row of Panel B. Panel B reports the results of numeric ML before and after bias correction process, respectively, with the subscript "B" and "A" the same meaning as those in Panel A. Any MLE process in this table has 300 observations (25 years) and there are 10,000 replications. The number of bootstrap resampling is 500.

Table 2.3 Parameter estimates based on Vasicek- and CIR-simulated zero coupon yields with three maturities

Panel A: Vasicek Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/12, n=300			
After Bias Correction: 3 Month			
Estimates	$\hat{\kappa}_A$	$\hat{\theta}_A$	$\hat{\sigma}_A$
Mean	0.1882	0.08487	0.01779
S.D.	0.2391	0.03619	0.00099
Median	0.1494	0.08476	0.01773
Skewness	1.0953	11.216	0.4677
Kurtosis	1.0024	256.04	1.3299
After Bias Correction: 10 Year			
Estimates	$\hat{\kappa}_A$	$\hat{\theta}_A$	$\hat{\sigma}_A$
Mean	0.1883	0.08475	0.01779
S.D.	0.2392	0.03618	0.00099
Median	0.1455	0.08457	0.01772
Skewness	1.0960	11.199	0.4675
Kurtosis	1.0046	255.53	1.3283
After Bias Correction: 20 Year			
Estimates	$\hat{\kappa}_A$	$\hat{\theta}_A$	$\hat{\sigma}_A$
Mean	0.1883	0.08466	0.01779
S.D.	0.2391	0.03611	0.00099
Median	0.1452	0.08443	0.01773
Skewness	1.0954	11.273	0.4677
Kurtosis	1.0026	257.71	1.3292
Panel B: CIR Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/12, n=300			
After Bias Correction: 3 Month			
Estimates	$\hat{\kappa}_A$	$\hat{\theta}_A$	$\hat{\sigma}_A$
Mean	0.2081	0.08961	0.06948
S.D.	0.2197	0.04221	0.00475
Median	0.1452	0.08475	0.06964
Skewness	1.2914	13.407	1.2447
Kurtosis	3.4181	303.37	4.3012
After Bias Correction: 10 Year			
Estimates	$\hat{\kappa}_A$	$\hat{\theta}_A$	$\hat{\sigma}_A$
Mean	0.2081	0.08963	0.06948
S.D.	0.2197	0.04220	0.00474
Median	0.1449	0.08475	0.06960
Skewness	1.2913	13.411	1.2440
Kurtosis	3.4145	303.37	4.3011
After Bias Correction: 20 Year			
Estimates	$\hat{\kappa}_A$	$\hat{\theta}_A$	$\hat{\sigma}_A$
Mean	0.2081	0.08966	0.06948
S.D.	0.2196	0.04221	0.00474
Median	0.1450	0.08477	0.06966
Skewness	1.2914	13.404	1.2448
Kurtosis	3.4180	303.36	4.3012

The Vasicek-MLE (Eq. (2.24)) and the CIR-MLE (Eq. (2.50)) are applied to both Vasicek-simulated and CIR-simulated zero bond yields whose parameter values are given in the first row of Panel A and Panel B, respectively. Both panels present the results of numeric MLE after bias correction process, which are applied to the 3 month, 10 year, and 20 year zero coupon yields. The subscript "A" indicates after the bias correction. Any MLE process in this table has 300 observations (25 years) and there are 10,000 replications. The number of bootstrap resampling is 500.

Table 2.4 Parameter estimates based on Vasicek- and CIR-simulated 6-month zero-coupon yields

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/12, n=300									
	MLE-Vasicek						**MLE-CIR**		
	Before Bias Correction						*Before Bias Correction*		
	Mean	S.D.	Bias	R. Bias(%)	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{B,Vas}$	0.2683	0.2108	0.1283	91.64	0.2446	$\hat{\kappa}_{B,CIR}$	0.3463	2.00	0.97
$\hat{\theta}_{B,Vas}$	0.09080	0.04679	0.0017	1.91	0.0467	$\hat{\theta}_{B,CIR}$	0.10172	-	-
$\hat{\sigma}_{B,Vas}$	0.01828	0.00128	0.0010	5.54	0.0015	$\hat{\sigma}_{B,CIR}$	0.06496	-	-
	After Bias Correction						*After Bias Correction*		
	Mean	S.D.	Bias	R. Bias(%)	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{A,Vas}$	0.1883	0.1941	0.0483	34.5	0.1999	$\hat{\kappa}_{A,CIR}$	0.2236	3.10	0.98
$\hat{\theta}_{A,Vas}$	0.08834	0.04102	-0.0008	-0.85	0.0405	$\hat{\theta}_{A,CIR}$	0.10197	-	-
$\hat{\sigma}_{A,Vas}$	0.01752	0.00141	0.0002	1.15	0.0022	$\hat{\sigma}_{A,CIR}$	0.06193	-	-
Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/12, n=300									
	MLE-CIR						**MLE-Vasicek**		
	Before Bias Correction						*Before Bias Correction*		
	Mean	S.D.	Bias	R. Bias(%)	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{B,CIR}$	0.3475	0.2006	0.1995	134.8	0.2815	$\hat{\kappa}_{B,Vas}$	0.3625	3.38	0.97
$\hat{\theta}_{B,CIR}$	0.09157	0.04297	0.0016	1.74	0.0048	$\hat{\theta}_{B,Vas}$	0.09472	-	-
$\hat{\sigma}_{B,CIR}$	0.07453	0.00511	0.0038	5.42	0.0065	$\hat{\sigma}_{B,Vas}$	0.02216	-	-
	After Bias Correction						*After Bias Correction*		
	Mean	S.D.	Bias	R. Bias(%)	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{A,CIR}$	0.2080	0.2150	0.0600	40.54	0.2319	$\hat{\kappa}_{A,Vas}$	0.2264	3.06	0.98
$\hat{\theta}_{A,CIR}$	0.09170	0.04722	0.0017	1.89	0.0051	$\hat{\theta}_{A,Vas}$	0.09499	-	-
$\hat{\sigma}_{A,CIR}$	0.07149	0.00579	0.0008	1.12	0.0060	$\hat{\sigma}_{A,Vas}$	0.02114	-	-

The Vasicek-MLE (Eq. (2.24)) and the CIR-MLE (Eq. (2.50)) are applied to both Vasicek-simulated and CIR-simulated 6-month zero bond prices whose parameter values are given in the first row of Panel A and Panel B, respectively. Then the bootstrap bias correction process is applied to each set of estimates. Each MLE process has 300 months of zero bond prices (T=25 years) with maturity of 6 months and $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. There are 10,000 replications. The number of bootstrap resampling is 500. Panel A presents the statistics of estimates based on the Vasicek-simulated data, with subscript "*B, Vas*" and "*A, Vas*" indicating before and after the bias corrected Vasicek model estimates, respectively, and with subscript "*B, CIR*" and "*A, CIR*" indicating before and after the bias corrected CIR model estimates, respectively. Panel B presents the corresponding statistics based on the CIR-simulated data. *Bias* is defined as the estimate minus true value. *R. Bias* is the relative bias. *S.D.* is the standard deviation. RMSE is the root mean square errors of the estimate. Half-Life is the defined as $-\ln 2/\kappa$. The autoregressive coefficient $\rho = e^{-\kappa \Delta t}$.

Table 2.5 Parameter estimates of non-linear diffusion and AR(1) models

Panel A: Vasicek Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/12, m=300					
			ND Parameter Estimates		
	Mean	S.D.	Median	Skew.	Kurtosis
Intercept $\hat{\alpha}_0$	-0.0680	0.1038	-0.0364	-3.3979	18.366
$\hat{\alpha}_1$	0.7372	1.0119	0.4472	3.0183	15.412
$\hat{\alpha}_2$	-2.7762	3.4497	-1.8274	-2.7287	13.506
$\hat{\alpha}_3$	0.0022	0.0037	0.0010	3.9227	23.893
			AR(1) Parameter Estimates		
	Mean	S.D.	Median	Skew.	Kurtosis
$\hat{\alpha}$	0.0025	0.0017	0.0021	1.4743	3.4578
$\hat{\beta}$	0.9715	0.0175	0.9753	-1.3379	2.7157
Panel B: CIR Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/12, m=300					
			ND Parameter Estimates		
	Mean	S.D.	Median	Skew.	Kurtosis
Intercept $\hat{\alpha}_0$	-0.0634	0.0929	-0.0342	-3.0372	17.069
$\hat{\alpha}_1$	0.6716	1.0499	0.6771	3.0081	21.204
$\hat{\alpha}_2$	-2.1932	5.4854	-1.8910	-1.6987	19.029
$\hat{\alpha}_3$	0.0028	0.0038	0.0010	3.6593	29.900
			AR(1) Parameter Estimates		
	Mean	S.D.	Median	Skew.	Kurtosis
$\hat{\alpha}$	0.0026	0.0016	0.0023	1.9037	9.5840
$\hat{\beta}$	0.9702	0.0173	0.9738	-1.2923	3.4932

The non-linear diffusion (ND) model of Eq. (2.58) and the autoregressive model with first order (AR(1)) of Eq. (2.61) are applied to Vasicek- and CIR-simulated 6-month zero bond yields whose parameter values are given in the first row of each panel, respectively. Each estimation process has 300 observations (25 years) and there are 10,000 replications. Panel A gives the descriptive statistics of the estimates of ND and AR(1) model when the data are simulated from the Vasicek model and Panel B reports the statistics results of the same models as in Panel A but the data are from the CIR model.

Table 2.6 ATSMs out-of-sample forecasts

		Panel A Vasicek Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/12, m=300						**Panel B** CIR Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/12, m=300					
Models	Forecast Horizon	True Short Rate r_t	Forecasted Short Rate $\hat{r}_t$	RMSE of $\hat{r}_t$	True Yield y_t	Forecasted Yield $\hat{y}_t$	RMSE of $\hat{y}_t$	True Short Rate r_t	Forecasted Short Rate $\hat{r}_t$	RMSE of $\hat{r}_t$	True Yield y_t	Forecasted Yield $\hat{y}_t$	RMSE of $\hat{y}_t$
Vasicek-B	1-month	0.0887 (0.0327)	0.0894 (0.0317)	0.0075	0.0867 (0.0316)	0.0895 (0.0309)	0.0072	0.0900 (0.0386)	0.0910 (0.0374)	0.0065	0.0897 (0.0370)	0.0914 (0.0353)	0.0066
	6-month	0.0889 (0.0324)	0.0904 (0.0312)	0.0152	0.0888 (0.0313)	0.0905 (0.0290)	0.0154	0.0902 (0.0386)	0.0912 (0.0346)	0.0196	0.0902 (0.0372)	0.0913 (0.0345)	0.0198
	12-month	0.0890 (0.0323)	0.0906 (0.0299)	0.0294	0.0889 (0.0312)	0.0907 (0.0281)	0.0303	0.0903 (0.0384)	0.0915 (0.0250)	0.0298	0.0902 (0.0372)	0.0916 (0.0361)	0.0323
Vasicek-A	1-month	0.0887 (0.0327)	0.0890 (0.0313)	0.0050	0.0867 (0.0316)	0.0892 (0.0303)	0.0052	0.0900 (0.0386)	0.0903 (0.0376)	0.0062	0.0897 (0.0370)	0.0902 (0.0357)	0.0065
	6-month	0.0889 (0.0324)	0.0893 (0.0280)	0.0129	0.0888 (0.0313)	0.0894 (0.0285)	0.0132	0.0902 (0.0386)	0.0906 (0.0349)	0.0156	0.0902 (0.0372)	0.0906 (0.0360)	0.0157
	12-month	0.0890 (0.0323)	0.0896 (0.0278)	0.0189	0.0889 (0.0312)	0.0897 (0.0274)	0.0195	0.0903 (0.0384)	0.0910 (0.0241)	0.0216	0.0902 (0.0372)	0.0912 (0.0344)	0.0214
CIR-B	1-month	0.0887 (0.0327)	0.0897 (0.0317)	0.0149	0.0867 (0.0316)	0.0899 (0.0323)	0.0156	0.0900 (0.0386)	0.0904 (0.0373)	0.0063	0.0897 (0.0370)	0.0906 (0.0345)	0.0065
	6-month	0.0889 (0.0324)	0.0914 (0.0246)	0.0250	0.0888 (0.0313)	0.0916 (0.0291)	0.0282	0.0902 (0.0386)	0.0910 (0.0327)	0.0179	0.0902 (0.0372)	0.0911 (0.0305)	0.0180
	12-month	0.0890 (0.0323)	0.0916 (0.0214)	0.0319	0.0889 (0.0312)	0.0917 (0.0215)	0.0381	0.0903 (0.0384)	0.0913 (0.0285)	0.0269	0.0902 (0.0372)	0.0913 (0.0268)	0.0286
CIR-A	1-month	0.0887 (0.0327)	0.0892 (0.0317)	0.0051	0.0867 (0.0316)	0.0894 (0.0309)	0.0055	0.0900 (0.0386)	0.0900 (0.0376)	0.0060	0.0897 (0.0370)	0.0899 (0.0354)	0.0062
	6-month	0.0889 (0.0324)	0.0901 (0.0295)	0.0142	0.0888 (0.0313)	0.0902 (0.0223)	0.0149	0.0902 (0.0386)	0.0903 (0.0340)	0.0154	0.0902 (0.0372)	0.0903 (0.0322)	0.0153
	12-month	0.0890 (0.0323)	0.0902 (0.0269)	0.0211	0.0889 (0.0312)	0.0903 (0.0216)	0.0225	0.0903 (0.0384)	0.0906 (0.0206)	0.0212	0.0902 (0.0372)	0.0906 (0.0291)	0.0208

MLE of ATSMs are applied to 25 year (300 months) of 6–month zero bond prices simulated from the Vasicek and CIR model whose true parameter values are given in the first row of Panel A and B, respectively. The corresponding four sets of model estimates are used to make the out-of-sample forecasts. There are 10,000 replications. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. For each estimate set of "*Vasicek-B*" and "*Vasicek-A*", the Eqs. (2.25) and (2.26) and the Eqs. (2.27) and (2.28) are used to forecast 1-, 6-, and 12-month ahead short rates and zero bond yields, respectively. The Eqs. (2.51) and (2.52) and the Eqs. (2.53) and (2.54) are applied to the set of "*CIR-B*" and "*CIR-A*" to forecast 1-, 6-, and 12-month out-of-sample short rates and zero bond yields. RMSE is the root mean square error of the forecasts. The number in the cell is the mean value and the standard deviation is in the parentheses. Panel A represents the forecasts when the data are simulated from the Vasicek model and Panel B reports the results when the data are from the CIR model.

Table 2.7 Comparison of models out-of-sample forecasting performance

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/12, n=300						
	1-month Forecasting		6-month Forecasting		12-month Forecasting	
	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Vasicek-B	0.0098	0.0082	0.0154	0.0128	0.0205	0.0164
Vasicek-A	0.0063	0.0050	0.0134	0.0109	0.0195	0.0154
CIR-B	0.0110	0.0095	0.0168	0.0142	0.0217	0.0174
CIR-A	0.0083	0.0070	0.0142	0.0117	0.0200	0.0159
ND	0.0357	0.0270	0.0390	0.0273	0.0416	0.0266
AR(1)	0.0396	0.0288	0.0425	0.0301	0.0433	0.0300
Martingale	0.2882	0.2297	0.7051	0.5639	0.9851	0.7873
Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/12, n=300						
	1-month Forecasting		6-month Forecasting		12-month Forecasting	
	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Vasicek-B	0.0109	0.0089	0.0186	0.0149	0.0234	0.0164
Vasicek-A	0.0092	0.0077	0.0169	0.0136	0.0217	0.0149
CIR-B	0.0099	0.0079	0.0181	0.0143	0.0220	0.0152
CIR-A	0.0072	0.0057	0.0161	0.0128	0.0210	0.0144
ND	0.0367	0.0279	0.0290	0.0294	0.0432	0.0257
AR(1)	0.0406	0.0296	0.0308	0.0321	0.0449	0.0291
Martingale	0.2879	0.2295	0.7050	0.5640	0.9847	0.7870

Seven alternative estimation techniques are applied to 25 year (300 months) of 6-month zero yields simulated from a Vasicek model and a CIR model whose true parameter values are given in the first row of Panel A and Panel B, respectively. The corresponding sets of model estimates are used to make out-of-sample forecasts. There are 10,000 replications. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model. "*AR(1)*" denotes the autoregressive model with the first order. "*Martingale*" is the martingale model. Denote RMSE as the root mean square error of the forecasted zero bond yields and denote MAE as the mean of the absolute error of the forecasted values.

Table 2.8 Models' relative predictive ability test based on simulated 6-month zero yields

Panel A: Vasicek Simulated Data with Parameter Values: **κ=0.140, θ=0.0891, σ=0.0173, Δt=1/12, n=300**							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	94.72	18.52	80.51	14.67	54.08	7.03
	CIR-B	95.49	85.50**	87.32	78.32**	73.18	62.05*
	CIR-A	85.33	19.55	73.84	17.55	65.60	16.79
	ND	99.78	98.81***	99.55	96.41***	98.78	94.96***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	97.45	83.42**	92.11	79.12**	64.72	53.12*
	CIR-A	84.71	57.54*	65.24	48.75	57.13	41.61
	ND	98.63	96.04***	97.43	94.84***	95.92	92.88***
	AR(1)	100	98.96***	100	98.80***	100	98.12***
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	96.88	65.47	84.11	5.45	62.35	4.46
	ND	99.32	98.92***	98.82	97.75***	98.78	96.92***
	AR(1)	100	99.96***	100	99.90***	100	99.78***
	Martingale	100	99.88***	100	99.88***	100	99.88***
CIR-A	ND	99.61	98.96***	99.28	97.80***	99.11	97.23***
	AR(1)	100	99.93***	100	99.90***	100	99.89***
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	81.29	51.97*	79.07	50.29*	72.50	49.12*
	Martingale	99.84	97.87***	99.56	96.96***	99.55	95.57***
AR(1)	Martingale	100	100***	100	100***	100	100***

Table 2.8 (Continued)

Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/12, n=300							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths of #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	99.96	17.67	99.10	6.93	93.28	6.80
	CIR-B	99.96	47.98	99.84	47.41	99.16	47.03
	CIR-A	99.88	29.76	99.84	25.74	96.66	24.97
	ND	100	100***	99.89	99.80***	98.60	97.31***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	99.75	70.79**	99.42	70.29**	96.39	62.42*
	CIR-A	98.71	45.62	97.15	45.59	95.88	45.08
	ND	99.63	99.32***	99.37	99.16***	98.80	98.10***
	AR(1)	99.96	99.90***	99.89	99.79***	99.79	99.76***
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	99.96	6.99	99.44	6.93	93.73	6.46
	ND	99.96	99.96***	99.91	99.90***	98.52	98.31***
	AR(1)	99.89	98.80***	99.63	98.04***	99.42	97.15***
	Martingale	100	100***	100	100***	100	100***
CIR-A	ND	99.89	99.03***	98.68	96.50***	97.15	94.32***
	AR(1)	99.64	99.27***	99.43	98.01***	99.37	97.80***
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	83.55	52.48*	82.75	52.29*	82.59	51.12*
	Martingale	100	100***	100	100***	100	100***
AR(1)	Martingale	100	100***	100	100***	100	100***

Seven alternative estimation techniques are applied to 25 year (300 months) of 6-month zero yields simulated from a Vasicek model and a CIR model whose true parameter values are given in the first row of Panel A and Panel B, respectively. The corresponding sets of model estimates are used to make 1- , 6-, and 12-month ahead forecasts. There are 10,000 replications. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR B*" and "*CIR A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model, "*AR(1)*" denotes the autoregressive model with the first order, and "*Martingale*" is the martingale model. The column "*% Paths #1 ≠ #2*" represents the proportion of 10,000 sample replications that the model #1 and the model #2 have NO equal predictive ability at the 5% significance level. The column "*% Paths #1 Beats #2*" stands for the proportion of 10,000 sample replications that the model #1 has greater predictive ability than the model #2 at the 5% significance level. Asterisks indicate that the proportion of sample replications that model #1 beats model #2 exceeds 50% (*), 70% (**), and 90% (***).

Table 2.9 Models' relative predictive ability test based on simulated 20-year zero yields

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/12, n=300							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	94.70	18.52	80.51	14.67	54.08	7.03
	CIR-B	95.49	85.50**	87.32	78.30**	73.15	62.05*
	CIR-A	85.12	19.54	73.84	17.55	65.60	16.79
	ND	99.80	98.79***	99.55	96.41***	98.78	94.96***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	97.45	83.42**	92.11	79.12**	64.72	53.12*
	CIR-A	84.75	57.54*	65.23	48.75	57.13	41.61
	ND	98.63	96.04***	97.43	94.84***	95.92	92.88***
	AR(1)	100	98.98***	100	98.81***	100	98.12***
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	96.88	65.49	84.09	5.45	62.35	4.46
	ND	99.29	98.93***	98.82	97.75***	98.78	96.92***
	AR(1)	100	99.99***	100	99.90***	100	99.78***
	Martingale	100	99.88***	100	99.88***	100	99.88***
CIR-A	ND	99.61	98.96***	99.33	97.80***	99.11	97.23***
	AR(1)	100	99.93***	100	99.91***	100	99.89***
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	81.29	51.98*	79.08	50.29*	72.50	49.12*
	Martingale	99.82	97.87***	99.56	96.96***	99.55	95.58***
AR(1)	Martingale	100	100***	100	100***	100	100***

Table 2.9 (Continued)

		Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/12, n=300					
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths of #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	99.96	17.68	99.13	6.89	93.28	6.80
	CIR-B	99.96	47.99	99.84	47.46	99.16	47.13
	CIR-A	99.88	29.81	99.83	26.11	96.67	24.99
	ND	100	100***	99.89	99.80***	98.60	97.31***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	99.75	70.77**	99.42	70.29**	96.39	62.39*
	CIR-A	98.71	45.64	97.15	45.63	95.88	45.11
	ND	99.63	99.32***	99.37	99.18***	98.80	98.10***
	AR(1)	99.96	99.90***	99.89	99.79***	99.80	99.76***
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	99.96	7.09	99.44	6.96	93.73	6.68
	ND	99.96	99.96***	99.91	99.90***	98.52	98.31***
	AR(1)	99.89	98.81***	99.63	98.04***	99.42	97.15***
	Martingale	100	100***	100	100***	100	100***
CIR-A	ND	99.89	99.11***	98.68	96.50***	97.16	95.00***
	AR(1)	99.64	99.29***	99.43	98.10***	99.37	97.80***
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	83.55	54.38*	82.75	53.09*	82.59	52.22*
	Martingale	100	100***	100	100***	100	100***
AR(1)	Martingale	100	100***	100	100***	100	100***

Seven alternative estimation techniques are applied to 25 year (300 months) of 20-year zero yields simulated from a Vasicek model and a CIR model whose true parameter values are given in the first row of Panel A and Panel B, respectively. The corresponding sets of model estimates are used to make 1- , 6-, and 12-month ahead forecasts. There are 10,000 replications. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR B*" and "*CIR A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model, "*AR(1)*" denotes the autoregressive model with the first order, and "*Martingale*" is the martingale model. The column "*% Paths #1 ≠ #2*" represents the proportion of 10,000 sample replications that the model #1 and the model #2 have NO equal predictive ability at the 5% significance level. The column "*% Paths #1 Beats #2*" stands for the proportion of 10,000 sample replications that the model #1 has greater predictive ability than the model #2 at the 5% significance level. Asterisks indicate that the proportion of sample replications that model #1 beats model #2 exceeds 50% (*), 70% (**), and 90% (***).

