

# Inference for Diffusion Processes using Combined Estimating Functions

A. Thavaneswaran<sup>1</sup>, You Liang<sup>1</sup> and N. Ravishanker<sup>2</sup>

<sup>1</sup>University of Manitoba, Canada

<sup>2</sup>University of Connecticut, U.S.A.

## ABSTRACT

*A class of martingale estimating functions provides a convenient framework for studying inference for nonlinear time series models. Further, when information about higher order conditional moments of the observed process is available, the estimation based on combined estimating functions becomes more informative. In this paper, a general framework is developed for estimating parameters of diffusion processes with discretely sampled data using combined estimating functions. The approach is used to study parameter estimation for diffusion models for asset pricing including the Black Scholes model, the Vasicek model, and the Cox-Ingersoll-Ross (CIR) model. Closed form expressions for the gain in information are also discussed in some detail.*

**Key words:** Combined Estimating Functions; Diffusion Processes; Information.

## 1. Introduction

For nonlinear time series models, Chandra and Taniguchi [1], Bera et al. [2], Merkouris [3], Ghahramani and Thavaneswaran [4], and more recently Liang *et al.* [5] among others have studied inference using estimating functions. For discretely sampled diffusion-type models, parameter estimation using estimating functions has been studied in Bibby and Sørensen [6], Sørensen [7], and Bibby *et al.* [8]. However, additional assumptions were made and constraints were imposed to obtain the estimates. Moreover, information issues related to the estimating function approach have not been sufficiently addressed in the literature. In this paper, we study combined *martingale* estimating functions and show that the *combined estimating functions are more informative* when the conditional mean and variance of the observed process depend on the

same parameter of interest. We then apply our approach to discretely sampled observations from diffusion models.

This paper is organized as follows. The rest of Section 1 presents the basics of estimating functions and information associated with estimating functions. Section 2 presents the general model framework for discretely sampled observations from a continuous process, and presents the form of the optimal combined estimating function. In Section 3, the theory is applied to three different diffusion models that are widely used in asset pricing.

Suppose that  $\{\mathbf{y}_t, t=1, \dots, n\}$  is a realization of a discrete-time stochastic process and its distribution depends on a vector parameter  $\boldsymbol{\theta}$  belonging to an open subset  $\Theta$  of the  $p$ -dimensional Euclidean space. Let  $(\Omega, \mathcal{F}, P_0)$  denote the underlying probability space, and let  $\mathcal{F}_t^y$  be the  $\sigma$ -field generated by  $\{\mathbf{y}_1, \dots, \mathbf{y}_t, t \geq 1\}$ . Let  $\mathbf{h}_t = \mathbf{h}_t(\mathbf{y}_1, \dots, \mathbf{y}_t, \boldsymbol{\theta}), 1 \leq t \leq n$  be specified  $q$ -dimensional vectors that are martingales. We consider the class  $\mathcal{M}$  of zero mean and square integrable  $p$ -dimensional martingale estimating functions of the form

$$\mathcal{M} = \left\{ \mathbf{g}_n(\boldsymbol{\theta}) : \mathbf{g}_n(\boldsymbol{\theta}) = \sum_{t=1}^n \mathbf{a}_{t-1} \mathbf{h}_t \right\},$$

where  $\mathbf{a}_{t-1}$  are  $p \times q$  matrices depending on  $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}, 1 \leq t \leq n$ . The estimating functions  $\mathbf{g}_n(\boldsymbol{\theta})$  are further assumed to be almost surely differentiable with respect to the components of  $\boldsymbol{\theta}$ , and such that  $E \left[ \frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{n-1}^y \right]$  and  $E[\mathbf{g}_n(\boldsymbol{\theta}) \mathbf{g}_n(\boldsymbol{\theta})' | \mathcal{F}_{n-1}^y]$  are non-singular for all  $\boldsymbol{\theta} \in \Theta$  and for each  $n \geq 1$ . These expectations are taken with respect to  $P_0$ . Estimators of  $\boldsymbol{\theta}$  can be obtained by solving estimating equation  $\mathbf{g}_n(\boldsymbol{\theta}) = \mathbf{0}$ . Furthermore, the  $p \times p$  matrix  $E[\mathbf{g}_n(\boldsymbol{\theta}) \mathbf{g}_n(\boldsymbol{\theta})' | \mathcal{F}_{n-1}^y]$  is assumed to be positive definite for all  $\boldsymbol{\theta} \in \Theta$ . Then in the class  $\mathcal{M}$  of all zero mean and square integrable martingale estimating functions, the optimal estimating function  $\mathbf{g}_n^*(\boldsymbol{\theta})$  which maximizes, in the partial order of nonnegative definite matrices, the information matrix

$$\mathbf{I}_{\mathbf{g}_n}(\boldsymbol{\theta}) = \left( E \left[ \frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{n-1}^y \right] \right)' \left( E[\mathbf{g}_n(\boldsymbol{\theta}) \mathbf{g}_n(\boldsymbol{\theta})' | \mathcal{F}_{n-1}^y] \right)^{-1} \left( E \left[ \frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{n-1}^y \right] \right)$$

is given by

$$\mathbf{g}_n^*(\boldsymbol{\theta}) = \sum_{t=1}^n \mathbf{a}_{t-1}^* \mathbf{h}_t = \sum_{t=1}^n \left( \mathbf{E} \left[ \frac{\partial \mathbf{h}_t}{\partial \boldsymbol{\theta}} \middle| \mathbf{F}_{t-1}^y \right] \right)' (\mathbf{E}[\mathbf{h}_t \mathbf{h}_t' | \mathbf{F}_{t-1}^y])^{-1} \mathbf{h}_t,$$

and the corresponding optimal information reduces to  $\mathbf{E}[\mathbf{g}_n^*(\boldsymbol{\theta}) \mathbf{g}_n^*(\boldsymbol{\theta})' | \mathbf{F}_{n-1}^y]$  (Godambe [9]). It follows from Lindsay ([10], page 916) that if we solve an unbiased estimating equation  $\mathbf{g}_n(\boldsymbol{\theta}) = \mathbf{0}$  to get an estimator, then the asymptotic variance of the resulting estimator is the inverse of the information  $\mathbf{I}_{\mathbf{g}_n}$ . Hence the estimator obtained from a more informative estimating equation is asymptotically more efficient.

