

# Inference for Interest Rate Models Using Milstein's Approximation

Theodoro Koulis, Aera Thavaneswaran

Department of Statistics, University of Manitoba, Winnipeg, Canada

Email: theo\_koulis@umanitoba.ca

Received October 9, 2012; revised November 24, 2012; accepted December 8, 2012

## ABSTRACT

A class of martingale estimating functions based on the first two moments of the observed process provides a convenient framework for estimating the parameters of diffusion processes [1]. In the Bayesian set up, combined estimating functions had been studied for diffusion processes in [2] with filtering applications. However, when the conditional mean and the conditional variance are functions of parameters of interest in a diffusion process model, the basic martingales generating components of quadratic estimating functions are such that one is an absolute continuous function with respect to the other [3, p. 94]. Hence, the combined martingale estimating functions cannot be constructed for continuous-time diffusion processes. In this paper, a general framework for *parameter estimation* of discretely observed interest rate models is developed by using the Milstein approximation and closed form expressions for the information gain are also obtained. The method is used to study the estimates of the parameters for an extended version of the Cox-Ingersoll-Ross interest rate model.

**Keywords:** Interest Rate Models; Combined Estimating Functions; Information; Diffusion Processes; Milstein Approximation

## 1. Introduction

Inference for discrete-time stochastic processes using estimating functions was discussed in [4]. In [1] and [5], estimation for semimartingales was studied using estimating functions. In addition, filtering and prediction problems were studied in [6] and [2] using estimating functions in the Bayesian context.

The standard method of estimation for parameters in the drift coefficient of interest rate models [7] involves the calculation of a likelihood ratio (Radon-Nikodym derivative) and hence the maximum likelihood estimator(s). This is less than straightforward for complicated models, and indeed it is *not available* at all because of the non-existence of the Radon-Nikodym derivative. The estimating function method, however, allows estimators to be obtained straightforwardly under very general conditions on the first two conditional moments [3, p. 131]. They can deal, in particular, with the situation in which the Brownian motion in a diffusion is replaced by a general square-integrable martingale as in [1]. The combined estimating function approach used in this paper, based on selection of an optimal estimating function from within a specified class of martingale estimating functions, involves assumptions on the first four conditional moments of the underlying process. In most realistic situations the

diffusions cannot be observed continuously, so discrete time approximation to stochastic integrals or a direct approach using the discrete time observations is required.

Recently in [8], among others, the estimating functions approach was used to study the estimation problems for some discretely observed interest rate models. However, these methods involve the closed form expressions for the first four conditional moments, obtained by Ito's approximations, and these are *not available* for general time-homogeneous diffusion process models and in particular for an extended CIR interest rate model.

In [9], the asymptotic theory of the maximum likelihood estimator for diffusion models was studied, first by using Milstein's approximation of diffusion processes [10], and further by approximating the conditional transition density by a normal density by ignoring the skewness and kurtosis.

In this paper, we study combined *martingale* estimating functions for interest rate models and show that the combined estimating functions are more informative when the conditional mean and variance of the observed process depend on the *parameter* of interest. This paper is organized as follows. The rest of Section 1 presents the basics of estimating functions and information associated with estimating functions for discrete-time stochastic processes. Section 2 presents combined estimating func-

tions for discretely observed continuous-time diffusion processes based on closed form expressions for the first four conditional moments via Itô's formula. In Section 3, the theory of combined estimating functions is applied to general diffusion processes.

Suppose that  $\{y_t, t = 1, \dots, n\}$  is a realization of a discrete-time stochastic process and its distribution depends on a vector parameter  $\theta$  belonging to an open subset  $\Theta$  of the  $p$ -dimensional Euclidean space. Let  $(\Omega, \mathcal{F}, P_\theta)$  denote the underlying probability space, and let  $\mathfrak{F}_t^y$  be the  $\sigma$ -field generated by  $\{y_1, \dots, y_t, t \geq 1\}$ . Let  $h_t = h_t(y_1, \dots, y_t, \theta)$ ,  $1 \leq t \leq n$  be specified  $q$ -dimensional vectors that are martingales. We consider the class  $\mathfrak{M}$  of zero mean and square integrable  $p$ -dimensional martingale estimating functions of the form

$$\mathfrak{M} = \left\{ g_n(\theta) : g_n(\theta) = \sum_{t=1}^n a_{t-1} h_t \right\}$$

where  $a_{t-1}$  are  $p \times q$  matrices depending on  $y_1, \dots, y_{t-1}$ ,  $1 \leq t \leq n$ . The estimating functions  $g_n(\theta)$  are further assumed to be almost surely differentiable with respect to the components of  $\theta$  and such that

$$E \left[ \frac{\partial g_n(\theta)}{\partial \theta} \middle| \mathfrak{F}_{n-1}^y \right] \text{ and } E \left[ g_n(\theta) g_n(\theta)' \middle| \mathfrak{F}_{n-1}^y \right] \text{ are non-}$$

singular for all  $\theta \in \Theta$  and for each  $n \geq 1$ . The expectations are always taken with respect to  $P_\theta$ . Estimators of  $\theta$  can be obtained by solving the estimating equation  $g_n(\theta) = 0$ . Furthermore, the  $p \times p$  matrix

$$E \left[ g_n(\theta) g_n(\theta)' \middle| \mathfrak{F}_{n-1}^y \right] \text{ is assumed to be positive definite}$$

for all  $\theta \in \Theta$ . Then in the class of all zero mean and square integrable martingale estimating functions  $\mathfrak{M}$ , the optimal estimating function  $g_n^*(\theta)$  which maximizes, in the partial order of nonnegative definite matrices, the information matrix