Table 2.10 Parameter estimates based on Vasicek- and CIR-simulated 6-month zero-coupon yields: Setting A when the mean reversion parameter is far away the unit root

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.858, θ=0.0891, σ=0.0173, Δt=1/12, n=300							
	MLE-Vasicek				**MLE-CIR**		
	Before Bias Correction				Before Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{B,Vas}$	1.0512	0.3356	0.3867	$\hat{\kappa}_{B,CIR}$	0.9643	0.72	0.93
$\hat{\theta}_{B,Vas}$	0.08905	0.00334	0.0033	$\hat{\theta}_{B,CIR}$	0.08905	-	-
$\hat{\sigma}_{B,Vas}$	0.01821	0.00182	0.1381	$\hat{\sigma}_{B,CIR}$	0.06028	-	-
	After Bias Correction				After Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{A,Vas}$	0.8679	0.3323	0.3363	$\hat{\kappa}_{A,CIR}$	0.7822	0.89	0.94
$\hat{\theta}_{A,Vas}$	0.08916	0.00334	0.0033	$\hat{\theta}_{A,CIR}$	0.08920	-	-
$\hat{\sigma}_{A,Vas}$	0.01729	0.00172	0.0017	$\hat{\sigma}_{A,CIR}$	0.05757	-	-
Panel B: CIR Simulated Data with Parameter Values: κ=0.892, θ=0.09, σ=0.0707, Δt=1/12, n=300							
	MLE-CIR				**MLE-Vasicek**		
	Before Bias Correction				Before Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{B,CIR}$	1.0993	0.3328	0.3881	$\hat{\kappa}_{B,Vas}$	1.1067	0.63	0.92
$\hat{\theta}_{B,CIR}$	0.08992	0.00383	0.0038	$\hat{\theta}_{B,Vas}$	0.08992	-	-
$\hat{\sigma}_{B,CIR}$	0.07451	0.00728	0.0083	$\hat{\sigma}_{B,Vas}$	0.02236	-	-
	After Bias Correction				After Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{A,CIR}$	0.9184	0.3330	0.3346	$\hat{\kappa}_{A,Vas}$	0.9131	0.76	0.93
$\hat{\theta}_{A,CIR}$	0.09011	0.00384	0.0038	$\hat{\theta}_{A,Vas}$	0.09006	-	-
$\hat{\sigma}_{A,CIR}$	0.07141	0.00700	0.0070	$\hat{\sigma}_{A,Vas}$	0.02123	-	-

The Vasicek-MLE (Eq. (2.24)) and the CIR-MLE (Eq. (2.50)) are applied to both Vasicek-simulated and CIR-simulated 6 month zero bond prices whose parameter values are given in the first row of Panel A and Panel B, respectively. Setting A: Change the model parameters: the Vasicek model $(\kappa,\theta,\sigma)=(0.858, 0.0891, 0.0173)$ and the CIR model $(\kappa,\theta,\sigma)=(0.892, 0.09, 0.0707)$. The mean reversion parameters of Setting A are 6 times the κ of the setting that the model is approaching the unit root. Setting A is designed to place the parameters far away from the unit root situation. Then the bootstrap bias correction process is applied to each set of estimates. Each MLE process has 300 months of zero bond prices (T=25 years) with maturity of 6 months and $\Delta t=\frac{1}{12}$ corresponds to monthly observations in an annualized basis. There are 10,000 replications. The number of bootstrap resampling is 500. Panel A presents the statistics of estimates based on the Vasicek-simulated data, with subscript "*B, Vas*" and "*A, Vas*" indicating before and after the bias corrected Vasicek model estimates, respectively, and with subscript "*B, CIR*" and "*A, CIR*" indicating before and after the bias corrected CIR model estimates, respectively. Panel B presents the corresponding statistics based on the CIR-simulated data. *S.D.* is the standard deviation. RMSE is the root mean square errors of the estimate. Half-Life is the defined as $-\ln 2/\kappa$. The autoregressive coefficient $\rho = e^{-\kappa\Delta t}$.

Table 2.11 Comparison of models out-of-sample forecasting performance: Setting A when the mean reversion parameter is far away the unit root

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.858, θ=0.0891, σ=0.0173, Δt=1/12, n=300						
	1-month Forecasting		6-month Forecasting		12-month Forecasting	
	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Vasicek-B	0.0041	0.0033	0.0085	0.0068	0.0101	0.0082
Vasicek-A	0.0041	0.0033	0.0084	0.0068	0.0099	0.0081
CIR-B	0.0041	0.0033	0.0085	0.0068	0.0101	0.0082
CIR-A	0.0041	0.0033	0.0085	0.0068	0.0100	0.0081
ND	0.0059	0.0041	0.0096	0.0074	0.0112	0.0090
AR(1)	0.0079	0.0051	0.0112	0.0096	0.0129	0.0109
Martingale	0.2881	0.2297	0.7050	0.5638	0.9848	0.7871
Panel B: CIR Simulated Data with Parameter Values: κ=0.892, θ=0.09, σ=0.0707, Δt=1/12, n=300						
	1-month Forecasting		6-month Forecasting		12-month Forecasting	
	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Vasicek-B	0.0051	0.0040	0.0104	0.0083	0.0123	0.0098
Vasicek-A	0.0051	0.0040	0.0104	0.0083	0.0124	0.0098
CIR-B	0.0051	0.0040	0.0104	0.0083	0.0124	0.0098
CIR-A	0.0050	0.0040	0.0104	0.0083	0.0123	0.0098
ND	0.0068	0.0038	0.0121	0.0087	0.0217	0.0115
AR(1)	0.0081	0.0053	0.0172	0.0112	0.0221	0.0156
Martingale	0.2879	0.2296	0.7049	0.5635	0.9846	0.7873

Seven alternative estimation techniques are applied to three settings of 6-month zero yields simulated from a Vasicek model (Panel A) and a CIR model (Panel B). The corresponding sets of model estimates are used to make out-of-sample forecasts. Setting A: Change the model parameters: for the Vasicek model $(\kappa, \theta, \sigma) = (0.858, 0.0891, 0.0173)$, and for the CIR model $(\kappa, \theta, \sigma) = (0.892, 0.09, 0.0707)$. The mean reversion parameters of Setting A are 6 times the κ of the setting that the model is approaching the unit root. Setting A is designed to place the parameters far away from the unit root situation. There are 10,000 replications and the number of bootstrap resampling is 500. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model. "*AR(1)*" denotes the autoregressive model with the first order. "*Martingale*" is the martingale model. Denote RMSE as the root mean square error of the forecasted zero bond yields and denote MAE as the mean of the absolute error of the forecasted values.

Table 2.12 Models' relative predictive ability test: Setting A when the mean reversion parameter is far away the unit root

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.858, θ=0.0891, σ=0.0173, Δt=1/12, n=300							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	23.01	6.61	22.81	4.14	18.78	2.24
	CIR-B	28.95	5.69	24.57	5.39	19.65	2.94
	CIR-A	24.76	4.69	22.51	3.79	18.16	2.81
	ND	26.97	25.27	19.35	15.25	12.67	6.93
	AR(1)	21.39	20.34	17.94	15.57	15.39	12.09
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	22.71	16.28	22.50	13.42	17.31	12.67
	CIR-A	19.90	9.35	19.37	6.65	16.10	5.20
	ND	25.49	23.03	17.82	10.68	12.59	7.00
	AR(1)	21.62	19.59	17.40	11.14	16.28	9.88
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	22.62	6.10	21.67	5.04	17.52	3.90
	ND	26.67	25.35	19.52	13.71	12.62	6.55
	AR(1)	21.27	19.78	17.40	13.75	10.77	7.96
	Martingale	100	100***	100	100***	100	100***
CIR-A	ND	26.57	24.62	18.28	9.65	12.67	6.93
	AR(1)	23.91	21.07	17.53	9.28	13.91	6.76
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	25.16	16.15	18.65	9.86	15.15	8.61
	Martingale	100	100***	100	100***	100	100***
AR(1)	Martingale	100	100***	100	100***	100	100***

Table 2.12 (Continued)

Panel B: CIR Simulated Data with Parameter Values: κ=0.892, θ=0.09, σ=0.0707, Δt=1/12, n=300							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths of #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	20.04	8.58	19.91	5.86	15.90	3.66
	CIR-B	20.14	11.07	19.95	9.49	16.17	7.73
	CIR-A	20.75	7.99	19.92	5.61	15.55	3.57
	ND	22.90	12.76	18.41	12.59	13.25	12.24
	AR(1)	22.14	15.69	18.35	12.76	13.48	12.19
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	20.85	14.41	19.62	12.55	16.31	12.06
	CIR-A	20.86	9.00	19.82	8.32	14.98	6.56
	ND	23.71	12.88	15.26	10.40	11.58	8.21
	AR(1)	22.41	12.99	16.48	11.20	12.93	9.02
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	20.12	7.92	19.87	5.64	15.64	3.59
	ND	22.84	13.22	16.56	11.84	12.51	11.06
	AR(1)	21.25	15.69	15.99	12.24	12.80	11.49
	Martingale	100	100***	100	100***	100	100***
CIR-A	ND	23.09	13.11	15.89	10.30	11.14	8.22
	AR(1)	23.01	12.88	17.17	11.01	12.16	8.60
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	23.49	13.11	16.85	9.58	3.82	2.20
	Martingale	100	100***	100	100***	100	100***
AR(1)	Martingale	100	100***	100	100***	100	100***

Seven alternative estimation techniques are applied to three settings of 6-month zero yields simulated from a Vasicek model (Panel A) and a CIR model (Panel B). The corresponding sets of model estimates are used to make 1- , 6-, and 12-month ahead forecasts. The corresponding sets of model estimates are used to make 1- , 6-, and 12-month ahead forecasts. Setting A: Change the model parameters: for the Vasicek model $(\kappa, \theta, \sigma) = (0.858, 0.0891, 0.0173)$, and for the CIR model $(\kappa, \theta, \sigma) = (0.892, 0.09, 0.0707)$. The mean reversion parameters of Setting A are 6 times the κ of the setting that the model is approaching the unit root. Setting A is designed to place the parameters far away from the unit root situation. There are 10,000 replications and the number of bootstrap resampling is 500. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model, "*AR(1)*" denotes the autoregressive model with the first order, and "*Martingale*" is the martingale model. The column "*% Paths #1 ≠ #2*" represents the proportion of 10,000 sample replications that the model #1 and the model #2 have NO equal predictive ability at the 5% significance level. The column "*% Paths #1 Beats #2*" stands for the proportion of 10,000 sample replications that the model #1 has greater predictive ability than the model #2 at the 5% significance level. Asterisks indicate that the proportion of sample replications that model #1 beats model #2 exceeds 50% (*), 70% (**), and 90% (***).

Table 2.13 Parameter estimates based on Vasicek-simulated 15-year zero-coupon yields when the value of κ increases

Parameter Estimates Based On One-Factor Vasicek Simulated 15-Year Zero Coupon Yields (θ=0.117, σ=0.043, λ=-0.705)												
	κ=0.160			κ=0.400			κ=0.700			κ=0.900		
	Mean	S.D.	R. Bias(%)	Mean	S.D.	R. Bias (%)	Mean	S.D.	R. Bias(%)	Mean	S.D.	R. Bias(%)
MLE-Vasicek												
$\hat{\kappa}_{B,Vas}$	0.3640	0.2022	127.50	0.5696	0.2052	42.40	0.7752	0.1512	10.74	0.9014	0.0963	0.15
$\hat{\theta}_{B,Vas}$	0.1198	0.0443	2.39	0.1181	0.0169	0.94	0.1173	0.0123	0.26	0.1127	0.0143	0.01
$\hat{\sigma}_{B,Vas}$	0.0511	0.0049	18.84	0.0615	0.0225	43.02	0.0476	0.0097	10.70	0.0403	0.0049	0.00
$\hat{\lambda}_{B,Vas}$	-0.1880	0.1643	-73.33	-0.2313	0.1646	-67.19	-0.3381	0.2131	-52.04	-0.3485	0.3217	0.32
$\hat{\kappa}_{A,Vas}$	0.2093	0.2569	30.81	0.4258	0.2832	6.45	0.7232	0.0052	3.31	0.9009	0.1489	0.10
$\hat{\theta}_{A,Vas}$	0.1192	0.0592	1.88	0.1139	0.0271	-2.65	0.1166	0.0226	-0.34	0.1132	0.0279	0.03
$\hat{\sigma}_{A,Vas}$	0.0519	0.0620	20.70	0.0508	0.0309	18.14	0.0467	0.0142	8.60	0.0416	0.0075	0.01
$\hat{\lambda}_{A,Vas}$	-0.0924	0.2519	-86.89	-0.1704	0.2705	-75.83	-0.3471	0.3981	-50.77	-0.3364	0.6276	0.63
MLE-CIR												
$\hat{\kappa}_{B,CIR}$	0.1647	0.1822		0.4268	0.1862		0.6411	0.1588		0.7417	0.1444	
$\hat{\theta}_{B,CIR}$	0.1576	0.2999		0.02552	0.0643		0.0225	0.0368		0.0333	0.0409	
$\hat{\sigma}_{B,CIR}$	0.0183	0.0158		0.0106	0.0133		0.0192	0.0232		0.02688	0.0261	
$\hat{\lambda}_{B,CIR}$	-0.4378	0.4047		-0.7062	0.2813		-0.7475	0.2371		-0.7140	0.2713	
$\hat{\kappa}_{A,CIR}$	0.2747	0.0743		0.3551	0.2474		0.6076	0.2148		0.7327	0.2048	
$\hat{\theta}_{A,CIR}$	0.2198	0.4645		0.01299	0.1091		0.0257	0.0595		0.0053	0.0690	
$\hat{\sigma}_{A,CIR}$	0.0870	0.2125		0.0067	0.0311		0.0140	0.0317		0.0061	0.0397	
$\hat{\lambda}_{A,CIR}$	-0.4410	0.5566		-0.7769	0.4774		-0.9137	0.3890		-0.9061	0.4703	

The Vasicek-MLE (Eq. (2.24)) and the CIR-MLE (Eq. (2.50)) are applied to Vasicek-simulated 15-year zero bond yields whose parameter values $(\theta, \sigma, \lambda)$ are given in the first row. The value of κ has increased from 0.160, 0.400, 0.600, up to 0.900. The purpose of this study is to investigate the impact of the increment of κ on the parameter estimates. Then the bootstrap bias correction process is applied to each set of estimates. Each MLE process has 300 months of zero bond prices (*T*=25 years) with maturity of 15 years and $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. There are 10,000 replications. The number of bootstrap resampling is 500. The numbers in this table present the statistics of estimates based on the Vasicek-simulated data, with subscript "*B, Vas*" and "*A, Vas*" indicating before and after the bias corrected Vasicek model estimates, respectively, and with subscript "*B, CIR*" and "*A, CIR*" indicating before and after the bias corrected CIR model estimates, respectively. *S.D.* is the standard deviation. *R. Bias (%)* is the relative bias of the estimate.

Table 2.14 Models' relative predictive ability test when the value of κ increases

Vasicek Simulated Data with Parameter Values: θ=0.117, σ=0.043, λ=-0.705									
		κ=0.160		κ=0.400		κ=0.700		κ=0.900	
Model #1	Model #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	54.07	6.96	33.29	7.36	22.58	5.40	18.77	2.18
	CIR-B	73.10	61.98*	52.20	29.53	23.53	6.16	19.59	2.87
	CIR-A	65.52	16.74	45.29	20.69	21.16	6.17	18.08	2.78
	ND	98.76	94.88***	78.38	20.50	15.79	10.49	12.58	6.89
	AR(1)	99.99	99.96***	79.12	23.34	19.30	15.21	15.35	11.99
	Martingale	99.97	99.90***	100	99.82***	100	99.33***	99.98	99.91***
Vasicek-A	CIR-B	64.63	53.12*	44.43	22.67	20.83	15.92	17.22	12.65
	CIR-A	57.09	41.61	36.90	9.95	19.47	8.47	16.01	5.12
	ND	95.91	92.87***	75.75	20.13	15.63	10.35	12.50	6.91
	AR(1)	99.91	98.11***	79.36	46.88	20.06	13.46	16.19	9.78
	Martingale	99.92	99.91***	99.60	99.19***	100	99.69***	100	99.94***
CIR-B	CIR-A	62.32	4.40	41.93	14.85	21.01	7.10	17.47	3.88
	ND	98.72	96.85***	78.06	31.14	16.50	9.89	12.57	6.52
	AR(1)	99.93	99.73***	79.03	30.58	14.01	11.22	10.74	7.89
	Martingale	99.91	99.86***	100	99.48***	100	99.79***	100	99.93***
CIR-A	ND	99.08	97.16***	78.34	25.79	15.77	10.33	12.61	6.86
	AR(1)	99.92	99.88***	79.48	27.82	17.79	9.79	13.84	6.68
	Martingale	100.00	99.94***	99.37	99.25***	99.83	99.66***	99.94	99.94***

Seven alternative estimation techniques are applied to four settings of 15-year zero yields simulated from a Vasicek model. The corresponding sets of model estimates are used to make 12-month ahead forecasts. The parameter values $(\theta, \sigma, \lambda)$ are given in the first row. The value of κ has increased from 0.160, 0.400, 0.600, up to 0.900. The purpose of this study is to investigate the impact of the increment of κ on the models' predictive ability. There are 10,000 replications and the number of bootstrap resampling is 500. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model, "*AR(1)*" denotes the autoregressive model with the first order, and "*Martingale*" is the martingale model. The column "*% Paths #1 ≠ #2*" represents the proportion of 10,000 sample replications that the model #1 and the model #2 have NO equal predictive ability at the 5% significance level. The column "*% Paths #1 Beats #2*" stands for the proportion of 10,000 sample replications that the model #1 has greater predictive ability than the model #2 at the 5% significance level. Asterisks indicate that the proportion of sample replications that model #1 beats model #2 exceeds 50% (*), 70% (**), and 90% (***).

Table 2.15 Parameter estimates based on Vasicek- and CIR-simulated 6-month zero-coupon yields: Setting B when increasing the sample size

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/365, n=9,125							
	MLE-Vasicek				**MLE-CIR**		
	Before Bias Correction				Before Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{B,Vas}$	0.3415	0.2033	0.2804	$\hat{\kappa}_{B,CIR}$	0.3917	1.77	0.99
$\hat{\theta}_{B,Vas}$	0.09095	0.00391	0.0043	$\hat{\theta}_{B,CIR}$	0.10618	-	-
$\hat{\sigma}_{B,Vas}$	0.01822	0.00093	0.0011	$\hat{\sigma}_{B,CIR}$	0.06486	-	-
	After Bias Correction				After Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{A,Vas}$	0.2096	0.2134	0.2251	$\hat{\kappa}_{A,CIR}$	0.2286	3.03	0.99
$\hat{\theta}_{A,Vas}$	0.09190	0.04154	0.0427	$\hat{\theta}_{A,CIR}$	0.10483	-	-
$\hat{\sigma}_{A,Vas}$	0.01747	0.00109	0.0011	$\hat{\sigma}_{A,CIR}$	0.09940	-	-
Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/365, n=9,125							
	MLE-CIR				**MLE-Vasicek**		
	Before Bias Correction				Before Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{B,CIR}$	0.3584	0.14349	0.2913	$\hat{\kappa}_{B,Vas}$	0.3632	1.91	0.99
$\hat{\theta}_{B,CIR}$	0.09393	0.06484	0.0646	$\hat{\theta}_{B,Vas}$	0.09514	-	-
$\hat{\sigma}_{B,CIR}$	0.07278	0.00270	0.0035	$\hat{\sigma}_{B,Vas}$	0.02218	-	-
	After Bias Correction				After Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{A,CIR}$	0.2124	0.1527	0.2650	$\hat{\kappa}_{A,Vas}$	0.2287	3.03	0.99
$\hat{\theta}_{A,CIR}$	0.09283	0.06064	0.0610	$\hat{\theta}_{A,Vas}$	0.09658	-	-
$\hat{\sigma}_{A,CIR}$	0.07009	0.00526	0.0054	$\hat{\sigma}_{A,Vas}$	0.02130	-	-

The Vasicek-MLE (Eq. (2.24)) and the CIR-MLE (Eq. (2.50)) are applied to both Vasicek-simulated and CIR-simulated 6-month zero bond prices whose parameter values are given in the first row of Panel A and Panel B, respectively. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size increases from n = 300 to n' = 9,125. The latter corresponds to high frequency data. The bootstrap bias correction process is applied to each set of estimates. There are 10,000 replications. The number of bootstrap resampling is 500. Panel A presents the statistics of estimates based on the Vasicek-simulated data, with subscript "*B, Vas*" and "*A, Vas*" indicating before and after the bias corrected Vasicek model estimates, respectively, and with subscript "*B, CIR*" and "*A, CIR*" indicating before and after the bias corrected CIR model estimates, respectively. Panel B presents the corresponding statistics based on the CIR-simulated data. *S.D.* is the standard deviation. RMSE is the root mean square errors of the estimate. Half-Life is the defined as $-\ln 2/\kappa$. The autoregressive coefficient $\rho = e^{-\kappa\Delta t}$.

Table 2.16 Comparison of models out-of-sample forecasting performance: Setting B when increasing the sample size

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/365, n=9,125						
	1-month Forecasting		6-month Forecasting		12-month Forecasting	
	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Vasicek-B	0.0056	0.0039	0.0114	0.0081	0.0168	0.0121
Vasicek-A	0.0033	0.0026	0.0074	0.0058	0.0156	0.0088
CIR-B	0.0060	0.0044	0.0128	0.0101	0.0208	0.0162
CIR-A	0.0051	0.0033	0.0105	0.0070	0.0160	0.0103
ND	0.0269	0.0140	0.0349	0.0171	0.0360	0.0242
AR(1)	0.0324	0.0219	0.0359	0.0297	0.0432	0.0315
Martingale	0.2871	0.2324	0.6998	0.5611	0.9810	0.7854
Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/365, n=9,125						
	1-month Forecasting		6-month Forecasting		12-month Forecasting	
	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Vasicek-B	0.0051	0.0039	0.0112	0.0089	0.0159	0.0121
Vasicek-A	0.0044	0.0033	0.0100	0.0077	0.0141	0.0108
CIR-B	0.0048	0.0037	0.0108	0.0085	0.0155	0.0114
CIR-A	0.0038	0.0029	0.0089	0.0068	0.0124	0.0103
ND	0.0253	0.0137	0.0322	0.0188	0.0390	0.0299
AR(1)	0.0342	0.0219	0.0358	0.0297	0.0421	0.0315
Martingale	0.2872	0.2326	0.7003	0.5615	0.9807	0.7858

Seven alternative estimation techniques are applied to three settings of 6-month zero yields simulated from a Vasicek model (Panel A) and a CIR model (Panel B). The corresponding sets of model estimates are used to make out-of-sample forecasts. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. There are 10,000 replications and the number of bootstrap resampling is 500. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model. "*AR(1)*" denotes the autoregressive model with the first order. "*Martingale*" is the martingale model. Denote RMSE as the root mean square error of the forecasted zero bond yields and denote MAE as the mean of the absolute error of the forecasted values.

