Enhancing Estimation for Interest Rate Diffusion Models with Bond Prices

Zou, Tao and Chen, Song Xi

Peking University, Peking University

January 2014

Enhancing Estimation for Interest Rate Diffusion Models with Bond Prices

Tao Zou and Song Xi Chen*

Department of Business Statistics and Econometrics

Guanghua School of Management and Center for Statistical Science

Peking University

Abstract

We consider improving estimating parameters of diffusion processes for interest rates by incorporating information in bond prices. This is designed to improve the estimation of the drift parameters, which are known to be subject to large estimation errors. It is shown that having the bond prices together with the short rates leads to more efficient estimation of all parameters for the interest rate models. It enhances the estimation efficiency of the maximum likelihood estimation based on the interest rate dynamics alone. The combined estimation based on the bond prices and the interest rate dynamics can also provide inference to the risk premium parameter. Simulation experiments were conducted to confirm the theoretical properties of the estimators concerned. We analyze the overnight Fed fund rates together with the U.S. Treasury bond prices.

JEL CLASSIFICATION: C50, C58.

Key words: Interest Rate Models; Affine Term Structure; Bond Prices; Market Price of Risk; Combined Estimation; Parameter Estimation.

*Corresponding Author: +86 1062750427; fax: +86 1062753624. Email: csx@gsm.pku.edu.cn

1 Introduction

Interest rate models especially those for the short rates, as basic financial instruments and measures for the risk-free assets, have attracted much attention in financial and econometric studies. Modeling the term structure of the interest rates is a focal point of these studies. Diffusion processes constitute a popular class of models for the interest rate dynamics. The Vasicek and the CIR diffusion models, introduced in Vasicek (1977) and Cox, Ingersoll and Ross (1985), and the much broader affine term structure models (Duffie and Kan, 1996; Dai and Singleton, 2000; Duffee, 2002) are the basic interest rate models for pricing the zero-coupon or coupon bearing bonds and interest rate derivatives.

It is known that the drift parameters of the diffusion processes are more difficult to estimate than the diffusion parameters, as shown in Phillips and Yu (2005) and Tang and Chen (2009). This is because that the drift part contains far less information since it is of a smaller order as compared to the diffusion part. Despite this understanding, the pricings of bonds and interest rate derivatives require better estimation of the drift parameters as well as parameters which define the risk premium process. The latter cannot be identified under the physical measure.

In this paper, we consider estimating parameters of interest rate diffusion processes by utilizing the interest rate data along with the bond prices. We first analyze the least squares estimation based on the converted zero-coupon bond prices only under the affine term structure models without using the interest rate data. Although it is known (Brown and Dybvig, 1986) that the least squares estimation cannot identify all the parameters due to a collinearity, we provide explicit descriptions on which linear combination of the original parameters can be identified, and propose a method that selects the largest number of equations from the redundant least squares estimating equations.

To utilize the bond prices, we propose a framework that combines the short rate data and the model with the bond prices to improve the parameter estimation of the short rate parameters. The combined estimation is designed to achieve two goals. One is to improve the estimation efficiency of the maximum likelihood estimator (MLE) based on the interest rate data only. The second goal is to identify all the parameters including those of the risk premium. Since the combined estimation has the extra bond prices and their model information, it enhances the estimation based on the interest rates only. This is attractive as it improves the MLE of the drift parameters which are known to have larger estimation errors (Tang and Chen, 2009). We analyze the Federal fund overnight rates together with the treasury bond prices from 1972 to 2012 to demonstrate our proposal.

The paper is structured as follows. Section 2 introduces the interest rate models and the associated bond pricing. Section 3 analyzes the issues for the least squares estimation based on the bond prices. Section 4 proposes the combined estimation approach that utilizes both the interest rates and bond prices, whose theoretical properties are given in Section 5. Numerical results from simulation experiments which compared different estimators are reported in Section 6. Section 7 analyzes the overnight rates of the Federal Reserve and the U.S. Treasury bond prices. A conclusion is made in Section 8. Assumptions and theoretical proofs are relegated to the Appendix.

2 Interest Rate Models and Bond Prices

Let $r(t)$ be the short rate at time t. Under the physical measure Q_0, the short rate follows a diffusion process

$$dr(t) = \mu_0 \{t, r(t); \beta\} \, dt + \sigma \{t, r(t); \beta\} \, dW_0(t), \tag{2.1}$$

where $\mu_0(\cdot)$ and $\sigma(\cdot)$ are respectively the drift and diffusion functions, β is a $q \times 1$ vector containing the model parameters and $W_0(t)$ is the standard Brownian motion under Q_0. The maximum likelihood estimation (MLE) has been a popular method for parameter estimation. Suppose that the short rates are stationary and we observe the short rates at equally spaced time interval δ: $r(0), r(\delta) \cdots, r(n\delta)$. To simplify notations we write $r(t\delta)$ as r_t by hiding δ.

Let $f_t(r_t|r_{t-1}, \delta; \beta)$ be the transition density of r_t given r_{t-1}, implied from (2.1) under the physical measure Q_0. The log-likelihood function of the parameter β is $\sum_{t=1}^{n} \ell_t(\beta)$, where

$$\ell_t(\beta) = \log f_t(r_t | r_{t-1}, \delta; \beta). \tag{2.2}$$

The MLE $\tilde{\beta}_n$ solves the score equation $\sum_{t=1}^{n} \frac{\partial \ell_t(\beta)}{\partial \beta} = 0$. If the diffusion process (2.1) is time homogeneous and stationary, the consistency and the asymptotic normality of the MLE have been well understood; see for instance Chang and Chen (2011). If the diffusion process (2.1) is time inhomogeneous, the likelihood score is still a sum of martingale differences, but the differences are no longer identically distributed. The asymptotic normality of the MLE can still be established based on the martingale central limit theorems (Hall and Heyde, 1980).

Popular models for the short rates include the Vasicek model (Vasicek, 1977) which is $dr(t) = \kappa \{\alpha - r(t)\} dt + \sigma dW_0(t)$, namely an Ornstein-Uhlenbeck process under Q_0. Cox et al. (1985) proposed using Feller (1951)'s square root diffusion process $dr(t) = \kappa \{\alpha - r(t)\} dt + \sigma \sqrt{r(t)} dW_0(t)$ to model the short rates with positive parameters κ, α and σ such that $2\kappa\alpha/\sigma^2 > 1$. The analytical forms of the transition densities for these two processes are known to be the densities of a normal and a non-central chi-squared ones, respectively, which facilitate the MLEs. For interest rate diffusion models whose transition densities are unknown, which is often the case, Aït-Sahalia (1999, 2002)'s approximate MLE can be employed.

Despite the MLE or the approximate MLE being consistent and asymptotically normal, the estimation for the drift parameters encounters a slower rate of convergence ($\sqrt{n\delta}$) and a large order of bias ($(n\delta)^{-1}$) as revealed in Tang and Chen (2009). In contrast, the convergence rate of the estimation for the diffusion parameter is $\sqrt{n}$ and the bias is of order n^{-1}, which are much smaller than those of the drift parameter.

A new initiative is needed to improve the parameter estimation as the pricing of bonds and interest rate derivatives requires more accurate estimation of the drift parameters as well as parameters which define the risk premium process. The latter cannot be identified in the short rate process under the physical measure. Our proposal is to bring in bond prices under an interest rate diffusion model to produce a more efficient combined estimation of the parameters.

Let us first outline the basics on the bond pricing framework. Let $P(t, s)$ be the price of a zero-coupon bond at time t that matures at a future time $s > t$. In order to discuss the bond pricing theory, the short rate given in (2.1) is considered under the risk-neutral measure Q_1:

$$\begin{aligned} dr(t) &= \mu_1 \{t, r(t); \theta\} dt + \sigma \{t, r(t); \theta\} dW_1(t) \\ &= [\mu_0 \{t, r(t); \beta\} + \sigma \{t, r(t); \beta\} \Lambda \{t, r(t); \lambda\}] dt + \sigma \{t, r(t); \beta\} dW_1(t), \end{aligned} \tag{2.3}$$

where $W_1(t)$ is the standard Brownian motion under Q_1, $\theta = (\beta', \lambda')'$ is a $(q + d) \times 1$ vector of parameters with a new d-dimensional parameter λ that defines the market price of risk, and $\Lambda(t) := \Lambda \{t, r(t); \lambda\}$ is the market price of risk process relying on the parameter λ. The two measures Q_0 and Q_1 are connected through the Girsanov change of measure.

If $r(t)$ follows an one-factor affine term structure model, namely

$$\mu_1 \{t, r(t); \theta\} = K_0(t; \theta) + K_1(t; \theta)r(t) \text{ and } \sigma^2 \{t, r(t); \theta\} = H_0(t; \theta) + H_1(t; \theta)r(t), \tag{2.4}$$

for some deterministic functions $K_0(\cdot), K_1(\cdot), H_0(\cdot)$ and $H_1(\cdot)$ of t and θ, respectively, then based on the no-arbitrage pricing theory, the bond price $P(t, s)$ is shown to satisfy (Duffie, 2001)

$$-\log P(t, s) = A(t, s; \theta) + B(t, s; \theta)r(t), \tag{2.5}$$

and the pricing functions $A(t, s; \theta)$ and $B(t, s; \theta)$ are determined by the Riccati differential equation

$$\frac{\partial B(t, s; \theta)}{\partial t} = \frac{1}{2} H_1(t; \theta)B^2(t, s; \theta) - K_1(t; \theta)B(t, s; \theta) - 1; \;\; B(s, s; \theta) = 0 \tag{2.6}$$

and an integral equation

$$A(t, s; \theta) = \int_t^s \left\{ K_0(u, \theta)B(u, s; \theta) - \frac{1}{2} H_0(u; \theta)B^2(u, s; \theta) \right\} du. \tag{2.7}$$

To illustrate the key ingredients in the affine term structure, we consider two specific affine models: the Vasicek and CIR models. Under the risk-neutral measure Q_1, the Vasicek model follows (Brigo and Mercurio, 2006) $dr(t) = [\kappa \{\alpha - r(t)\} + \sigma\lambda r(t)] dt + \sigma dW_1(t)$, where λ is the univariate market price of risk parameter, while the CIR model admits $dr(t) = [\kappa \{\alpha - r(t)\} + \sigma\lambda r(t)] dt + \sigma\sqrt{r(t)}dW_1(t)$. Both the Vasicek and CIR models have explicit expressions of $B(t, s; \theta)$ and $A(t, s; \theta)$ by solving (2.6) and (2.7) (given in Brigo and Mercurio, 2006). For other affine term structure models which do not have explicit $B(\cdot)$ and $A(\cdot)$, numerical solutions to the differential equation (2.6) and the integral equation (2.7) can be attained, for example using the Runge-Kutta discretization method in Hairer, Nøersett and Wanner (2006).

3 Generalized Least Squares Estimation

It is worth noting that the observed bonds in a fixed income market are most likely coupon-bearing. There are methods to convert coupon-bearing bond prices to the zero-coupon bond prices, such as the bootstrap method of Hull (2009), the parametric method of Nelson and Siegel (1987) and Svensson (1994), and the spline method of McCulloch (1975, 1993) and Vasicek and Fong (1982).