$$I_g(\theta) = \left( \sum_{t=1}^n a_{t-1} E \left[ \frac{\partial h_t}{\partial \theta} \middle| \mathfrak{F}_{t-1}^y \right] \right)' \left( \sum_{t=1}^n E \left[ (a_{t-1} h_t)(a_{t-1} h_t)' \middle| \mathfrak{F}_{t-1}^y \right] \right)^{-1} \left( \sum_{t=1}^n a_{t-1} E \left[ \frac{\partial h_t}{\partial \theta} \middle| \mathfrak{F}_{t-1}^y \right] \right)$$

is given by

$$g_n^*(\theta) = \sum_{t=1}^n a_{t-1}^* h_t = \sum_{t=1}^n \left( E \left[ \frac{\partial h_t}{\partial \theta} \middle| \mathfrak{F}_{t-1}^y \right] \right)' \left( E \left[ h_t h_t' \middle| \mathfrak{F}_{t-1}^y \right] \right)^{-1} h_t$$

and the corresponding optimal information reduces to

$$E \left[ g_n^*(\theta) g_n^*(\theta)' \middle| \mathfrak{F}_{n-1}^y \right].$$

The function  $g_n^*(\theta)$  is also called the *optimal esti-*

*mating function* and has properties similar to those of a score function in the sense that  $E[g_n^*(\theta)] = 0$  and  $E \left[ g_n^*(\theta) g_n^*(\theta)' \right] = -E \left[ \frac{\partial g_n^*(\theta)}{\partial \theta'} \right]$ . This is a more general

result in the sense that for its validity we do not need to assume that the true underlying distribution belongs to the exponential family of distributions. Moreover, it follows from [11, p. 916] that if we solve an unbiased estimating equation  $g_n(\theta) = 0$  to get an estimator, then the asymptotic variance of the resulting estimator is the inverse of the information  $I_{g_n}$ . Hence the estimator obtained from a more informative estimating equation is asymptotically more efficient.

## 2. Combined Estimating Functions for Discretely Observed Diffusions

In this section, we discuss the discrete time results on combining estimating functions and obtain the closed form expression for the gain in information. Assume the real-valued continuous time process  $\{y_t\}$  is recorded discretely at the time points  $h, 2h, \dots$ , where  $h$  is the discrete interval of observations of  $\{y_t\}$ . Now we consider the observable discrete-time process  $\{y_{th}, t = 1, 2, \dots\}$  with conditional moments

$$\mu_t(\theta) = E \left[ y_{th} \middle| \mathfrak{F}_{(t-1)h}^y \right] \tag{2.1}$$

$$\sigma_t^2(\theta) = \text{Var} \left[ y_{th} \middle| \mathfrak{F}_{(t-1)h}^y \right] \tag{2.2}$$

$$\gamma_t(\theta) = E \left[ (y_{th} - \mu_t(\theta))^3 \middle| \mathfrak{F}_{(t-1)h}^y \right] \tag{2.3}$$

$$\kappa_t(\theta) = E \left[ (y_{th} - \mu_t(\theta))^4 \middle| \mathfrak{F}_{(t-1)h}^y \right] \tag{2.4}$$

where  $\mathfrak{F}_{(t-1)h}^y$  is the  $\sigma$ -field generated by  $\{y_{1h}, \dots, y_{(t-1)h}, t \geq 1\}$ . That is, we assume that the third and the fourth moments of  $y_t$  do not contain any additional parameters. In order to estimate the parameter  $\theta$  based on the observations  $y_{1h}, \dots, y_{nh}$ , we consider two classes of martingale differences

$\{m_t(\theta) = y_{th} - \mu_t(\theta), t = 1, \dots, n\}$  and  $\{M_t(\theta) = m_t^2(\theta) - \sigma_t^2(\theta), t = 1, \dots, n\}$ , where the quadratic variation and covariation of  $m_t$  and  $M_t$  are

$$\langle m \rangle_t = E \left[ m_t^2 \middle| \mathfrak{F}_{(t-1)h}^y \right] = \sigma_t^2,$$

$$M_t = E \left[ M_t^2 \middle| \mathfrak{F}_{(t-1)h}^y \right] = \kappa_t - \sigma_t^4,$$

and

$$\langle m, M \rangle_t = E \left[ m_t M_t \middle| \mathfrak{F}_{(t-1)h}^y \right] = \gamma_t,$$

respectively. The optimal estimating functions based on the martingale differences  $m_t$  and  $M_t$ , and the corre-

sponding information are given by

$$g_m^*(\theta) = -\sum_{t=1}^n \frac{\partial \mu_t}{\partial \theta} \frac{m_t}{\langle m \rangle_t}, \tag{2.5}$$

$$I_{g_m^*}^*(\theta) = \sum_{t=1}^n \frac{\partial \mu_t}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \frac{1}{\langle m \rangle_t},$$

$$g_M^*(\theta) = -\sum_{t=1}^n \frac{\partial \sigma_t^2}{\partial \theta} \frac{M_t}{\langle M \rangle_t}, \tag{2.6}$$

$$I_{g_M^*}^*(\theta) = \sum_{t=1}^n \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{1}{\langle M \rangle_t}.$$

The following theorem provides optimality of the combined estimating function based on martingales  $m_t$  and  $M_t$  for the multi-parameter case.