Table 2.17 Models' relative predictive ability test: Setting B when increasing the sample size

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/365, n=9,125							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	51.80	4.96	45.19	4.08	33.38	2.64
	CIR-B	52.84	6.24	48.60	4.60	35.32	3.15
	CIR-A	51.25	5.17	47.03	3.55	35.49	2.80
	ND	88.04	79.84**	79.52	65.68*	65.96	62.36*
	AR(1)	90.64	81.08**	84.20	66.52*	77.24	63.52*
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	37.13	19.98	35.94	19.95	28.85	15.96
	CIR-A	46.67	11.96	34.56	11.22	28.64	8.96
	ND	80.68	78.76**	71.44	65.12*	62.44	61.64*
	AR(1)	84.72	79.64**	74.92	65.24*	64.48	63.48*
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	39.94	14.82	39.16	13.09	32.28	12.16
	ND	85.92	78.36**	74.16	64.32*	54.08	33.08
	AR(1)	89.12	79.04**	77.24	65.16*	61.84	39.53
	Martingale	100	100***	100	100***	100	100***
CIR-A	ND	77.32	62.81*	55.60	37.47	36.32	23.29
	AR(1)	79.48	64.66*	67.44	45.48	58.00	37.00
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	76.00	59.00*	73.08	58.16*	67.88	56.48*
	Martingale	100	100***	100	100***	100	100***
AR(1)	Martingale	100	100***	100	100***	100	100***

Table 2.17 (Continued)

Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/365, n=9,125							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths of #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	46.44	18.06	43.72	16.14	31.74	11.70
	CIR-B	51.94	22.95	44.66	20.68	35.03	16.45
	CIR-A	50.03	18.48	42.19	17.50	32.44	12.16
	ND	79.56	65.03*	67.64	51.95*	50.96	30.58
	AR(1)	80.24	65.52*	70.36	55.04*	64.32	38.01
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	47.86	34.78	46.96	28.98	36.22	23.40
	CIR-A	45.27	20.16	44.28	19.60	33.32	13.86
	ND	80.88	72.86**	75.36	58.15*	55.32	31.42
	AR(1)	84.40	73.65**	76.64	58.29*	66.84	55.86*
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	50.76	23.90	47.11	21.15	38.64	15.20
	ND	78.68	72.52**	63.76	48.53	45.00	24.17
	AR(1)	80.52	73.73**	73.00	54.82*	53.92	28.95
	Martingale	100	100***	100	100***	100	100***
CIR-A	ND	85.32	74.50**	57.56	46.21	38.56	22.55
	AR(1)	86.80	75.86**	59.88	47.31	41.32	23.71
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	77.96	58.13*	63.80	38.28	53.56	27.38
	Martingale	100	100***	100	100***	100	100***
AR(1)	Martingale	100	100***	100	100***	100	100***

Seven alternative estimation techniques are applied to three settings of 6-month zero yields simulated from a Vasicek model (Panel A) and a CIR model (Panel B). The corresponding sets of model estimates are used to make 1- , 6-, and 12-month ahead forecasts. The corresponding sets of model estimates are used to make 1- , 6-, and 12-month ahead forecasts. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. There are 10,000 replications and the number of bootstrap resampling is 500. "*Vasicek- B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model, "*AR(1)*" denotes the autoregressive model with the first order, and "*Martingale*" is the martingale model. The column "*% Paths #1 ≠ #2*" represents the proportion of 10,000 sample replications that the model #1 and the model #2 have NO equal predictive ability at the 5% significance level. The column "*% Paths #1 Beats #2*" stands for the proportion of 10,000 sample replications that the model #1 has greater predictive ability than the model #2 at the 5% significance level. Asterisks indicate that the proportion of sample replications that model #1 beats model #2 exceeds 50% (*), 70% (**), and 90% (***).

Table 2.18 Parameter estimates based on Vasicek- and CIR-simulated 6-month zero-coupon yields: Setting C when reducing the time span

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/365, n=300							
	MLE-Vasicek				**MLE-CIR**		
	Before Bias Correction				Before Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{B,Vas}$	1.9362	0.3971	1.8395	$\hat{\kappa}_{B,CIR}$	2.1262	0.33	0.99
$\hat{\theta}_{B,Vas}$	0.09002	0.00397	0.00408	$\hat{\theta}_{B,CIR}$	0.09113	-	-
$\hat{\sigma}_{B,Vas}$	0.02397	0.00201	0.00695	$\hat{\sigma}_{B,CIR}$	0.06001	-	-
	After Bias Correction				After Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{A,Vas}$	1.0402	0.4074	0.9880	$\hat{\kappa}_{A,CIR}$	1.2515	0.55	0.99
$\hat{\theta}_{A,Vas}$	0.09013	0.00405	0.00417	$\hat{\theta}_{A,CIR}$	0.09110	-	-
$\hat{\sigma}_{A,Vas}$	0.02304	0.00202	0.00607	$\hat{\sigma}_{A,CIR}$	0.05993	-	-
Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/365, n=300							
	MLE-CIR				**MLE-Vasicek**		
	Before Bias Correction				Before Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{B,CIR}$	1.7352	0.3952	1.6356	$\hat{\kappa}_{B,Vas}$	2.0442	0.34	0.99
$\hat{\theta}_{B,CIR}$	0.09212	0.00386	0.00441	$\hat{\theta}_{B,Vas}$	0.09313	-	-
$\hat{\sigma}_{B,CIR}$	0.07194	0.00491	0.00506	$\hat{\sigma}_{B,Vas}$	0.02996	-	-
	After Bias Correction				After Bias Correction		
	Mean	S.D.	RMSE		Mean	Half-Life	ρ
$\hat{\kappa}_{A,CIR}$	0.9587	0.3931	0.9009	$\hat{\kappa}_{A,Vas}$	1.1482	0.60	0.99
$\hat{\theta}_{A,CIR}$	0.09210	0.00389	0.00442	$\hat{\theta}_{A,Vas}$	0.09302	-	-
$\hat{\sigma}_{A,CIR}$	0.07101	0.00501	0.00502	$\hat{\sigma}_{A,Vas}$	0.02707	-	-

The Vasicek-MLE (Eq. (2.24)) and the CIR-MLE (Eq. (2.50)) are applied to both Vasicek-simulated and CIR-simulated 6-month zero bond prices whose parameter values are given in the first row of Panel A and Panel B, respectively. Setting C: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$) while keeping the sample size n = 300. Correspondingly, the time span is reduced from $T = 25$ years to $T' = n \times \Delta t' = 300 \times \frac{1}{365} = 0.82$ years. The bootstrap bias correction process is applied to each set of estimates. There are 10,000 replications. The number of bootstrap resampling is 500. Panel A presents the statistics of estimates based on the Vasicek-simulated data, with subscript "*B, Vas*" and "*A, Vas*" indicating before and after the bias corrected Vasicek model estimates, respectively, and with subscript "*B, CIR*" and "*A, CIR*" indicating before and after the bias corrected CIR model estimates, respectively. Panel B presents the corresponding statistics based on the CIR-simulated data. *S.D.* is the standard deviation. RMSE is the root mean square errors of the estimate. Half-Life is the defined as $-\ln 2/\kappa$. The autoregressive coefficient $\rho = e^{-\kappa \Delta t}$.

Table 2.19 Comparison of models out-of-sample forecasting performance: Setting C when reducing the time span

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/365, n=300

	1-month Forecasting		6-month Forecasting		12-month Forecasting	
	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Vasicek-B	0.0072	0.0063	0.0151	0.0129	0.0295	0.0243
Vasicek-A	0.0052	0.0047	0.0132	0.0106	0.0180	0.0143
CIR-B	0.0148	0.0107	0.0257	0.0234	0.0299	0.0257
CIR-A	0.0055	0.0053	0.0138	0.0109	0.0185	0.0147
ND	0.0441	0.0202	0.0633	0.0314	0.0666	0.0448
AR(1)	0.0552	0.0438	0.0650	0.0528	0.0712	0.0572
Martingale	0.2861	0.2281	0.7032	0.5623	0.9843	0.7869

Panel B: CIR Simulated Data with Parameter Values: κ=0.148, θ=0.09, σ=0.0707, Δt=1/365, n=300

	1-month Forecasting		6-month Forecasting		12-month Forecasting	
	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Vasicek-B	0.0066	0.0059	0.0188	0.0153	0.0232	0.0189
Vasicek-A	0.0065	0.0051	0.0157	0.0124	0.0214	0.0157
CIR-B	0.0065	0.0056	0.0180	0.0146	0.0229	0.0183
CIR-A	0.0062	0.0048	0.0153	0.0119	0.0209	0.0141
ND	0.0239	0.0129	0.0303	0.0234	0.0409	0.0310
AR(1)	0.0263	0.0150	0.0318	0.0273	0.0413	0.0341
Martingale	0.2863	0.2283	0.7037	0.5625	0.9847	0.7872

Seven alternative estimation techniques are applied to three settings of 6-month zero yields simulated from a Vasicek model (Panel A) and a CIR model (Panel B). The corresponding sets of model estimates are used to make out-of-sample forecasts. Setting C: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$) while keeping the sample size n = 300. Correspondingly, the time span is reduced from $T = 25$ years to $T' = n \times \Delta t' = 300 \times \frac{1}{365} = 0.82$ years. There are 10,000 replications and the number of bootstrap resampling is 500. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model. "*AR(1)*" denotes the autoregressive model with the first order. "*Martingale*" is the martingale model. Denote RMSE as the root mean square error of the forecasted zero bond yields and denote MAE as the mean of the absolute error of the forecasted values.

Table 2.20 Models' relative predictive ability test: Setting C when reducing the time span

Panel A: Vasicek Simulated Data with Parameter Values: κ=0.140, θ=0.0891, σ=0.0173, Δt=1/365, n=300							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	86.92	36.72	76.84	34.88	47.96	30.96
	CIR-B	81.76	54.56*	71.56	51.80*	52.00	48.07
	CIR-A	78.52	43.52	68.84	37.00	59.04	32.00
	ND	99.76	99.11***	99.12	98.34***	99.10	98.08***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	77.60	60.92*	67.36	55.36*	39.12	20.78
	CIR-A	68.88	53.44*	57.72	46.84*	37.32	19.90
	ND	100	100***	99.96	99.95***	99.91	99.90***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	78.76	46.04	58.04	44.76	39.36	28.92
	ND	99.99	99.96***	99.98	99.94***	99.97	99.91***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
CIR-A	ND	99.99	99.96***	99.90	99.75***	99.85	98.99***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	37.72	22.11	18.29	11.70	16.37	10.88
	Martingale	100	100***	100	100***	100	100***
AR(1)	Martingale	100	100***	100	100***	100	100***

Table 2.20 (Continued)

Panel B: CIR Simulated Data with Parameter Values: *κ*=0.148, *θ*=0.09, *σ*=0.0707, *Δt*=1/365, *n*=300							
		1-month Forecasting		6-month Forecasting		12-month Forecasting	
Model #1	Model #2	% Paths of #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2	% Paths #1≠ #2	% Paths #1Beats #2
Vasicek-B	Vasicek-A	78.92	40.24	68.28	36.85	56.08	31.94
	CIR-B	80.60	46.17	70.36	44.84	59.12	41.50
	CIR-A	78.64	35.49	66.12	32.26	45.32	28.25
	ND	99.88	99.20***	99.86	98.01***	99.60	97.88***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
Vasicek-A	CIR-B	78.76	61.74*	68.00	43.96	55.36	30.95
	CIR-A	76.08	44.89	67.56	30.56	51.68	22.88
	ND	100	100***	100	100***	100	100***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
CIR-B	CIR-A	77.08	34.36	56.64	30.53	34.16	22.87
	ND	100	100***	100	100***	100	100***
	AR(1)	100	100***	100	100***	100	100***
	Martingale	100	100***	100	100***	100	100***
CIR-A	ND	99.96	99.96***	99.90	99.80***	98.89	98.00***
	AR(1)	99.32	98.80***	99.04	98.21***	99.00	98.12***
	Martingale	100	100***	100	100***	100	100***
ND	AR(1)	38.62	25.08	18.11	10.44	17.38	10.03
	Martingale	100	100***	100	100***	100	100***
AR(1)	Martingale	100	100***	100	100***	100	100***

Seven alternative estimation techniques are applied to three settings of 6-month zero yields simulated from a Vasicek model (Panel A) and a CIR model (Panel B). The corresponding sets of model estimates are used to make 1- , 6-, and 12-month ahead forecasts. The corresponding sets of model estimates are used to make 1- , 6-, and 12-month ahead forecasts. Setting C: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$) while keeping the sample size $n = 300$. Correspondingly, the time span is reduced from $T = 25$ years to $T' = n \times \Delta t' = 300 \times \frac{1}{365} =$ 0.82 years. There are 10,000 replications and the number of bootstrap resampling is 500. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. "*ND*" stands for the Non-linear diffusion model, "*AR(1)*" denotes the autoregressive model with the first order, and "*Martingale*" is the martingale model. The column "*% Paths #1 ≠ #2*" represents the proportion of 10,000 sample replications that the model #1 and the model #2 have NO equal predictive ability at the 5% significance level. The column "*% Paths #1 Beats #2*" stands for the proportion of 10,000 sample replications that the model #1 has greater predictive ability than the model #2 at the 5% significance level. Asterisks indicate that the proportion of sample replications that model #1 beats model #2 exceeds 50% (*), 70% (**), and 90% (***).

2.6 References

Andrew, D.W.K., 1993, Exactly median-unbiased estimation of first order autoregressive/unit root models, Econometrica 61, 139-166.

Ait-Sahalia, Yacine, 1996, Nonparametric pricing of interest rate derivative securities, Econometrica 64, 527-560.

Arnold, Ludwig, 1974, Stochastic differential equations: Theory and applications (Wiley, New York).

Bauer, M.D., G.D. Rudebusch, and J.C. Wu, 2012, Correcting estimation bias in dynamics term structure models, Journal of Business and Economic Statistics 30(3): 454-467.

Brigo, Damiano and Fabio Mercurio, 2006, Interest rate models -theory and practice with smile, inflation and credit (Springer).

Broadie, M. and Ozgur Kaya, 2006, Exact simulation of stochastic volatility and other affine jump diffusion processes, Operations Research 54, 217-231.

Chan K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, Journal of Finance 47(3), 1209-1227.

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

Dixit, A.K. and Pindyck R.S., 1994, Investment under uncertainty (Princeton University Press, Princeton, New Jersey).

Duffie, D. and R. Kan, 1996, A yield-factor model of interest rates, Mathematical Finance 6, 379-406.

Giacomini, R. and H. White, 2006, Tests of coditional predictive ability, Econometrica 74, 1545-1578.

Pearson, N.D. and T.S. Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the cox, ingersoll, and ross model, Journal of Finance 49, 1279-1304.

Phillips, P.C., 1972, The structural estimation of a stochastic differential equation system, Econometrica 40, 1021-1041.

Phillips, P.C. and J. Yu, 2009, Maximum likelihood and gaussian estimation of continuous time models in finance, Handbook of Financial Time Series 497-530.

Smith, A.A., 1993, Estimating nonlinear time-series models using simulated vector autoregressions, Journal of Applied Econometrics 8, S63-S84.

Tang, C.Y and S.X. Chen, 2009, Parameter estimation and bias correction for diffusion processes, Journal of Econometrics 149, 65-81.

Vasicek, Oldrich A., 1977, An equilibrium characterization of the term structure, Journal of Financial Economics 5, 177-188.

Yu, Jun and P.C. Phillips, 2001, A gaussian approach for estimating continuous time models of short term interest rates, The Econometrics Journal 4, 211-225.

Yu, Jun, 2012, Bias in the estimation of the mean reversion parameter in continous time models, Journal of Econometrics 169, 114-122.

CHAPTER III

THE VALUE OF KNOWING THE TRUTH AND THE COST OF BEING WRONG: AN APPLICATION TO MULTI-FACTOR AFFINE TERM STRUCTURE MODELS

3.1 Introduction

The objective of my dissertation is to discuss under what circumstances that we can distinguish between different ATSMs, given the data are from a particular type of ATSMs and observed without errors. I focus on the discussion of one-factor ATSMs in Chapter II. Those results show that if the data are simulated from a high mean reversion process with a large sample size and with a high sampling frequency, the models' forecasting accuracy can be improved, in contrast, the power to distinguish between different ATSMs will be reduced. In this chapter I extend the study of the question to the multi-factor ATSMs and examine whether the results based on the one-factor models can be applied to the multi-factor models. In particular, I will study under what circumstances we can distinguish between different multi-factor ATSMs. Once again I will focus on the power and size of the out-of-sample forecasts using Giacomini and White's test. For instance, if the data is generated by a two factor CIR model, is it more likely the data will reject a one factor CIR model or a two factor Vasicek model? Similarly, how frequently will we reject the two factor CIR model as the correct model? To simplify the data simulation process and focus on the questions of interest, I use the two-factor Vasicek and CIR models as a representation of the multi-factor affine models without any loss of generality.

In this chapter I use the same Monte Carlo study as that in the last chapter. That is, I simulate the data from both the 2-factor Vasicek and CIR models, and then estimate all models using that data, including the 1- and 2-factor Vasicek and CIR models. Same as the last chapter, I employ the bootstrap bias correction process to reduce the bias of the mean reversion estimates. This means there are two sets of parameter estimates for each ATSM. In the following sections, I use "*B*" to stand for the estimates before the bias correction and "*A*" to represent the estimates after the bias correction. Last, I compare the models' performance based on the out-of-sample forecasts over a 12-month horizon. The Giacomini and White (2006) test is employed to examine the models' relative predictive ability as well as the model's power to statistically distinguish between different ATSMs. The model's forecasting accuracy is measured by the mean squared error (MSE). The standardized difference of two models' MSEs measures the model's power to statistically distinguish between different ATSMs. The larger the absolute value of the difference of two models' MSEs is, the stronger the power to distinguish one model from another one is. The smaller MSE represents more accurate forecasts and the higher relative predictive ability as well.

In terms of the model estimation method, I follow Pearson and Sun (1994) to derive the likelihood functions for the 1- and 2-factor Vasicek and CIR models and then apply the maximum likelihood estimation to all models. Since two series of the bond prices with different maturities are needed to estimate the 2-factor ATSMs, I use the two simulated factors to construct three groups of bond prices with each group consisting of two bond series with two maturities. Based on the length difference of the time to maturities, three groups of the bond series are 2 shorts: 0.5 and 1 year; 2 mediums: 1 and 20 years; and 2 longs: 20 and 30 years. Such a category is for the study of the influence of the time to maturities on the models' predictive ability.

The basic simulation setting (Base Setting) is the same as the initial setting in Chapter II, except for the true parameters, which are point estimates from fitting the 2-factor Vasicek and CIR models to the month-end price quotes for treasury bonds. Comparing the out-of-sample forecasts of the 1- and 2-factor Vasicek and CIR models, I find that relative to the one-factor models, the multi-factor models exhibit little improvement in the models' predictive ability as well as the models' power to differentiate themselves from the one-factor models. The 2-factor counterparts of the true models have the least predictive ability and are the most distinguishable from other 1- or 2-factor models. For example, given the 2-factor Vasicek model is the true model, the probability that the 2-factor Vasicek model can be differentiated from 2-factor CIR model at the 12-month ahead horizon is over 99%, while the probability to distinguish the true model from other 1-factor ATSMs is about 50% and the number is less than 30% in differentiating between two 1-factor ATSMs. This means that having more factors in the ATSMs increases models' power to distinguish between models. In addition, the MSEs of the out-of-sample forecasts of three groups of the bond series have no much difference. This implies that the time to maturities of the bond have no much impact on the models' out-of-sample performance.

The conclusion in Chapter II states that the sample size and the time span of the sample can influence the one-factor ATSMs' predictive ability and models' power to distinguish between the one-factor models as well. In this chapter, I want to investigate whether the sample size and the sample's time span have the similar impact on multi-factor ATSMs. Therefore, I simulate the data using the same two alternative settings as those in the last chapter and then repeat the estimation, forecasting, and testing processes. The two alternative settings are Setting B and Setting C. The former is in correspondence with the increase of the sample size by the way of the increase in the sampling frequency within a fixed time span. Setting C refers to the shorter time span by the way of an increase in the sampling frequency within a fixed sample size.

The results based on the simulated data from Settings B and C show the similar results as those from the Base Setting. That is, adding more factors does not benefit the ATSMs in improving models' predictive accuracy. But having more factors in the models increases the models' power to distinguish the one from another. If the sample size rises, the out-of-sample performance of all models' are generally improved, and the 1-factors models cluster together and are difficult to differentiate between each other. The probability to differentiate between two one-factor ATSMs drops to below 10% in the case of Setting B. However, the two-factor models can be easily distinguishable from the one-factor models. The results based on Setting C show that the shorter time span, by the way of the increase in the sampling frequency, makes the one-factor models more easily distinguishable. The two-factor models, however, become less distinguishable from the 1-factor models in the case of Setting C.

In summary, having more factors in the ATSMs does not improve models' predictive ability. But it increases the models' power to distinguish between each other.

The multi-factor ATSMs with larger sample size and longer time span will have more predictive ability and stronger power to differentiate between models.

The rest of this chapter is organized as follows: Section 3.2 describes the 2-factor Vasicek and CIR models and then introduce the estimation methods. The detail data simulation procedure is stated in Section 3.3. Section 3.4 examines the models' out-of-sample performance and investigates the power and size of the two-factor ATSMs as well. Section 3.5 concludes the paper.

3.2 Model description and estimation methods

In this section, I first introduce the two-factor Vasicek and CIR models, followed by the description of the method of the maximum likelihood estimation of those two models.

To simplify the data simulation process and focus on the question of interest, I assume the state factors are independent. Accordingly, the dynamics of two-factor Vasicek model under the physical measure *P* are:

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix} \begin{bmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} dW_{1t}^P \\ dW_{2t}^P \end{bmatrix} \tag{3.1}$$

The dynamics of two-factor CIR model under the physical measure *P* are

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix} \begin{bmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_1\sqrt{X_{1t}} & 0 \\ 0 & \sigma_2\sqrt{X_{2t}} \end{bmatrix} \begin{bmatrix} dW_{1t}^P \\ dW_{2t}^P \end{bmatrix} \tag{3.2}$$

where the parameters are $(\kappa_1, \theta_1, \sigma_1, \kappa_2, \theta_2, \sigma_2)$, (X_{1t}, X_{2t}) are two state factors, and (W_{1t}^P, W_{2t}^P) are Wiener processes under the physical measure *P*.