Suppose that by one of the above conversion methods, at a date t we have M zero-coupon bonds with time to maturities $\tau_1, \tau_2, \cdots, \tau_M$ which do not depend on t. Let $p_{it} = -\log P(t\delta, t\delta + \tau_i)$ be the transformed zero-coupon bond price at time t with maturity τ_i. As consequences of the conversion procedures to get the zero-coupon bond prices and the uncertainty with the models and the randomness in the observed prices, measurement errors are inevitably present in the observed bond prices. Hence, p_{it} can deviate from (2.5) such that

$$p_{it} = A_{it}(\theta_0) + B_{it}(\theta_0) r_t + u_{0it}, \text{ for } i = 1, \cdots, M; \ t = 1, \cdots, n; \tag{3.1}$$

where u_{0it} denotes the pricing error, $A_{it}(\theta) := A(t\delta, t\delta + \tau_i; \theta_0)$, $B_{it}(\theta) := B(t\delta, t\delta + \tau_i; \theta_0)$ and θ_0 is the true parameter. In the above, i is a shorter version of τ_i. Model (3.1) has been considered in literatures (Pearson and Sun, 1994; Duffee, 2002; Cheridito, Filipović and Kimmel, 2007; Aït-Sahalia and Kimmel, 2010) and the number of bonds M used is rather smaller than the sample size n. Hence we assume M is fixed.

Let $p_t = (p_{1t}, \cdots, p_{Mt})', A_t(\theta) = (A_{1t}(\theta), \cdots, A_{Mt}(\theta))', B_t(\theta) = (B_{1t}(\theta), \cdots, B_{Mt}(\theta))'$ and $u_{0t} = (u_{01t}, \cdots, u_{0Mt})'$. Then, (3.1) can be written as

$$p_t = A_t(\theta_0) + B_t(\theta_0) r_t + u_{0t}. \tag{3.2}$$

Suppose the measurement errors $\{u_{0t}\}$ form a martingale difference sequence with respect to the filtration $\{\mathcal{G}_t\}$, where $\mathcal{G}_t$ is the σ-algebra generated by $\{(r_{l+1}, u'_{0l})'\}_{l \le t}$. The generalized least squares (GLS) estimator of θ can be attained by minimizing

$$\sum_{t=1}^{n} \{p_t - A_t(\theta) - B_t(\theta) r_t\}' W \{p_t - A_t(\theta) - B_t(\theta) r_t\}, \tag{3.3}$$

for an $M \times M$ positive definite weighting matrix W. Then the GLS estimator solves

$$\sum_{t=1}^{n} g_t(\theta; W) = 0, \tag{3.4}$$

where $g_t(\theta; W) = \left\{ \frac{\partial A_t(\theta)}{\partial \theta'} + \frac{\partial B_t(\theta)}{\partial \theta'} r_t \right\}' W u_t(\theta)$ for $u_t(\theta) = p_t - A_t(\theta) - B_t(\theta) r_t$.

However, (3.4) cannot identify all of the parameters in θ. This is because the dynamics under the risk neutral measure can be specified with a parameter transformation $\vartheta = \vartheta(\theta)$ whose dimension is less than that of θ. This implies that Model (2.3) under the risk neutral measure Q_1 can be written as

$$dr(t) = \tilde{\mu}_1 \{t, r(t); \vartheta\} + \tilde{\sigma} \{t, r(t); \vartheta\} dW_1(t). \tag{3.5}$$

For instance, the Vasicek and CIR models in Section 2 can be expressed with three parameters under the risk neutral measure via a new parameterization: $b = \kappa - \sigma\lambda$, $a = \frac{\kappa\alpha}{\kappa - \sigma\lambda}$ and σ, rather than the four parameters in $\theta = (\kappa, \alpha, \sigma, \lambda)'$. As a result, the pricing functions $A_t(\theta)$ and $B_t(\theta)$ may be written via the smaller set ϑ so that $B_t(\theta) = \tilde{B}_t(\vartheta)$ and $A_t(\theta) = \tilde{A}_t(\vartheta)$ in (2.6) and (2.7). This means that (3.4) has redundant equations. The redundancy has been noticed in the literatures, for instance in Brown and Dybvig (1986). However, what has not been considered in the literatures is the selection of the non-redundant equations in (3.4) and how to use them to improve the estimation of parameters. We will investigate these issues in the following section.

4 Combined Estimation

In this section, we propose an estimation method that combines the MLE approach based on the interest rate dynamics, with the use of the non-redundant equations in the GLS estimation based on the bond prices. This combination of two data sources and model information can improve the estimation efficiency of the MLE revealed in Section 2, and enable the estimation of the risk premium parameter, as well as repair the identification issue of the GLS estimation.

If the short rate under (2.1) and (2.3) is time homogenous, then one can show that the transition density $f_t(\cdot)$ as well as $A_t(\cdot)$ and $B_t(\cdot)$ are time homogeneous too such that $f_t(r_t|r_{t-1}, \delta; \beta) = f(r_t|r_{t-1}, \delta; \beta)$, $A_t(\theta) = A(\theta)$ and $B_t(\theta) = B(\theta)$. We focus on the time homogeneous case in the following. Extension to the time inhomogeneous case can be made with more involved notations and technical details, which is discussed in Section 8.

To start with, we select a set of non-redundant equations in (3.4), denoted by $\mathcal{E}' g_t(\theta; W)$ where $\mathcal{E}$ is a matrix consisting of $q^\dagger$ columns of the identity matrix I_{q+d} and $q^\dagger < q + d$ is the maximum number of non-redundant equations. As $\sum_{t=1}^{n} \mathcal{E}' g_t(\theta; W) = 0$ can not identify θ, we combine it with the likelihood score to form a combined generalized method of moment (GMM) equations

$$h_t(\theta; \mathcal{E}, W) = \begin{pmatrix} \frac{\partial \ell_t(\beta)}{\partial \beta} \\ \mathcal{E}' g_t(\theta; W) \end{pmatrix}, \tag{4.1}$$

which has $q + q^\dagger$ moment conditions for $q + d$ unknown parameters. It is noted that, at θ_0,

$$\mathbb{E}\left\{h_t(\theta_0; \mathcal{E}, W)\right\} = 0. \tag{4.2}$$

Let $V_0 = \mathbb{E}\left(u_{0t} u_{0t}' | \mathcal{G}_{t-1}\right) =: (v_{jk})_{M \times M}$ be the conditional covariance matrix of the measurement errors, which is assumed to be of full rank. For a given W, the optimal GMM estimation utilizes a weighting matrix which is the inverse of the long-run covariance

$$\Sigma_h(\theta_0; \mathcal{E}, W) =: \lim_{n \to \infty} n \mathrm{Var}\left\{\frac{1}{n} \sum_{t=1}^{n} h_t(\theta_0; \mathcal{E}, W)\right\}. \tag{4.3}$$

This implies that $q^\dagger$ and $\mathcal{E}$ should be chosen properly to make $\Sigma_h(\theta_0; \mathcal{E}, W)$ invertible.

Let $\mathcal{I}_0(\delta)$ be the Fisher information matrix associated with the likelihood score for β,

$$\psi(\theta_0) := \begin{pmatrix} \frac{\partial A(\theta_0)}{\partial \theta'} + \frac{\partial B(\theta_0)}{\partial \theta'} \mathbb{E}(r_t) \\ \frac{\partial B(\theta_0)}{\partial \theta'} \sqrt{\mathrm{Var}(r_t)} \end{pmatrix}$$

and $\Xi_0(\mathcal{E}, W) = \mathcal{E}' \psi(\theta_0)' \left\{I_2 \otimes (W V_0 W)\right\} \psi(\theta_0) \mathcal{E}$.

Proposition 4.1. *Under Assumptions 1 - 5 given in Appendix, and for any $\delta \in (0, \Delta_1]$ where $\Delta_1 > 0$ is a finite constant, $\Sigma_h(\theta_0; \mathcal{E}, W) = diag\{\mathcal{I}_0(\delta), \Xi_0(\mathcal{E}, W)\}$.*

The proposition implies that $\Sigma_h(\theta_0; \mathcal{E}, W)$ is invertible if and only if both $\mathcal{I}_0(\delta)$ and $\Xi_0(\mathcal{E}, W)$ are invertible. Since $\mathcal{I}_0(\delta)$, V_0 and W are nonsingular, we only require $\psi(\theta_0)\mathcal{E}$ to be of full rank for the largest possible $q^\dagger$.

We select $\mathcal{E}$ from the following set

$$\{\mathcal{E} : \psi(\theta_0)\mathcal{E} \text{ form a largest collection of linearly independent columns of } \psi(\theta_0)\} \tag{4.4}$$

where $q^\dagger := \text{rank}\{\psi(\theta_0)\} = \text{rank}\{\psi(\theta_0)\mathcal{E}\}$. As (4.4) has more than one element, different $\mathcal{E}$s in (4.4) select different non-redundant equations in the score $g_t(\theta; W)$. Theorem 5.2 will show that the combined estimators attain the same asymptotic efficiency despite using different $\mathcal{E}$s in (4.4).

For both the Vasicek and CIR models, it is illustrated in the supplementary material that if there is only one bond available at each t, namely $M = 1$, $\mathcal{E}$ can be any two columns of I_4 with $q^\dagger = 2$; and if there are at least two bonds, namely $M \geq 2$, $\mathcal{E}$ can consist of three columns of I_4 which must has the third column of I_4, with $q^\dagger = 3$. We note that the third column corresponds to the diffusion parameter σ for $\theta = (\kappa, \alpha, \sigma, \lambda)'$. Since $q^\dagger$ is at least 2 which is larger than $d = 1$ for the Vasicek and CIR models, the proposed combined estimation is able to identify all the parameters including the market price of risk parameter.

In order to carry out the GMM estimation, an initial estimation of θ is needed to estimate the weighting matrix $\Sigma_h^{-1}(\theta_0; \mathcal{E}, W)$, which is

$$\bar{\theta}_n(\mathcal{E}, W) = \arg\min_{\theta\in\Theta} \left\{\frac{1}{n}\sum_{t=1}^{n} h_t(\theta; \mathcal{E}, W)\right\}' \left\{\frac{1}{n}\sum_{t=1}^{n} h_t(\theta; \mathcal{E}, W)\right\}.$$

Let $\hat{\mathcal{I}}_n(\beta) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial \ell_t(\beta)}{\partial \beta}\frac{\partial \ell_t(\beta)}{\partial \beta'}$ and $\hat{\Xi}_n(\theta; \mathcal{E}, W) = \frac{1}{n}\sum_{t=1}^{n} \mathcal{E}' g_t(\theta; W) g_t(\theta; W)'\mathcal{E}$. Write $\bar{\theta}_n(\mathcal{E}, W) = \left(\bar{\beta}_n(\mathcal{E}, W)', \bar{\lambda}_n(\mathcal{E}, W)'\right)'$. Define the estimated weighting matrix

$$\hat{\mathscr{W}}_n(\mathcal{E}, W) := \text{diag}\left\{\hat{\mathcal{I}}_n^{-1}\left(\bar{\beta}_n(\mathcal{E}, W)\right), \hat{\Xi}_n^{-1}\left(\bar{\theta}_n(\mathcal{E}, W); \mathcal{E}, W\right)\right\}.$$

The proposed combined (GMM) estimator for θ, consisting of both the interest rate parameter β and the risk premium parameter λ, is

$$\hat{\theta}_n(\mathcal{E}, W) = \arg\min_{\theta\in\Theta} \left\{ \frac{1}{n} \sum_{t=1}^{n} h_t(\theta; \mathcal{E}, W) \right\}' \hat{\mathscr{W}}_n(\mathcal{E}, W) \left\{ \frac{1}{n} \sum_{t=1}^{n} h_t(\theta; \mathcal{E}, W) \right\}. \tag{4.5}$$