**Theorem 1.** For a discretely observed process, in the class of all combined estimating functions of the form

$$g_c = \left\{ g_c(\theta) : g_c(\theta) = \sum_{t=1}^n (a_{t-1} m_t + b_{t-1} M_t) \right\},$$

(a) The optimal estimating function is given by

$$g_c^*(\theta) = \sum_{t=1}^n (a_{t-1}^* m_t + b_{t-1}^* M_t), \text{ where}$$

$$a_{t-1}^* = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \left( -\frac{\partial \mu_t}{\partial \theta} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \right)$$

and

$$b_{t-1}^* = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \left( \frac{\partial \mu_t}{\partial \theta} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} - \frac{\partial \sigma_t^2}{\partial \theta} \frac{1}{\langle m \rangle_t} \right);$$

(b) the information  $I_{g_c^*}^*(\theta)$  is given by

$$I_{g_c^*}^*(\theta) = \sum_{t=1}^n \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \left[ \frac{\partial \mu_t}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{1}{\langle M \rangle_t} + \left( \frac{\partial \mu_t}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \right) \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \right];$$

(c) the gain in information  $I_{g_c^*}^*(\theta) - I_{g_m^*}^*(\theta)$  is given by

$$\sum_{t=1}^n \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \left[ \frac{\partial \mu_t}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{1}{\langle m \rangle_t} + \left( \frac{\partial \mu_t}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \right) \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \right].$$

**Example 1** (Combined Estimating Functions for Cox-Ingersoll-Ross Model). Recently there has been a growing interest in studying inference for interest rate models. In most realistic situations, the diffusion cannot be observed continuously, so discrete time approximations to stochastic integrals or a direct approach using discrete time observations is required. As a concrete illustration of the methodology, we shall discuss the estimation for the Cox, Ingersoll and Ross [12] short-term interest rate model of the form

$$dy_t = -k(y_t - \theta)dt + \eta\sqrt{y_t}dW_t, \tag{2.7}$$

with  $y_0 > 0$ ,  $k > 0$ ,  $\theta > 0$ ,  $\eta > 0$  and  $W_t$  is the standard Brownian motion. The unknown parameters of interest are  $\theta = (k, \theta, \eta^2)'$ . Let

$\alpha_t(\lambda) = \alpha_t(\lambda, \theta, k) = (\theta - \lambda y_{(t-1)h})e^{-\lambda kh}$ . It is of interest to note that in this example we can obtain the closed form expressions for the first four conditional moments by using Itô's formula for  $y_t^u$ ,  $u = 2, 3, 4$ . The first four conditional moments of  $y_t$  are calculated as (see Appendix A for the details):

$$\mu_t(\theta) = \theta - \alpha_t(1), \tag{2.8}$$

$$\sigma_t^2(\theta) = \frac{\eta^2}{2k}(\theta - 2\alpha_t(1) + \alpha_t(2)), \tag{2.9}$$

$$\gamma_t(\theta) = \frac{\eta^4}{2k^2}(\theta - 3\alpha_t(1) + 3\alpha_t(2) - \alpha_t(3)), \tag{2.10}$$

and

$$\kappa_t(\theta) = \frac{3\eta^6}{4k^3}(\theta - 4\alpha_t(1) + 6\alpha_t(2) - 4\alpha_t(3) + \alpha_t(4)) + 3\sigma_t^4(\theta). \tag{2.11}$$

Then based on the discretely observed observations  $y_0, y_h, y_{2h}, \dots, y_{nh}$ , the martingale differences are

$m_t = y_{th} - \mu_t(\theta)$ , and  $M_t = m_t^2 - \sigma_t^2(\theta)$ . Also,

$m_t = \sigma_t^2(\theta), M_t = \kappa_t(\theta) - \sigma_t^4(\theta)$ , and

$\langle m, M \rangle_t = \gamma_t(\theta)$ . The derivatives are given by

$$\frac{\partial \mu_t(\theta)}{\partial \theta} = (\alpha_t(1)h, 1 - e^{-kh}, 0)', \tag{2.12}$$

$$\frac{\partial \sigma_t^2(\theta)}{\partial \theta} = \begin{pmatrix} -\frac{\eta^2}{2k^2}(\theta + 2(1-hk)\alpha_t(1) - (1-2hk)\alpha_t(2)) \\ \frac{\eta^2(1-e^{-kh})^2}{2k} \\ \frac{1}{2k}(\theta - 2\alpha_t(1) + \alpha_t(2)) \end{pmatrix} \tag{2.13}$$

Hence, the optimal estimating functions based on the martingale differences  $m_t$  and the corresponding infor-

mation matrix are given by

$$g_m^*(\theta) = \begin{pmatrix} -he^{-kh} \sum_{t=1}^n \frac{(\theta - y_{(t-1)h}) m_t}{\sigma_t^2(\theta)} \\ -(1 - e^{-kh}) \sum_{t=1}^n \frac{m_t}{\sigma_t^2(\theta)} \\ 0 \end{pmatrix}, \quad (2.14)$$

$$I_{g_m^*(\theta)} = \begin{pmatrix} I_{11}^m & I_{12}^m & 0 \\ I_{12}^m & I_{22}^m & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where

$$I_{11}^m = h^2 \sum_{t=1}^n \frac{\alpha_t^2(1)}{\sigma_t^2(\theta)},$$

$$I_{12}^m = h(1 - e^{-kh}) \sum_{t=1}^n \frac{\alpha_t(1)}{\sigma_t^2(\theta)},$$

$$I_{22}^m = (1 - e^{-kh})^2 \sum_{t=1}^n \frac{1}{\sigma_t^2(\theta)}.$$

Similarly, the optimal estimating functions based on the martingale differences  $M_t$  and the corresponding information matrix are given by