In order to price the bonds based on the state factors, we have to identify the price of risk. This paper assumes the price of risk is completed. This means the price of risk (λ_1, λ_2) are two constants. Therefore, after adjusting the interest rate risk, the dynamics of two models under the equivalent martingale measure Q have the expressions:

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix} \begin{bmatrix} \theta_1 - \frac{\sigma_1 \lambda_1}{\kappa_1} - X_{1t} \\ \theta_2 - \frac{\sigma_2 \lambda_2}{\kappa_2} - X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} dW_{1t}^Q \\ dW_{2t}^Q \end{bmatrix} \tag{3.3}$$

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix} \begin{bmatrix} \theta_1 - \frac{\kappa_1 + \lambda_1}{\kappa_1} X_{1t} \\ \theta_2 - \frac{\kappa_2 + \lambda_2}{\kappa_2} X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_1 \sqrt{X_{1t}} & 0 \\ 0 & \sigma_2 \sqrt{X_{2t}} \end{bmatrix} \begin{bmatrix} dW_{1t}^Q \\ dW_{2t}^Q \end{bmatrix} \tag{3.4}$$

Based on the dynamics of the state factors, we can explicitly express the zero bond prices as

$$P_t(\tau) = e^{\sum_{i=1}^{2}[A_i(\tau) - B_i(\tau) X_{it}]} \tag{3.5}$$

with the Vasicek model's

$$A_i(\tau) = \left(\theta_i - \frac{\sigma_i \lambda_i}{\kappa_i} - \frac{\sigma_i^2}{2\kappa_i^2}\right)(B_i(\tau) - \tau) - \frac{\sigma_i^2 B_i^2(\tau)}{4\kappa_i} \tag{3.6}$$

$$B_i(\tau) = \frac{1}{\kappa_i}(1 - e^{-\kappa_i \tau}) \tag{3.7}$$

and the CIR model's

$$A_i(\tau) = \frac{2\kappa_i \theta_i}{\sigma_i^2} ln\left[\frac{2\gamma_i e^{\frac{\tau}{2}(\kappa_i + \lambda_i + \gamma_i)}}{(\kappa_i + \lambda_i + \gamma_i)(e^{\gamma_i \tau} - 1) + 2\gamma_i}\right] \tag{3.8}$$

$$B_i(\tau) = \frac{2(\mathrm{e}^{\gamma_i \tau} - 1)}{(\kappa_i + \lambda_i + \gamma_i)(\mathrm{e}^{\gamma_i \tau} - 1) + 2\gamma_i} \tag{3.9}$$

$$\gamma_i = \sqrt{(\kappa_i + \lambda_i)^2 + 2\sigma_i^2} \tag{3.10}$$

Since we know the transition density functions of one-factor Vasicek and CIR models, it is easy to extend them to the two-factor models when we assume the two factors are independent. Still, we use the method of Pearson and Sun (2004) to map the state factors to bond prices and then derive the price density functions.

To derive the transition density functions of the bond prices of the 2-factor ATSMs, we need at least two series of bond data with different maturities, among which two series of bond prices are assumed to be no measurement errors and the rest can have errors. In this chapter we do not consider any measurement errors in order to focus on the questions of interest. So only two series of bond prices without errors are considered.

Denote $P_t(\tau_1)$ and $P_t(\tau_2)$ as two bond prices with maturities τ_1 and τ_2, respectively. Then we can derive the Jacobian as

$$J = \begin{vmatrix} \frac{\partial X_{1t}}{\partial P_t(\tau_1)} & \frac{\partial X_{1t}}{\partial P_t(\tau_2)} \\ \frac{\partial X_{2t}}{\partial P_t(\tau_1)} & \frac{\partial X_{2t}}{\partial P_t(\tau_2)} \end{vmatrix}$$

$$= \frac{1}{[B_1(\tau_1)B_2(\tau_2) - B_2(\tau_1)B_1(\tau_2)]^2} \begin{vmatrix} \frac{-B_2(\tau_2)}{P_t(\tau_1)} & \frac{B_2(\tau_1)}{P_t(\tau_2)} \\ \frac{B_1(\tau_2)}{P_t(\tau_1)} & \frac{-B_1(\tau_1)}{P_t(\tau_2)} \end{vmatrix}$$

$$= \frac{1}{[B_1(\tau_1)B_2(\tau_2) - B_2(\tau_1)B_1(\tau_2)]P_t(\tau_1)P_t(\tau_2)} \qquad (3.11)$$

where the definitions of $B_i(\tau)$ and $A_i(\tau)$ of Vasicek and CIR models are referred to Eq. (3.5).

Therefore, the bond price transition density functions of 2-factor ATSMs are obtained by multiplying the transition density of the state factors with the absolute value of Jacobian. Denote the density functions of the independent state vector as $h_V(X_t|X_{t-1})$

for the Vasicek model and $h_{CIR}(X_t|X_{t-1})$ for the CIR model. The density functions of bond prices of 2-factor ATSMs are:

2-factor Vasicek model:

$$h_V\big(P_t(\tau_1), P_t(\tau_2)\big|P_{t-1}(\tau_1), P_{t-1}(\tau_2)\big)$$

$$= |J| \cdot h_V\big(X_{1t}, X_{2t}\big|X_{1,t-1}, X_{2,t-1}\big)$$

$$= \frac{\prod_{i=1}^{2} h_V(X_{i,t}|X_{i,t-1})}{|B_1(\tau_1)B_2(\tau_2) - B_2(\tau_1)B_1(\tau_2)|P_t(\tau_1)P_t(\tau_2)} \tag{3.12}$$

2-factor CIR model:

$$h_{CIR}\big(P_t(\tau_1), P_t(\tau_2)\big|P_{t-1}(\tau_1), P_{t-1}(\tau_2)\big)$$

$$= \frac{\prod_{i=1}^{2} h_{CIR}(X_{i,t}|X_{i,t-1})}{|B_1(\tau_1)B_2(\tau_2) - B_2(\tau_1)B_1(\tau_2)|P_t(\tau_1)P_t(\tau_2)} \tag{3.13}$$

By dividing the time interval [0, *T*] into *n* subintervals of equal width, we have *n* time knots $t = 1, 2, \dots, n$. If denoting Δt as the time increment, the time span is $T = n\Delta t$. Assume the zero bond prices at each time knot are $P_1, P_2, \cdots, P_n$. The log-likelihood functions of 2-factor Vasicek and CIR models are

$$\text{Vasicek: } \mathcal{L}(\boldsymbol{\alpha}|\boldsymbol{P}(\tau_1), \boldsymbol{P}(\tau_2)) = \ln \prod_{t=2}^{n} h_V\big(P_t(\tau_1), P_t(\tau_2)\big|P_{t-1}(\tau_1), P_{t-1}(\tau_2)\big)$$

$$= \sum_{t=2}^{n} \ln h_V\big(P_t(\tau_1), P_t(\tau_2)\big|P_{t-1}(\tau_1), P_{t-1}(\tau_2)\big) \tag{3.14}$$

$$\text{CIR: } \mathcal{L}\big(\boldsymbol{\alpha}\big|\boldsymbol{P}(\tau_1), \boldsymbol{P}(\tau_2)\big) = \sum_{t=2}^{n} \ln h_{CIR}\big(P_t(\tau_1), P_t(\tau_2)\big|P_{t-1}(\tau_1), P_{t-1}(\tau_2)\big) \tag{3.15}$$

where

$$\boldsymbol{\alpha} = (\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2) \tag{3.16}$$

and

$$\boldsymbol{P}(\tau_k) = (P_1(\tau_k), P_2(\tau_k), \cdots, P_n(\tau_k))', k = 1, 2. \tag{3.17}$$

Once getting the parameter estimates, $\widehat{\boldsymbol{\alpha}} = (\hat{\kappa}_1, \hat{\theta}_1, \hat{\sigma}_1, \hat{\lambda}_1, \hat{\kappa}_2, \hat{\theta}_2, \hat{\sigma}_2, \hat{\lambda}_2)$, I can forecast the zero bond yields. The *h*-step ahead zero bond yields forecasts can be expressed as:

$$E\left(X_{j,t+h}|X_{j,t}\right) = \hat{\theta}_j\left(1 - e^{-\hat{\kappa}_j h\Delta t}\right) + e^{-\hat{\kappa}_j h\Delta t} X_{j,t} \tag{3.18}$$

$$\hat{y}_{t+h}(\tau) = -\frac{\sum_{j=1}^{2}\left[\hat{A}_j(\tau) - \hat{B}_j(\tau)\hat{X}_{j,t+h}\right]}{\tau} \tag{3.19}$$

3.3 Data simulation procedure

Since the state factors are assumed to be uncorrelated and each model is an independent one-factor ATSM, I can simulate two uncorrelated series of state factors, with each one based on the data generating process of the one-factor ATSM. The detail simulation processes of one-factor Vasicek and CIR models are described in Chapter II.

The true parameters used for the 2-factor Vasicek model are $(\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2)$ = (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045) and the parameters for the 2-factor CIR model are $(\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2)$ = (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Those true parameters are the estimates of fitting the models to the month-end price quotes for treasury issues, which are taken from the CRSP Government Bond files. The data cover the time span from January 1970 to December 1995. Callable bonds and bills under one month to maturity are excluded from the sample.

We know that two series of the bond prices with different maturities are needed to estimate the 2-factor ATSMs. So I simulate three groups of the bond data from the two-factor ATSMs, with each group consisting of two bond series with different maturities. Based on the length difference of the time to maturities, three groups of the bond series are 2 shorts: 0.5 and 1 year, 2 mediums: 1 and 20 years, and 2 longs: 20 and 30 years. Such a category is for the study of the influence of the time to maturities on the models' predictive ability.

The basic simulation setting (Base Setting) is that each series of the simulated data contains 300 months of zero bond yields (T=25 years). $\Delta t = \frac{1}{12}$ corresponds to the monthly observations in an annualized basis. The Monte Carlo simulation repeats 20,000 times. The bootstrap bias correction process is based on 500 resamples. The time horizons of the out-of-sample forecasts are 1 through 12 months.

The conclusion in Chapter II states that the sample size and the time span of the sample can influence the one-factor ATSMs' predictive ability and the power to distinguish between the one-factor models as well. In this chapter, I want to investigate whether the sample size and the sample's time span have the similar impact on multi-factor ATSMs. Therefore, I simulate the data using two same alternative settings as those in Chapter II and then repeat the estimation, forecasting and testing processes. The two alternative settings are Setting B in correspondence with the increase of the sample size and Setting C that refers to the shorter time span. In detail,

Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data.

Setting C: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$) while keeping the sample size n = 300.

Correspondingly, the time span is reduced from $T = 25$ years to $T' = n \times \Delta t' = 300 \times \frac{1}{365} = 0.82$ years.

3.4 Results

This paper studies the 2-factor Vasicek and CIR models. The main purpose is to investigate whether the 2-factor ATSMs can be differentiated from each other and whether the 2-factor ATSMs can be distinguished from the one-factor ATSMs, given the data are simulated from 2-factor models. In terms of the bias of the parameter estimates, I apply the bootstrap bias correction process to each ATSM.

There are four subsections in this section. Section 3.4.1 states the results of the model estimation. Section 3.4.2 shows the models' out-of-sample forecast performance and Section 2.4.3 reports the results on the evaluation of the models' forecasts. The last subsection discusses the power and size of the models' to distinguish between different ATSMs.

3.4.1 Model estimation results

Table 3.1 shows the statistics of the parameters of the true models. Panel A provides the summary of the estimates from the two-factor Vasicek model and Panel B is the corresponding results based on the data simulated from the two-factor CIR model. In each panel, the first column shows the true parameter values used to generate the data, followed by the estimates based on different simulation settings. Since I apply the bootstrap bias correction process to each ATSM, there are two sets of estimates for each group of the bond series, with the values before the bias correction on the upper part of each panel and the values after the bias correction on the lower part.

Check the estimates in Panel A, we can see that all estimates are close to the true values, except for the mean reversion estimates. Before running the bias correction, the biases of $(\hat{\kappa}_1, \hat{\kappa}_2)$ are about 0.053 and 0.011, respectively, and the relative biases of $(\hat{\kappa}_1, \hat{\kappa}_2)$ are 11.21% and 25.58%, respectively. After the bias correction process, the biases and the relative biases of $(\hat{\kappa}_1, \hat{\kappa}_2)$ are reduced to (0.011, 0.003) and (2.33%, 6.98%), respectively. The similar results can be found in Panel B. The biases of the mean reversion estimates are (0.015, 0.057) before the bias correction and (0.005, 0.002) after the bias correction. Accordingly, the relative biases of $(\hat{\kappa}_1, \hat{\kappa}_2)$ are (2.29%, 105.56%) before the bias correction and (0.76%, 3.70%) after the bias correction. This result implies that the bias of the mean reversion estimates exhibits by the maximum likelihood estimation of two-factor models. The bootstrap bias correction process works well in the reduction of the mean reversion estimate biases.

Except for the estimation of the true models, I also estimate those non-data-generating models using the simulated data. Those estimates report in Table 3.2 and Table 3.3. In Tables 3.2 and 3.3, the estimates are not reported directly but showed with the values of the long-run yield, the half-life of the shock, and the unconditional standard deviation. The aim of such a report is to help the readers better compare the different estimates because there are both one- and two-factor Vasicek and CIR models and the parameters of all models are not in the same scale. Also, those numbers in Tables 3.2 and 3.3 can reveal the readers the economic interpretations of the estimates of the models'. The long-run yield is the bond yield when the time approaches infinity. The half-life of the shock to the interest rate is defined as the $-\ln 0.5/\kappa$. The larger the speed of the mean reversion is, the shorter the half-life of a shock is, then the faster the model will reach its long-run mean. The unconditional standard deviation is the σ in the Vasicek model and is calculated as $\sigma\sqrt{\frac{\theta}{2\kappa}}$ in the CIR model.

In Table 3.2, Panel A reports the estimates of 1-year bond-yield data from the two-factor Vasicek model and Panel B are the estimates from the two-factor CIR model. For each two-factor model, I report two values of the half-life of the shock and two unconditional standard deviations. From the Table 3.2, we can see that the estimates of the data-generating models are very close to the true values. This is consistent with the results in Table 3.1. Another finding is that the estimates of the one-factor models are almost the same if the simulation setting is the same, no matter what the group of the bond data is. For example, in Panel A, if we compare the values of the one-factor Vasicek and CIR models in '*2 Shorts*' with the corresponding values in '*2 Mediums*' under Base Setting, we can find that the long-run yield, the half-life of the shock, and the unconditional standard deviation of the same models in the different group of the bond data are almost the same. This result can be found in Panel B. This implies that the one-factor models are not easily to be differentiated from each other once the data are simulated from a two-factor model. Such a result also implies that the time to maturity of the bond-yield data have no impact on the estimates of the one-factor models. Table 3.3 reports the estimates of 20-year bond-yield data from the two-factor Vasicek model (Panel A) and from the two-factor CIR model (Panel B). The results in Table 3.3 are the same as in Table 3.2. This has again proved that the time to maturity has no influence on the estimation of the ATSMs.

3.4.2 Models' out-of sample forecasts

To compare the predictive ability of the models, I construct one-year-ahead bond yield forecasts and compute the RMSEs and MAEs (mean absolute errors) of the forecasts, which are showed in Tables 3.4 and 3.5. Panel A reports the RMSEs and MAEs of the forecasts when the data are simulated from the two-factor Vasicek model and Panel B are the results when the data are simulated from the two-factor CIR model. "Vasicek-B" and "Vasicek-A" represent the sets of MLE-Vasicek estimates, with "B" indicating before the bias correction and "A" indicating after the bias correction. Similarly, "CIR-B" and "CIR-A" are the corresponding sets of estimates for the MLE-CIR.

The results in Table 3.4 show that for all models, the RMSEs and MAEs of the bond yield forecasts are increasing as the forecasting horizon increases from 1 month to 6 months, and up to 12 months. In Panel A, although the two-factor Vasicek models are the data-generating models, they do not have better out-of-sample performance than other non-data-generating models. For example, if we compare the RMSEs and MAEs of all models under Base Setting in Panel A, we can see that the RMSEs and MAEs of the two-factor Vasicek model are very close to the values of the one-factor Vasicek and CIR models. As the forecasting horizon becomes longer, the difference of the RMSEs of two models becomes larger. Also, the results show that the time to maturities have no impact on the out-of-sample performance of the forecasts. An interesting finding in Panel A is that the two-factor CIR models have relatively worst forecast performance. This shows that adding more factors in ATSMs does not improve the models predictive ability, even though it is the true model. If the models are the wrong models, adding more factors reduces models' out-of-sample performance. This same result can be found in Panel B. Table 3.5 reports the similar results but for the 20-year bond yield data. The results in Table 3.5 have no much big difference from those in Table 3.4. So the time to maturity does not impact the ATSMs' performance.

3.4.3 Evaluation of out-of-sample forecasting performance

To observe the models out-of-sample performance straightforward, I draw graphs based on RMSEs of the one-year-ahead bond yield forecasts. Figures 3.1 and 3.2 show the forecasts of 1-year and 20-year bond yields, respectively, when the data are simulated from two-factor Vasicek model. Figures 3.3 and 3.4 exhibit the forecasts of 1-year and 20-year bond yields, respectively, while the data are from two-factor CIR model. Each figure have six charts. The charts are arranged by columns. They are results based on Base Setting, Setting B and Setting C from left to right. In each column, the chart on the top shows the results with a group of bond data with shorter maturities and the chart on the bottom is the results with a group of bond data with longer maturities. The results of Giacomini and White tests are reported on Tables 3.6 and 3.7. The numbers in those two tables are the percentage of sample paths that two competing models (Model #1 and Model #2) are not statistically equal at a 12-month ahead horizon. Panel A shows the testing results when the data are from the two-factor Vasicek model and Panel B are results based on two-factor CIR model.

Checking the values in both Panels A of Tables 3.6 and 3.7, I find that the two-factor CIR models, both before and after the bias correction, are the most distinguishable models. The probability to differentiate between two-factor Vasicek models and two-factor CIR models is over 99%. In contrast, the possible differentiation between two-factor Vasicek models and one-factor ATSMs is below 60%. Furthermore, the number to distinguish a one-factor model from another one-factor model drops below 30%. Those numbers tell us that ATSMs with multi-factors do not improve models' relative predictive ability, instead making the wrong models more distinguishable. This result is showed in the figures. In Figure 3.1 and Figure 3.2, we see that although the two-factor Vasicek models are the data-generating models, two-factor Vasicek-A and two-factor Vasicek-B do not have better forecast performance than the one-factor Vasicek or CIR models. This result shows that relative to the one-factor models, the two-factor models have no much advantage to improve the models' predictive ability as well as to increase the models' power to differentiate themselves from the one-factor models. In addition, we can see that two-factor CIR-A and two-factor CIR-B can be easily distinguished from other models. They do the worst in the out-of-sample forecasts among all eight models. This implies that having more factors in the ATSMs increases models' power to distinguish between different models. If we check the RMSEs of the out-of-sample forecasts of three groups of the bond series, we find there are no much difference among three groups. This implies that the time to maturities of the bond have no much impact on the models' out-of-sample performance.

Figures 3.3 and 3.4 show the out-of-sample forecasts when the data are simulated from the two-factor CIR model. The corresponding testing results are in Panels B in Tables 3.6 and 3.7. We get the similar results as those in Panels A of the two tables. The two-factor Vasicek-A and two-factor Vasicek-B are the worst models to predict the yields among all models. But they are the most easily distinguishable models. Given the two-factor CIR models are true models, it is hard to differentiate between the two-factor CIR models with other one-factor ATSMs in the graphs. Therefore, the multi-factor ATSMs do not show benefits in the improvement of the models predictive ability, comparing to the one-factor models. However, adding factors can increase models' power to distinguish between each other.

Tables 3.8 and 3.9 report the percentage of the 20,000 replications that Model #1 has greater predictive ability than Model #2 at a 12-month ahead horizon. The smaller the model's MSE is, the greater predictive ability the model has. If Model #1 has smaller MSE than Model #2 at the 5% significance level, then Model #1 beats Model #2 and Model #1 has greater predictive ability at the 5% significance level. Panel A shows the testing results when the data are from the two-factor Vasicek model and Panel B are results based on two-factor CIR model.

The values in both Panels A of Tables 3.8 and 3.9 tell us that the two-factor CIR models, both before and after the bias correction, are the most distinguishable models. The probability that the two-factor Vasicek models beat the two-factor CIR models is over 99%. In contrast, the probability that the two-factor Vasicek models beat the other one-factor ATSMs is below 30%. In addition, the probability that the two-factor CIR models beat other one-factor ATSMs is less than 10%. The numbers in Panel B of both tables represent the similar results, although the two-factor CIR models are the date-generating models. Those results show that having more factors in ATSMs makes the models' predictive ability even worse if the models are the wrong models.

3.4.4 Discussion of power and size of the ATSMs

The conclusion in Chapter II states that the sample size and the time span of the sample can influence the one-factor ATSMs' predictive ability and the power to distinguish between the one-factor models as well. In this section, I want to investigate whether the sample size and the sample's time span have the similar impact on multi-factor ATSMs. Therefore, I simulate the data using two same alternative settings as those in the last chapter and then repeat the estimation, forecasting and testing processes. The two alternative settings are Setting B in correspondence with the increase of the sample size by the way of the increase in the sampling frequency within a fixed time span and Setting C that refers to the shorter time span by the way of the increase in the sampling frequency within a fixed sample size.

3.4.4.1 The impact of sample size n on models out-of-sample forecasting performance

To investigate the impact of the sample size on ATSMs' relative predictive ability, I increase the size of the estimation window by the way of increasing the sampling frequency while keeping the time span unchanged. That is, I rerun the same Monte Carlo simulation as that in the prior sections but increase the frequency from the monthly basis ($\Delta t = \frac{1}{12}$) to the daily basis ($\Delta t' = \frac{1}{365}$). Correspondingly, the sample size n increases from 300 months to 9,125 days when the time span T is 25 years. The latter corresponds to high frequency data.

In Table 3.1, the estimates of the true models based on the simulated data from Setting B show the similar results as those from Base Setting. The biases of the mean reversion estimates are relatively larger than those of other parameter estimates. After running the bias correction process, the biases of the mean reversion estimates are reduced. Tables 3.2 and 3.3 report the estimates of all eight models. Under Setting B, the true models have the estimates closest to the true values. No matter what the time to maturity is, the one-factor models have very close estimates. In Table 3.4 and Table 3.5, we see that the RMSEs and MAEs of all models under Setting B are smaller than the corresponding values under Base Setting. This finding shows that the models' predictive ability can be improved if the sample size rises.

Checking the Figures 3.1 and 3.2, we find that the one-factors models cluster together and it is hardly to differentiate one from another. In Tables 3.6 and 3.7, the probability to differentiate one one-factor ATSM from another one-factor ATSM decreases from over 20% in Base Setting to below 10% in Setting B. However, the two-factor CIR models can be easily distinguishable from the one-factor models, and the probability to differentiate the two-factor CIR model from the one-factor models is over 70%. Besides these findings, I get the similar results with those based on Base Setting. That is, having more factors in the models increase the models' power to distinguish the one from another.

3.4.4.2 The impact of time span *T* on models out-of-sample forecasting performance

Tang and Chen (2009) states that the bias of the mean reversion estimator is not a function of the number of observations but is a decreasing function of the time span. In details, the bias of the mean reversion estimator is a function of T^{-1}. The results in the last section prove the former part of their statement. This section will study the impact of the change in the time span T on ATSMs relative predictive ability. Therefore, I use the same Monte Carlo study but changing the sampling frequency from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$) while keeping the sample size n = 300 unchanged. Correspondingly, the time span is reduced from $T = 25$ years to $T' = n \times \Delta t' = 300 \times \frac{1}{365} = 0.82$ years.

Comparing the out-of-sample forecast performance based on Setting C with those based on Base Setting, we find that the two settings generate very similar results. The RMSEs and MAEs of all models based on those two settings are very similar. Therefore, the models' predictive ability are almost the same under these two settings. That is, adding more factor does not benefit the ATSMs in improving models' predictive accuracy. But having more factors in the models increase the models' power to distinguish the one from another.

If comparing the out-of-sample forecast performance based on Setting C with those based on Setting B, the results based on Setting C show that the shorter time span by the way of the increase in the sampling frequency makes the one-factor models more easily distinguishable. But the two-factor models become the less distinguishable from the one-factor models in the case of Setting C.