5 Theoretical Results

The theoretical properties of the combined estimator for θ are presented in this section. We first need to define a few matrices to convey the asymptotic normality of the combined estimator. Let

$$G_0(\mathcal{E}, W) := \mathbb{E}\left\{ \frac{\partial \mathcal{E}' g_t(\theta_0; W)}{\partial \theta'} \right\} = -\mathcal{E}'\psi(\theta_0)' \left(I_2 \otimes W\right) \psi(\theta_0), \tag{5.1}$$

$$\mathcal{H}_0(\delta; \mathcal{E}, W) := \mathbb{E}\left\{ \frac{\partial h_t(\theta_0; \mathcal{E}, W)}{\partial \theta'} \right\} = \begin{pmatrix} (-\mathcal{I}_0(\delta), 0_{q\times d}) \\ G_0(\mathcal{E}, W) \end{pmatrix} \text{ and}$$

$$\begin{aligned} \mathcal{Q}_0(\delta; \mathcal{E}, W) &:= \mathcal{H}_0(\delta; \mathcal{E}, W)' \mathrm{diag}\left\{ \mathcal{I}_0^{-1}(\delta), \Xi_0^{-1}(\mathcal{E}, W) \right\} \mathcal{H}_0(\delta; \mathcal{E}, W) \\ &= \begin{pmatrix} \mathcal{I}_0(\delta) & 0_{q\times d} \\ 0_{d\times q} & 0_{d\times d} \end{pmatrix} + G_0(\mathcal{E}, W)' \Xi_0^{-1}(\mathcal{E}, W) G_0(\mathcal{E}, W) \\ &=: \begin{pmatrix} \mathcal{Q}_{11,0}(\delta) & \mathcal{Q}_{12,0} \\ \mathcal{Q}_{21,0} & \mathcal{Q}_{22,0} \end{pmatrix}, \text{ say.} \end{aligned} \tag{5.2}$$

Furthermore, let

$$\Omega(\delta; \mathcal{E}, W) := \mathcal{Q}_{11,0}(\delta) - \mathcal{Q}_{12,0}\mathcal{Q}_{22,0}^{-1}\mathcal{Q}_{21,0}, \tag{5.3}$$

where $\mathcal{Q}_{22,0}$ is invertible under Assumption 6. It can be checked that $\mathcal{Q}_0(\delta; \mathcal{E}, W)$ is invertible based on Lemma A.4 in the Appendix for any $\delta \in (0, \Delta_1]$ and

$$\mathcal{Q}_0^{-1}(\delta; \mathcal{E}, W) = \begin{pmatrix} \Omega^{-1}(\delta; \mathcal{E}, W) & -\Omega^{-1}(\delta; \mathcal{E}, W)\mathcal{Q}_{12,0}\mathcal{Q}_{22,0}^{-1} \\ -\mathcal{Q}_{22,0}^{-1}\mathcal{Q}_{21,0}\Omega^{-1}(\delta; \mathcal{E}, W) & \mathcal{Q}_{22,0}^{-1} + \mathcal{Q}_{22,0}^{-1}\mathcal{Q}_{21,0}\Omega^{-1}(\delta; \mathcal{E}, W)\mathcal{Q}_{12,0}\mathcal{Q}_{22,0}^{-1} \end{pmatrix}. \tag{5.4}$$

Theorem 5.1. *Under Assumptions 1 - 6 given in Appendix, for any $\delta \in (0, \Delta_1]$ as $n \to \infty$,*

$$\sqrt{n}\mathcal{Q}_0^{1/2}(\delta; \mathcal{E}, W) \left(\hat{\theta}_n(\mathcal{E}, W) - \theta_0 \right) \xrightarrow{d} N\left(0, I_{q+d}\right).$$

Theorem 5.1 implies that the asymptotic variance (Avar) of $\hat{\theta}_n(\mathcal{E}, W)$ is $n^{-1}\mathcal{Q}_0^{-1}(\delta; \mathcal{E}, W)$. It is known that the asymptotic variance of the MLE $\tilde{\beta}_n$ based on the short rates only is $n^{-1}\mathcal{I}_0^{-1}(\delta)$. Write $\hat{\theta}_n(\mathcal{E}, W) = \left(\hat{\beta}_n(\mathcal{E}, W)', \hat{\lambda}_n(\mathcal{E}, W)'\right)'$ where $\hat{\beta}_n(\mathcal{E}, W)$ is the new estimator of β by the proposed combined estimation and $\hat{\lambda}_n(\mathcal{E}, W)$ is the estimator of the risk premium parameter. From (5.4),

$$\mathrm{Avar}\left(\hat{\beta}_n(\mathcal{E}, W)\right) = n^{-1}\Omega^{-1}(\delta; \mathcal{E}, W) \text{ and}$$

$$\mathrm{Avar}\left(\hat{\lambda}_n(\mathcal{E}, W)\right) = n^{-1}\mathcal{Q}_{22,0}^{-1} + n^{-1}\mathcal{Q}_{22,0}^{-1}\mathcal{Q}_{21,0}\Omega^{-1}(\delta; \mathcal{E}, W)\mathcal{Q}_{12,0}\mathcal{Q}_{22,0}^{-1} = O\left(n^{-1}\delta^{-1}\right).$$

The following corollary shows that the combined inference for β is at least as efficient as the MLE $\tilde{\beta}_n$ based on the interest rates. The bond information indeed enhances the estimation.

Corollary 5.1. *Under Assumptions 1 - 6 given in Appendix, for any positive definite W, and $\mathcal{E}$ satisfying (4.4), then $Avar\left(\hat{\beta}_n(\mathcal{E}, W)\right) \leq Avar\left(\tilde{\beta}_n\right)$ for any $\delta \in (0, \Delta_1]$.*

Recall that different $\mathcal{E}$s in set (4.4) select different GLS moment restrictions in $g_t(\theta; W)$. The following theorem shows that different $\mathcal{E}$s lead to the same asymptotic efficiency as long as they satisfy (4.4).

Theorem 5.2. *Under Assumptions 1 - 6 given in Appendix, for any two $\mathcal{E}_1 \neq \mathcal{E}_2$ satisfying (4.4), $Avar\left(\hat{\theta}_n(\mathcal{E}_1, W)\right) = Avar\left(\hat{\theta}_n(\mathcal{E}_2, W)\right)$.*

Since usually the number of the moment conditions $q + q^\dagger > q + d$ (e.g., in both Vasicek and CIR models introduced in Section 2), we can perform the over-identification test (the J-test, Hansen, 1982) to check on the appropriateness of

$$\mathrm{H}_0 : \mathbb{E}\left\{h_t(\theta_0; \mathcal{E}, W)\right\} = 0 \quad \text{versus} \quad \mathrm{H}_1 : \mathbb{E}\left\{h_t(\theta_0; \mathcal{E}, W)\right\} \neq 0. \tag{5.5}$$

The J-statistic is

$$J_n = n\left\{\frac{1}{n}\sum_{t=1}^{n} h_t\left(\hat{\theta}_n(\mathcal{E}, W); \mathcal{E}, W\right)\right\}' \mathscr{W}_n(\mathcal{E}, W)\left\{\frac{1}{n}\sum_{t=1}^{n} h_t\left(\hat{\theta}_n(\mathcal{E}, W); \mathcal{E}, W\right)\right\},$$

which can be shown to converge to $\chi^2_{q^\dagger - d}$ in distribution under H_0 based on Theorem 5.1. Specifically, $q^\dagger - d = 1$ as $M = 1$ and $q^\dagger - d = 2$ as $M \geq 2$ for the Vasicek and CIR models. If the

J-test (5.5) is rejected, then the moment condition (4.2), which is decided by different affine term structure models and the exogeneity of the measurement errors, is not appropriately specified. Hence the J-test (5.5) provides a model selection criterion to decide which of the affine models (e.g., Vasicek or CIR) is preferred for the data.

Now let us consider the role of W.

Corollary 5.2. *Under Assumptions 1 - 6 given in Appendix, for any* $\delta \in (0, \Delta_1]$, *as* $n \to \infty$, $Avar\left(\hat{\theta}_n(\mathcal{E}, V_0^{-1})\right) \leq Avar\left(\hat{\theta}_n(\mathcal{E}, W)\right)$ *for any positive definite* W.

Corollary 5.2 implies that choosing $W = V_0^{-1}$ leads to the efficient estimator of θ for each given $\mathcal{E}$. This is consistent with the theory of the GLS method and implies that $g_t(\theta; V_0^{-1})$ should be used if we have the knowledge of V_0.

It is noted that the efficiency lower bound is n^{-1} times the inverse of

$$\mathcal{Q}_0(\delta; \mathcal{E}, V_0^{-1}) = \begin{pmatrix} \mathcal{I}_0(\delta) & 0_{q\times d} \\ 0_{d\times q} & 0_{d\times d} \end{pmatrix} + \psi(\theta_0)' \left(I_2 \otimes V_0^{-1}\right) \psi(\theta_0). \tag{5.6}$$

Hence, the accuracy of the combined estimation is adversely influenced by V_0, the variance of the measurement errors, although it is still more accurate than the MLEs for β.

We now consider the impact of M, the number of bonds used in the inference.

Corollary 5.3. *Under Assumptions 1 - 6 given in Appendix, when the number of bonds* M *increases,* $Avar\left(\hat{\theta}_n(\mathcal{E}, V_0^{-1})\right)$ *does not increase for any* V_0.

Corollary 5.3 means that if $W = V_0^{-1}$, the more bonds we include in the estimating procedure, the more efficient the estimators are. However, this may not be true if $W \neq V_0^{-1}$, which is confirmed in our numerical study reported later, namely the bond prices can improve the efficiency only if we consider the measurement error structure of the new information.

As V_0 is unknown in practice, we consider the following "sample covariance" estimator

$$\hat{V}_n(\mathcal{E}) = \frac{1}{n} \sum_{t=1}^{n} u_t\left(\hat{\theta}_n(\mathcal{E}, I_M)\right) u_t\left(\hat{\theta}_n(\mathcal{E}, I_M)\right). \tag{5.7}$$

It can be shown that under Assumptions 1 - 6 given in Appendix, for any $\delta \in (0, \Delta_1]$

$$\hat{V}_n(\mathcal{E}) \xrightarrow{p} V_0 \text{ as } n \to \infty. \tag{5.8}$$

With $\hat{V}_n(\mathcal{E})$, we get $\hat{\theta}_n(\mathcal{E}, \hat{V}_n^{-1}(\mathcal{E}))$ which we call the feasible combined estimator of θ. It can be shown that $\hat{\theta}_n(\mathcal{E}, \hat{V}_n^{-1}(\mathcal{E}))$ attains the same asymptotic efficiency as $\hat{\theta}_n(\mathcal{E}, V_0^{-1})$.