$$g_M^*(\theta) = \begin{pmatrix} \frac{\eta^2 \sum_{t=1}^n (\theta + 2(1 - hk)\alpha_t(1) - (1 - 2hk)\alpha_t(2)) M_t}{2k^2 \sum_{t=1}^n \langle M \rangle_t} \\ -\frac{\eta^2 (1 - e^{-kh})^2 \sum_{t=1}^n M_t}{2k \sum_{t=1}^n \langle M \rangle_t} \\ -\frac{1 \sum_{t=1}^n (\theta - 2\alpha_t(1) + \alpha_t(2)) M_t}{2k \sum_{t=1}^n \langle M \rangle_t} \end{pmatrix}, \quad (2.15)$$

$$I_{g_M^*(\theta)} = \begin{pmatrix} I_{11}^M & I_{12}^M & I_{13}^M \\ I_{12}^M & I_{22}^M & I_{23}^M \\ I_{13}^M & I_{13}^M & I_{33}^M \end{pmatrix},$$

where

$$I_{11}^M = \frac{\eta^4 \sum_{t=1}^n (\theta + 2(1 - hk)\alpha_t(1) - (1 - 2hk)\alpha_t(2))^2}{4k^4 \sum_{t=1}^n \langle M \rangle_t},$$

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

$$I_{13}^M = -\frac{\eta^2 \sum_{t=1}^n (\theta + 2(1 - hk)\alpha_t(1) - (1 - 2hk)\alpha_t(2))(\theta - 2\alpha_t(1) + \alpha_t(2))}{4k^3 \sum_{t=1}^n \langle M \rangle_t},$$

$$I_{22}^M = \frac{\eta^4 (1 - e^{-kh})^4 \sum_{t=1}^n 1}{4k^2 \sum_{t=1}^n \langle M \rangle_t}, \quad I_{23}^M = \frac{\eta^2 (1 - e^{-kh})^2 \sum_{t=1}^n \theta - 2\alpha_t(1) + \alpha_t(2)}{4k^2 \sum_{t=1}^n \langle M \rangle_t}, \quad I_{33}^M = \sum_{t=1}^n \frac{1}{4k^2} \frac{(\theta - 2\alpha_t(1) + \alpha_t(2))^2}{\langle M \rangle_t}.$$

The optimal combined estimating function using  $m_t$  and  $M_t$  is given by

$$g_C^*(\theta) = \sum_{t=1}^n (a_{t-1}^* m_t + b_{t-1}^* M_t), \quad (2.16)$$

where

$$a_{t-1}^* = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \begin{pmatrix} -\frac{h\alpha_t(1)}{\langle m \rangle_t} - \frac{\eta^2 (\theta + 2(1 - hk)\alpha_t(1) - (1 - 2hk)\alpha_t(2)) \langle m, M \rangle_t}{2k^2 \langle m \rangle_t \langle M \rangle_t} \\ -\frac{1 - e^{-kh}}{\langle m \rangle_t} + \frac{\eta^2 (1 - e^{-kh})^2 \langle m, M \rangle_t}{2k \langle m \rangle_t \langle M \rangle_t} \\ \frac{(\theta - 2\alpha_t(1) + \alpha_t(2)) \langle m, M \rangle_t}{2k \langle m \rangle_t \langle M \rangle_t} \end{pmatrix}$$

and

$$b_{t-1}^* = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \begin{pmatrix} \frac{h\alpha_t(1)\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} + \frac{\eta^2(\theta + 2(1-hk)\alpha_t(1) - (1-2hk)\alpha_t(2))}{2k^2 \langle M \rangle_t} \\ \frac{1 - e^{-kh} \langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} - \frac{\eta^2(1 - e^{-kh})^2}{2k \langle M \rangle_t} \\ - \frac{(\theta - 2\alpha_t(1) + \alpha_t(2))}{2k \langle M \rangle_t} \end{pmatrix}.$$

Further, let  $W_t(\theta)$  be the matrix

$$W_t(\theta) = \begin{pmatrix} \frac{\partial \mu_t(\theta)}{\partial \theta} & \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} & \frac{\partial \sigma_t^2(\theta)}{\partial \theta} & \frac{\partial \mu_t(\theta)}{\partial \theta'} \\ w_{11} & w_{12} & w_{13} \\ w_{12} & w_{22} & w_{23} \\ w_{13} & w_{23} & w_{33} \end{pmatrix}, \tag{2.17}$$

where

$$w_{11} = -\frac{\eta^2 h \alpha_t(1)}{k^2} [\theta + 2(1-hk)\alpha_t(1) - (1-2hk)\alpha_t(2)],$$

$$w_{12} = \frac{\eta^2 h (1 - e^{-kh})^2 \alpha_t(1)}{2k}$$

$$- \frac{\eta^2 (1 - e^{-kh})}{2k^2} [\theta + 2(1-hk)\alpha_t(1) - (1-2hk)\alpha_t(2)],$$

$$w_{22} = \frac{h(1 - e^{-kh})^3}{k} (\theta - y_{(t-1)h}),$$

$$w_{23} = \frac{1 - e^{-kh}}{2k} [\theta - 2\alpha_t(1) + \alpha_t(2)],$$

$$w_{33} = 0.$$

The information associated with the optimal combined estimating function is

$$I_{g_C^*(\theta)} = \sum_{i=1}^n \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t} \right)^{-1} \begin{pmatrix} \frac{\partial \mu_t(\theta)}{\partial \theta} & \frac{\partial \mu_t(\theta)}{\partial \theta'} & \frac{1}{\langle m \rangle_t} \\ + \frac{\partial \sigma_t^2(\theta)}{\partial \theta} & \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} & \frac{1}{\langle M \rangle_t} - W_t(\theta) \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \end{pmatrix}.$$

Note: If we allow  $\eta$  in (2.7) to be a function of  $k$ , then the estimating function  $g_M^*(\theta)$  and the combined estimating function  $g_C^*(\theta)$  become intractable.