3.5 Conclusion

This chapter extends the study of the one-factor ATSMs to the multi-factor models. The results show that having more factors in the ATSMs does not improve models' predictive ability. But it increase the models' power to distinguish between each other. The multi-factor ATSMs with larger sample size and longer time span have more predictive ability and stronger power to differentiate between models.

By far, all the work are based on the model simulated data with no error embedded. However, it is necessary to consider data errors, regarding the market are not frictionless and so the market data are observed with errors. Thus, my next step is about to investigate the role of data errors on the model's power to distinguish between models. In addition, in this paper I only consider the multi-factor ATSMs with a relative simple case, which is the state factors are independent and the price of risk is constant. In my future research, I will study the change of models' relative predictive ability if relaxing the models' assumptions.

Table 3.1 Parameter estimates based on data from two-factor ATSMs

	Panel A: Estimates from 2-factor Vasicek Model								
	Base Setting			Setting B			Setting C		
True Values	2 Shorts (τ=0.5&1)	2 Mediums (τ=1&20)	2 Longs (τ=20&30)	2 Shorts (τ=0.5&1)	2 Mediums (τ=1&20)	2 Longs (τ=20&30)	2 Shorts (τ=0.5&1)	2 Mediums (τ=1&20)	2 Longs (τ=20&30)
	Before Bias Correction								
κ_1= 0.473	0.526	0.525	0.526	0.525	0.525	0.525	0.526	0.526	0.526
	(0.038)	(0.047)	(0.030)	(0.029)	(0.031)	(0.031)	(0.036)	(0.045)	(0.032)
θ_1= 0.046	0.046	0.045	0.045	0.046	0.045	0.045	0.046	0.046	0.045
	(0.006)	(0.009)	(0.007)	(0.003)	(0.005)	(0.004)	(0.009)	(0.011)	(0.009)
σ_1= 0.087	0.087	0.087	0.088	0.087	0.087	0.087	0.088	0.088	0.088
	(0.004)	(0.008)	(0.005)	(0.002)	(0.004)	(0.002)	(0.005)	(0.009)	(0.003)
λ_1= -0.107	-0.110	-0.109	-0.109	-0.109	-0.108	-0.109	-0.110	-0.109	-0.109
	(0.041)	(0.059)	(0.047)	(0.035)	(0.039)	(0.033)	(0.052)	(0.061)	(0.047)
κ_2= 0.043	0.054	0.052	0.054	0.053	0.052	0.053	0.054	0.054	0.054
	(0.026)	(0.030)	(0.032)	(0.018)	(0.016)	(0.020)	(0.029)	(0.029)	(0.027)
θ_2= 0.019	0.020	0.019	0.021	0.019	0.019	0.019	0.020	0.020	0.021
	(0.009)	(0.009)	(0.005)	(0.003)	(0.006)	(0.002)	(0.005)	(0.008)	(0.006)
σ_2= 0.024	0.026	0.025	0.025	0.025	0.025	0.026	0.026	0.026	0.025
	(0.008)	(0.011)	(0.005)	(0.002)	(0.005)	(0.002)	(0.007)	(0.008)	(0.006)
λ_2= -0.045	-0.047	-0.046	-0.047	-0.046	-0.046	-0.047	-0.047	-0.046	-0.047
	(0.029)	(0.031)	(0.025)	(0.024)	(0.029)	(0.023)	(0.033)	(0.036)	(0.037)

Table 3.1 (Continued)

					After Bias Correction				
κ_1= 0.473	0.483	0.482	0.482	0.482	0.482	0.482	0.483	0.483	0.484
	(0.037)	(0.044)	(0.029)	(0.029)	(0.037)	(0.026)	(0.033)	(0.046)	(0.033)
θ_1= 0.046	0.045	0.046	0.046	0.046	0.046	0.046	0.047	0.046	0.045
	(0.005)	(0.006)	(0.007)	(0.002)	(0.003)	(0.005)	(0.005)	(0.011)	(0.009)
σ_1= 0.087	0.086	0.088	0.088	0.086	0.087	0.088	0.089	0.088	0.089
	(0.008)	(0.010)	(0.008)	(0.002)	(0.004)	(0.003)	(0.007)	(0.011)	(0.009)
λ_1= -0.107	-0.108	-0.108	-0.108	-0.108	-0.107	-0.108	-0.109	-0.109	-0.109
	(0.045)	(0.047)	(0.044)	(0.033)	(0.040)	(0.034)	(0.042)	(0.043)	(0.046)
κ_2= 0.043	0.046	0.045	0.046	0.044	0.044	0.045	0.046	0.045	0.046
	(0.027)	(0.031)	(0.034)	(0.014)	(0.018)	(0.019)	(0.024)	(0.029)	(0.023)
θ_2= 0.019	0.020	0.019	0.020	0.019	0.019	0.019	0.020	0.020	0.020
	(0.006)	(0.009)	(0.005)	(0.002)	(0.006)	(0.002)	(0.006)	(0.010)	(0.007)
σ_2= 0.024	0.025	0.024	0.024	0.024	0.024	0.025	0.025	0.025	0.025
	(0.007)	(0.012)	(0.004)	(0.005)	(0.008)	(0.002)	(0.007)	(0.009)	(0.004)
λ_2= -0.045	-0.044	-0.045	-0.045	-0.045	-0.045	-0.045	-0.044	-0.045	-0.045
	(0.023)	(0.028)	(0.024)	(0.020)	(0.023)	(0.022)	(0.025)	(0.029)	(0.028)

Table 3.1 (Continued)

Panel B: Estimates from 2-factor CIR Model									
	Base Setting			Setting B			Setting C		
True Values	2 Shorts (τ=0.5&1)	2 Mediums (τ=1&20)	2 Longs (τ=20&30)	2 Shorts (τ=0.5&1)	2 Mediums (τ=1&20)	2 Longs (τ=20&30)	2 Shorts (τ=0.5&1)	2 Mediums (τ=1&20)	2 Longs (τ=20&30)
Before Bias Correction									
κ_1= 0.654	0.669	0.667	0.668	0.667	0.668	0.668	0.668	0.669	0.669
	(0.024)	(0.030)	(0.026)	(0.020)	(0.022)	(0.022)	(0.038)	(0.042)	(0.039)
θ_1= 0.038	0.037	0.036	0.037	0.036	0.035	0.036	0.038	0.038	0.038
	(0.006)	(0.007)	(0.005)	(0.002)	(0.002)	(0.001)	(0.008)	(0.012)	(0.010)
σ_1= 0.150	0.152	0.152	0.151	0.152	0.151	0.152	0.152	0.152	0.152
	(0.005)	(0.009)	(0.007)	(0.003)	(0.004)	(0.002)	(0.005)	(0.008)	(0.005)
λ_1= -0.126	-0.124	-0.125	-0.125	-0.125	-0.125	-0.125	-0.124	-0.124	-0.124
	(0.034)	(0.039)	(0.036)	(0.022)	(0.028)	(0.025)	(0.035)	(0.042)	(0.038)
κ_2= 0.054	0.111	0.111	0.111	0.111	0.111	0.112	0.112	0.113	0.112
	(0.022)	(0.032)	(0.022)	(0.029)	(0.034)	(0.028)	(0.042)	(0.045)	(0.036)
θ_2= 0.022	0.025	0.026	0.025	0.025	0.026	0.026	0.024	0.024	0.024
	(0.007)	(0.012)	(0.005)	(0.004)	(0.006)	(0.005)	(0.008)	(0.010)	(0.008)
σ_2= 0.041	0.041	0.043	0.042	0.043	0.042	0.043	0.041	0.041	0.040
	(0.010)	(0.012)	(0.009)	(0.005)	(0.010)	(0.005)	(0.006)	(0.015)	(0.009)
λ_2= -0.048	-0.047	-0.048	-0.048	-0.048	-0.048	-0.048	-0.047	-0.047	-0.048
	(0.038)	(0.043)	(0.034)	(0.025)	(0.044)	(0.026)	(0.033)	(0.041)	(0.038)

Table 3.1 (Continued)

				After Bias Correction					
κ_1= 0.654	0.658 (0.028)	0.658 (0.035)	0.658 (0.027)	0.659 (0.019)	0.658 (0.025)	0.658 (0.023)	0.659 (0.041)	0.658 (0.047)	0.659 (0.037)
θ_1= 0.038	0.038 (0.007)	0.038 (0.010)	0.039 (0.008)	0.038 (0.003)	0.038 (0.003)	0.038 (0.002)	0.038 (0.008)	0.039 (0.011)	0.038 (0.006)
σ_1= 0.150	0.150 (0.005)	0.151 (0.011)	0.150 (0.008)	0.150 (0.004)	0.151 (0.005)	0.151 (0.003)	0.152 (0.004)	0.152 (0.010)	0.151 (0.006)
λ_1= -0.126	-0.126 (0.028)	-0.126 (0.034)	-0.125 (0.029)	-0.126 (0.025)	-0.125 (0.026)	-0.126 (0.024)	-0.127 (0.038)	-0.127 (0.051)	-0.126 (0.034)
κ_2= 0.054	0.058 (0.024)	0.056 (0.034)	0.057 (0.023)	0.056 (0.027)	0.057 (0.038)	0.055 (0.032)	0.056 (0.044)	0.056 (0.053)	0.056 (0.034)
θ_2= 0.022	0.022 (0.005)	0.022 (0.012)	0.022 (0.009)	0.022 (0.002)	0.022 (0.005)	0.022 (0.004)	0.022 (0.003)	0.021 (0.008)	0.022 (0.007)
σ_2= 0.041	0.041 (0.009)	0.041 (0.017)	0.042 (0.008)	0.041 (0.005)	0.041 (0.013)	0.041 (0.006)	0.041 (0.005)	0.041 (0.012)	0.041 (0.011)
λ_2= -0.048	-0.048 (0.039)	-0.048 (0.057)	-0.047 (0.032)	-0.048 (0.022)	-0.049 (0.037)	-0.048 (0.031)	-0.047 (0.038)	-0.047 (0.043)	-0.048 (0.041)

The two-factor Vasicek-MLE and CIR-MLE are applied to two-factor Vasicek-simulated and two-factor CIR-simulated zero bond prices, respectively. The true parameters $(\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2)$ are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to $T' = n \times \Delta t'$=300×1/365=0.82 years. The bootstrap bias correction process is applied to each set of estimates. There are 20,000 replications. The number of bootstrap resampling is 500. Panel A presents the statistics of estimates based on the two-factor Vasicek-simulated data. Panel B presents the corresponding statistics based on the two-factor CIR-simulated data. The means of estimates are showed in cells with standard deviation included in the parentheses.

Table 3.2 Estimates of 1-year zero yields from two-factor ATSMs

Panel A: Estimates of 1-year yield from the 2-factor Vasicek Model							
			Long-run Yield	**Half life of Shock**		**Unconditional S.D.**	
		Estimation Models	True Value = 0.0685	Factor 1= 1.47	Factor 2= 16.12	Factor 1= 0.0870	Factor 2= 0.0240
Base Setting	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0697	1.32	12.84	0.0871	0.0264
		2-factor Vasicek-A	0.0685	1.44	15.07	0.0863	0.0248
		2-factor CIR-B	0.0731	1.32	13.49	0.0670	0.0258
		2-factor CIR-A	0.0726	1.26	14.88	0.0651	0.0256
		1-factor Vasick-B	0.0714	1.33		0.0945	
		1-factor Vasick-A	0.0701	1.60		0.0920	
		1-factor CIR-B	0.0718	1.21		0.0987	
		1-factor CIR-A	0.0711	1.51		0.0806	
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0696	1.32	12.84	0.0876	0.0251
		2-factor Vasicek-A	0.0686	1.44	15.40	0.0882	0.0240
		2-factor CIR-B	0.0730	1.33	13.80	0.0654	0.0289
		2-factor CIR-A	0.0726	1.30	12.58	0.0668	0.0246
		1-factor Vasicek-B	0.0714	1.33		0.0945	
		1-factor Vasicek-A	0.0701	1.60		0.0920	
		1-factor CIR-B	0.0718	1.21		0.0987	
		1-factor CIR-A	0.0711	1.51		0.0806	
Setting B	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0686	1.32	13.08	0.0871	0.0249
		2-factor Vasicek-A	0.0685	1.43	15.75	0.0861	0.0244
		2-factor CIR-B	0.0720	1.47	12.59	0.0628	0.0284
		2-factor CIR-A	0.0717	1.34	15.13	0.0651	0.0249
		1-factor Vasicek-B	0.0701	1.18		0.0970	
		1-factor Vasicek-A	0.0693	1.61		0.0895	
		1-factor CIR-B	0.0709	1.01		0.0843	
		1-factor CIR-A	0.0700	1.10		0.0812	
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0690	1.32	13.33	0.0869	0.0252
		2-factor Vasicek-A	0.0685	1.44	15.75	0.0870	0.0239
		2-factor CIR-B	0.0721	1.35	13.61	0.0614	0.0262
		2-factor CIR-A	0.0717	1.34	12.39	0.0671	0.0248
		1-factor Vasicek-B	0.0701	1.18		0.0970	
		1-factor Vasicek-A	0.0693	1.61		0.0895	
		1-factor CIR-B	0.0709	1.01		0.0843	
		1-factor CIR-A	0.0700	1.10		0.0812	
Setting C	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0706	1.32	12.84	0.0876	0.0257
		2-factor Vasicek-A	0.0697	1.44	15.07	0.0892	0.0252
		2-factor CIR-B	0.0729	1.44	13.28	0.0639	0.0273
		2-factor CIR-A	0.0722	1.36	14.50	0.0647	0.0252
		1-factor Vasicek-B	0.0709	1.80		0.0991	
		1-factor Vasicek-A	0.0703	1.82		0.0872	
		1-factor CIR-B	0.0717	1.09		0.0863	
		1-factor CIR-A	0.0710	1.13		0.0836	
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0708	1.32	12.84	0.0882	0.0261
		2-factor Vasicek-A	0.0694	1.44	15.40	0.0877	0.0253
		2-factor CIR-B	0.0732	1.36	13.10	0.0625	0.0331
		2-factor CIR-A	0.0725	1.30	12.69	0.0681	0.0213
		1-factor Vasicek-B	0.0709	1.80		0.0991	
		1-factor Vasicek-A	0.0703	1.82		0.0872	
		1-factor CIR-B	0.0717	1.09		0.0863	
		1-factor CIR-A	0.0710	1.13		0.0836	

Table 3.2 (Continued)

		Panel B: Estimates of 1-year yield from the 2-factor CIR Model					
			Long-run Yield	**Half life of Shock**		**Unconditional S.D.**	
		Estimation Models	True Value = 0.0624	Factor 1= 1.06	Factor 2= 12.84	Factor 1= 0.0256	Factor 2= 0.0185
Base Setting	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0672	0.82	11.48	0.0336	0.0041
		2-factor Vasicek-A	0.0666	0.88	13.59	0.0301	0.0139
		2-factor CIR-B	0.0644	1.04	6.24	0.0253	0.0138
		2-factor CIR-A	0.0624	1.05	11.95	0.0255	0.0179
		1-factor Vasicek-B	0.0643	1.07		0.0307	
		1-factor Vasicek-A	0.0610	1.19		0.0303	
		1-factor CIR-B	0.0650	1.04		0.0249	
		1-factor CIR-A	0.0643	1.20		0.0260	
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0671	0.84	11.93	0.0249	0.0068
		2-factor Vasicek-A	0.0664	0.90	12.08	0.0304	0.0082
		2-factor CIR-B	0.0644	1.04	6.24	0.0250	0.0147
		2-factor CIR-A	0.0624	1.05	12.38	0.0257	0.0182
		1-factor Vasicek-B	0.0643	1.07		0.0307	
		1-factor Vasicek-A	0.0610	1.19		0.0303	
		1-factor CIR-B	0.0650	1.04		0.0249	
		1-factor CIR-A	0.0643	1.20		0.0260	
Setting B	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0659	0.80	12.84	0.0333	0.0041
		2-factor Vasicek-A	0.0652	0.90	14.15	0.0310	0.0120
		2-factor CIR-B	0.0634	1.04	6.24	0.0250	0.0142
		2-factor CIR-A	0.0625	1.05	12.38	0.0255	0.0182
		1-factor Vasicek-B	0.0617	0.94		0.0323	
		1-factor Vasicek-A	0.0610	1.19		0.0303	
		1-factor CIR-B	0.0643	0.90		0.0263	
		1-factor CIR-A	0.0628	0.91		0.0266	
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0660	0.82	11.82	0.0337	0.0070
		2-factor Vasicek-A	0.0654	0.89	12.29	0.0302	0.0088
		2-factor CIR-B	0.0634	1.04	6.24	0.0244	0.0144
		2-factor CIR-A	0.0624	1.05	12.16	0.0257	0.0180
		1-factor Vasicek-B	0.0617	0.94		0.0323	
		1-factor Vasicek-A	0.0610	1.11		0.0305	
		1-factor CIR-B	0.0643	0.90		0.0263	
		1-factor CIR-A	0.0627	0.91		0.0266	
Setting C	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0678	0.80	11.80	0.0331	0.0041
		2-factor Vasicek-A	0.0668	0.88	13.59	0.0308	0.0104
		2-factor CIR-B	0.0644	1.04	6.19	0.0256	0.0134
		2-factor CIR-A	0.0624	1.05	12.38	0.0258	0.0182
		1-factor Vasicek-B	0.0654	0.81		0.0333	
		1-factor Vasicek-A	0.0586	0.85		0.0333	
		1-factor CIR-B	0.0662	0.83		0.0354	
		1-factor CIR-A	0.0660	0.89		0.0253	
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0678	0.82	12.67	0.0334	0.0091
		2-factor Vasicek-A	0.0667	0.88	13.49	0.0308	0.0088
		2-factor CIR-B	0.0645	1.04	6.13	0.0260	0.0134
		2-factor CIR-A	0.0625	1.05	12.38	0.0262	0.0178
		1-factor Vasicek-B	0.0654	0.81		0.0333	
		1-factor Vasicek-A	0.0587	0.85		0.0333	
		1-factor CIR-B	0.0662	0.83		0.0354	
		1-factor CIR-A	0.0660	0.90		0.0253	

Table 3.2 (Continued)

Eight estimation models are applied to two-factor Vasicek-simulated and two-factor CIR-simulated 1-year zero bond prices, respectively. The true parameters ($\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2$) are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to T' = $n \times \Delta t'$=300×1/365=0.82 years. The bootstrap bias correction process is applied to each set of estimates. There are 20,000 replications. The number of bootstrap resampling is 500. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. Panel A presents the statistics of estimates based on the two-factor Vasicek-simulated data with one year maturity. Panel B presents the corresponding statistics based on the two-factor CIR-simulated data with one year maturity.

Table 3.3 Estimates of 20-year zero yields from two-factor ATSMs

Panel A: Estimates of 20-year yield from the 2-factor Vasicek Model								
			Long-run Yield	Half life of Shock		Unconditional S.D.	
		Estimation Models	True Value = 0.0705	Factor 1= 1.47	Factor 2= 16.12	Factor 1= 0.0870	Factor 2= 0.0240
Base Setting	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0716	1.32	12.84	0.0876	0.0251
		2-factor Vasicek-A	0.0706	1.44	15.40	0.0882	0.0240
		2-factor CIR-B	0.0750	2.91	13.80	0.0454	0.0289
		2-factor CIR-A	0.0746	4.79	12.58	0.0868	0.0246
		1-factor Vasicek-B	0.0734	2.70		0.0990	
		1-factor Vasicek-A	0.0720	5.75		0.0528	
		1-factor CIR-B	0.0738	2.13		0.0947	
		1-factor CIR-A	0.0730	2.80		0.0903	
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0712	1.32	12.84	0.0876	0.0251
		2-factor Vasicek-A	0.0704	1.44	15.07	0.0882	0.0243
		2-factor CIR-B	0.0747	2.74	19.81	0.0415	0.0423
		2-factor CIR-A	0.0745	4.58	11.14	0.0534	0.0239
		1-factor Vasicek-B	0.0734	2.70		0.0990	
		1-factor Vasicek-A	0.0720	5.75		0.0528	
		1-factor CIR-B	0.0738	2.13		0.0947	
		1-factor CIR-A	0.0730	2.80		0.0903	
Setting B	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0709	1.32	13.33	0.0869	0.0252
		2-factor Vasicek-A	0.0705	1.44	15.75	0.0870	0.0239
		2-factor CIR-B	0.0740	2.96	14.10	0.0481	0.0331
		2-factor CIR-A	0.0736	4.47	12.69	0.0842	0.0213
		1-factor Vasicek-B	0.0720	2.85		0.0322	
		1-factor Vasicek-A	0.0712	4.77		0.0550	
		1-factor CIR-B	0.0728	5.27		0.0147	
		1-factor CIR-A	0.0719	2.20		0.0553	
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0709	1.32	13.08	0.0874	0.0257
		2-factor Vasicek-A	0.0704	1.44	15.40	0.0880	0.0249
		2-factor CIR-B	0.0740	2.05	15.27	0.0459	0.0495
		2-factor CIR-A	0.0736	4.83	13.41	0.0574	0.0235
		1-factor Vasicek-B	0.0720	2.85		0.0322	
		1-factor Vasicek-A	0.0712	4.77		0.0550	
		1-factor CIR-B	0.0728	5.27		0.0147	
		1-factor CIR-A	0.0719	2.20		0.0553	
Setting C	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0717	1.32	12.84	0.0882	0.0261
		2-factor Vasicek-A	0.0703	1.44	15.40	0.0877	0.0253
		2-factor CIR-B	0.0751	2.98	13.61	0.0419	0.0362
		2-factor CIR-A	0.0744	4.54	12.39	0.0873	0.0248
		1-factor Vasicek-B	0.0728	2.94		0.0302	
		1-factor Vasicek-A	0.0722	3.77		0.0451	
		1-factor CIR-B	0.0736	2.55		0.0548	
		1-factor CIR-A	0.0729	2.21		0.0543	
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0715	1.32	12.84	0.0879	0.0246
		2-factor Vasicek-A	0.0705	1.43	15.07	0.0894	0.0251
		2-factor CIR-B	0.0748	2.05	15.86	0.0436	0.0420
		2-factor CIR-A	0.0741	4.44	12.45	0.0561	0.0237
		1-factor Vasicek-B	0.0728	2.94		0.0302	
		1-factor Vasicek-A	0.0722	3.77		0.0451	
		1-factor CIR-B	0.0736	2.55		0.0548	
		1-factor CIR-A	0.0729	2.21		0.0543	

Table 3.3 (Continued)

Panel B: Forecasts of 20-year yield from the 2-factor CIR Model							
			Long-run Yield	**Half life of Shock**		**Unconditional S.D.**	
		Estimation Models	True Value = 0.0744	Factor 1= 1.06	Factor 2= 12.84	Factor 1= 0.0256	Factor 2= 0.0185
Base Setting	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0795	0.87	10.93	0.0307	0.0088
		2-factor Vasicek-A	0.0789	0.90	12.08	0.0304	0.0082
		2-factor CIR-B	0.0751	1.04	6.24	0.0250	0.0147
		2-factor CIR-A	0.0742	1.05	12.38	0.0257	0.0182
		1-factor Vasicek-B	0.0774	2.17		0.0358	
		1-factor Vasicek-A	0.0765	3.87		0.0200	
		1-factor CIR-B	0.0788	2.60		0.0134	
		1-factor CIR-A	0.0768	5.27		0.0236	
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0798	0.89	23.50	0.0396	0.0070
		2-factor Vasicek-A	0.0784	0.92	23.80	0.0389	0.0070
		2-factor CIR-B	0.0751	1.04	6.24	0.0251	0.0141
		2-factor CIR-A	0.0745	1.05	12.16	0.0258	0.0185
		1-factor Vasicek-B	0.0774	2.17		0.0358	
		1-factor Vasicek-A	0.0765	3.87		0.0200	
		1-factor CIR-B	0.0788	2.60		0.0134	
		1-factor CIR-A	0.0768	5.27		0.0236	
Setting B	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0789	0.86	11.82	0.0307	0.0090
		2-factor Vasicek-A	0.0774	0.89	12.29	0.0302	0.0088
		2-factor CIR-B	0.0741	1.04	6.24	0.0244	0.0144
		2-factor CIR-A	0.0743	1.05	12.60	0.0257	0.0183
		1-factor Vasicek-B	0.0775	1.63		0.0491	
		1-factor Vasicek-A	0.0746	3.65		0.0219	
		1-factor CIR-B	0.0778	3.40		0.0211	
		1-factor CIR-A	0.0774	3.12		0.0207	
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0787	0.89	23.51	0.0403	0.0068
		2-factor Vasicek-A	0.0773	0.91	23.44	0.0389	0.0068
		2-factor CIR-B	0.0751	1.04	6.19	0.0250	0.0147
		2-factor CIR-A	0.0743	1.06	12.73	0.0257	0.0184
		1-factor Vasicek-B	0.0775	1.63		0.0491	
		1-factor Vasicek-A	0.0746	3.65		0.0219	
		1-factor CIR-B	0.0778	3.40		0.0211	
		1-factor CIR-A	0.0774	3.12		0.0207	
Setting C	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0791	0.82	12.67	0.0304	0.0091
		2-factor Vasicek-A	0.0783	0.84	13.49	0.0308	0.0088
		2-factor CIR-B	0.0747	1.04	6.13	0.0256	0.0134
		2-factor CIR-A	0.0739	1.05	12.38	0.0262	0.0178
		1-factor Vasicek-B	0.0774	0.82		0.0979	
		1-factor Vasicek-A	0.0746	0.86		0.0951	
		1-factor CIR-B	0.0792	2.40		0.0289	
		1-factor CIR-A	0.0780	3.03		0.0253	
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0792	0.92	23.62	0.0384	0.0069
		2-factor Vasicek-A	0.0780	1.02	23.88	0.0353	0.0068
		2-factor CIR-B	0.0750	1.04	6.19	0.0256	0.0131
		2-factor CIR-A	0.0742	1.05	12.38	0.0256	0.0182
		1-factor Vasicek-B	0.0774	0.82		0.0979	
		1-factor Vasicek-A	0.0746	0.86		0.0951	
		1-factor CIR-B	0.0792	2.40		0.0289	
		1-factor CIR-A	0.0780	3.03		0.0253	

Table 3.3 (Continued)

Eight estimation models are applied to two-factor Vasicek-simulated and two-factor CIR-simulated 20-year zero bond prices, respectively. The true parameters ($\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2$) are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to $T' = n \times \Delta t'$=300×1/365=0.82 years. The bootstrap bias correction process is applied to each set of estimates. There are 20,000 replications. The number of bootstrap resampling is 500. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. Panel A presents the statistics of estimates based on the two-factor Vasicek-simulated data with twenty year maturity. Panel B presents the corresponding statistics based on the two-factor CIR-simulated data with twenty year maturity.