## 2. Estimating Function Approach for a Discretely Sampled Continuous Stochastic Process

Assume that a real-valued continuous-time process  $\{y_t\}$  is recorded discretely at the time points  $0, h, 2h, \dots$ , where  $h$  is the discrete interval of observations of  $\{y_t\}$ . Consider the observable discrete-time process  $\{y_{th}, t = 0, 1, 2, \dots\}$  with conditional moments

$$\mu_t(\boldsymbol{\theta}) = \mathbf{E}[y_{th} | \mathbf{F}_{(t-1)h}^y], \quad (2.1)$$

$$\sigma_t^2(\boldsymbol{\theta}) = \text{Var}(y_{th} | \mathbf{F}_{(t-1)h}^y), \quad (2.2)$$

$$\gamma_t(\boldsymbol{\theta}) = \mathbf{E}[(y_{th} - \mu_t(\boldsymbol{\theta}))^3 | \mathbf{F}_{(t-1)h}^y], \text{ and} \quad (2.3)$$

$$\kappa_t(\boldsymbol{\theta}) = \mathbf{E}[(y_{th} - \mu_t(\boldsymbol{\theta}))^4 | \mathbf{F}_{(t-1)h}^y], \quad (2.4)$$

where  $\mathbf{F}_{(t-1)h}^y$  is the  $\sigma$ -field generated by  $\{y_h, \dots, y_{(t-1)h}, t \geq 0\}$ . We assume that the third and fourth moments of  $y_t$  do not contain any additional parameter. In order to estimate the parameter  $\boldsymbol{\theta}$  based on the observations  $y_0, y_h, \dots, y_{nh}$ , we consider two classes of martingale differences, viz.,  $\{m_t(\boldsymbol{\theta}) = y_{th} - \mu_t(\boldsymbol{\theta}), t = 1, \dots, n\}$  and  $\{M_t(\boldsymbol{\theta}) = m_t^2(\boldsymbol{\theta}) - \sigma_t^2(\boldsymbol{\theta}), t = 1, \dots, n\}$ . The quadratic variations of  $m_t$  and  $M_t$ , and the quadratic covariation of  $m_t$  and  $M_t$  are respectively  $\langle m \rangle_t = \mathbf{E}[m_t^2 | \mathbf{F}_{(t-1)h}^y] = \sigma_t^2$ ,  $\langle M \rangle_t = \mathbf{E}[M_t^2 | \mathbf{F}_{(t-1)h}^y] = \kappa_t - \sigma_t^4$ , and  $\langle m, M \rangle_t = \mathbf{E}[m_t M_t | \mathbf{F}_{(t-1)h}^y] = \gamma_t$ . Then the optimal estimating function and the

corresponding information based on these martingale differences  $m_t$  and  $M_t$  are respectively

$$\mathbf{g}_m^*(\boldsymbol{\theta}) = -\sum_{t=1}^n \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{m_t}{\langle m \rangle_t}; \mathbf{I}_{\mathbf{g}_m^*}(\boldsymbol{\theta}) = \sum_{t=1}^n \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{\partial \mu_t}{\partial \boldsymbol{\theta}'} \frac{1}{\langle m \rangle_t}, \text{ and} \quad (2.5)$$

$$\mathbf{g}_M^*(\boldsymbol{\theta}) = -\sum_{t=1}^n \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{M_t}{\langle M \rangle_t}; \mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta}) = \sum_{t=1}^n \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'} \frac{1}{\langle M \rangle_t}. \quad (2.6)$$

For the discretely observed process  $\{y_{th}\}$ , the following theorem provides optimality of the combined estimating function based on martingales  $m_t$  and  $M_t$ .

**Theorem 1.** *For the general model in (2.1)-(2.4), in the class of all combined estimating functions of the form*

$$\mathbf{G}_C = \left\{ \mathbf{g}_C(\boldsymbol{\theta}) : \mathbf{g}_C(\boldsymbol{\theta}) = \sum_{t=1}^n (\mathbf{a}_{t-1} m_t + \mathbf{b}_{t-1} M_t) \right\}$$

(a) *the optimal estimating function is given by  $\mathbf{g}_C^*(\boldsymbol{\theta}) = \sum_{t=1}^n (\mathbf{a}_{t-1}^* m_t + \mathbf{b}_{t-1}^* M_t)$ ,*

where

$$\mathbf{a}_{t-1}^* = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \left( -\frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \right) \text{ and} \quad (2.7)$$

$$\mathbf{b}_{t-1}^* = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \left( \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} - \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{1}{\langle M \rangle_t} \right); \quad (2.8)$$

(b) *the information  $\mathbf{I}_{\mathbf{g}_C^*}(\boldsymbol{\theta})$  is given by*

$$\mathbf{I}_{\mathbf{g}_C^*}(\boldsymbol{\theta}) = \sum_{t=1}^n \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \left( \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{\partial \mu_t}{\partial \boldsymbol{\theta}'} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'} \frac{1}{\langle M \rangle_t} - \left( \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'} + \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{\partial \mu_t}{\partial \boldsymbol{\theta}'} \right) \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \right). \quad (2.9)$$

**Proof.** Proof of Theorem 1 is similar to that of Theorem 2.1 in Liang *et al.* (2011) and is omitted.

### 3. Examples

In the three examples provided in this section, we assume that  $W_t$  is a Wiener process.