### 3. Combined Estimating Functions for General Models

For extended versions of the CIR model, closed form

expressions for the first four conditional moments cannot be obtained easily by using Itô's formula, as was done for the CIR model. Recently, the Milstein's approximation was used in [9] to obtain the first two conditional moments of the diffusion. In this section, we use Milstein's approximation to obtain the first four conditional moments and construct the optimal estimating functions.

Consider the diffusion process given by the time-homogeneous stochastic differential equation of the form

$$dy_t = a(\alpha, y_t)dt + b(\beta, y_t)dW_t. \tag{3.1}$$

Where  $a$  and  $b$  are the drift and diffusion functions, respectively, and  $W_t$  is the standard Brownian motion.

A special case of (3.1) is the Brownian motion with constant drift and diffusion:

$$dy_t = \alpha dt + \beta dW_t,$$

where  $\beta > 0$ . In this case, the conditional distribution of  $y_t$  given  $y_0 = y$  is a normal with mean  $y + \alpha t$  and variance  $\beta^2 t$ . If we consider the geometric Brownian motion given by

$$dy_t = \nu y_t dt + \omega y_t dW_t,$$

with  $\omega > 0$ , then  $\log(y_t)$  becomes a Brownian motion with drift with  $\alpha = \nu - \omega^2/2$  and  $\beta = \omega$ . In this case, the conditional distribution of  $\log(y_t)$  given  $\log(y_0) = \log(y)$  is also normal. The CIR process can be re-parameterized to the following form:

$$dy_t = (\alpha_1 + \alpha_2 y_t)dt + \beta \sqrt{y_t} dW_t.$$

In this case, we have computed the first four conditional moments of the process to use in an estimating function framework. Extended versions of the CIR process model have been proposed for modeling interest rate processes. For example, some consider the constant elasticity of variance process of the form

$$dy_t = (\alpha_1 + \alpha_2 y_t)dt + \beta y_t^{\beta_2} dW_t$$

or the nonlinear drift diffusion process [13] given by

$$dy_t = (\alpha_1 + \alpha_2 y_t + \alpha_3 y_t^2 + \alpha_4 y_t^{-1})dt + \sqrt{\beta_1 + \beta_2 y_t + \beta_3 y_t^{\beta_3}} dW_t.$$

For more general extended models, the diffusion is a function of the observation  $y_t$  and hence, closed form expressions of the conditional distributions, as well as closed form expressions for the conditional moments cannot be easily obtained by solving differential equations obtained by repeated application Itô's formula. However, Milstein's approximation can be used to obtain the first four conditional moments.

Milstein's approximation applied to (3.1) produces

$$y_{th} = y_{(t-1)h} + a(\alpha, y_{(t-1)h})h + b(\beta, y_{(t-1)h})\sqrt{h}\epsilon_t + \frac{1}{2}b(\beta, y_{(t-1)h})b_y(\beta, y_{(t-1)h})(\epsilon_t^2 - 1)h, \tag{3.2}$$

where  $b_y = \frac{\partial b}{\partial y}$  and  $\epsilon_t \sim N(0,1)$ , i.i.d. Unlike the Euler approximation for diffusion processes, the Milstein approximation does not yield a conditional normal distribution for  $y_{th} - y_{(t-1)h}$ . The distribution implied by the Milstein approximation is a mixture of a normal and chi-square distributions. By using (3.2), the first four conditional moments of  $y_{th}$  given  $y_{(t-1)h}$  are approximated by

$$\mu_t(\alpha) = y_{(t-1)h} + a(\alpha, y_{(t-1)h})h, \tag{3.3}$$

$$\sigma_t^2(\beta) = b^2(\beta, y_{(t-1)h})h + \frac{1}{2}b^2(\beta, y_{(t-1)h})b_y^2(\beta, y_{(t-1)h})h^2, \tag{3.4}$$

$$\gamma_t(\beta) = 3b^3(\beta, y_{(t-1)h})b_y(\beta, y_{(t-1)h})h^2 + b^3(\beta, y_{(t-1)h})b_y^3(\beta, y_{(t-1)h})h^3, \tag{3.5}$$

$$\kappa_t(\beta) = 15b^4(\beta, y_{(t-1)h})b_y^2(\beta, y_{(t-1)h})h^3 + \frac{15}{4}b^4(\beta, y_{(t-1)h})b_y^4(\beta, y_{(t-1)h})h^4 + 3b^4(\beta, y_{(t-1)h})h^2. \tag{3.6}$$