Table 3.4 Comparison of models out-of-sample forecasts of 1-year zero yields

Panel A: Forecast Errors of 1-year yield from the 2-factor Vasicek Model

			1-month		**6-month**		**12-month**	
		Estimation Models	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Base Setting	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0069	0.0055	0.0157	0.0124	0.0202	0.0161
		2-factor Vasicek-A	0.0069	0.0055	0.0156	0.0123	0.0200	0.0159
		2-factor CIR-B	0.0110	0.0093	0.0196	0.0161	0.0246	0.0202
		2-factor CIR-A	0.0090	0.0073	0.0176	0.0141	0.0226	0.0182
		1-factor Vasicek-B	0.0070	0.0055	0.0163	0.0129	0.0216	0.0173
		1-factor Vasicek-A	0.0069	0.0056	0.0159	0.0127	0.0205	0.0164
		1-factor CIR-B	0.0070	0.0056	0.0166	0.0132	0.0219	0.0176
		1-factor CIR-A	0.0070	0.0056	0.0161	0.0128	0.0210	0.0169
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0070	0.0056	0.0158	0.0126	0.0201	0.0160
		2-factor Vasicek-A	0.0070	0.0055	0.0158	0.0125	0.0201	0.0160
		2-factor CIR-B	0.0096	0.0079	0.0191	0.0154	0.0249	0.0202
		2-factor CIR-A	0.0081	0.0064	0.0176	0.0139	0.0234	0.0187
		1-factor Vasicek-B	0.0070	0.0056	0.0166	0.0132	0.0219	0.0176
		1-factor Vasicek-A	0.0070	0.0056	0.0159	0.0127	0.0205	0.0164
		1-factor CIR-B	0.0070	0.0056	0.0167	0.0133	0.0224	0.0181
		1-factor CIR-A	0.0070	0.0056	0.0161	0.0129	0.0211	0.0170
Setting B	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0058	0.0045	0.0147	0.0116	0.0196	0.0155
		2-factor Vasicek-A	0.0058	0.0044	0.0147	0.0116	0.0196	0.0155
		2-factor CIR-B	0.0095	0.0074	0.0182	0.0143	0.024	0.0192
		2-factor CIR-A	0.0083	0.0062	0.0170	0.0131	0.0231	0.0180
		1-factor Vasicek-B	0.0059	0.0045	0.0156	0.0124	0.0214	0.0171
		1-factor Vasicek-A	0.0058	0.0045	0.0152	0.0120	0.0207	0.0165
		1-factor CIR-B	0.0058	0.0044	0.0155	0.0119	0.0212	0.0162
		1-factor CIR-A	0.0058	0.0044	0.0154	0.0118	0.0211	0.0160
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0059	0.0045	0.0148	0.0117	0.0197	0.0156
		2-factor Vasicek-A	0.0058	0.0044	0.0147	0.0116	0.0196	0.0155
		2-factor CIR-B	0.0093	0.0075	0.0181	0.0144	0.0239	0.0190
		2-factor CIR-A	0.0080	0.0062	0.0168	0.0131	0.0226	0.0177
		1-factor Vasicek-B	0.0059	0.0045	0.0156	0.0124	0.0214	0.0171
		1-factor Vasicek-A	0.0059	0.0045	0.0154	0.0122	0.0211	0.0169
		1-factor CIR-B	0.0058	0.0044	0.0158	0.0121	0.0217	0.0167
		1-factor CIR-A	0.0058	0.0044	0.0155	0.0119	0.0212	0.0163
Setting C	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0070	0.0056	0.0160	0.0128	0.0206	0.0165
		2-factor Vasicek-A	0.0070	0.0056	0.0160	0.0128	0.0205	0.0164
		2-factor CIR-B	0.0087	0.0070	0.0179	0.0146	0.0238	0.0196
		2-factor CIR-A	0.0087	0.0070	0.0176	0.0144	0.0232	0.0191
		1-factor Vasicek-B	0.0070	0.0056	0.0164	0.0131	0.0219	0.0175
		1-factor Vasicek-A	0.0070	0.0056	0.0162	0.0130	0.0216	0.0173
		1-factor CIR-B	0.0070	0.0056	0.0167	0.0134	0.0227	0.0181
		1-factor CIR-A	0.0070	0.0056	0.0166	0.0133	0.0225	0.0179
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0070	0.0056	0.0159	0.0127	0.0203	0.0162
		2-factor Vasicek-A	0.0069	0.0056	0.0158	0.0126	0.0202	0.0161
		2-factor CIR-B	0.0089	0.0070	0.0181	0.0141	0.0244	0.0192
		2-factor CIR-A	0.0085	0.0065	0.0176	0.0136	0.0235	0.0181
		1-factor Vasicek-B	0.0070	0.0056	0.0162	0.0130	0.0216	0.0172
		1-factor Vasicek-A	0.0070	0.0056	0.0160	0.0128	0.0210	0.0168
		1-factor CIR-B	0.0069	0.0056	0.0163	0.0130	0.0220	0.0173
		1-factor CIR-A	0.0069	0.0056	0.0162	0.0129	0.0216	0.0172

Table 3.4 (Continued)

		Panel B: Forecast Errors of 1-year yield from the 2-factor CIR Model						
			1-month		**6-month**		**12-month**	
		Estimation Models	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Base Setting	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0083	0.0067	0.0168	0.0127	0.0224	0.0189
		2-factor Vasicek-A	0.0079	0.0064	0.0160	0.0129	0.0205	0.0174
		2-factor CIR-B	0.0066	0.0050	0.0144	0.0110	0.0180	0.0138
		2-factor CIR-A	0.0065	0.0050	0.0144	0.0109	0.0180	0.0137
		1-factor Vasicek-B	0.0065	0.0050	0.0149	0.0114	0.0189	0.0144
		1-factor Vasicek-A	0.0065	0.0050	0.0146	0.0113	0.0182	0.0140
		1-factor CIR-B	0.0066	0.0051	0.0151	0.0117	0.0197	0.0153
		1-factor CIR-A	0.0065	0.0050	0.0148	0.0112	0.0186	0.0140
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0099	0.0086	0.0179	0.0151	0.0219	0.0186
		2-factor Vasicek-A	0.0091	0.0078	0.0165	0.0136	0.0199	0.0163
		2-factor CIR-B	0.0066	0.0052	0.0145	0.0114	0.0180	0.0142
		2-factor CIR-A	0.0065	0.0051	0.0144	0.0112	0.0178	0.0139
		1-factor Vasicek-B	0.0065	0.0050	0.0149	0.0114	0.0189	0.0144
		1-factor Vasicek-A	0.0065	0.0050	0.0146	0.0113	0.0182	0.0140
		1-factor CIR-B	0.0066	0.0051	0.0151	0.0117	0.0192	0.0154
		1-factor CIR-A	0.0065	0.0050	0.0148	0.0113	0.0186	0.0140
Setting B	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0075	0.0064	0.0159	0.0132	0.0220	0.0187
		2-factor Vasicek-A	0.0063	0.0047	0.0147	0.0117	0.0205	0.0171
		2-factor CIR-B	0.0056	0.0040	0.0131	0.0099	0.0170	0.0129
		2-factor CIR-A	0.0056	0.0040	0.0131	0.0099	0.0170	0.0129
		1-factor Vasicek-B	0.0057	0.0041	0.0139	0.0106	0.0186	0.0143
		1-factor Vasicek-A	0.0057	0.0041	0.0137	0.0104	0.0184	0.0139
		1-factor CIR-B	0.0057	0.0041	0.0140	0.0107	0.0188	0.0146
		1-factor CIR-A	0.0057	0.0041	0.0139	0.0107	0.0188	0.0144
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0075	0.0065	0.0158	0.0135	0.0220	0.0188
		2-factor Vasicek-A	0.0063	0.0046	0.0147	0.0117	0.0205	0.0173
		2-factor CIR-B	0.0056	0.0040	0.0131	0.0099	0.0170	0.0129
		2-factor CIR-A	0.0056	0.0040	0.0131	0.0099	0.0170	0.0129
		1-factor Vasicek-B	0.0057	0.0041	0.0139	0.0106	0.0186	0.0143
		1-factor Vasicek-A	0.0057	0.0041	0.0137	0.0104	0.0184	0.0139
		1-factor CIR-B	0.0057	0.0041	0.0140	0.0107	0.0188	0.0146
		1-factor CIR-A	0.0057	0.0041	0.0139	0.0107	0.0188	0.0144
Setting C	2 Shorts (τ=0.5&1)	2-factor Vasicek-B	0.0100	0.0086	0.0183	0.0154	0.0233	0.0199
		2-factor Vasicek-A	0.0089	0.0071	0.0172	0.0139	0.0225	0.0179
		2-factor CIR-B	0.0067	0.0051	0.0146	0.0112	0.0180	0.0140
		2-factor CIR-A	0.0066	0.0050	0.0145	0.0111	0.0179	0.0139
		1-factor Vasicek-B	0.0066	0.0050	0.0150	0.0116	0.0196	0.0153
		1-factor Vasicek-A	0.0067	0.0051	0.0149	0.0114	0.0194	0.0149
		1-factor CIR-B	0.0066	0.0051	0.0154	0.0117	0.0198	0.0150
		1-factor CIR-A	0.0067	0.0051	0.0149	0.0115	0.0186	0.0140
	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0099	0.0086	0.0179	0.0151	0.0219	0.0209
		2-factor Vasicek-A	0.0091	0.0078	0.0172	0.0139	0.0199	0.0179
		2-factor CIR-B	0.0066	0.0051	0.0145	0.0114	0.0180	0.0142
		2-factor CIR-A	0.0066	0.0050	0.0144	0.0113	0.0180	0.0138
		1-factor Vasicek-B	0.0067	0.0051	0.0150	0.0114	0.0196	0.0153
		1-factor Vasicek-A	0.0066	0.0051	0.0149	0.0113	0.0194	0.0149
		1-factor CIR-B	0.0067	0.0051	0.0151	0.0119	0.0197	0.0154
		1-factor CIR-A	0.0067	0.0051	0.0149	0.0115	0.0187	0.0144

Table 3.4 (Continued)

Eight estimation methods are applied to one-year zero yields simulated from a 2-factor Vasicek model and a 2-factor CIR model. The true parameters ($\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2$) are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to T' = $n \times \Delta t'$=300×1/365=0.82 years. The sets of model estimates are used to make out-of-sample forecasts. There are 10,000 replications. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. Denote RMSE as the root mean square error of the zero bond yield forecasts and denote MAE as the mean of the absolute error of the forecasts.

Table 3.5 Comparison of models out-of-sample forecasts of 20-year zero yields

Panel A: Forecast Errors of 20-year yield from the 2-factor Vasicek Model								
			1-month		6-month		12-month	
		Estimation Models	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Base Setting	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0070	0.0057	0.0158	0.0125	0.0201	0.0160
		2-factor Vasicek-A	0.0070	0.0057	0.0158	0.0125	0.0201	0.0160
		2-factor CIR-B	0.0096	0.0079	0.0191	0.0154	0.0249	0.0202
		2-factor CIR-A	0.0081	0.0064	0.0176	0.0139	0.0234	0.0187
		1-factor Vasicek-B	0.0070	0.0056	0.0166	0.0132	0.0219	0.0176
		1-factor Vasicek-A	0.0070	0.0056	0.0159	0.0127	0.0205	0.0164
		1-factor CIR-B	0.0070	0.0056	0.0167	0.0133	0.0225	0.0182
		1-factor CIR-A	0.0070	0.0056	0.0163	0.010	0.0214	0.0173
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0069	0.0055	0.0157	0.0124	0.0202	0.0161
		2-factor Vasicek-A	0.0069	0.0055	0.0156	0.0123	0.0200	0.0159
		2-factor CIR-B	0.0110	0.0093	0.0196	0.0161	0.0246	0.0202
		2-factor CIR-A	0.0090	0.0073	0.0176	0.0141	0.0226	0.0182
		1-factor Vasicek-B	0.0069	0.0055	0.0163	0.0129	0.0216	0.0173
		1-factor Vasicek-A	0.0069	0.0055	0.0159	0.0127	0.0205	0.0164
		1-factor CIR-B	0.0070	0.0056	0.0167	0.0133	0.0220	0.0177
		1-factor CIR-A	0.0070	0.0056	0.0162	0.0130	0.0213	0.0172
Setting B	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0059	0.0045	0.0148	0.0117	0.0197	0.0156
		2-factor Vasicek-A	0.0058	0.0044	0.0147	0.0116	0.0196	0.0155
		2-factor CIR-B	0.0093	0.0075	0.0181	0.0144	0.0239	0.0190
		2-factor CIR-A	0.0080	0.0062	0.0168	0.0131	0.0226	0.0177
		1-factor Vasicek-B	0.0059	0.0045	0.0156	0.0124	0.0214	0.0171
		1-factor Vasicek-A	0.0059	0.0045	0.0154	0.0122	0.0211	0.0169
		1-factor CIR-B	0.0058	0.0044	0.0158	0.0122	0.0218	0.0168
		1-factor CIR-A	0.0058	0.0044	0.0156	0.0120	0.0215	0.0166
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0058	0.0045	0.0147	0.0116	0.0196	0.0155
		2-factor Vasicek-A	0.0058	0.0045	0.0147	0.0116	0.0196	0.0155
		2-factor CIR-B	0.0095	0.0074	0.0182	0.0143	0.0234	0.0192
		2-factor CIR-A	0.0083	0.0062	0.0170	0.0131	0.0231	0.0180
		1-factor Vasicek-B	0.0059	0.0045	0.0156	0.0124	0.0214	0.0171
		1-factor Vasicek-A	0.0059	0.0045	0.0155	0.0123	0.0212	0.0170
		1-factor CIR-B	0.0058	0.0044	0.0156	0.0119	0.0213	0.0163
		1-factor CIR-A	0.0058	0.0044	0.0156	0.0119	0.0213	0.0163
Setting C	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0071	0.0056	0.0159	0.0127	0.0203	0.0162
		2-factor Vasicek-A	0.0070	0.0055	0.0158	0.0126	0.0202	0.0161
		2-factor CIR-B	0.0089	0.0070	0.0181	0.0141	0.0244	0.0192
		2-factor CIR-A	0.0085	0.0065	0.0176	0.0136	0.0235	0.0181
		1-factor Vasicek-B	0.0070	0.0056	0.0162	0.0130	0.0216	0.0172
		1-factor Vasicek-A	0.0070	0.0056	0.0160	0.0128	0.0210	0.0168
		1-factor CIR-B	0.0070	0.0056	0.0165	0.0132	0.0223	0.0176
		1-factor CIR-A	0.0069	0.0055	0.0164	0.0130	0.0219	0.0175
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0070	0.0056	0.0160	0.0128	0.0206	0.0164
		2-factor Vasicek-A	0.0070	0.0056	0.0160	0.0128	0.0205	0.0164
		2-factor CIR-B	0.0087	0.0070	0.0179	0.0146	0.0238	0.0196
		2-factor CIR-A	0.0087	0.0070	0.0176	0.0144	0.0232	0.0191
		1-factor Vasicek-B	0.0070	0.0056	0.0164	0.0131	0.0219	0.0175
		1-factor Vasicek-A	0.0070	0.0056	0.0162	0.0130	0.0216	0.0173
		1-factor CIR-B	0.0070	0.0056	0.0167	0.0134	0.0228	0.0182
		1-factor CIR-A	0.0070	0.0056	0.0167	0.0134	0.0228	0.0182

Table 3.5 (Continued)

Panel B: Forecast Errors of 20-year yield from the 2-factor CIR Model								
			1-month		6-month		12-month	
		Estimation Models	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$	RMSE of $\hat{y}_t$	MAE of $\hat{y}_t$
Base Setting	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0100	0.0086	0.0180	0.0152	0.0222	0.0186
		2-factor Vasicek-A	0.0092	0.0078	0.0168	0.0139	0.0206	0.0170
		2-factor CIR-B	0.0066	0.0052	0.0145	0.0114	0.0180	0.0142
		2-factor CIR-A	0.0065	0.0051	0.0144	0.0112	0.0178	0.0139
		1-factor Vasicek-B	0.0065	0.0050	0.0149	0.0114	0.0189	0.0144
		1-factor Vasicek-A	0.0065	0.0050	0.0146	0.0113	0.0182	0.0140
		1-factor CIR-B	0.0066	0.0051	0.0152	0.0120	0.0193	0.0155
		1-factor CIR-A	0.0066	0.0051	0.0151	0.0115	0.0192	0.0146
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0101	0.0086	0.0184	0.0156	0.0235	0.0201
		2-factor Vasicek-A	0.0080	0.0062	0.0175	0.0133	0.0231	0.0175
		2-factor CIR-B	0.0066	0.0051	0.0146	0.0113	0.0181	0.0139
		2-factor CIR-A	0.0065	0.0050	0.0145	0.0111	0.0179	0.0137
		1-factor Vasicek-B	0.0065	0.0050	0.0149	0.0114	0.0189	0.0144
		1-factor Vasicek-A	0.0065	0.0050	0.0146	0.0113	0.0182	0.0140
		1-factor CIR-B	0.0066	0.0051	0.0154	0.0118	0.0200	0.0151
		1-factor CIR-A	0.0065	0.0051	0.0151	0.0115	0.0192	0.0146
Setting B	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0084	0.0070	0.0168	0.0140	0.0223	0.0190
		2-factor Vasicek-A	0.0075	0.0061	0.0158	0.0130	0.0212	0.0179
		2-factor CIR-B	0.0056	0.0040	0.0131	0.0100	0.0170	0.0129
		2-factor CIR-A	0.0056	0.0040	0.0131	0.0099	0.0170	0.0129
		1-factor Vasicek-B	0.0057	0.0041	0.0139	0.0106	0.0186	0.0143
		1-factor Vasicek-A	0.0057	0.0041	0.0137	0.0104	0.0184	0.0139
		1-factor CIR-B	0.0057	0.0041	0.0140	0.0108	0.0189	0.0147
		1-factor CIR-A	0.0057	0.0041	0.0142	0.0110	0.0194	0.0150
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0075	0.0064	0.0161	0.0133	0.0223	0.0190
		2-factor Vasicek-A	0.0063	0.0048	0.0150	0.0120	0.0212	0.0177
		2-factor CIR-B	0.0056	0.0040	0.0131	0.0099	0.0170	0.0129
		2-factor CIR-A	0.0056	0.0040	0.0131	0.0099	0.0170	0.0129
		1-factor Vasicek-B	0.0057	0.0041	0.0139	0.0106	0.0186	0.0143
		1-factor Vasicek-A	0.0057	0.0041	0.0137	0.0104	0.0184	0.0139
		1-factor CIR-B	0.0057	0.0041	0.0140	0.0108	0.0189	0.0147
		1-factor CIR-A	0.0056	0.0040	0.0135	0.0103	0.0180	0.0136
Setting C	2 Mediums (τ=1&20)	2-factor Vasicek-B	0.0083	0.0066	0.0172	0.0136	0.0225	0.0180
		2-factor Vasicek-A	0.0075	0.0055	0.0160	0.0119	0.0207	0.0152
		2-factor CIR-B	0.0066	0.0051	0.0144	0.0110	0.0180	0.0138
		2-factor CIR-A	0.0066	0.0050	0.0143	0.0109	0.0180	0.0137
		1-factor Vasicek-B	0.0067	0.0051	0.0150	0.0116	0.0196	0.0152
		1-factor Vasicek-A	0.0067	0.0051	0.0149	0.0114	0.0194	0.0149
		1-factor CIR-B	0.0067	0.0051	0.0151	0.0117	0.0198	0.0154
		1-factor CIR-A	0.0067	0.0051	0.0152	0.0118	0.0199	0.0156
	2 Longs (τ=20&30)	2-factor Vasicek-B	0.0083	0.0067	0.0169	0.0132	0.0226	0.0174
		2-factor Vasicek-A	0.0079	0.0064	0.0162	0.0128	0.0209	0.0173
		2-factor CIR-B	0.0066	0.0050	0.0144	0.0110	0.0181	0.0138
		2-factor CIR-A	0.0066	0.0050	0.0144	0.0109	0.0180	0.0137
		1-factor Vasicek-B	0.0067	0.0051	0.0154	0.0119	0.0203	0.0159
		1-factor Vasicek-A	0.0067	0.0051	0.0150	0.0115	0.0196	0.0151
		1-factor CIR-B	0.0067	0.0051	0.0151	0.0117	0.0198	0.0154
		1-factor CIR-A	0.0067	0.0051	0.0149	0.0114	0.0192	0.0149

Table 3.5 (Continued)

Eight estimation methods are applied to 20-year zero yields simulated from a 2-factor Vasicek model and a 2-factor CIR model. The true parameters $(\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2)$ are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to $T' = n \times \Delta t'$=300×1/365=0.82 years. The sets of model estimates are used to make out-of-sample forecasts. There are 10,000 replications. "*Vasicek-B*" and "*Vasicek-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. Denote RMSE as the root mean square error of the zero bond yield forecasts and denote MAE as the mean of the absolute error of the forecasts.