### 3.1. Geometric Brownian Motion with Volatility as a Function of Drift

Consider the Black and Scholes model (Black and Scholes [11]) of the form

$$dX_t = \theta X_t dt + \sigma(\theta) X_t dW_t.$$

We estimate the unknown parameter  $\theta$  appearing simultaneously in the conditional mean and variance. The first four conditional moments of  $X_{th} | X_{(t-1)h}$  are given by

$$\mu_t(\theta) = e^{\theta h} X_{(t-1)h}, \quad \sigma_t^2(\theta) = \left( e^{\sigma^2(\theta)h} - 1 \right) e^{2\theta h} X_{(t-1)h}^2,$$

$$\gamma_t(\theta) = \left( e^{\sigma^2(\theta)h} - 1 \right)^2 \left( e^{\sigma^2(\theta)h} + 2 \right) X_{(t-1)h}^3,$$

$$\kappa_t(\theta) = \left( e^{4\sigma^2(\theta)h} + 2e^{3\sigma^2(\theta)h} + 3e^{2\sigma^2(\theta)h} - 3 \right) \left( e^{\sigma^2(\theta)h} - 1 \right)^2 e^{4\theta h} X_{(t-1)h}^4.$$

Based on the discretely sampled observations  $X_0, X_h, X_{2h}, \dots, X_{nh}$ , we consider two martingale differences  $m_t = X_{th} - e^{\theta h} X_{(t-1)h}$  and

$M_t = X_{th}^2 - 2e^{\theta h} X_{th} X_{(t-1)h} - \left( e^{\sigma^2(\theta)h} - 2 \right) e^{2\theta h} X_{(t-1)h}^2$ , then

$$\langle m \rangle_t = \left( e^{\sigma^2(\theta)h} - 1 \right) e^{2\theta h} X_{(t-1)h}^2,$$

$$\langle M \rangle_t = \left( e^{\sigma^2(\theta)h} - 1 \right)^2 e^{4\theta h} \left( e^{4\sigma^2(\theta)h} + 2e^{3\sigma^2(\theta)h} + 3e^{2\sigma^2(\theta)h} - 4 \right) X_{(t-1)h}^4,$$

$$\langle m, M \rangle_t = \left( e^{\sigma^2(\theta)h} - 1 \right)^2 \left( e^{\sigma^2(\theta)h} + 2 \right) e^{3\theta h} X_{(t-1)h}^3.$$

The optimal estimating function based on the martingale difference  $m_t$  is given by

$$g_m^*(\theta) = -\frac{h}{\left( e^{\sigma^2(\theta)h} - 1 \right) e^{\theta h}} \sum_{t=1}^n \left( \frac{X_{th}}{X_{(t-1)h}} - e^{\theta h} \right),$$

which gives an estimator for  $\theta$  of the form  $\hat{\theta}_n = \frac{1}{h} \ln \frac{1}{n} \sum_{t=1}^n \frac{X_{th}}{X_{(t-1)h}}$ .

Similarly, the optimal estimating function based on the martingale difference  $M_t$  is given by

$$g_M^*(\theta) = -\frac{2h\left(\left(\frac{1}{2}\frac{\partial\sigma^2(\theta)}{\partial\theta}+1\right)e^{\sigma^2(\theta)h}-1\right)}{\left(e^{\sigma^2(\theta)h}-1\right)^2 e^{2\theta h}\left(e^{2\sigma^2(\theta)h}+2e^{3\sigma^2(\theta)h}+3e^{2\sigma^2(\theta)h}-4\right)} \\ \times \sum_{t=1}^n \left( \frac{X_{th}^2}{X_{(t-1)h}^2} - 2e^{\theta h} \frac{X_{th}}{X_{(t-1)h}} - \left(e^{\sigma^2(\theta)h}-2\right)e^{2\theta h} \right).$$

The corresponding information associated with  $\mathbf{g}_m^*(\theta)$  and  $\mathbf{g}_M^*(\theta)$  are respectively

$$I_{\mathbf{g}_m^*}(\theta) = \frac{nh^2}{e^{\sigma^2(\theta)h}-1}, \text{ and } I_{\mathbf{g}_M^*}(\theta) = \frac{4nh^2\left(\left(\frac{1}{2}\frac{\partial\sigma^2(\theta)}{\partial\theta}+1\right)e^{\sigma^2(\theta)h}-1\right)^2}{\left(e^{\sigma^2(\theta)h}-1\right)^2\left(e^{4\sigma^2(\theta)h}+2e^{3\sigma^2(\theta)h}+3e^{2\sigma^2(\theta)h}-4\right)}.$$

Moreover,  $I_{\mathbf{g}_m^*}(\theta) \rightarrow 2n/\sigma^4(\theta)$  and  $I_{\mathbf{g}_M^*}(\theta) \rightarrow n\left(\partial\sigma^2(\theta)/\partial\theta\right)^2/\sigma^4(\theta)$  as  $h \rightarrow 0$ .