Then based on the discretely observed observations  $y_0, y_h, y_{2h}, \dots, y_{nh}$ , the martingale differences are  $m_t = y_{th} - \mu_t(\alpha)$ , and  $M_t = m_t^2 - \sigma_t^2(\beta)$ . In this case, we have

$$\langle m \rangle_t = \sigma_t^2(\alpha), \tag{3.8}$$

$$\begin{aligned} \langle M \rangle_t &= \kappa_t(\beta) - \sigma_t^4(\beta) \\ &= 14b^4(\beta, y_{(t-1)h})b_y^2(\beta, y)h^3 \\ &\quad + \frac{7}{2}b^4(\beta, y_{(t-1)h})b_y^4(\beta, y_{(t-1)h})h^4 \\ &\quad + 2b^4(\beta, y_{(t-1)h})h^2 \end{aligned} \tag{3.9}$$

$$\langle m, M \rangle_t = \gamma_t(\beta). \tag{3.10}$$

In addition, if we let  $\rho_t^2 = \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t}$ , then

$$\begin{aligned} \rho_t^2 &= \frac{4hb_y^2(\beta, y_{(t-1)h})}{\left[2 + b_y^2(\beta, y_{(t-1)h})h\right]} \\ &\quad \times \frac{\left[9 + 6b_y^2(\beta, y_{(t-1)h})h + b_y^4(\beta, y_{(t-1)h})h^2\right]}{\left[4 + 28b_y^2(\beta, y_{(t-1)h})h + 7b_y^4(\beta, y_{(t-1)h})h^2\right]}. \end{aligned} \tag{3.11}$$

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

$$g_m^*(\alpha) = -\sum_{t=1}^n \frac{\partial \mu_t}{\partial \alpha} \frac{m_t}{\langle m \rangle_t}, \tag{3.12}$$

$$I_{g_m^*}(\alpha) = \sum_{t=1}^n \frac{\partial \mu_t}{\partial \alpha} \frac{\partial \mu_t}{\partial \alpha'} \frac{1}{\langle m \rangle_t},$$

$$g_M^*(\beta) = -\sum_{t=1}^n \frac{\partial \sigma_t^2}{\partial \beta} \frac{M_t}{\langle M \rangle_t}, \tag{3.13}$$

$$I_{g_M^*}(\beta) = \sum_{t=1}^n \frac{\partial \sigma_t^2}{\partial \beta} \frac{\partial \sigma_t^2}{\partial \beta'} \frac{1}{\langle M \rangle_t},$$

The combined estimating function and the corresponding information follow from Theorem 1 by taking  $\theta = (\alpha', \beta')$ .

**Example 2** (NLD Process). *The nonlinear drift (NLD) diffusion process for modeling interest rates was introduced in [13]. Here we consider the following NLD*

$$dy_t = (\alpha_1 + \alpha_2 y_t^{-1})dt + \sqrt{\beta_1 + \beta_2 y_t}dW_t \tag{3.14}$$

where  $\beta_1 > 0$ ,  $\beta_2 > 0$ ,  $0 < \alpha_1 < \beta_2/2$ , and  $\alpha_2 > \beta_1/2$ . These parameter ranges are chosen to guarantee a positive recurrent solution to the SDE. For this process,  $a(\alpha, y) = (\alpha_1 + \alpha_2/y)$ ,  $B(\alpha, y) = \sqrt{\beta_1 + \beta_2 y}$ , and  $b_y(\alpha, y) = 1/2(\beta_1 + \beta_2 y)^{-1/2}$ . The Milstein approximation gives the following discretized version of the process:

$$y_{th} = y_{(t-1)h} + (\alpha_1 + \alpha_2 y_{(t-1)h}^{-1})h + \sqrt{\beta_1 + \beta_2 y_{(t-1)h}}\sqrt{h}\epsilon_t + \frac{1}{4}\beta_2(\epsilon_t^2 - 1)h. \tag{3.15}$$

In this case, we have the following for the first four conditional moments of the discretized process:

$$\mu_t(\alpha) = y_{(t-1)h} + (\alpha_1 + \alpha_2 y_{(t-1)h}^{-1})h, \tag{3.16}$$

$$\langle m \rangle_t = (\beta_1 + \beta_2 y_{(t-1)h})h + \frac{1}{8}\beta_2^2 h^2, \tag{3.17}$$

$$\begin{aligned} \langle M \rangle_t &= \frac{7}{2} \beta_2^2 h^3 \beta_1 + \frac{7}{2} \beta_2^3 h^3 y_{(t-1)h} \\ &+ \frac{7}{32} \beta_2^4 h^4 + 2h^2 \beta_1^2 + 4h^2 \beta_1 \beta_2 y_{(t-1)h} \\ &+ 2h^2 \beta_2^2 y_{(t-1)h}^2, \end{aligned} \tag{3.18}$$

$$\begin{aligned} \langle m, M \rangle_t &= \gamma_t(\beta) \\ &= \frac{3}{2} (\beta_1 + \beta_2 y_{(t-1)h}) \beta_2 h^2 + \frac{1}{8} \beta_2^3 h^3. \end{aligned} \tag{3.19}$$

The estimating function and corresponding information based on  $m_t$  are given by

$$g_m^*(\alpha) = \begin{pmatrix} -h \sum_{t=1}^n \frac{m_t}{\sigma_t^2(\beta)} \\ -h \sum_{t=1}^n \frac{m_t}{y_{(t-1)h} \sigma_t^2(\beta)} \end{pmatrix}, \tag{3.20}$$

$$I_{g_m^*}(\alpha) = \begin{pmatrix} I_{11}^m & I_{12}^m \\ I_{12}^m & I_{22}^m \end{pmatrix},$$

where

$$\begin{aligned} I_{11}^m &= h^2 \sum_{t=1}^n \frac{1}{\sigma_t^2(\beta)}, \\ I_{12}^m &= h^2 \sum_{t=1}^n \frac{1}{y_{(t-1)h} \sigma_t^2(\beta)}, \\ I_{22}^m &= h^2 \sum_{t=1}^n \frac{1}{y_{(t-1)h}^2 \sigma_t^2(\beta)}. \end{aligned}$$