We can also estimate the asymptotic variance $n^{-1}\mathcal{Q}_0^{-1}(\delta; \mathcal{E}, W)$ upon given $\mathcal{E}$ and W. From (5.1),

$$\begin{aligned}
& G_0(\mathcal{E}, W) \\
= \;& -\mathcal{E}' \left[\frac{\partial A(\theta_0)'}{\partial \theta} W \frac{\partial A(\theta_0)}{\partial \theta'} + \mathbb{E}(r_t) \left\{ \frac{\partial A(\theta_0)'}{\partial \theta} W \frac{\partial B(\theta_0)}{\partial \theta'} + \frac{\partial B(\theta_0)'}{\partial \theta} W \frac{\partial A(\theta_0)}{\partial \theta'} \right\} + \mathbb{E}(r_t^2) \frac{\partial B(\theta_0)'}{\partial \theta} W \frac{\partial B(\theta_0)}{\partial \theta'} \right] \\
=: \;& G\left\{\theta_0, \mathbb{E}(r_t), \mathbb{E}(r_t^2); \mathcal{E}, W\right\}, \quad \text{say.}
\end{aligned}$$

Let

$$\begin{aligned}
\hat{G}_n(\mathcal{E}, W) &= G\left\{ \hat{\theta}_n(\mathcal{E}, W), \frac{1}{n}\sum_{t=1}^n r_t, \frac{1}{n}\sum_{t=1}^n r_t^2; \mathcal{E}, W \right\} \quad \text{and} \\
\hat{\mathcal{Q}}_n(\mathcal{E}, W) &= \begin{pmatrix} \hat{\mathcal{I}}_n\left(\hat{\beta}_n(\mathcal{E}, W)\right) & 0_{q\times d} \\ 0_{d\times q} & 0_{d\times d} \end{pmatrix} + \hat{G}_n(\mathcal{E}, W)' \hat{\Xi}_n^{-1}\left(\hat{\theta}_n(\mathcal{E}, W); \mathcal{E}, W\right) \hat{G}_n(\mathcal{E}, W).
\end{aligned} \tag{5.9}$$

It can be shown by a routine derivation that $\hat{\mathcal{Q}}_n(\mathcal{E}, W)$ is a consistent estimator of $\mathcal{Q}_0(\delta; \mathcal{E}, W)$, which can be used in forming confidence intervals and testing hypothesis for each parameter.

Let us summarize the key steps in carrying out the proposed combined estimation. After having the interest rate data $\{r_t\}_{t=0}^n$ and the bond prices $\{p_t\}_{t=1}^n$, a model from the affine term structure models (e.g., Vasicek or CIR) is identified, followed by finding the matrix $\mathcal{E} \in \mathbb{R}^{(q+d)\times q^\dagger}$ from (4.4) to select the maximum non-redundant estimating equations in (3.4). We then carry out the GMM estimation $\hat{\theta}_n(\mathcal{E}, I_M)$ in (4.5) with the initial weight matrix $W = I_M$. Finally, we obtain the efficient GMM estimator $\hat{\theta}_n(\mathcal{E}, \hat{V}_n^{-1}(\mathcal{E}))$ in (4.5) and its estimated standard error via (5.9) by letting $W = \hat{V}_n^{-1}(\mathcal{E})$.

6 Simulation Studies

We report results of simulation experiments which were designed to confirm the theoretical findings in the previous section. We specifically want to check on the efficiency gain of the combined estimators by comparing with the MLE or the approximate MLE in the context of the Vasicek and CIR models. The full MLE was employed for the Vasicek model, while the approximate MLE based on a two-term expansion (Aït-Sahalia, 1999, 2002) was employed for the CIR. The latter was to evaluate the approximate MLE in our context, though the full MLE for the CIR can be conducted. The parameters used for both models were $(\kappa, \alpha, \sigma, \lambda) = (0.892, 0.09, \sqrt{0.033}, 0.1)$, with the monthly sampling interval $\delta = 1/12$. The sample size n was 300, 500, 1000 and 2000, respectively. All the simulation results were base on 2000 simulations.

The simulated short rates were generated from both processes via their known transition distributions with the initial value from their known stationary distributions, respectively. In the simulation of the bond prices, we considered two designs for the maturity. One had fifteen bonds ($M = 15$) with the time to maturity ranging from 6 months to 7.5 years; and the other had five bonds ($M = 5$) which have the time to maturity ranging from 6 months to 2.5 years. Both settings had six months between two adjacent maturities. The bond prices were generated according to (3.2) with the measurement errors $\{u_{0t}\} \overset{iid}{\sim} N(0, V_0)$. Following the analysis of Cheridito et al. (2007) and Aït-Sahalia and Kimmel (2010), we designed $V_0 = \text{diag}\{v_1^2, v_2^2, \cdots, v_M^2\}$, and let $v_i = 0.001 \times 3^{\tau_i}$ for $i = 1, \cdots, M$. The specification above implies that the measurement errors were independent of r_t and was homogeneous with respect to the time, and the standard deviations of the errors increased exponentially with respect to τ_i. As a consequence of the diagonal form of V_0, we only need to estimate the diagonal elements instead of the estimation in (5.7), namely

$$\hat{V}_n(\mathcal{E}) = \text{diag}\left\{ \frac{1}{n} \sum_{t=1}^{n} u_t\left(\hat{\theta}_n(\mathcal{E}, I_M)\right) u_t\left(\hat{\theta}_n(\mathcal{E}, I_M)\right) \right\}.$$

On the other hand, since we had more than two bonds, we chose $\mathcal{E}_1 = (e_2, e_3, e_4)$ and $\mathcal{E}_2 = (e_1, e_2, e_3)$ to select the moments for the proposed combined estimators, where the four dimension identity matrix $I_4 := (e_1, e_2, e_3, e_4)$.

We evaluated the MLEs and the approximate MLEs for β, and the combined estimators for $\theta = (\beta', \lambda)'$. Figures 1 - 3 display the standard deviation and the averaged absolute bias of the estimates for the CIR model. The results for the Vasicek model were largely similar and are given in the supplementary material. We did not report the bias for α and σ since they are of much smaller order (Tang and Chen, 2009). The most striking feature emerged from these figures are (i) the feasible combined estimators (with $W = \hat{V}_n^{-1}(\mathcal{E})$) offered much improvement in the standard deviations of κ, α and σ, and in the bias of κ, over those of the MLEs/approximate MLEs. The amount of improvement offered by the feasible combined estimator varied among the parameters, with the most improvement registered for κ in both the standard deviation and the bias. This is very encouraging since the mean reverting parameter is the most difficult to estimate. Another feature conveyed from these figures is that the combined estimates without using the optimal weight, namely $W = I_M$, may not be able to produce the best possible performance. However, when the sample size were increased, the combined estimation with both forms of W were better than the MLEs. The estimation error for σ is known (Tang and Chen, 2009) to be much smaller than the bias of the drift parameters k and α. These were clearly reflected in the vertical scales of the respective panels for the three parameters.

The scale of estimation errors for λ was much larger than those of the other parameters despite that λ was much smaller than κ and was only slightly larger than α. The reduction in the estimation errors for λ was quite slow as n was increased. These confirmed the well known challenge in the estimation of the risk premium parameter. We observed that using $M = 15$ bonds produced smaller standard deviations than those of using $M = 5$ bonds for the feasible combined estimator with $W = \hat{V}_n^{-1}(\mathcal{E})$. This was not necessarily the case for the combined estimator with $W = I_M$. The latter suggested we need to use the feasible estimator to ensure the quality of the combined estimator when more bond prices are brought into the inference as they may be subject to more errors along with the increased maturity. We also note that, as the sample size went large, the standard deviations of the combined estimators by choosing different $\mathcal{E}$s were almost the same, which was consistent with Theorem 5.2.

7 Case Study

We analyze the US short interest rates in conjunction with the treasury bond prices, and demonstrate the proposed combined estimation approach. We used the Federal funds overnight rates as proxies to the short rates. The data series was between January 1972 and December 2012, sampled at monthly frequency. The source of the overnight rates is the H.15 Federal Reserve Statistical Release Series. Part of the series (with a different time range) was analyzed in Aït-Sahalia (1999) who also used the overnight rates as proxies for the short rates. The zero-coupon bond prices were obtained from the monthly zero-coupon yields over the same time period as the Federal funds overnight rate series, constructed by Gürkaynak, Sack and Wright (2007) who have been updating the bond yield data on the Federal Reserve Board Finance and Economics Discussion Series. There were $n = 491$ bond prices with the time to maturity ranging from 1 to 15 years. We grouped the bonds to five categories according to the maturity: 1-3 years, 1-5 years, 1-10 years, 1-15 years and 6-15 years. The combined estimators for each category were conducted to gain insight on the impact of the maturity on the parameters.

We estimated the parameters of the Vasicek and CIR models introduced in Section 2 respectively. The proposed combined estimator with $W = I_M$ and $W = \hat{V}_n^{-1}(\mathcal{E})$ were considered. The MLEs for Vasicek and the approximate MLEs for CIR were computed to serve as the benchmarks of estimation. For the combined estimation, we used $\mathcal{E}_1 = (e_2, e_3, e_4)$ (Moment Conditions I) and $\mathcal{E}_2 = (e_1, e_2, e_3)$ (Moment Conditions II). The standard errors of the combined estimates were obtained by estimating the asymptotic variance via (5.9) in Section 5, and those of the MLEs/approximate MLEs were obtained by the estimated Fisher information matrices. The estimated Fisher information matrix for the approximate MLE under the CIR model was based on Theorem 4 in Chang and Chen (2011). We also obtained the estimated measurement errors (estimated residuals) and the covariance of the measurement errors according to (5.7).

Table 1 reports the p-values of the J-tests (5.5) for the Vasicek and CIR models with respect to the five categories of maturity based on $W = I_M$ and Moment Conditions I and II. The J-tests overwhelmingly rejected the Vasicek model for all the categories of time to maturity. The tests found empirical support to the CIR model for shorter maturity of 1-3 years and 1-5 years, as reflected by the quite large p-values. However, as the maturity range was expanded to more than 10 years, the p-values of the CIR model became quite small, indicating that the model was no longer reasonable. We recall the work of Chen, Gao and Tang (2008), which conducted goodness-of-fit tests of the Vasicek and CIR models for the same series of short rates with a different time range (without considering the bond prices). They found that while the Vasicek model was severely mis-specified, there was quite some empirical support to the CIR model. This finding was also consistent with the market segmentation theory that there is a liquidity premium attached to the bonds with long maturities in additional to the risk premium (Fama, 1976; Langetieg, 1980). We also report the parameter estimates and their standard errors (in parentheses) for the Vasicek and CIR models. According to Table 1, the parameter estimates given in Table 2 for the 1-3 years and 1-5 years maturity under the CIR model were more credible than the other estimates reported in the same table and the results under the Vasicek model. The results of the Vasicek model are given in the supplementary material.

The estimates in Table 2 were based on $W = \hat{V}_n^{-1}(\mathcal{E}_1)$ with the moment selection $\mathcal{E}_1 = (e_2, e_3, e_4)$. The results for the other moment selection were very similar and hence are not reported. It is observed from the table that there were quite variations among the parameter estimates across different categories of the time to maturity under the CIR model; and the standard errors of the combined estimates were smaller than those of the MLEs. Compared with the estimates for κ, the combined estimates for α and σ were less varying than the MLEs. The combined estimates of λ varied the most and had the largest standard errors, which were consistent with the simulation results.

Regarding the Vasicek model's result in the supplementary material, the combined estimates for κ tended to be larger than the corresponding MLE for each category of maturity. We would like to recall the analysis reported in Tang and Chen (2009) which showed that the MLEs based on the short rates only over-estimated κ under both Vasicek and CIR models. Hence, the fact that the more accurate combined estimates were larger than the MLEs (rather than smaller) under the Vasicek model was another indication that the Vasicek model was mis-specified. In contrast, the combined estimates for κ under the CIR model with a shorter range of maturity tended to be smaller than the MLEs, which were consistent with the findings of Tang and Chen (2009) under the CIR model. This indicated that the CIR was a better model than the Vasicek for the data. The over-identification test reported in Table 1 shortly lends some support to this belief too.