It follows from Theorem 1 that the optimal combined estimating function based on  $m_t$  and  $M_t$  has the form

$$g_C^*(\theta) = \frac{e^{4\sigma^2(\theta)h}+2e^{3\sigma^2(\theta)h}+3e^{2\sigma^2(\theta)h}-4}{e^{3\sigma^2(\theta)h}\left(e^{\sigma^2(\theta)h}+1\right)} \\ \times \sum_{t=1}^n \left( \frac{-he^{\sigma^2(\theta)h}\left(e^{3\sigma^2(\theta)h}+2e^{2\sigma^2(\theta)h}-\left(\frac{\partial\sigma^2(\theta)}{\partial\theta}-1\right)e^{\sigma^2(\theta)h}-2\left(\frac{\partial\sigma^2(\theta)}{\partial\theta}+1\right)\right)}{e^{\theta h}\left(e^{\sigma^2(\theta)h}-1\right)\left(e^{4\sigma^2(\theta)h}+2e^{3\sigma^2(\theta)h}+3e^{2\sigma^2(\theta)h}-4\right)} \right) \\ \times \frac{X_{th}-e^{\theta h}X_{t(h-1)}}{X_{t(h-1)}} + \frac{he^{\sigma^2(\theta)h}\left(e^{\sigma^2(\theta)h}-\left(\frac{\partial\sigma^2(\theta)}{\partial\theta}+1\right)\right)}{\left(e^{\sigma^2(\theta)h}-1\right)^2 e^{2\theta h}\left(e^{4\sigma^2(\theta)h}+2e^{3\sigma^2(\theta)h}+3e^{2\sigma^2(\theta)h}-4\right)} \\ \times \left( \frac{X_{th}^2-2e^{\theta h}X_{th}X_{(t-1)h}-\left(e^{\sigma^2(\theta)h}-2\right)e^{2\theta h}X_{(t-1)h}^2}{X_{t(h-1)}^2} \right)$$

$$\begin{aligned}
&= \frac{h}{e^{2\sigma^2(\theta)h+\theta h} \left( e^{\sigma^2(\theta)h} - 1 \right) \left( e^{\sigma^2(\theta)h} + 1 \right)} \\
&\times \sum_{t=1}^n \left( - \left( e^{3\sigma^2(\theta)h} + 2e^{2\sigma^2(\theta)h} - \left( \frac{\partial \sigma^2(\theta)}{\partial \theta} - 1 \right) e^{\theta^2 h} - 2 \left( \frac{\partial \sigma^2(\theta)}{\partial \theta} + 1 \right) \right) \left( \frac{X_{th}}{X_{t(h-1)}} - e^{\theta h} \right) \right. \\
&\left. + \frac{\left( e^{\sigma^2(\theta)h} - \left( \frac{\partial \sigma^2(\theta)}{\partial \theta} + 1 \right) \right)}{\left( e^{\sigma^2(\theta)h} - 1 \right) e^{\theta h}} \left( \frac{X_{th}^2}{X_{(t-1)h}^2} - 2e^{\theta h} \frac{X_{th}}{X_{(t-1)h}} - \left( e^{\sigma^2(\theta)h} - 2 \right) e^{2\theta h} \right) \right),
\end{aligned}$$

with information

$$I_{g_c^*}(\theta) = \frac{nh^2 \left( e^{5\sigma^2(\theta)h} + e^{4\sigma^2(\theta)h} - \left( 2 \frac{\partial \sigma^2(\theta)}{\partial \theta} + 3 \right) e^{3\sigma^2(\theta)h} + \left( \frac{\partial \sigma^2(\theta)}{\partial \theta} + 1 \right)^2 e^{2\sigma^2(\theta)h} \right)}{e^{3\sigma^2(\theta)h} \left( e^{\sigma^2(\theta)h} - 1 \right)^2 \left( e^{\sigma^2(\theta)h} + 1 \right)},$$

which also approaches  $n \left( \partial \sigma^2(\theta) / \partial \theta \right)^2 / \sigma^4(\theta)$  as  $h \rightarrow 0$ .

### 3.2. Ornstein-Uhlenbeck Model

Consider the Ornstein-Uhlenbeck model

$$dX_t = \alpha(\mu - X_t)dt + \sigma dW_t,$$

which was discussed by Vasicek [12] to describe the movement of short term interest rates. The unknown parameters of interest are  $\boldsymbol{\theta} = (\alpha, \mu, \sigma^2)'$  where the parameter  $\alpha$  appears both in the conditional mean and variance. The conditional distribution of  $X_{th} | X_{(t-1)h}$  for a fixed  $h$  follows a normal distribution

$$N\left(\mu\left(1 - e^{-\alpha h}\right) + e^{-\alpha h} X_{(t-1)h}, \sigma^2\left(1 - e^{-2\alpha h}\right) / (2\alpha)\right),$$

with the first four conditional moments

$$\begin{aligned}
\mu_t(\boldsymbol{\theta}) &= \mu\left(1 - e^{-\alpha h}\right) + e^{-\alpha h} X_{(t-1)h}, \sigma_t^2(\boldsymbol{\theta}) = \sigma^2\left(1 - e^{-2\alpha h}\right) / (2\alpha), \\
\gamma_t(\boldsymbol{\theta}) &= 0, \kappa_t(\boldsymbol{\theta}) = 3\sigma^4\left(1 - e^{-2\alpha h}\right)^2 / (4\alpha^2).
\end{aligned}$$

Based on the discretely observed observations  $X_0, X_h, X_{2h}, \dots, X_{nh}$ , the martingale differences are derived as

$$m_t = X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h}),$$

$$M_t = \left( X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h}) \right)^2 - \sigma^2(1 - e^{-2\alpha h}) / (2\alpha),$$

while the quadratic variations and covariation are

$$\langle m \rangle_t = \sigma^2(1 - e^{-2\alpha h}) / (2\alpha), \langle M \rangle_t = \sigma^4(1 - e^{-2\alpha h})^2 / (2\alpha^2), \langle m, M \rangle_t = 0.$$