Moreover, the estimating function and corresponding information based on  $M_t$  are given by

$$g_M^*(\beta) = \begin{pmatrix} -h \sum_{t=1}^n \frac{M_t}{\langle M \rangle_t} \\ -h \sum_{t=1}^n \frac{(y_{(t-1)h} + 1/4 \beta_2 h) M_t}{\langle M \rangle_t} \end{pmatrix}, \tag{3.21}$$

$$I_{g_M^*}(\beta) = \begin{pmatrix} I_{11}^M & I_{12}^M \\ I_{12}^M & I_{22}^M \end{pmatrix}$$

where

$$\begin{aligned} I_{11}^M &= h^2 \sum_{t=1}^n \frac{1}{\langle M \rangle_t}, \\ I_{12}^M &= h^2 \sum_{t=1}^n \frac{(y_{(t-1)h} + 1/4 \beta_2 h)}{\langle M \rangle_t}, \\ I_{22}^M &= h^2 \sum_{t=1}^n \frac{(y_{(t-1)h} + 1/4 \beta_2 h)^2}{\langle M \rangle_t} \end{aligned}$$

In this case,  $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2)'$  and the optimal

combined estimating function using  $m_t$  and  $M_t$  is given by  $g_c^*(\theta) = \sum_{t=1}^n (a_{t-1}^* m_t + b_{t-1}^* M_t)$ , where

$$\begin{aligned} a_{t-1}^* &= (1 - \rho_t^2)^{-1} \left( -\frac{\partial \mu_t}{\partial \theta} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \right), \\ b_{t-1}^* &= (1 - \rho_t^2)^{-1} \left( \frac{\partial \mu_t}{\partial \theta} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} - \frac{\partial \sigma_t^2}{\partial \theta} \frac{1}{\langle m \rangle_t} \right), \end{aligned}$$

with

$$\frac{\partial \mu_t}{\partial \theta} = (h, h y_{(t-1)h}^{-1}, 0, 0)'$$

and

$$\frac{\partial \sigma_t^2}{\partial \theta} = (0, 0, h, h y_{(t-1)h} + 1/4 \beta_2 h^2)'$$

The information for the combined estimating function  $I_{g_c^*}(\theta)$  is given by

$$\begin{aligned} I_{g_c^*}(\theta) &= \sum_{t=1}^n (1 - \rho_t^2)^{-1} \left[ \frac{\partial \mu_t}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{1}{\langle M \rangle_t} \right. \\ &\quad \left. + \left( \frac{\partial \mu_t}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \right) \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} \right]. \end{aligned}$$

### 4. Conclusion

For discretely observed general interest rate models, the combined estimating function method allows estimators to be obtained straightforwardly under very general conditions on the first four conditional moments. In this paper, we have studied inference for interest rate models, first by using the Milstein approximation, and then combining estimating functions using martingale differences and have obtained the closed form expression for the information gain.

### REFERENCES

- [1] A. Thavaneswaran and M. E. Thompson, "Optimal Estimation for Semimartingales," *Journal of Applied Probability*, Vol. 23, No. 2, 1986, pp. 409-417. [doi:10.2307/3214183](https://doi.org/10.2307/3214183)
- [2] A. Thavaneswaran and M. E. Thompson, "A Criterion for Filtering in Semimartingale Models," *Stochastic Processes and Their Applications*, Vol. 28, No. 2, 1988, pp. 259-265. [doi:10.1016/0304-4149\(88\)90099-3](https://doi.org/10.1016/0304-4149(88)90099-3)
- [3] C. Heyde, "Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation," Springer Series in Statistics, Springer, 1997. [doi:10.1007/b98823](https://doi.org/10.1007/b98823)
- [4] V. P. Godambe, "The Foundations of Finite Sample Estimation in Stochastic Processes," *Biometrika*, Vol. 72, No. 2, 1985, pp. 419-428. [doi:10.1093/biomet/72.2.419](https://doi.org/10.1093/biomet/72.2.419)
- [5] J. E. Hutton and P. I. Nelson, "Quasilikelihood Estimation

- for Semimartingales,” *Stochastic Processes and Their Applications*, Vol. 22, No. 2, 1986, pp. 245-257. [doi:10.1016/0304-4149\(86\)90004-9](https://doi.org/10.1016/0304-4149(86)90004-9)
- [6] U. V. Naik-Nimbalkar and M. B. Rajarshi, “Filtering and Smoothing via Estimating Functions,” *Journal of the American Statistical Association*, Vol. 90, No. 429, 1995, pp. 301-306. [doi:10.1080/01621459.1995.10476513](https://doi.org/10.1080/01621459.1995.10476513)
- [7] A. Paseka, T. Koulis and A. Thavaneswaran, “Interest Rate Models,” *Journal of Mathematical Finance*, Vol. 2, No. 2, 2012, pp. 141-158. [doi:10.4236/jmf.2012.22016](https://doi.org/10.4236/jmf.2012.22016)
- [8] A. Thavaneswaran, Y. Liang and N. Ravishanker, “Inference for Diffusion Processes Using Combined Estimating Functions,” *Sri Lankan Journal of Applied Statistics*, Vol. 12, 2011, pp. 145-160.
- [9] M. Jeong and J. Y. Park, “Asymptotic Theory of Maximum Likelihood Estimator for Diffusion Model,” Working Paper, Indiana University, Bloomington, 2010.
- [10] O. Elerian, “A Note on the Existence of a Closed form Conditional Transition Density for the Milstein Scheme,” Economics Discussion Paper 1998-W18, Nuffield College, Oxford, 1998.
- [11] B. G. Lindsay, “Using Empirical Partially Bayes Inference for Increased Efficiency,” *The Annals of Statistics*, Vol. 13, No. 3, 1985, pp. 914-931. [doi:10.1214/aos/1176349646](https://doi.org/10.1214/aos/1176349646)
- [12] J. C. Cox, J. E. Ingersoll Jr. and S. A. Ross, “A Theory of the Term Structure of Interest Rates,” *Econometrica*, Vol. 53, No. 2, 1985, pp. 385-407. [doi:10.2307/1911242](https://doi.org/10.2307/1911242)
- [13] Y. Ait-Sahalia, “Testing Continuous-Time Models of The Spot Interest Rate,” *Review of Financial Studies*, Vol. 9, No. 2, 1996, pp. 385-426. [doi:10.1093/rfs/9.2.385](https://doi.org/10.1093/rfs/9.2.385)