As the MLE and approximate MLE cannot identify the risk premium parameter λ, the proposed combined estimates based on the short rates and the bond prices offered viable estimates. Given the J-test results discussed above, we would pay more attention on the two estimates under the CIR with the bond maturity of 1-3 and 1-5 years. The standard errors of the λ-estimates were quite large relative to the estimates, rendering insignificance for λ being zero versus being positive. This reflects an often encountered situation regarding the inference for the presence of the risk premium, for instance in Aït-Sahalia and Kimmel (2007) and Aït-Sahalia and Kimmel (2010).

In Section 3, we consider a parameter transformation $\vartheta = (b, a, \sigma)' = \vartheta(\theta)$ for both of the Vasicek and CIR models such that $b = \kappa - \sigma\lambda$ and $a = \frac{\kappa\alpha}{\kappa - \sigma\lambda}$ are the drift parameters under the risk neutral measure, which are more direct to the bond price. Let

$$\frac{\partial \vartheta(\theta)}{\partial \theta'} = \begin{pmatrix} 1 & 0 & -\lambda & -\sigma \\ \frac{-\alpha\sigma\lambda}{(\kappa-\sigma\lambda)^2} & \frac{\kappa}{\kappa-\sigma\lambda} & \frac{\kappa\alpha\lambda}{(\kappa-\sigma\lambda)^2} & \frac{\kappa\alpha\sigma}{(\kappa-\sigma\lambda)^2} \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

According to Theorem 5.1 and the delta-method, the plug-in estimator $\hat{\vartheta}_n(\mathcal{E}, W) = \vartheta\left(\hat{\theta}_n(\mathcal{E}, W)\right)$ is asymptotic normal in that

$$\sqrt{n}\left(\hat{\vartheta}_n(\mathcal{E}, W) - \vartheta_0\right) \xrightarrow{d} N\left(0, \frac{\partial \vartheta(\theta_0)}{\partial \theta'} \mathcal{Q}_0^{-1}(\delta; \mathcal{E}, W) \frac{\partial \vartheta(\theta_0)'}{\partial \theta}\right).$$

Table 2 also reports the estimates of a and b under the risk neutral measure. It reveals that despite the rather volatile estimates for the parameters under the physical measure, the parameter estimates under the risk neutral measure (a, b and σ) were largely stable. This means that although the estimated λ may incur large errors, the estimates for parameters which directly influence the bond pricing were more reliable.

Table 3 reports the estimated correlation matrix (standardized $\hat{V}_n(\mathcal{E}_1)$) of the pricing errors under the CIR model by using the maturity group of 1-5 years, which has been shown to fit the CIR model quite well in Table 1. Before standardizing $\hat{V}_n(\mathcal{E}_1)$, we found that the estimated standard deviations of the pricing errors were increasing along with the time to maturity almost linearly. We observe from Table 3 that (i) the dependence in the pricing errors was quite persistent with the correlation coefficients decaying very gradually as the gap between the maturities was increased; and (ii) the correlations were consistently positive. The results reveal that it may be too simplistic to specify a diagonal form for V_0.

The sample covariance $\hat{V}_n(\mathcal{E}_1)$ has been used to estimate $V_0 \in \mathbb{R}^{M\times M}$ to obtain the feasible combined estimators in Table 2. Suggested by a referee, we implemented an alternative covariance estimator suitable for high dimensions. Specifically, we consider the non-negative covariance estimator proposed in Rothman (2012). It is noted that although in our current setting the dimension M (the number of bonds) is fixed and is smaller than the sample size n, experimenting the estimator of Rothman (2012) provides insights for higher dimensional situations. Denote Rothman (2012)'s covariance estimator as $\bar{V}_n(\mathcal{E}_1)$ which was computed using an R package "PDSCE". The combined estimates for the CIR model based on $\bar{V}_n(\mathcal{E}_1)$ are reported in the supplementary material. The results also contain the spectral norm of $\bar{V}_n(\mathcal{E}_1) - \hat{V}_n(\mathcal{E}_1)$, which indicates the two covariance estimators were generally close to each other. Comparing Table 2 with the combined estimates based on $\bar{V}_n(\mathcal{E}_1)$, we observe that although there were some differences in the parameter estimates using the two covariance estimators, the differences were not significant when considered in terms of the standard errors. And more importantly, the insights found in Table 2 as discussed above were largely maintained.

8 Conclusion

Despite the interest rate models are such basics in the modern financial theory and practice, getting proper models and estimating their parameters have been challenging. A key aspect of the challenge is rooted in the fact that the short (instantaneous) rates are not directly observable. There have been two approaches to find approximations to the short rates. One is to use rates with shorter maturities as proxies to the short rates, as adopted in Chan, Karolyi, Longstaff and Sanders (1992), Nowman (1997), Aït-Sahalia (1996) and Aït-Sahalia (1999). The other approach, which we call the implied state variable approach, is to calibrate the short rates via the bond prices by assuming that one or a few bond prices follow exactly (2.5) without errors whereas the other bonds are subject to errors; see Chen and Scott (1993), Duffee (2002), Cheridito et al. (2007), Aït-Sahalia and Kimmel (2010), Joslin, Singleton and Zhu (2011) and Hamilton and Wu (2014). It is fair to say that both approaches use certain type of proxies to approximate the short rates. Indeed, while the first approach assumes that there are quality proxies to the short rates, the implied state variable approach assumes certain numbers of the bond prices are observed accurately.

We use the first approach in our analysis in this paper. Although we used the over-night Fed fund rates as proxies to the short rates in the case study, interest rates with other maturities can be used to avoid the micro-structures of the over-night rates as noted in Filipović (2009). For instance, Aït-Sahalia (1996) used the 7-day Eurodollar deposit spot rate, bid-ask midpoint as the proxy of the short rate.

The proposed approach can be viewed as a further development of the bond return method used in Brown and Dybvig (1986) and Gibbons and Ramaswamy (1993) by considering cross-sectional prices of bonds as well as the conditional model information of the short rate processes. The combination of the short rate dynamic information and the bond prices allows for enhancement of estimation beyond the MLE based on the short rates only and identification of all parameters. The proposed combined estimation is semiparametric with respect to the measurement errors. The nonparametric specification on the pricing errors is more adaptive to the underlying pricing error structure given our findings in the case study that the size of the pricing error was largely influenced by the maturity, and that there was substantial dependence between errors of different maturities. The dependence became larger as the time to maturity increased, which indicates that it would be too simplistic to assume a diagonal form for V_0 as assumed in some of the implementation of the implied state variable approach. The proposed approach avoids directly specifying the covariance structure of the pricing errors while still achieving good efficiency in the estimation.

The reason why we focus on the time homogenous affine term structure modeling in this paper is mainly driven by real applications of interest rates and bond prices (Aït-Sahalia, 2002; Cheridito et al., 2007; Aït-Sahalia and Kimmel, 2010). Proposed by an anonymous referee, we discuss the time inhomogeneous scenario in the following. The MLE discussed in Section 2 can be shown to be asymptotically normal using the martingale convergence theorems (Hall and Heyde, 1980). Based on the similar technics, we can show that

$$\left[\mathrm{Var}\left\{\frac{1}{n}\sum_{t=1}^{n} h_t(\theta_0;\mathcal{E},W)\right\}\right]^{-1/2}\frac{1}{n}\sum_{t=1}^{n} h_t(\theta_0;\mathcal{E},W)$$

is asymptotically normal if we impose the similar conditions of Assumptions 1 and 2 in Hall and Heyde (1980, p. 160). Then our combined estimation approach can be carried out and the asymptotic normality of the GMM estimator is still valid. However, the range of the time inhomogeneous processes adapted to the conditions needs to be further investigated in the future study.

We have considered in this paper one-factor models in our attempt to utilize the bond prices to enhance the estimation of the parameters of the interest rate processes. Extensions may be made to the multi-factor affine term structure models, which would require the filtering techniques to be used. While we leave this extension to future consideration, we note that despite a set of multi-factor models have been proposed, almost all these models are rejected in the empirical testing. Chen et al. (2008) did find some empirical support to the univariate CIR model.

A Appendix: Technical Details

Throughout the appendix, we use Δ to denote a finite positive constant, and denote the spectral norm of a matrix $A = (a_{ij})_{q\times p}$ as $\|A\|_2 = \sqrt{\lambda_{\max}(A'A)}$, where $\lambda_{\max}(A'A)$ is the largest eigenvalue of $A'A$, and the Frobenius norm $\|A\| = \sqrt{tr(A'A)} = \sqrt{\sum_{i=1}^{q}\sum_{j=1}^{p} a_{ij}^2}$. For a stationary matrix process $\left\{F_t(\theta) = (F_{ij,t}(\theta))_{q\times p}\right\}$ relying on a finite dimension vector θ, where p and q are also finite, $\mathbb{P}_n F_t(\theta) := \frac{1}{n}\sum_{t=1}^{n} F_t(\theta)$. For the first and second derivatives, $\dot{\ell}_t(\beta) := \frac{\partial \ell_t(\beta)}{\partial \beta}$, $\ddot{\ell}_t(\beta) := \frac{\partial^2 \ell_t(\beta)}{\partial \beta \partial \beta'}$, $\dot{A}(\theta) := \frac{\partial A(\theta)}{\partial \theta'}$, $\dot{B}(\theta) := \frac{\partial B(\theta)}{\partial \theta'}$ and $\dot{h}_t(\theta) := \frac{\partial h_t(\theta)}{\partial \theta'}$. We suppress the expression $\mathcal{E}$ and W in $\bar{\theta}_n(\mathcal{E}, W)$ and $h_t(\theta; \mathcal{E}, W)$, and write them as $\bar{\theta}_n$ and $h_t(\theta)$ whenever doing so would not cause confusion.

We firstly present the assumptions needed in our analysis. Assumptions used for the approximate MLE of Aït-Sahalia (1999, 2002) are presented as well, as we have used an approach that can lead to results for both the combined estimation using either the full likelihood scores or the approximated likelihood scores. Discussions to the assumptions including comparison with the conditions in the extant literatures are given in the supplementary material.

Assumption 1. (i) $\theta = (\beta', \lambda')' = (\theta_1, \cdots, \theta_{q+d})' \in \Theta$ which is a compact set in $\mathbb{R}^{q+d}$, where $\beta = (\beta_1, \cdots, \beta_q)' \in \mathcal{B} \subset \mathbb{R}^q$. (ii) The true value $\theta_0 \in \Theta$ is an interior point. (iii) β_0 is the unique root of $\mathbb{E}\left\{\frac{\partial \ell_t(\beta)}{\partial \beta}\right\} = 0$ for every δ. (iv) θ_0 is the unique root of $\mathbb{E}\left\{h_t(\theta; I_{q+d}, W)\right\} = 0$ for every δ.

Assumption 2. (i) The short rate $r(t)$ follows the time homogeneous diffusion processes (2.1) and (2.3) under measures Q_0 and Q_1. Assumption 1 in Aït-Sahalia and Mykland (2004) is satisfied under the measure Q_0, and (2.4) holds under the measure Q_1. (ii) The pricing functions in (2.6) and (2.7) are three times differentiable with respect to θ. (iii) For fixed $\tau_1, \cdots, \tau_M$ and fixed M, the time homogeneous pricing functions $A(\theta) = (A_1(\theta), \cdots, A_M(\theta))'$ and $B(\theta) = (B_1(\theta), \cdots, B_M(\theta))'$ satisfy

$$\sup_{\theta\in\Theta} |A_i(\theta)| \le M_1, \sup_{\theta\in\Theta} |B_i(\theta)| \le M_1, \sup_{\theta\in\Theta} \left|\frac{\partial^l A_i(\theta)}{\partial \theta_{j_1} \cdots \partial \theta_{j_l}}\right| \le M_1 \text{ and } \sup_{\theta\in\Theta} \left|\frac{\partial^l B_i(\theta)}{\partial \theta_{j_1} \cdots \partial \theta_{j_l}}\right| \le M_1,$$

for a fixed positive constant $M_1 > 0$ and any $i = 1, 2, \cdots, M$, $l = 1, 2, 3$ and $j_1, j_2, j_3 \in \{1, 2, \cdots, q+d\}$.