The optimal estimating functions based on the martingale differences  $m_t$  and  $M_t$  are respectively given by

$$\mathbf{g}_m^*(\boldsymbol{\theta}) = \begin{pmatrix} \frac{2\alpha h e^{-\alpha h}}{\sigma^2(1 - e^{-2\alpha h})} \sum_{t=1}^n (X_{(t-1)h} - \mu) (X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h})) \\ - \frac{2\alpha(1 - e^{-\alpha h})}{\sigma^2(1 - e^{-2\alpha h})} \sum_{t=1}^n (X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h})) \\ 0 \end{pmatrix} \text{ and}$$

$$\mathbf{g}_M^*(\boldsymbol{\theta}) = \begin{pmatrix} - \frac{(2\alpha h e^{-2\alpha h} + e^{-2\alpha h} - 1)}{\sigma^2(1 - e^{-2\alpha h})^2} \sum_{t=1}^n \left( (X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h}))^2 - \frac{\sigma^2(1 - e^{-2\alpha h})}{2\alpha} \right) \\ 0 \\ - \frac{\alpha}{\sigma^4(1 - e^{-2\alpha h})} \sum_{t=1}^n \left( (X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h}))^2 - \frac{\sigma^2(1 - e^{-2\alpha h})}{2\alpha} \right) \end{pmatrix}.$$

Moreover, the information matrices associated with  $\mathbf{g}_m^*(\boldsymbol{\theta})$  and  $\mathbf{g}_M^*(\boldsymbol{\theta})$  are



$$\mathbf{I}_{\mathbf{g}_m^*(\theta)} = \begin{pmatrix} \frac{2\alpha h^2 e^{-2\alpha h} \sum_{t=1}^n (X_{(t-1)h} - \mu)^2}{\sigma^2 (1 - e^{-2\alpha h})} & -\frac{2\alpha h e^{-\alpha h} (1 - e^{-\alpha h}) \sum_{t=1}^n (X_{(t-1)h} - \mu)}{\sigma^2 (1 - e^{-2\alpha h})} & 0 \\ 0 & \frac{2n\alpha (1 - e^{-\alpha h})^2}{\sigma^2 (1 - e^{-2\alpha h})} & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and}$$

$$\mathbf{I}_{\mathbf{g}_M^*(\theta)} = \begin{pmatrix} \frac{n(2\alpha h e^{-2\alpha h} + e^{-2\alpha h} - 1)^2}{2\alpha^2 (1 - e^{-2\alpha h})^2} & 0 & \frac{n(2\alpha h e^{-2\alpha h} + e^{-2\alpha h} - 1)}{2\alpha \sigma^2 (1 - e^{-2\alpha h})} \\ 0 & 0 & 0 \\ \frac{n(2\alpha h e^{-2\alpha h} + e^{-2\alpha h} - 1)}{2\alpha \sigma^2 (1 - e^{-2\alpha h})} & 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

It follows from Theorem 1 that the optimal combined estimating functions based on  $m_t$  and  $M_t$  for  $\alpha$ ,  $\mu$  and  $\sigma^2$  have the following forms:

$$g_C^*(\alpha) = \frac{2\alpha h e^{-\alpha h}}{\sigma^2 (1 - e^{-2\alpha h})} \sum_{t=1}^n (X_{(t-1)h} - \mu) (X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h}))$$

$$- \frac{(2\alpha h e^{-2\alpha h} + e^{-2\alpha h} - 1)}{\sigma^2 (1 - e^{-2\alpha h})^2} \sum_{t=1}^n \left( (X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h}))^2 - \frac{\sigma^2 (1 - e^{-2\alpha h})}{2\alpha} \right)$$

$$g_C^*(\mu) = -\frac{2\alpha (1 - e^{-\alpha h})}{\sigma^2 (1 - e^{-2\alpha h})} \sum_{t=1}^n (X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h}))$$

$$g_C^*(\sigma^2) = -\frac{\alpha}{\sigma^4 (1 - e^{-2\alpha h})} \sum_{t=1}^n \left( (X_{th} - e^{-\alpha h} X_{(t-1)h} - \mu(1 - e^{-\alpha h}))^2 - \frac{\sigma^2 (1 - e^{-2\alpha h})}{2\alpha} \right).$$

The above estimating functions are set equal to zero and can be solved simultaneously to obtain the estimators for  $\alpha$ ,  $\mu$  and  $\sigma^2$  as

$$\hat{\alpha}_n = \frac{1}{h} \ln \left( \frac{n \left( \sum_{t=1}^n X_{(t-1)h}^2 \right) - \left( \sum_{t=1}^n X_{(t-1)h} \right)^2}{n \left( \sum_{t=1}^n X_{(t-1)h} X_{th} \right) - \left( \sum_{t=1}^n X_{th} \right) \left( \sum_{t=1}^n X_{(t-1)h} \right)} \right),$$

$$\hat{\mu}_n = \frac{\left( \sum_{t=1}^n X_{(t-1)h}^2 \right) \left( \sum_{t=1}^n X_{th} \right) - \left( \sum_{t=1}^n X_{(t-1)h} \right) \left( \sum_{t=1}^n X_{(t-1)h} X_{th} \right)}{n \left( \sum_{t=1}^n X_{(t-1)h}^2 \right) - \left( \sum_{t=1}^n X_{(t-1)h} \right)^2 - n \left( \sum_{t=1}^n X_{(t-1)h} X_{th} \right) + \left( \sum_{t=1}^n X_{th} \right) \left( \sum_{t=1}^n X_{(t-1)h} \right)},$$

$$\hat{\sigma}_n^2 = \frac{2\hat{\alpha}_n}{n(1 - e^{-2\hat{\alpha}_n h})} \sum_{t=1}^n \left( X_{th} - e^{-\hat{\alpha}_n h} X_{(t-1)h} - \hat{\mu}_n (1 - e^{-\hat{\alpha}_n h}) \right)^2.$$