Table 3.6 Models' relative predictive ability test of 1-year zero yields

Panel A: Giacomini and White Test of 1-year yield from the 2-factor Vasicek Model

(%Path of #1 ≠ #2 out of 20,000 simulations over 12-month ahead forecasts)

		Base Setting		**Setting B**		**Setting C**	
Model #1	Model #2	2 Shorts	2 Mediums	2 Shorts	2 Mediums	2 Shorts	2 Mediums
2f-Vas-B	2f-Vas-A	29.91	26.00	16.96	17.31	29.84	20.03
	2f-CIR-B	99.91	99.35	99.64	99.74	99.97	99.63
	2f-CIR-A	99.85	99.32	99.01	99.30	99.59	99.12
	1f-Vas-B	55.35	59.64	54.95	55.57	56.64	55.38
	1f-Vas-A	50.03	52.79	54.80	55.33	54.83	54.75
	1f-CIR-B	58.54	56.29	56.63	56.59	55.81	55.36
	1f-CIR-A	52.95	52.90	56.81	56.45	51.20	53.96
2f-Vas-A	2f-CIR-B	99.27	99.28	99.65	99.69	99.96	99.80
	2f-CIR-A	99.08	99.02	99.62	99.57	99.91	99.76
	1f-Vas-B	59.11	59.80	54.91	55.35	56.24	56.37
	1f-Vas-A	53.54	51.32	54.63	55.32	54.42	54.89
	1f-CIR-B	58.86	57.32	56.77	56.87	58.14	57.31
	1f-CIR-A	58.90	57.18	56.76	56.57	56.12	56.15
2f-CIR-B	2f-CIR-A	21.20	27.99	44.54	46.82	41.38	47.78
	1f-Vas-B	76.71	77.89	86.61	88.28	83.64	85.20
	1f-Vas-A	79.68	71.96	86.70	89.08	84.11	87.30
	1f-CIR-B	70.10	73.52	87.38	89.45	85.29	86.70
	1f-CIR-A	75.22	77.12	87.37	89.47	83.04	81.24
2f-CIR-A	1f-Vas-B	77.40	77.21	76.89	78.12	72.50	77.60
	1f-Vas-A	75.60	72.51	77.33	79.17	73.30	78.15
	1f-CIR-B	71.29	74.81	77.04	77.12	76.00	78.85
	1f-CIR-A	75.93	70.19	77.00	77.13	78.17	74.06
1f-Vas-B	1f-Vas-A	26.45	24.92	10.64	9.12	29.98	29.98
	1f-CIR-B	27.93	27.65	9.06	9.82	29.41	29.41
	1f-CIR-A	26.77	22.55	9.08	9.83	28.06	28.06
1f-Vas-A	1f-CIR-B	27.10	27.07	9.27	9.41	26.98	26.98
	1f-CIR-A	28.51	27.31	9.26	9.05	29.44	29.44
1f-CIR-B	1f-CIR-A	21.18	20.90	8.80	9.81	28.80	28.80

Table 3.6 (Continued)

Panel B: Giacomini and White Test of 1-year yield from the 2-factor CIR Model

(%Path of #1 ≠ #2 out of 20,000 simulations over 12-month ahead forecasts)

		Base Setting		**Setting B**		**Setting C**	
Model #1	Model #2	2 Shorts	2 Mediums	2 Shorts	2 Mediums	2 Shorts	2 Mediums
2f-Vas-B	2f-Vas-A	23.06	29.21	46.49	43.82	48.97	41.28
	2f-CIR-B	99.20	99.72	99.58	99.03	99.43	99.80
	2f-CIR-A	99.95	99.80	99.46	99.03	99.34	99.51
	1f-Vas-B	70.27	73.31	76.11	71.25	79.73	79.60
	1f-Vas-A	70.21	78.55	75.96	71.44	79.40	78.61
	1f-CIR-B	79.74	72.06	76.23	70.73	70.50	73.88
	1f-CIR-A	71.92	75.90	76.11	71.12	79.44	78.87
2f-Vas-A	2f-CIR-B	99.09	99.13	99.97	99.41	99.06	99.47
	2f-CIR-A	99.52	99.96	99.91	99.39	99.15	99.85
	1f-Vas-B	77.06	79.49	70.87	72.01	79.26	76.62
	1f-Vas-A	71.43	73.34	70.87	71.73	78.44	76.18
	1f-CIR-B	75.84	78.56	70.01	71.72	71.21	74.52
	1f-CIR-A	79.10	71.51	70.53	71.87	77.89	76.48
2f-CIR-B	2f-CIR-A	25.30	26.10	18.87	18.87	20.25	19.60
	1f-Vas-B	54.73	54.09	55.01	55.13	58.36	53.93
	1f-Vas-A	50.59	59.30	55.11	55.13	58.28	54.73
	1f-CIR-B	53.53	52.81	55.43	55.57	52.57	54.12
	1f-CIR-A	58.07	56.96	55.18	55.47	58.64	54.96
2f-CIR-A	1f-Vas-B	53.23	54.40	55.09	55.01	53.70	53.16
	1f-Vas-A	58.05	59.77	55.01	55.08	53.95	54.25
	1f-CIR-B	52.10	53.11	55.71	55.49	53.93	53.99
	1f-CIR-A	55.96	57.60	55.41	55.25	53.09	53.91
1f-Vas-B	1f-Vas-A	20.85	20.85	8.65	9.65	21.58	91.58
	1f-CIR-B	24.43	24.40	9.24	9.24	26.14	26.14
	1f-CIR-A	28.53	28.53	9.21	9.21	21.78	21.78
1f-Vas-A	1f-CIR-B	21.30	21.30	9.52	9.52	25.06	25.06
	1f-CIR-A	22.63	22.63	9.25	9.25	29.59	29.59
1f-CIR-B	1f-CIR-A	28.29	28.29	9.27	9.43	25.27	25.27

Eight estimation models are applied to two-factor Vasicek-simulated and two-factor CIR-simulated 1-year zero bond prices, respectively. The true parameters ($\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2$) are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to T' = $n \times \Delta t'$=300×1/365=0.82 years. The bootstrap bias correction process is applied to each set of estimates. There are 20,000 replications. The number of bootstrap resampling is 500. The corresponding sets of model estimates are used to make 12-month ahead forecasts. "*Vas-B*" and "*Vas-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. The numbers in cells are the proportion of 20,000 sample replications that the model #1 and the model #2 have NO equal predictive ability at the 5% significance level.

Table 3.7 Models' relative predictive ability test of 20-year zero yields

Panel A: Giacomini and White Test of 20-year yield from the 2-factor Vasicek Model

(%Path of #1 ≠ #2 out of 20,000 simulations over 12-month ahead forecasts)

		Base Setting		Setting B		Setting C	
Model #1	Model #2	2 Mediums	2 Longs	2 Mediums	2 Longs	2 Mediums	2 Longs
2f-Vas-B	2f-Vas-A	27.66	28.86	20.36	20.25	24.09	20.98
	2f-CIR-B	100.00	99.84	100.00	99.02	100.00	99.12
	2f-CIR-A	100.00	99.45	99.13	99.90	100.00	99.03
	1f-Vas-B	53.04	55.87	57.78	57.89	55.40	52.85
	1f-Vas-A	58.20	50.52	58.44	57.38	55.86	52.90
	1f-CIR-B	54.40	54.53	57.69	57.96	52.87	52.94
	1f-CIR-A	55.13	59.00	54.01	53.30	59.24	56.21
2f-Vas-A	2f-CIR-B	100.00	99.37	100.00	99.61	100.00	99.26
	2f-CIR-A	100.00	99.22	99.09	99.98	100.00	99.99
	1f-Vas-B	59.72	55.82	57.15	57.35	54.37	52.92
	1f-Vas-A	55.95	50.68	57.07	57.09	54.68	53.03
	1f-CIR-B	54.11	54.18	58.13	58.02	54.80	53.15
	1f-CIR-A	54.88	59.09	53.64	53.22	58.00	56.46
2f-CIR-B	2f-CIR-A	20.00	23.65	49.99	44.59	40.04	48.53
	1f-Vas-B	76.12	77.02	71.20	72.99	74.53	79.89
	1f-Vas-A	74.70	78.28	79.21	74.85	77.09	79.89
	1f-CIR-B	75.29	74.52	78.15	74.11	79.61	79.86
	1f-CIR-A	72.50	73.60	77.31	79.45	77.78	74.14
2f-CIR-A	1f-Vas-B	79.14	78.86	74.83	78.49	79.84	70.92
	1f-Vas-A	72.55	74.21	76.27	79.30	78.56	70.74
	1f-CIR-B	73.11	72.13	74.17	77.79	76.74	77.30
	1f-CIR-A	77.17	73.98	75.99	74.71	78.14	78.45
1f-Vas-B	1f-Vas-A	28.80	28.79	9.56	9.56	28.86	28.86
	1f-CIR-B	28.05	29.49	7.49	9.49	27.06	27.06
	1f-CIR-A	26.97	26.32	8.40	8.40	29.76	29.76
1f-Vas-A	1f-CIR-B	27.72	28.41	7.72	7.72	27.29	27.29
	1f-CIR-A	28.46	29.11	8.46	8.46	29.81	29.81
1f-CIR-B	1f-CIR-A	28.14	28.05	9.69	9.69	29.47	29.47

Table 3.7 (Continued)

Panel B: Giacomini and White Test of 20-year yield from the 2-factor CIR Model (%Path of #1 ≠ #2 out of 20,000 simulations over 12-month ahead forecasts)							
		Base Setting		Setting B		Setting C	
Model #1	Model #2	2 Mediums	2 Longs	2 Mediums	2 Longs	2 Mediums	2 Longs
2f-Vas-B	2f-Vas-A	26.20	23.46	49.10	48.67	49.27	48.97
	2f-CIR-B	99.71	99.55	99.11	99.22	99.99	99.27
	2f-CIR-A	99.00	99.49	99.11	99.21	100.00	99.49
	1f-Vas-B	71.80	73.27	78.29	77.95	72.40	77.59
	1f-Vas-A	76.99	77.03	77.95	77.99	77.19	70.72
	1f-CIR-B	76.48	79.20	78.07	78.09	74.63	71.25
	1f-CIR-A	76.90	73.40	77.97	78.50	71.53	70.75
2f-Vas-A	2f-CIR-B	99.17	99.48	99.63	99.59	99.98	99.67
	2f-CIR-A	99.92	99.73	99.68	99.61	100.00	99.24
	1f-Vas-B	73.40	75.07	78.51	76.83	77.62	70.35
	1f-Vas-A	71.59	70.76	78.66	76.85	78.50	76.92
	1f-CIR-B	77.49	75.26	77.75	77.03	70.08	78.64
	1f-CIR-A	79.64	74.06	78.73	76.28	72.30	72.25
2f-CIR-B	2f-CIR-A	22.43	27.01	17.96	18.35	15.89	12.73
	1f-Vas-B	54.58	55.51	59.03	59.03	52.40	55.66
	1f-Vas-A	56.95	50.89	56.08	55.94	50.73	54.89
	1f-CIR-B	55.59	55.28	56.47	56.46	58.33	55.54
	1f-CIR-A	53.32	53.28	56.64	58.90	59.70	59.68
2f-CIR-A	1f-Vas-B	59.44	55.44	59.07	59.04	56.73	53.67
	1f-Vas-A	56.48	50.89	56.03	55.97	55.59	52.72
	1f-CIR-B	59.70	55.46	56.48	56.50	59.12	52.84
	1f-CIR-A	53.07	53.19	56.65	58.89	52.74	58.40
1f-Vas-B	1f-Vas-A	28.88	28.88	5.40	9.40	29.85	29.85
	1f-CIR-B	24.38	24.38	9.75	9.75	25.88	25.88
	1f-CIR-A	20.72	20.72	9.95	9.87	24.44	24.44
1f-Vas-A	1f-CIR-B	25.75	25.75	6.61	6.61	24.06	24.06
	1f-CIR-A	24.94	24.94	7.58	8.30	25.84	25.84
1f-CIR-B	1f-CIR-A	24.29	24.29	9.53	28.42	27.76	27.76

Eight estimation models are applied to two-factor Vasicek-simulated and two-factor CIR-simulated 20-year zero bond prices, respectively. The true parameters ($\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2$) are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to T' = $n \times \Delta t'$=300×1/365=0.82 years. The bootstrap bias correction process is applied to each set of estimates. There are 20,000 replications. The number of bootstrap resampling is 500. The corresponding sets of model estimates are used to make 12-month ahead forecasts. "*Vas-B*" and "*Vas-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. The numbers in cells are the proportion of 20,000 sample replications that the model #1 and the model #2 have NO equal predictive ability at the 5% significance level.

Table 3.8 Comparison of models' relative predictive ability test of 1-year zero yields

Panel A: Giacomini and White Test of 1-year yield from the 2-factor Vasicek Model

(%Path of #1 Beats #2 out of 20,000 simulations over 12-month ahead forecasts)

		Base Setting		**Setting B**		**Setting C**	
Model #1	Model #2	2 Shorts	2 Mediums	2 Shorts	2 Mediums	2 Shorts	2 Mediums
2f-Vas-B	2f-Vas-A	15.18	13.13	8.55	8.70	14.99	10.16
	2f-CIR-B	99.86	99.31	99.60	99.71	99.93	99.59
	2f-CIR-A	99.81	99.28	98.98	99.26	99.55	99.09
	1f-Vas-B	28.18	30.19	27.59	28.14	28.50	28.17
	1f-Vas-A	25.15	26.61	27.74	28.06	27.47	27.86
	1f-CIR-B	29.32	28.69	28.75	28.46	28.31	27.98
	1f-CIR-A	26.62	26.53	28.43	28.68	25.73	27.12
2f-Vas-A	2f-CIR-B	99.23	99.24	99.61	99.65	99.92	99.77
	2f-CIR-A	99.04	98.99	99.58	99.52	99.87	99.71
	1f-Vas-B	29.95	30.11	27.80	27.85	28.28	28.37
	1f-Vas-A	27.21	25.78	27.69	28.02	27.73	27.86
	1f-CIR-B	29.86	29.02	28.63	28.98	29.55	29.15
	1f-CIR-A	29.57	28.64	28.79	28.55	28.30	28.53
2f-CIR-B	2f-CIR-A	10.72	14.14	22.38	23.58	20.82	24.05
	1f-Vas-B	8.33	8.17	9.45	9.07	8.54	9.02
	1f-Vas-A	8.63	7.55	8.77	9.22	9.11	9.55
	1f-CIR-B	7.09	7.64	9.60	9.27	8.61	8.95
	1f-CIR-A	8.15	8.40	8.85	9.68	8.67	8.80
2f-CIR-A	1f-Vas-B	8.10	8.32	7.80	8.02	7.42	7.81
	1f-Vas-A	7.64	7.97	7.89	7.96	7.45	7.90
	1f-CIR-B	7.31	7.79	8.23	8.15	8.04	8.52
	1f-CIR-A	8.00	7.25	8.33	7.96	8.27	7.83
1f-Vas-B	1f-Vas-A	13.28	12.67	5.42	4.60	15.03	15.26
	1f-CIR-B	14.02	13.98	4.60	5.00	14.86	14.71
	1f-CIR-A	13.41	11.30	4.57	4.97	14.28	14.13
1f-Vas-A	1f-CIR-B	13.77	13.71	4.64	4.75	13.59	13.59
	1f-CIR-A	14.38	13.76	4.64	4.57	14.73	14.88
1f-CIR-B	1f-CIR-A	10.61	10.46	4.47	4.98	14.48	14.41

Table 3.8 (Continued)

		Panel B: Giacomini and White Test of 1-year yield from the 2-factor CIR Model (%Path of #1 Beats #2 out of 20,000 simulations over 12-month ahead forecasts)					
		Base Setting		**Setting B**		**Setting C**	
Model #1	Model #2	2 Shorts	2 Mediums	2 Shorts	2 Mediums	2 Shorts	2 Mediums
2f-Vas-B	2f-Vas-A	11.66	14.73	23.29	21.99	24.79	20.65
	2f-CIR-B	2.91	2.11	2.18	2.08	2.31	2.47
	2f-CIR-A	2.39	2.77	2.14	2.42	2.59	2.03
	1f-Vas-B	7.64	7.66	7.96	7.24	7.98	8.27
	1f-Vas-A	7.65	8.12	8.33	7.71	8.06	8.10
	1f-CIR-B	8.07	7.77	7.98	7.36	7.29	7.70
	1f-CIR-A	7.66	7.67	7.99	7.80	8.24	8.67
2f-Vas-A	2f-CIR-B	2.89	2.11	2.22	2.36	2.72	2.15
	2f-CIR-A	2.05	2.42	2.24	2.38	2.53	2.72
	1f-Vas-B	7.89	8.15	7.15	7.42	8.64	7.96
	1f-Vas-A	7.46	7.53	7.17	7.83	7.99	7.93
	1f-CIR-B	8.23	8.08	7.05	7.19	7.55	7.67
	1f-CIR-A	7.98	7.41	7.63	7.27	8.51	8.17
2f-CIR-B	2f-CIR-A	12.79	13.13	9.51	9.48	10.25	9.98
	1f-Vas-B	27.54	27.09	27.65	27.86	29.27	27.24
	1f-Vas-A	25.31	29.96	27.75	27.74	29.58	27.58
	1f-CIR-B	26.81	26.67	28.07	27.79	26.73	27.24
	1f-CIR-A	29.34	28.71	27.60	28.05	29.36	27.94
2f-CIR-A	1f-Vas-B	27.02	27.45	27.96	28.02	27.07	26.61
	1f-Vas-A	29.07	30.08	27.95	27.77	27.20	27.22
	1f-CIR-B	26.23	26.98	28.35	28.10	27.41	27.43
	1f-CIR-A	28.22	29.11	28.13	27.89	26.76	27.22
1f-Vas-B	1f-Vas-A	10.54	10.50	4.33	4.88	10.98	46.42
	1f-CIR-B	12.45	12.36	4.63	4.68	13.28	13.09
	1f-CIR-A	14.28	14.28	4.62	4.63	10.97	10.93
1f-Vas-A	1f-CIR-B	10.68	10.69	4.78	4.77	12.58	12.65
	1f-CIR-A	11.42	11.44	4.71	4.67	14.89	15.08
1f-CIR-B	1f-CIR-A	14.28	14.30	4.70	4.80	12.68	12.66

Eight estimation models are applied to two-factor Vasicek-simulated and two-factor CIR-simulated 1-year zero bond prices, respectively. The true parameters $(\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2)$ are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to T' = $n \times \Delta t'$=300×1/365=0.82 years. The bootstrap bias correction process is applied to each set of estimates. There are 20,000 replications. The number of bootstrap resampling is 500. The corresponding sets of model estimates are used to make 12-month ahead forecasts. "*Vas-B*" and "*Vas-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. The numbers in cells are the proportion of 20,000 sample replications that the model #1 has greater predictive ability than the model #2 at the 5% significance level.

Table 3.9 Comparison of models' relative predictive ability test of 20-year zero yields

Panel A: Giacomini and White Test of 20-year yield from the 2-factor Vasicek Model

(%Path of #1 Beats #2 out of 20,000 simulations over 12-month ahead forecasts)

		Base Setting		**Setting B**		**Setting C**	
Model #1	Model #2	2 Mediums	2 Longs	2 Mediums	2 Longs	2 Mediums	2 Longs
2f-Vas-B	2f-Vas-A	13.89	14.49	10.25	10.19	12.17	10.54
	2f-CIR-B	99.97	99.80	99.97	98.99	99.96	99.07
	2f-CIR-A	99.96	99.42	99.09	99.86	99.96	98.98
	1f-Vas-B	26.97	28.46	29.08	29.45	28.02	26.47
	1f-Vas-A	29.32	25.27	29.73	28.90	28.39	26.56
	1f-CIR-B	27.36	27.76	29.08	29.42	26.71	26.68
	1f-CIR-A	27.89	29.58	27.43	27.13	30.15	28.57
2f-Vas-A	2f-CIR-B	99.96	99.33	99.96	99.56	99.95	99.22
	2f-CIR-A	99.96	99.19	99.04	99.94	99.96	99.95
	1f-Vas-B	30.22	28.27	29.12	28.93	27.45	26.47
	1f-Vas-A	28.04	25.51	28.78	28.55	27.42	26.79
	1f-CIR-B	27.51	27.12	29.53	29.18	27.44	26.89
	1f-CIR-A	27.98	29.77	27.05	26.71	29.24	28.34
2f-CIR-B	2f-CIR-A	10.12	11.94	25.47	22.57	20.07	24.59
	1f-Vas-B	7.62	8.27	7.71	7.32	7.59	8.27
	1f-Vas-A	7.95	8.46	8.22	7.58	8.42	8.54
	1f-CIR-B	8.09	8.04	8.31	7.89	8.75	8.22
	1f-CIR-A	7.54	7.80	8.01	8.50	7.93	7.55
2f-CIR-A	1f-Vas-B	8.23	8.12	7.92	8.44	8.59	7.54
	1f-Vas-A	7.52	7.77	7.68	8.14	8.40	7.35
	1f-CIR-B	7.93	7.88	8.11	7.96	8.03	8.37
	1f-CIR-A	8.07	8.04	7.90	8.10	8.41	8.59
1f-Vas-B	1f-Vas-A	14.62	14.43	4.86	4.87	14.53	14.52
	1f-CIR-B	14.26	14.88	3.78	4.82	13.58	13.73
	1f-CIR-A	13.55	13.18	4.25	4.25	14.91	14.98
1f-Vas-A	1f-CIR-B	13.98	14.31	3.86	3.88	13.75	13.87
	1f-CIR-A	14.45	14.73	4.30	4.30	15.05	15.08
1f-CIR-B	1f-CIR-A	14.20	14.09	4.86	4.87	14.85	15.00

Table 3.9 (Continued)