Assumption 3. (i) $\frac{\partial}{\partial \beta}\int f(r_t|r_{t-1}, \delta; \beta) dr_t = \int \frac{\partial}{\partial \beta} f(r_t|r_{t-1}, \delta; \beta) dr_t$ and $\frac{\partial}{\partial \beta}\int \frac{\partial}{\partial \beta} f(r_t|r_{t-1}, \delta; \beta) dr_t =$

$\int \frac{\partial^2}{\partial\beta\partial\beta'} f(r_t|r_{t-1},\delta;\beta)dr_t$, which imply the Fisher information matrix $\mathbb{E}\left\{-\frac{\partial^2 \ell_t(\beta_0)}{\partial\beta\partial\beta'}\right\} =: \mathcal{I}_0(\delta)$. (ii) For any nonrandom $\delta > 0$, $\mathcal{I}_0(\delta)$ is invertible and $\left\|\delta^{1/2}\mathcal{I}_0^{-1/2}(\delta)\right\|_2 = O(1)$.

Assumption 4. The J-term expansion to the log of transition density $\ell_t(\beta)$ in (2.2) is

$$\ell_t^{(J)}(\beta) = -\log\sqrt{2\pi\delta} + \mathcal{A}_1(r_t|r_{t-1},\delta;\beta) + \mathcal{A}_2(r_t|r_{t-1},\delta;\beta) + \mathcal{A}_3^{(J)}(r_t|r_{t-1},\delta;\beta), \tag{A.1}$$

where $\mathcal{A}_3^{(J)}(x|x_0,\delta;\beta) = \log\left\{\sum_{j=0}^J c_j\left(\gamma(x;\beta)|\gamma(x_0;\beta);\beta\right)\frac{\delta^j}{j!}\right\}$ for $J \geq 1$, and the expressions of the functions $\mathcal{A}_1, \mathcal{A}_2, \gamma$ and c_j can be found in Aït-Sahalia (1999). Let

$$h_t^{(J)}(\theta;\mathcal{E},W) = \begin{pmatrix} \frac{\partial \ell_t^{(J)}(\beta)}{\partial\beta} \\ \mathcal{E}' g_t(\theta;W) \end{pmatrix} \tag{A.2}$$

as a approximate for $h_t(\theta;\mathcal{E},W)$ in (4.1) to establish our proposed estimator. Assumptions (A.3), (A.6) and (A.7) in Chang and Chen (2011) hold and there exist finite positive constants ν_k for $k = 0,1,2,3$, and M_2 such that $\nu_0 > 3, \nu_2 > \nu_1 > 3, \nu_3 > 1$ and for any $i_1,\cdots,i_3 \in \{1,\cdots,q\}$, $\delta \in (0,\Delta]$,

$$\mathbb{E}\left[\sup_{\beta\in\mathcal{B}}\left\{\sum_{l=0}^{\infty}\left|\frac{\partial^k c_l\left(\gamma(r_t;\beta)|\gamma(r_{t-1};\beta);\beta\right)}{\partial\beta_{i_1}\cdots\partial\beta_{i_k}}\right|\frac{\delta^l}{l!}\right\}^{2\nu_k}\right] \leq M_2. \tag{A.3}$$

Assumption 5. (i) The measurement error $\{u_{0t}\}$ in (3.2) is a martingale difference array with respect to the filtration $\{\mathcal{G}_t\}$, where $\mathcal{G}_t$ is the σ-algebra generated by $\{(r_{l+1},u_{0l}')'\}_{l\leq t}$. (ii) The short rate and measurement error process $\{(r_t,u_{0t}')'\}$ is stationary and satisfies (3.1). (iii) We assume $\{(r_t,u_{0t}')'\}$ is ρ-mixing with the ρ-mixing coefficient

$$\rho(k) := \sup_{Z_1\in L^2(\mathcal{F}_{-\infty}^0), Z_2\in L^2(\mathcal{F}_k^{+\infty})} |\mathrm{Corr}(Z_1,Z_2)| \leq C_1 e^{-C_2 k}, \tag{A.4}$$

where $\mathcal{F}_{-\infty}^0$ is the σ-algebra generated by $\{(r_l,u_{0l}')'\}_{l\leq 0}$, $\mathcal{F}_k^{\infty}$ is the σ-algebra generated by $\{(r_l,u_{0l}')'\}_{l\geq k}$, and C_1, C_2 are positive constants. (iv) $\mathbb{E}(r_t^4) < \infty$ and $\mathbb{E}(u_{0it}^4) < \infty$. (v) $\mathbb{E}\left(u_{0t}u_{0t}'|\mathcal{G}_{t-1}\right) =: V_0 = (v_{jk})_{M\times M}$ is of full rank and W is an $M \times M$ nonrandom positive definite matrix.

Assumption 6. (i) $q^\dagger \geq d$. (ii) Based on the definition of $\mathcal{E}$ in (4.4), there exists a $q^\dagger \times (q+d)$ nonrandom matrix $Z(\theta_0)$ which depends only on θ_0 and $\tau_1\cdots,\tau_M$, such that

$$\psi(\theta_0) = \psi(\theta_0)\mathcal{E}Z(\theta_0). \tag{A.5}$$

The $q^\dagger \times d$ matrix $Z_\lambda(\theta_0) := Z(\theta_0)(0_{d\times q}, I_d)'$ satisfies $\operatorname{rank}\{Z_\lambda(\theta_0)\} = d$.

In the following we present the lemmas as well as the proofs of the propositions and theorems by using these lemmas. The proofs of the lemmas and some corollaries are left in the supplementary material.

Lemma A.1. *Under Assumptions 1, 2, 4 and 5, there exist positive constants M_{01}, $M_{02} < \infty$ and Δ_1, such that for any J, where J can be infinity, $l = 1, 2$, $\delta \in (0, \Delta_1]$, $d \le K \le q + d$, and $i_1, i_2 \in \{1, 2, \cdots, q + d\}, j \in \{1, 2, \cdots, q + K\}$,*

$$\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|h_{tj}^{(J)}(\theta)\right|^2\right\} \le M_{01} \vee M_{02} \text{ and } \mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\frac{\partial^l h_{tj}^{(J)}(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_l}}\right|^2\right\} \le M_{01} \vee M_{02},$$

where $h_{tj}^{(\infty)}(\theta) = h_{tj}(\theta)$, $\frac{\partial^l h_{tj}^{(\infty)}(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_l}} = \frac{\partial^l h_{tj}(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_l}}$ and $h_t^{(J)}(\theta) = \left(h_{t1}^{(J)}(\theta), \cdots, h_{t(q+K)}^{(J)}(\theta)\right)'$.

Lemma A.2. *For every i, j, there exists a constant M_{31} such that $\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\frac{\partial F_{ij,t}(\theta)}{\partial\theta_k}\right|\right\} \le M_{31} < \infty$ for every component θ_k of vector θ. Besides, for each $\theta \in \Theta$, $\mathbb{P}_n F_{ij,t}(\theta) - \mathbb{E}\{F_{ij,t}(\theta)\} \xrightarrow{p} 0$. Then $\sup_{\theta\in\Theta}\|\mathbb{P}_n F_t(\theta) - \mathbb{E}\{F_t(\theta)\}\|_2 \xrightarrow{p} 0$ as $n \to \infty$.*

Lemma A.3. *Under Assumptions 1, 2 and 4,*

$$\mathbb{P}_n\dot{\ell}_t(\beta) - \mathbb{E}\left\{\dot{\ell}_t(\beta)\right\} = O_p\left((n\delta)^{-1/2}\right) \text{ for each } \beta \in \mathcal{B} \text{ and } \sup_{\beta\in\mathcal{B}}\left\|\mathbb{P}_n\dot{\ell}_t(\beta) - \mathbb{E}\left\{\dot{\ell}_t(\beta)\right\}\right\|_2 \xrightarrow{p} 0,$$

for $\delta \in (0, \Delta_1]$, $n \to \infty$.

Proof of Proposition 4.1: According to the stationary, (A.4) and Lemma A.1, by Theorem 16.3.8 in Athreya and Lahiri (2006), the long-run covariance matrix $\lim_{n\to\infty} n\operatorname{Var}\{\mathbb{P}_n h_t(\theta_0)\} = \Gamma(0) + \sum_{k=1}^\infty\{\Gamma(k) + \Gamma(k)'\}$ converges, where

$$\Gamma(k) = \mathbb{E}\begin{pmatrix} \dot{\ell}_t(\beta_0)\dot{\ell}_{t-k}(\beta_0)' & \dot{\ell}_t(\beta_0)g_{t-k}(\theta_0; W)'\mathcal{E} \\ \mathcal{E}'g_t(\theta_0; W)\dot{\ell}_{t-k}(\beta_0)' & \mathcal{E}'g_t(\theta_0; W)g_{t-k}(\theta_0; W)'\mathcal{E} \end{pmatrix}.$$

It is worth noting that we have $\mathbb{E}\left\{g_t(\theta_0; W)\middle|\mathcal{G}_{t-1}\right\} = 0$ by Assumption 5. Then the proposition can be proved together with Assumption 3. $\square$

Lemma A.4. *Under Assumptions 3, 5 and 6, $\Omega(\delta;\mathcal{E},W)$ defined in (5.3) and $\mathcal{Q}_0(\delta;\mathcal{E},W)$ defined in (5.2) are invertible for any $\delta\in(0,\Delta_1]$, and $\left\|\delta^{1/2}\mathcal{Q}_0^{-1/2}(\delta;\mathcal{E},W)\right\|_2=O(1)$.*