The corresponding information matrix for  $\boldsymbol{\theta} = (\alpha, \mu, \sigma^2)'$  is

$$\mathbf{I}_{\mathbf{g}_C^*(\boldsymbol{\theta})} = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\mu} & I_{\alpha\sigma^2} \\ I_{\alpha\mu} & I_{\mu\mu} & 0 \\ I_{\alpha\sigma^2} & 0 & I_{\sigma^2\sigma^2} \end{pmatrix},$$

where

$$I_{\alpha\alpha} = \frac{2\alpha h^2 e^{-2\alpha h} \sum_{t=1}^n (X_{(t-1)h} - \mu)^2}{\sigma^2 (1 - e^{-2\alpha h})} + \frac{n(2\alpha h e^{-2\alpha h} + e^{-2\alpha h} - 1)^2}{2\alpha^2 (1 - e^{-2\alpha h})^2},$$

$$I_{\alpha\mu} = -\frac{2\alpha h e^{-\alpha h} (1 - e^{-\alpha h}) \sum_{t=1}^n (X_{(t-1)h} - \mu)}{\sigma^2 (1 - e^{-2\alpha h})},$$

$$I_{\alpha\sigma^2} = \frac{n(2\alpha h e^{-2\alpha h} + e^{-2\alpha h} - 1)}{2\alpha\sigma^2 (1 - e^{-2\alpha h})}, I_{\mu\mu} = \frac{2n\alpha (1 - e^{-\alpha h})^2}{\sigma^2 (1 - e^{-2\alpha h})}, I_{\sigma^2\sigma^2} = \frac{n}{2\sigma^4}.$$

### 3.3. Cox-Ingersoll-Ross Model

Consider the CIR model (Cox *et al.*, [13]) of the form

$$dX_t = -k(X_t - \theta)dt + \eta\sqrt{X_t}dW_t,$$

which is widely used to model short term interest rates. The unknown parameters of interest are  $\boldsymbol{\theta} = (k, \theta, \eta^2)'$ . By using Itô's formula for  $X_t^u, u = 2, 3, 4$ , the first four conditional moments of  $X_{th} | X_{(t-1)h}$  are directly calculated as

$$\mu_t(\boldsymbol{\theta}) = \theta - (\theta - X_{(t-1)h})e^{-kh},$$

$$\sigma_t^2(\boldsymbol{\theta}) = \frac{\eta^2}{2k} \left( \theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh} \right),$$

$$\gamma_t(\boldsymbol{\theta}) = \frac{\eta^4}{2k^2} \left( \theta - 3(\theta - X_{(t-1)h})e^{-kh} + 3(\theta - 2X_{(t-1)h})e^{-2kh} - (\theta - 3X_{(t-1)h})e^{-3kh} \right),$$

$$\begin{aligned} \kappa_t(\boldsymbol{\theta}) &= \frac{3\eta^6}{4k^3} \left( \theta - 4(\theta - X_{(t-1)h})e^{-kh} + 6(\theta - 2X_{(t-1)h})e^{-2kh} - 4(\theta - 3X_{(t-1)h})e^{-3kh} \right. \\ &\quad \left. + (\theta - 4X_{(t-1)h})e^{-4kh} \right) + 3\sigma_t^4(\boldsymbol{\theta}). \end{aligned}$$

Based on the discretely observed observations  $X_0, X_h, X_{2h}, \dots, X_{nh}$ ,

$$m_t = X_{th} - e^{-kh} X_{(t-1)h} - \theta(1 - e^{-kh}),$$

$$\begin{aligned} M_t &= \left( X_{th} - e^{-kh} X_{(t-1)h} \right)^2 - 2\theta(1 - e^{-kh}) X_{th} + e^{-kh} (1 - e^{-kh}) \left( 2\theta - \frac{\eta^2}{k} \right) X_{(t-1)h} \\ &\quad + \theta \left( \theta - \frac{\eta^2}{2k} \right) (1 - e^{-kh})^2, \end{aligned}$$

and

$$\langle m \rangle_t = \frac{\eta^2}{2k} \left( \theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh} \right),$$

$$\begin{aligned} \langle M \rangle_t &= \frac{3\eta^6}{4k^3} \left( \theta - 4(\theta - X_{(t-1)h})e^{-kh} + 6(\theta - 2X_{(t-1)h})e^{-2kh} - 4(\theta - 3X_{(t-1)h})e^{-3kh} \right. \\ &\quad \left. + (\theta - 4X_{(t-1)h})e^{-4kh} \right) + 2\sigma_t^4(\boldsymbol{\theta}), \end{aligned}$$

$$\langle m, M \rangle_t = \frac{\eta^4}{2k^2} \left( \theta - 3(\theta - X_{(t-1)h})e^{-kh} + 3(\theta - 2X_{(t-1)h})e^{-2kh} - (\theta - 3X_{(t-1)h})e^{-3kh} \right).$$