Panel B: Giacomini and White Test of 20-year yield from the 2-factor CIR Model (%Path of #1 Beats #2 out of 20,000 simulations over 12-month ahead forecasts)							
		Base Setting		**Setting B**		**Setting C**	
Model #1	Model #2	2 Mediums	2 Longs	2 Mediums	2 Longs	2 Mediums	2 Longs
2f-Vas-B	2f-Vas-A	13.16	11.80	24.77	24.52	24.71	24.68
	2f-CIR-B	2.02	2.77	2.08	2.08	2.81	2.87
	2f-CIR-A	2.06	2.03	2.28	2.55	2.32	2.48
	1f-Vas-B	7.88	7.39	8.10	8.27	7.42	7.96
	1f-Vas-A	7.79	8.25	7.96	8.56	7.77	7.50
	1f-CIR-B	8.26	8.07	7.89	7.83	7.48	7.71
	1f-CIR-A	7.96	8.02	8.55	8.20	7.46	7.24
2f-Vas-A	2f-CIR-B	2.88	2.51	2.75	2.93	2.55	2.95
	2f-CIR-A	2.39	2.67	2.34	2.61	2.23	2.07
	1f-Vas-B	7.36	7.86	8.55	7.80	7.87	7.44
	1f-Vas-A	7.50	7.08	8.37	7.93	8.26	8.42
	1f-CIR-B	8.38	7.67	8.15	8.05	7.38	8.23
	1f-CIR-A	8.04	8.11	7.98	8.30	7.25	7.26
2f-CIR-B	2f-CIR-A	11.37	13.77	9.07	9.22	8.03	6.37
	1f-Vas-B	27.82	28.02	29.85	29.98	26.35	28.23
	1f-Vas-A	28.50	25.63	28.17	28.25	25.56	27.67
	1f-CIR-B	27.97	28.07	28.30	28.36	29.70	28.19
	1f-CIR-A	27.00	27.01	28.33	29.91	30.36	29.96
2f-CIR-A	1f-Vas-B	29.80	28.20	30.10	29.70	28.49	27.19
	1f-Vas-A	28.66	25.64	28.12	28.31	27.81	26.69
	1f-CIR-B	30.07	27.99	28.78	28.40	29.88	26.75
	1f-CIR-A	27.07	26.96	28.88	29.72	26.65	29.66
1f-Vas-B	1f-Vas-A	14.46	14.64	2.70	4.73	14.96	15.02
	1f-CIR-B	12.43	12.42	4.90	4.88	13.11	12.99
	1f-CIR-A	10.49	10.46	4.98	5.01	12.33	12.24
1f-Vas-A	1f-CIR-B	13.04	12.95	3.32	3.33	12.11	12.15
	1f-CIR-A	12.51	12.58	3.86	4.18	12.96	13.06
1f-CIR-B	1f-CIR-A	12.30	12.24	4.81	14.28	13.95	13.96

Eight estimation models are applied to two-factor Vasicek-simulated and two-factor CIR-simulated 20-year zero bond prices, respectively. The true parameters ($\kappa_1, \theta_1, \sigma_1, \lambda_1, \kappa_2, \theta_2, \sigma_2, \lambda_2$) are a) 2-factor Vasicek model: (0.473, 0.046, 0.087, -0.107, 0.043, 0.019, 0.024, -0.045); b) 2-factor CIR model: (0.654, 0.038, 0.150, -0.126, 0.054, 0.022, 0.041, -0.048). Base Setting: Each MLE process has 300 months of zero bond prices (T=25 years) with two maturities. $\Delta t = \frac{1}{12}$ corresponds to monthly observations in an annualized basis. Setting B: Change the sampling frequency of the data from the monthly basis ($\Delta t = 1/12$) to the daily basis ($\Delta t' = 1/365$), while keeping the time span $T = 25$ years. Thus, the sample size is increased from n = 300 to n' = 9,125. The latter corresponds to high frequency data. Setting C: Change the sampling frequency of the data from the monthly basis (Δt=1/12) to the daily basis ($\Delta t'$=1/365) while keeping the sample size n = 300. Correspondingly, the time span is reduced from T=25 years to T' = $n \times \Delta t'$=300×1/365=0.82 years. The bootstrap bias correction process is applied to each set of estimates. There are 20,000 replications. The number of bootstrap resampling is 500. The corresponding sets of model estimates are used to make 12-month ahead forecasts. "*Vas-B*" and "*Vas-A*" represent the sets of MLE-Vasicek estimates, with "*B*" indicating before the bias correction and "*A*" indicating after the bias correction. Similarly, "*CIR-B*" and "*CIR-A*" are the corresponding sets of estimates for the MLE-CIR. The numbers in cells are the proportion of 20,000 sample replications that the model #1 has greater predictive ability than the model #2 at the 5% significance level.

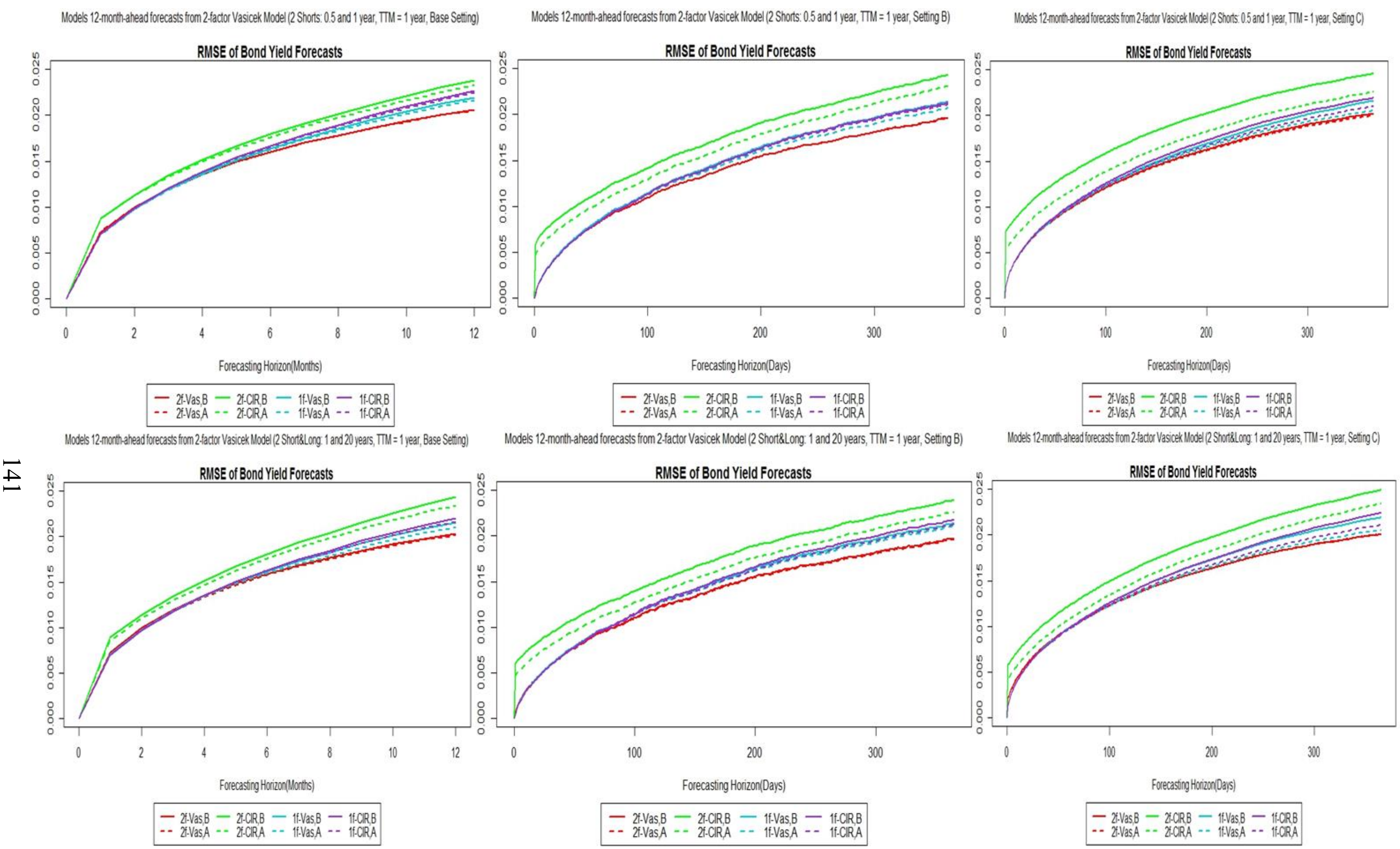

Figure 3.1 12-month-ahead forecasts of 1-year bond yields from two-factor Vasicek model

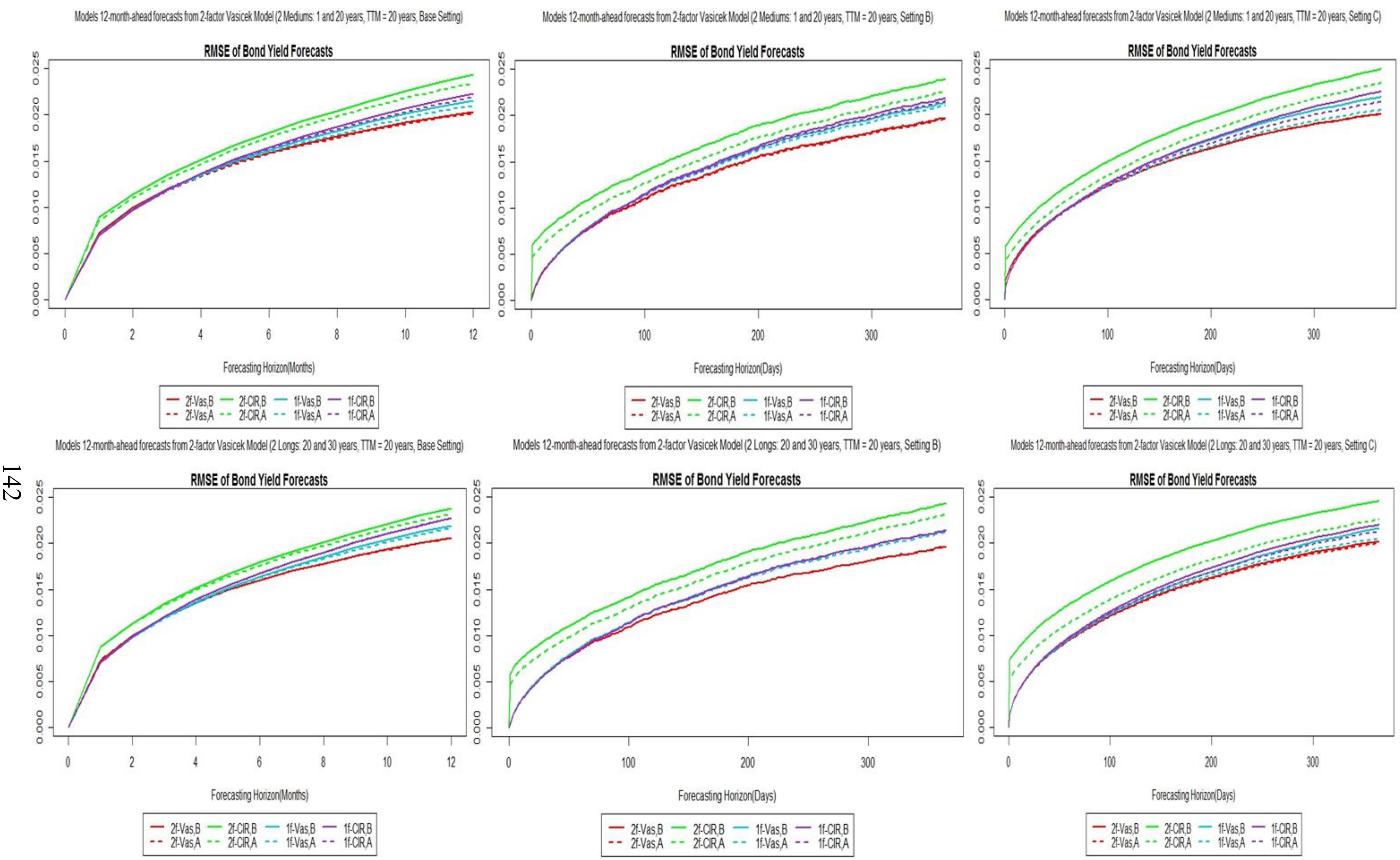

Figure 3.2 12-month-ahead forecasts of 20-year bond yields from two-factor Vasicek model

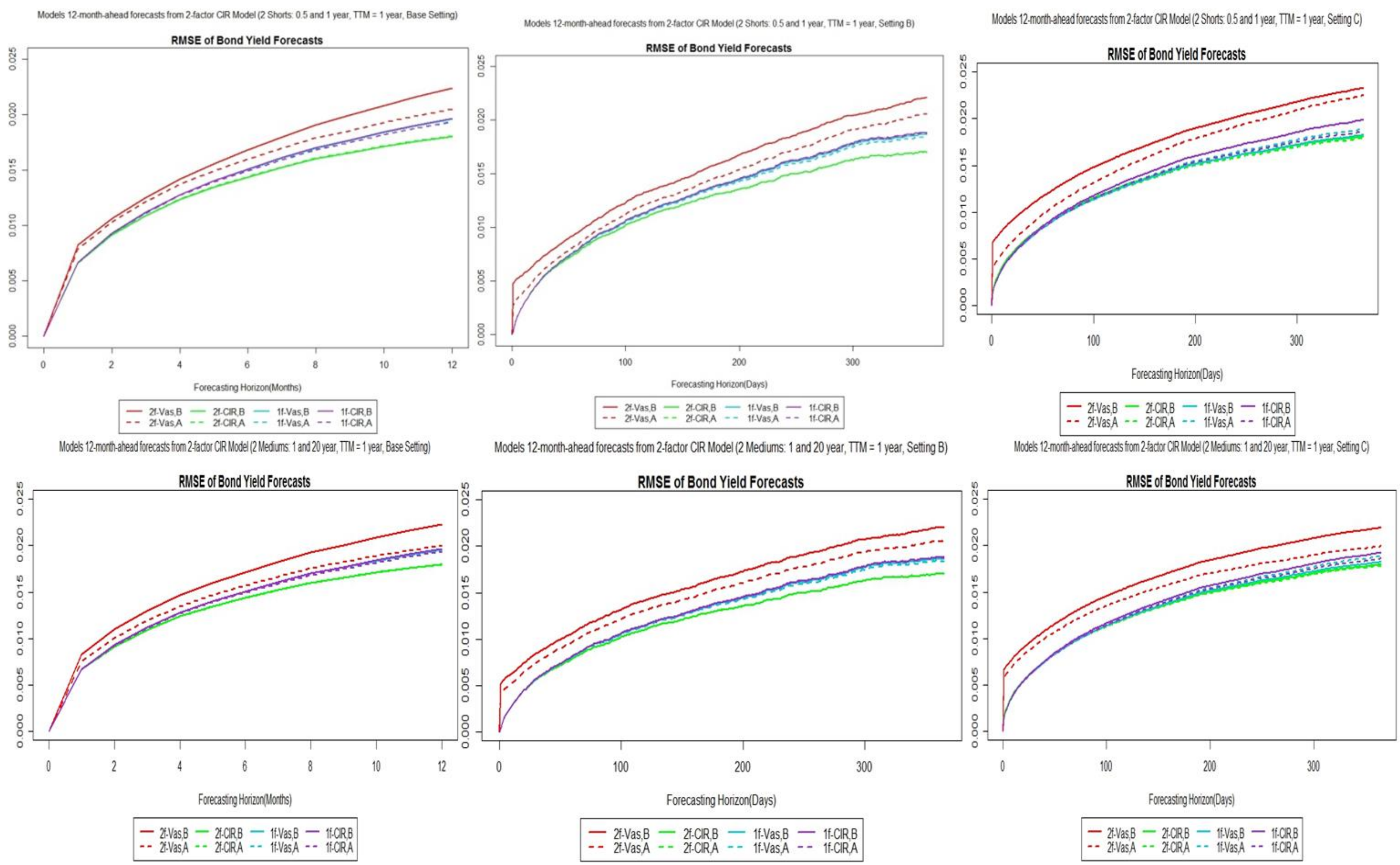

Figure 3.3 12-month-ahead forecasts of 1-year bond yields from two-factor CIR model

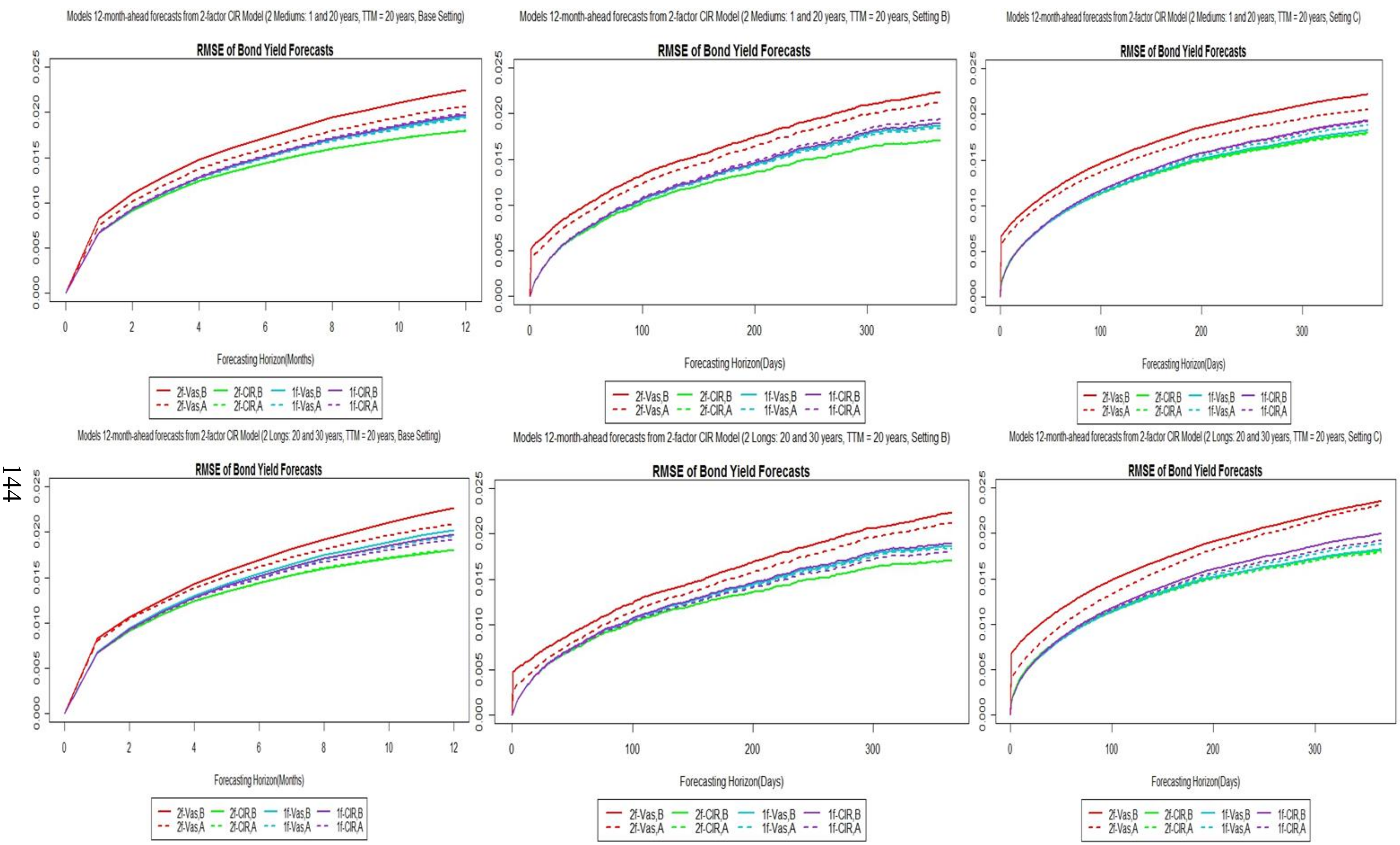

Figure 3.4 12-month-ahead forecasts of 20-year bond yields from two-factor CIR model

3.6 References

Andrew, D.W.K., 1993, Exactly median-unbiased estimation of first order autoregressive/unit root models, *Econometrica* 61, 139-166.

Ait-Sahalia, Yacine, 1996, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527-560.

Arnold, Ludwig, 1974. Stochastic differential equations: Theory and applications (Wiley, New York).

Ball, C.A. and W.N. Torous, 1996, Unit roots and the estimation of interest rate dynamics, *Journal of Empirical Finance* 3, 215-238.

Bliss, R.R., 1997, Testing term structure estimation methods, *Advances in Futures and Options Research* 9, 191-231.

Brandt, M.W. and Pedro Santa-Clara, 2002, Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets, *Journal of Financial Economics* 63, 161-210.

Brigo, Damiano and Fabio Mercurio, 2006. Interest rate models -theory and practice with smile, inflation and credit (Springer).

Broadie, M. and Ozgur Kaya, 2006, Exact simulation of stochastic volatility and other affine jump diffusion processes, *Operations Research* 54, 217-231.

Brown, S.J. and Philip H. Dybvig, 1986, The empirical implications of the cox, ingersoll, ross theory of the term structure of interest rates, *Journal of Finance* 41, 617-630.

Chan, K.C., G.A.Karolyi, F.A.Longstaff, and A.B.Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209-1227.

Chen, R.R. and L. Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14-31.

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

Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.

Dixit, A.K. and Pindyck R.S., 1994. *Investment under uncertainty* (Princeton University Press, Princeton, New Jersey).

Duarte, Jefferson, 2004, Evaluating an alternative risk preference in affine term structure models, *Review of Financial Studies* 17, 379-404.

Duffee, G.R., 2002, Term permia and interest rate forecasts in affine models, *Journal of Finance* 57, 405-443.

Duffie, D. and R. Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379-406.

Duffie, J.D. and Kenneth J. Singleton, 1993, Simulated moments estimation of markov models of asset prices, *Econometrica* 61, 929-952.

Fama, E.F., and R.R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680-692.

Fisher, M., D. Nychka, and D. Zervos, 1995, Fitting the term structure of interest rates with smoothing splines, *95-1, Finance and Economics Discussion Series, Federal Reserve Board*.

Gallant, A. R. and G. Tauchen, 1996, Which moments to match?, *Econometric Theory* 12, 657-681.

Giacomini, R. and H. White, 2006, Tests of coditional predictive ability, *Econometrica* 74, 1545-1578.

Gibbons, M. and K. Ramaswamy, 1993, A test of the cox, ingersoll, and ross model of the term structure, *Review of Financial Studies* 6(3), 619-658.

Longstaff, F.A., 1989, A nonlinear general equilibrium model of the term structure of interest rates, *Journal of Financial Economics* 23, 195-224.

McCulloch, J.H., 1975, The tax-adjusted yield curve, *Journal of Finance* 30, 811-830.

Nelson, C.R., and A.F. Siegel, 1987, Parsimonious modeling of yield curves, *Journal of Business* 60, 473-489.

Pasquariello, Paolo, Jennifer Roush and Clara Vega, 2013, Government intervention and strategic trading in the U.S. Treasury market, *Working paper*.

Pearson, N.D. and T.S. Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the cox, ingersoll, and ross model, *Journal of Finance* 49, 1279-1304.

Pedersen, A.R., 1995, A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics* 22, 55-71.

Phillips, P.C., 1972, The structural estimation of a stochastic differential equation system, *Econometrica* 40, 1021-1041.

Phillips, P.C. and J. Yu, 2009, Maximum likelihood and gaussian estimation of continuous time models in finance, *Handbook of Financial Time Series* 497-530.

Smith, A.A., 1993, Estimating nonlinear time-series models using simulated vector autoregressions, *Journal of Applied Econometrics* 8, S63-S84.

Tang, C.Y and S.X. Chen, 2009, Parameter estimation and bias correction for diffusion processes, *Journal of Econometrics* 149, 65-81.

Vasicek, Oldrich A., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

Yu, Jun and P.C. Phillips, 2001, A gaussian approach for estimating continuous time models of short term interest rates, *The Econometrics Journal* 4, 211-225.