Proof of Theorem 5.1: Let $\mathscr{W}=\mathrm{diag}\left\{\mathcal{I}_0^{-1}(\delta),\Xi_0^{-1}\right\}$. Taking the Taylor expansion at θ_0 on the first oder condition to the minimization, $\left\{\mathbb{P}_n\dot{h}_t(\theta_0)\right\}'\hat{\mathscr{W}}_n\left\{\mathbb{P}_nh_t(\theta_0)\right\}+\mathcal{Q}_n\left(\hat{\theta}_n-\theta_0\right)=0$ and

$$\mathcal{Q}_n:=\mathcal{Q}(\hat{\theta}):=\left\{\mathbb{P}_n\dot{h}_t(\hat{\theta})\right\}'\hat{\mathscr{W}}_n\left\{\mathbb{P}_n\dot{h}_t(\hat{\theta})\right\}+\left[\left\{\mathbb{P}_nh_t(\hat{\theta})\right\}'\hat{\mathscr{W}}_n\otimes I_{q+d}\right]\left[\mathbb{P}_n\frac{\partial\ \mathrm{vec}\left\{\dot{h}_t(\hat{\theta})'\right\}}{\partial\theta'}\right],\quad (\mathrm{A.6})$$

where $\hat{\theta}=\theta_0+z_n(\hat{\theta}_n-\theta_0)$ and $0\le z_n\le 1$. Then according to Lemma A.1 and Lemma A.3,

$$\begin{aligned}
&-\sqrt{n}\mathcal{Q}_0^{1/2}(\delta)\left(\hat{\theta}_n-\theta_0\right)=\sqrt{n}\mathcal{Q}_0^{-1/2}(\delta)\left\{\mathbb{P}_n\dot{h}_t(\theta_0)\right\}'\hat{\mathscr{W}}_n\left\{\mathbb{P}_nh_t(\theta_0)\right\}\\
&+\mathcal{Q}_0^{-1/2}(\delta)\left\{\mathcal{Q}_n-\mathcal{Q}_0(\delta)\right\}\mathcal{Q}_0^{-1/2}(\delta)\left\{\sqrt{n}\mathcal{Q}_0^{1/2}(\delta)\left(\hat{\theta}_n-\theta_0\right)\right\}\\
=\quad&\sqrt{n}\mathbb{P}_n\mathcal{Q}_0^{-1/2}(\delta)\mathcal{H}_0(\delta)'\mathscr{W}h_t(\theta_0)+o_p\left(\left\|\sqrt{n}\mathcal{Q}_0^{1/2}(\delta)(\hat{\theta}_n-\theta_0)\right\|_2\right)+o_p(1)\xrightarrow{d}N(0,I_{q+d})
\end{aligned}$$

since $\mathbb{E}\left\|\mathcal{Q}_0^{-1/2}(\delta)\mathcal{H}_0(\delta)'\mathscr{W}h_t(\theta_0)\right\|^2=q+d<\infty$, the long-run covariance matrix

$$\lim_{n\to\infty}n\mathrm{Var}\left\{\mathbb{P}_n\mathcal{Q}_0^{-1/2}(\delta)\mathcal{H}_0(\delta)'\mathscr{W}h_t(\theta_0)\right\}=I_{q+d}$$

by Proposition 4.1, the stationarity, (A.4) and Theorem 16.3.8 in Athreya and Lahiri (2006). □

Proof of Theorem 5.2: Note that $\psi(\theta_0)\mathcal{E}_1$ and $\psi(\theta_0)\mathcal{E}_2$ are both of full rank. From (A.5), we have $\psi(\theta_0)=\psi(\theta_0)\mathcal{E}_1Z_1(\theta_0)=\psi(\theta_0)\mathcal{E}_2Z_2(\theta_0)$, namely the columns of the matrices $\psi(\theta_0)\mathcal{E}_1$ and $\psi(\theta_0)\mathcal{E}_2$ form a basis to the columns of $\psi(\theta_0)$. Then there exists a full rank nonrandom $q^\dagger\times q^\dagger$ matrix $\mathcal{Z}(\theta_0)$ such that $\psi(\theta_0)\mathcal{E}_1=\psi(\theta_0)\mathcal{E}_2\mathcal{Z}(\theta_0)$. Hence, by (5.2) we have $\mathcal{Q}_0(\delta;\mathcal{E}_1,W)=\begin{pmatrix}\mathcal{I}_0(\delta) & 0_{q\times d}\\ 0_{d\times q} & 0_{d\times d}\end{pmatrix}+$ $Z_1(\theta_0)'\left\{\mathcal{E}_1'\psi(\theta_0)'\left(I_2\otimes W\right)\psi(\theta_0)\mathcal{E}_1\right\}\left\{\mathcal{E}_1'\psi(\theta_0)'\left[I_2\otimes(WV_0W)\right]\psi(\theta_0)\mathcal{E}_1\right\}^{-1}\left\{\mathcal{E}_1'\psi(\theta_0)'\left(I_2\otimes W\right)\psi(\theta_0)\mathcal{E}_1\right\}Z_1(\theta_0)$ $=Q_0(\delta;\mathcal{E}_2,W)$ considering the invertible $\mathcal{Z}(\theta_0)$. Then the theorem is proved by Theorem 5.1. □

Acknowledgements: We thank the Editor, the AE and two referees for helpful comments and suggestions which have improved the presentation of the paper. The research was partially supported by Natural Science Foundation of China grants 11131002 and 71371016.

References

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 2, 385–426.

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *The Journal of Finance*, 54, 1361–1395.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70, 223–262.

Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83, 413–452.

Aït-Sahalia, Y. and Kimmel, R. L. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions. *Journal of Financial Economics*, 98, 113–144.

Aït-Sahalia, Y. and Mykland, P. A. (2004). Estimators of diffusions with randomly spaced discrete observations: A general theory. *The Annals of Statistics*, 32, 2186–2222.

Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*. Springer Verlag.

Brigo, D. and Mercurio, F. (2006). *Interest Rate Models: Theory and Practice*. Springer Verlag.

Brown, S. J. and Dybvig, P. H. (1986). The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates. *The Journal of Finance*, 41, 617–630.

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

Chang, J. and Chen, S. X. (2011). On the approximate maximum likelihood estimation for diffusion processes. *The Annals of Statistics*, 39, 2820–2851.

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

Chen, S. X., Gao, J., and Tang, C. Y. (2008). A test for model specification of diffusion processes. *The Annals of Statistics*, 36, 167–198.

Cheridito, P., Filipović, D., and Kimmel, R. L. (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics*, 83, 123–170.

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

Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance*, 55, 1943–1978.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57, 405–443.

Duffie, D. (2001). *Dynamic Asset Pricing Theory*. Princeton University Press.

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

Fama, E. F. (1976). Forward rates as predictors of future spot rates. *Journal of Financial Economics*, 3, 361–377.

Feller, W. (1951). Two singular diffusion problems. *The Annals of Mathematics*, 54, 173–182.

Filipović, D. (2009). *Term-Structure Models: A Graduate Course*. Springer Verlag.

Gibbons, M. R. and Ramaswamy, K. (1993). A test of the Cox, Ingersoll, and Ross model of the term structure. *Review of Financial Studies*, 6, 619–658.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The US treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54, 2291–2304.

Hairer, E., Nøersett, S. P., and Wanner, G. (2006). *Solving Ordinary Differential Equations*. Springer Verlag.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

Hamilton, J. D. and Wu, J. C. (2014). Testable implications of affine term structure models. *Journal of Econometrics*, 178, 231–242.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50, 1029–1054.

Hull, J. (2009). *Options, Futures, and Other Derivatives*. Pearson.

Joslin, S., Singleton, K. J., and Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. *Review of Financial Studies*, 24, 926–970.

Langetieg, T. C. (1980). A multivariate model of the term structure. *The Journal of Finance*, 35, 71–97.

McCulloch, J. H.and Kwon, H.-C. (1993). U.S. term structure data, 1947-1991. *Ohio State Working Paper*, 93–6.

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

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

Nowman, K. B. (1997). Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *The Journal of Finance*, 52, 1695–1706.

Pearson, N. D. and Sun, T.-S. (1994). Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model. *The Journal of Finance*, 49, 1279–1304.

Phillips, P. C. and Yu, J. (2005). Jackknifing bond option prices. *Review of Financial Studies*, 18, 707–742.

Rothman, A. J. (2012). Positive definite estimators of large covariance matrices. *Biometrika*, 99, 733–740.

Svensson, L. E. O. (1994). Estimating and interpreting forward rates: Sweden 1992-4. *National Bureau of Economic Research Working Paper*, page 4871.

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

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177–188.

Vasicek, O. A. and Fong, H. G. (1982). Term structure modeling using exponential splines. *The Journal of Finance*, 37, 339–348.

Table 1: J-tests for the Vasicek and CIR models based on the Federal fund rates and bond prices.

		GMM: $\mathcal{E}_1 = (e_2, e_3, e_4)$ (Moment Conditions I)				
		$1-3$	$1-5$	$1-10$	$1-15$	$6-15$
Vasicek	J-statistic	10.2720	30.8105	11.4163	30.6171	32.3894
	p-value	0.0059	0.0000	0.0033	0.0000	0.0000
CIR	J-statistic	0.8903	1.4591	9.0660	26.4127	28.3110
	p-value	0.6407	0.4821	0.0107	0.0000	0.0000
		GMM: $\mathcal{E}_2 = (e_1, e_2, e_3)$ (Moment Conditions II)				
		$1-3$	$1-5$	$1-10$	$1-15$	$6-15$
Vasicek	J-statistic	10.3928	10.0876	6.8914	7.9270	7.9914
	p-value	0.0055	0.0064	0.0319	0.0190	0.0184
CIR	J-statistic	0.7671	1.4826	11.4109	48.9833	50.4626
	p-value	0.6814	0.4765	0.0033	0.0000	0.0000

Table 2: Combined estimates based on the monthly Federal fund rates and bond prices for the CIR model ($W = \hat{V}_n^{-1}(\mathcal{E}_1)$). Figures inside the parentheses are the standard errors of the estimates above.

	MLE	$1-3$	$1-5$	$1-10$	$1-15$	$6-15$
κ	0.1875	0.1705	0.1635	0.2454	0.1504	0.1836
	(0.0337)	(0.0239)	(0.0243)	(0.0331)	(0.0295)	(0.0295)
α	0.0751	0.0645	0.0696	0.0574	0.0857	0.0726
	(0.0133)	(0.0090)	(0.0103)	(0.0077)	(0.0167)	(0.0117)
σ	0.0641	0.0544	0.0548	0.0647	0.0604	0.0607
	(0.0007)	(0.0004)	(0.0004)	(0.0007)	(0.0005)	(0.0005)
λ	-	0.6589	0.4772	1.2344	0.0608	0.5713
	-	(0.4390)	(0.4417)	(0.5127)	(0.4867)	(0.4866)
b	-	0.1347	0.1374	0.1655	0.1467	0.1489
	-	(0.0058)	(0.0054)	(0.0046)	(0.0043)	(0.0043)
a	-	0.0817	0.0829	0.0852	0.0878	0.0895
	-	(0.0015)	(0.0010)	(0.0009)	(0.0010)	(0.0010)

Table 3: Estimated correlation matrix of the measurement errors for 1-5 years maturity under the CIR model.