Now the optimal estimating function based on the martingale difference  $m_t$  is given by

$$\mathbf{g}_m^*(\boldsymbol{\theta}) = \begin{pmatrix} -\frac{2hke^{-kh}}{\eta^2} \sum_{t=1}^n \frac{(\theta - X_{(t-1)h})(X_{th} - e^{-kh}X_{(t-1)h} - \theta(1 - e^{-kh}))}{\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh}} \\ -\frac{2k(1 - e^{-kh})}{\eta^2} \sum_{t=1}^n \frac{X_{th} - e^{-kh}X_{(t-1)h} - \theta(1 - e^{-kh})}{\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh}} \\ 0 \end{pmatrix},$$

with associated information matrix

$$\mathbf{I}_{\mathbf{g}_m^*(\boldsymbol{\theta})} = \sum_{t=1}^n \mathbf{U}_t(\boldsymbol{\theta}) = \sum_{t=1}^n \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{\partial \mu_t}{\partial \boldsymbol{\theta}'} \frac{1}{\langle m \rangle_t} = \begin{pmatrix} I_{11}^m & I_{12}^m & 0 \\ I_{12}^m & I_{22}^m & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where

$$I_{11}^m = \frac{2kh^2 e^{-2kh}}{\eta^2} \sum_{t=1}^n \frac{(\theta - X_{(t-1)h})^2}{\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh}},$$

$$I_{12}^m = \frac{2khe^{-kh}(1 - e^{-kh})}{\eta^2} \sum_{t=1}^n \frac{\theta - X_{(t-1)h}}{\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh}},$$

$$I_{22}^m = \frac{2k(1 - e^{-kh})^2}{\eta^2} \sum_{t=2}^n \frac{1}{\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh}}.$$

The optimal estimating function based on the martingale difference  $M_t$  is given

by

$$\mathbf{g}_M^*(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\eta^2 \sum_{t=1}^n (\theta + 2(1 - hk)(\theta - X_{(t-1)h})e^{-kh} - (1 - 2hk)(\theta - 2X_{(t-1)h})e^{-2kh}) \left( (X_{th} - e^{-kh}X_{(t-1)h} - \theta(1 - e^{-kh}))^2 - \sigma_t^2(\boldsymbol{\theta}) \right)}{2k^2 \sum_{t=1}^n \langle M \rangle_t} \\ -\frac{\eta^2 (1 - e^{-kh})^2 \sum_{t=1}^n (X_{th} - e^{-kh}X_{(t-1)h} - \theta(1 - e^{-kh}))^2 - \sigma_t^2(\boldsymbol{\theta})}{2k \sum_{t=1}^n \langle M \rangle_t} \\ -\frac{1}{2k} \sum_{t=1}^n \frac{(\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh}) \left( (X_{th} - e^{-kh}X_{(t-1)h} - \theta(1 - e^{-kh}))^2 - \sigma_t^2(\boldsymbol{\theta}) \right)}{\langle M \rangle_t} \end{pmatrix}$$

and the associated information is

$$\mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta}) = \sum_{t=1}^n \mathbf{V}_t(\boldsymbol{\theta}) = \sum_{t=1}^n \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'} \frac{1}{\langle M \rangle_t} = \begin{pmatrix} I_{11}^M & I_{12}^M & I_{13}^M \\ I_{12}^m & I_{22}^m & I_{23}^m \\ I_{13}^M & I_{13}^M & I_{33}^M \end{pmatrix},$$

$$I_{11}^M = \frac{\eta^4}{4k^4} \sum_{t=1}^n \frac{\left( \theta + 2(1-hk)(\theta - X_{(t-1)h})e^{-kh} - (1-2hk)(\theta - 2X_{(t-1)h})e^{-2kh} \right)^2}{\langle M \rangle_t},$$

$$I_{12}^M = -\frac{\eta^4 (1-e^{-kh})^2}{4k^3} \sum_{t=1}^n \frac{\theta + 2(1-hk)(\theta - X_{(t-1)h})e^{-kh} - (1-2hk)(\theta - 2X_{(t-1)h})e^{-2kh}}{\langle M \rangle_t},$$

$$I_{13}^M = -\frac{\eta^2}{4k^3} \sum_{t=1}^n \left( \frac{\left( \theta + 2(1-hk)(\theta - X_{(t-1)h})e^{-kh} - (1-2hk)(\theta - 2X_{(t-1)h})e^{-2kh} \right)}{\langle M \rangle_t} \right. \\ \left. \times \left( \theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh} \right) \right),$$

$$I_{22}^M = \frac{\eta^4 (1-e^{-kh})^4}{4k^2} \sum_{t=1}^n \frac{1}{\langle M \rangle_t},$$

$$I_{23}^M = \frac{\eta^2 (1-e^{-kh})^2}{4k^2} \sum_{t=1}^n \frac{\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh}}{\langle M \rangle_t},$$

$$I_{33}^M = \sum_{t=1}^n \frac{1}{4k^2} \frac{\left( \theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh} \right)^2}{\langle M \rangle_t}.$$