τ_i	2	3	4	5
1	0.950	0.879	0.811	0.749
2		0.983	0.949	0.911
3			0.991	0.971
4				0.994

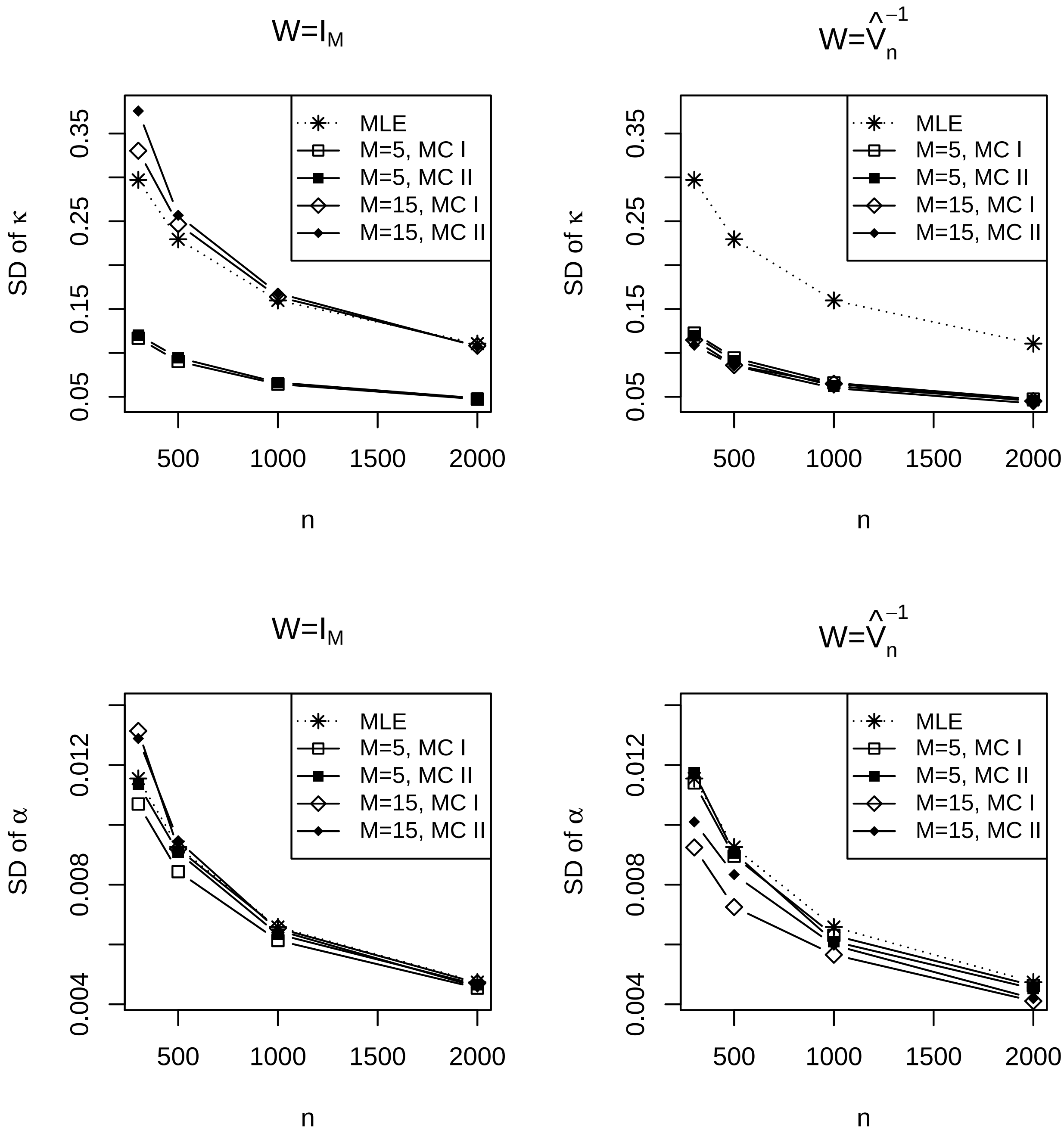

Figure 1: Simulated standard deviation (SD) of the MLEs ($\tilde{\beta}_n$), and the combined estimators ($\hat{\theta}_n$) with two moment selection matrices $\mathcal{E}_1$ (MC I) and $\mathcal{E}_2$ (MC II) for κ and α in the CIR model with $(\kappa, \alpha, \sigma, \lambda) = (0.892, 0.09, \sqrt{0.033}, 0.1))$; M denotes the number of bonds.

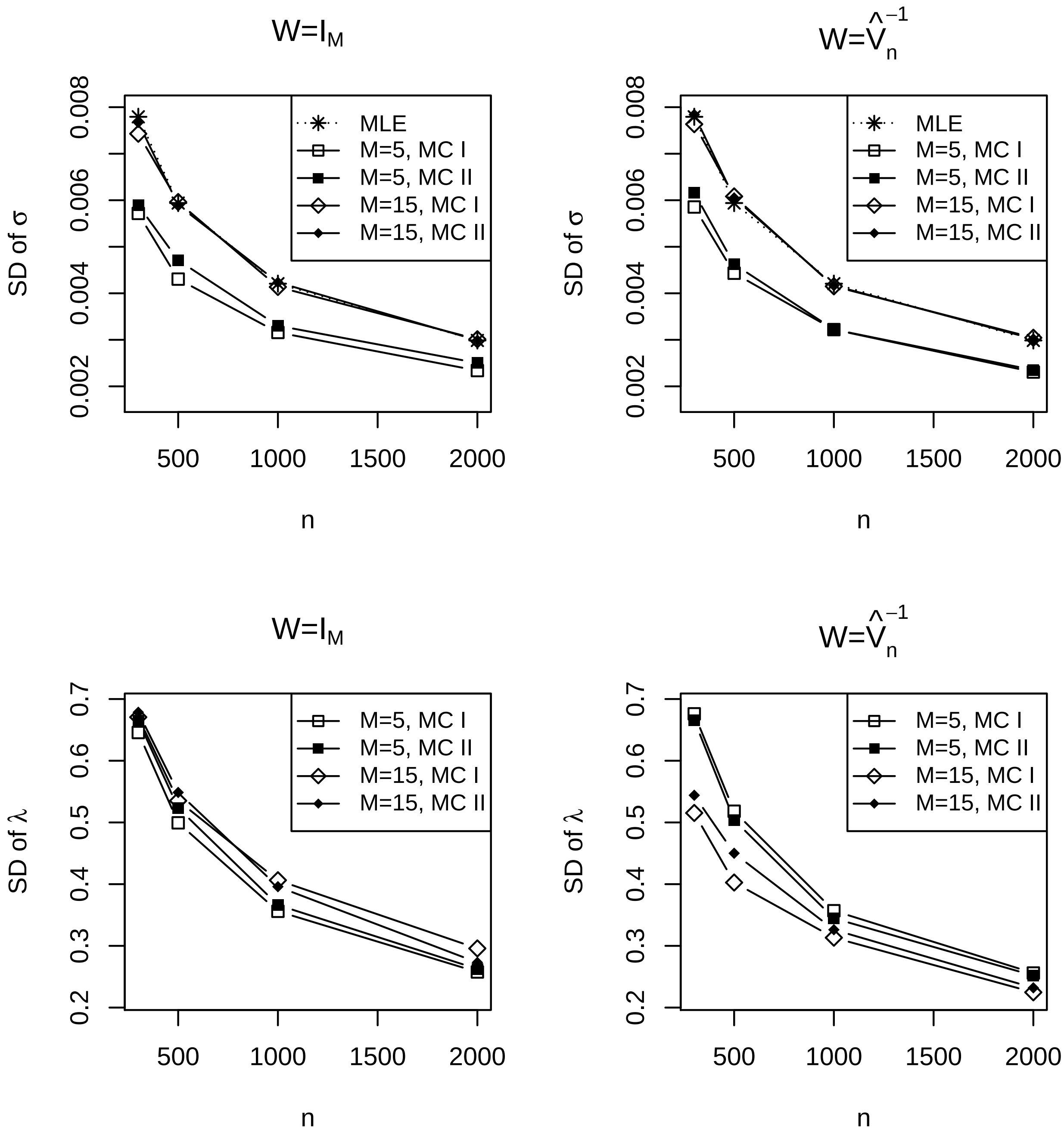

Figure 2: Simulated standard deviation (SD) of the MLEs ($\tilde{\beta}_n$), and the combined estimators ($\hat{\theta}_n$) with two moment selection matrices $\mathcal{E}_1$ (MC I) and $\mathcal{E}_2$ (MC II) for σ and λ in the CIR model with $(\kappa, \alpha, \sigma, \lambda) = (0.892, 0.09, \sqrt{0.033}, 0.1))$; M denotes the number of bonds.

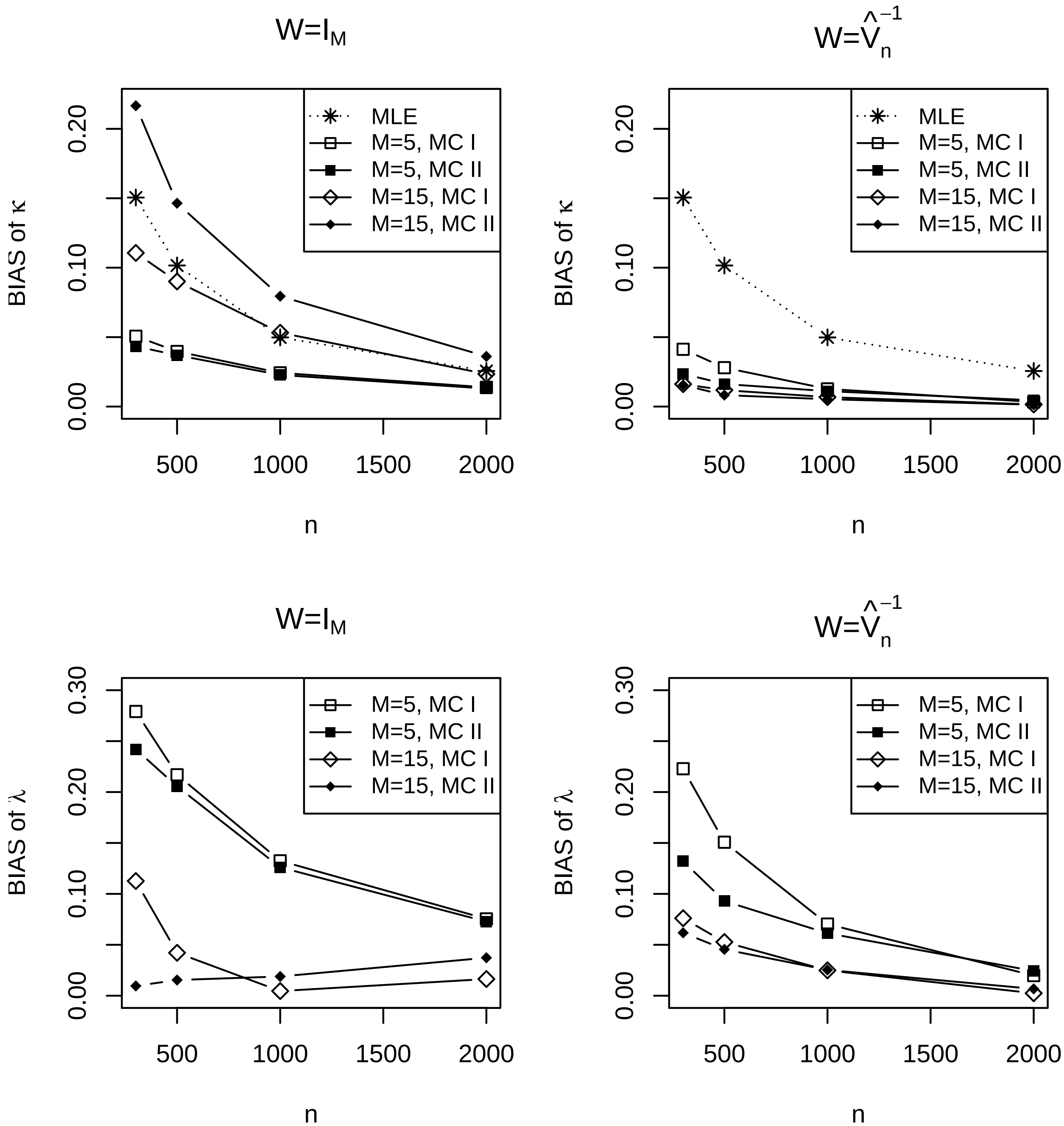

Figure 3: Averaged absolute bias (BIAS) of the MLEs ($\tilde{\beta}_n$), and the combined estimators ($\hat{\theta}_n$) with two moment selection matrices $\mathcal{E}_1$ (MC I) and $\mathcal{E}_2$ (MC II) for κ and λ in the CIR model with $(\kappa, \alpha, \sigma, \lambda) = (0.892, 0.09, \sqrt{0.033}, 0.1))$; M denotes the number of bonds.