Based on  $m_t$  and  $M_t$ , the optimal combined estimating function is given by

$$\mathbf{g}_C^*(\boldsymbol{\theta}) = \sum_{t=1}^n \left( \mathbf{a}_{t-1}^* m_t + \mathbf{b}_{t-1}^* M_t \right),$$

where

$$\mathbf{a}_{t-1}^* = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \times \left( \begin{array}{l} \frac{h(\theta - X_{(t-1)h})e^{-kh}}{\langle m \rangle_t} - \frac{\eta^2 (\theta + 2(1-hk)(\theta - X_{(t-1)h})e^{-kh} - (1-2hk)(\theta - 2X_{(t-1)h})e^{-2kh}) \langle m, M \rangle_t}{2k^2 \langle m \rangle_t \langle M \rangle_t} \\ \frac{1 - e^{-kh}}{\langle m \rangle_t} + \frac{\eta^2 (1 - e^{-kh})^2 \langle m, M \rangle_t}{2k \langle m \rangle_t \langle M \rangle_t} \\ \frac{(\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh}) \langle m, M \rangle_t}{2k \langle m \rangle_t \langle M \rangle_t} \end{array} \right),$$

and

$$\mathbf{b}_{t-1}^* = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \times \left( \begin{array}{l} \frac{h(\theta - X_{(t-1)h})e^{-kh} \langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} + \frac{\eta^2 (\theta + 2(1-hk)(\theta - X_{(t-1)h})e^{-kh} - (1-2hk)(\theta - 2X_{(t-1)h})e^{-2kh})}{2k^2 \langle M \rangle_t} \\ \frac{1 - e^{-kh} \langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} - \frac{\eta^2 (1 - e^{-kh})^2}{2k \langle M \rangle_t} \\ - \frac{(\theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh})}{2k \langle M \rangle_t} \end{array} \right).$$

Further let  $W_t(\boldsymbol{\theta})$  be

$$\mathbf{W}_t(\boldsymbol{\theta}) = \left( \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{12} & w_{22} & w_{23} \\ w_{13} & w_{23} & w_{33} \end{pmatrix},$$

$$\begin{aligned}
W_{11} &= -\frac{\eta^2 h e^{-kh}}{k^2} (\theta - X_{(t-1)h}) \left( \theta + 2(1-hk)(\theta - X_{(t-1)h})e^{-kh} - (1-2hk)(\theta - 2X_{(t-1)h})e^{-2kh} \right), \\
W_{12} &= \frac{\eta^2 h e^{-kh} (1 - e^{-kh})^2}{2k} (\theta - X_{(t-1)h}) \\
&\quad - \frac{\eta^2 (1 - e^{-kh})}{2k^2} \left( \theta + 2(1-hk)(\theta - X_{(t-1)h})e^{-kh} - (1-2hk)(\theta - 2X_{(t-1)h})e^{-2kh} \right), \\
W_{13} &= \frac{h e^{-kh}}{2k} (\theta - X_{(t-1)h}) \left( \theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh} \right), \\
W_{22} &= \frac{h(1 - e^{-kh})^3}{k} (\theta - X_{(t-1)h}), W_{23} = \frac{1 - e^{-kh}}{2k} \left( \theta - 2(\theta - X_{(t-1)h})e^{-kh} + (\theta - 2X_{(t-1)h})e^{-2kh} \right), W_{33} = 0.
\end{aligned}$$

Then, the information matrix associated with the optimal combined estimating function is

$$\mathbf{I}_{\mathbf{g}_c^*(\boldsymbol{\theta})} = \sum_{t=1}^n \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \left( \mathbf{U}_t(\boldsymbol{\theta}) + \mathbf{V}_t(\boldsymbol{\theta}) - \mathbf{W}_t(\boldsymbol{\theta}) \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \right).$$

#### 4. Conclusions

In this paper, the combined estimating function approach described in Liang *et al.* [5] is applied to study estimation for discretely sampled diffusion models which are widely used in finance. Specifically, estimation details based on the combined estimation functions are provided for the Black Scholes model under a geometric Brownian motion, the Ornstein-Uhlenbeck model, and the Cox-Ingersoll-Ross model, which are commonly used in finance. Information associated with the combined estimating function is also discussed in detail for each case.

#### References:

1. Chandra, S. A. and Taniguchi, M. (2001). Estimating functions for nonlinear time series models. *Annals of the Institute of Statistical Mathematics*, **53**, 125 - 141.
2. Bera, A. K., Biliyas, Y. and Simlai, P. (2006). Estimating functions and equations: an essay on historical developments with applications to

- economics. In *Palgrave Handbook of Econometrics*, T. C. Mills and K. Patterson, eds., Palgrave Macmillan: London, Vol I, 427 - 476.
3. Merkouris, T. (2007). Transform martingale estimating functions. *The Annals of Statistics*, **35**, 1975 - 2000.
  4. Ghahramani, M. and Thavaneswaran, A. (2009). Combining estimating functions for volatility. *Journal of Statistical Planning and Inference*, **139**, 1449 - 1461.
  5. Liang, Y., Thavaneswaran, A. and Abraham, B. Joint estimation using quadratic estimating functions. *Journal of Probability and Statistics*, Volume 2011 (2011), Article ID 372512, 14 pages.
  6. Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17-39.
  7. Sørensen, M. (2009). Parametric inference for discretely sampled stochastic differential equations. In *Handbook of Financial Time Series*. T. G. Andersen, R. A. Davis, and J. P. Kreis, eds., Springer: Heidelberg, 531-553.
  8. Bibby, B. M., Jacobsen, M. and Sørensen, M. (2010). Estimating functions for discretely sampled diffusion-type models. In *Handbook of Financial Economics*. Y. Ait-Sahalia and L. P. Hansen, eds., Amsterdam: North Holland, 203-268.
  9. Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika*, **72**, 319 - 328.
  10. Lindsay, B. G. (1985). Using empirical Bayes inference for increased efficiency. *Annals of Statistics*, **13**, 914-931.
  11. Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81 (3)**, 637-654.
  12. Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5 (2)**, 177-188.
  13. Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, **53 (2)**, 385-407.
  14. Thavaneswaran, A. and Abraham, B. (1988). Estimation of nonlinear time series models using estimating functions. *Journal of Time Series Analysis*, **9**, 99 - 108.