**Moments of the CIR Process**

The first four conditional moments of the CIR process

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

may be obtained by using Itô's formula on successive powers of the process. This gives

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

$$dy_t^2 = 2y_t dy_t + \frac{1}{2}2\eta^2 y_t dt,$$

$$dy_t^3 = 3y_t^2 dy_t + 3\eta^2 y_t^2 dt,$$

$$dy_t^4 = 4y_t^3 dy_t + 6\eta^2 y_t^3 dt,$$

with  $y(0) = x$  as an initial condition. We let

$$m_1(t) = \theta - e^{-kt}(\theta - x),$$

$$m_2(t) = \frac{(2k\theta + \eta^2)\theta}{2k} - \frac{(2k\theta + \eta^2)(\theta - x)}{k}e^{-kt} + \frac{(2k\theta + \eta^2)(\theta - 2x) + 2kx^2}{2k}e^{-2kt},$$

$$m_3(t) = \frac{(2k\theta + \eta^2)(3k\theta + 3\eta^2)\theta}{6k^2} - \frac{(2k\theta + \eta^2)(3k\theta + 3\eta^2)(\theta - x)}{2k^2}e^{-kt} + \frac{(3k\theta + 3\eta^2)((2k\theta + \eta^2)((\theta - 2x) + 2kx^2))}{2k^2}e^{-2kt} - \frac{(3k\theta + 3\eta^2)((2k\theta + \eta^2)(\theta - 3x) + 6kx^2)}{6k^2}e^{-3kt} - x^3e^{-3kt},$$

$$m_4(t) = \frac{(2k\theta + \eta^2)(3k\theta + 3\eta^2)(4k\theta + 6\eta^2)\theta}{24k^3} - \frac{(2k\theta + \eta^2)(3k\theta + 3\eta^2)(4k\theta + 6\eta^2)(\theta - x)}{6k^3}e^{-kt} + \frac{(3k\theta + 3\eta^2)(4k\theta + 6\eta^2)((2k\theta + \eta^2)(\theta - 2x) + 2kx^2)}{4k^3}e^{-2kt} - \frac{(3k\theta + 3\eta^2)(4k\theta + 6\eta^2)((2k\theta + \eta^2)(\theta - 3x) + 6kx^2 - 6k^2x^3)}{6k^3}e^{-3kt} + \frac{(4k\theta + 6\eta^2)((3k\theta + 3\eta^2)((2k\theta + \eta^2)(\theta - 4x) + 12kx^2) - 24k^2x^3)}{24k^3}e^{-4kt} + x^4e^{-4kt}.$$

Hence, the first four conditional centered moments are given as

$$\begin{aligned} E[(y_t - \mu)^2] &= \sigma_t^2(\theta) \\ &= m_2(t) - m_1^2(t) \\ &= \frac{\eta^2}{2k}(\theta - 2(\theta - x)e^{-kt} + (\theta - 2x)e^{-2kt}), \\ E[(y_t - \mu)^3] &= \gamma_t^3(\theta) \\ &= m_3(t) - 3m_1(t)m_2(t) + 2m_1^3(t) \\ &= \frac{\eta^4}{2k^2}(\theta - 3(\theta - x)e^{-kt} + 3(\theta - 2x)e^{-2kt} \\ &\quad - (\theta - 3x)e^{-3kt}), \end{aligned}$$

$m_j(t) = E(y_t^j | y_0 = x)$ , so that the first four conditional moments of  $y_t$  satisfy the following differential equations:

$$\frac{d}{dt}m_1(t) = -k(m_1(t) - \theta),$$

$$\frac{d}{dt}m_2(t) = -2km_2(t) + (2k\theta + \eta^2)m_1(t),$$

$$\frac{d}{dt}m_3(t) = -3km_3(t) + (3k\theta + 3\eta^2)m_2(t),$$

$$\frac{d}{dt}m_4(t) = -4km_4(t) + (4k\theta + 6\eta^2)m_3(t),$$

with initial conditions  $m_1(0) = x, m_2(0) = x^2, m_3(0) = x^3$ , and  $m_4(0) = x^4$ . Solving the differential equations in turn yields:

$$\begin{aligned} E[(y_t - \mu)^4] &= \kappa_t^4(\theta) \\ &= m_4(t) - 4m_1(t)m_3(t) + 6m_1^2(t)m_2(t) - 3m_1^4(t) \\ &= \frac{3\eta^6}{4k^3}(\theta - 4(\theta - x)e^{-kt} + 6(\theta - 2x)e^{-2kt} \\ &\quad - 4(\theta - 3x)e^{-3kt} + (\theta - 4x)e^{-4kt}) + 3\sigma_t^4(\theta). \end{aligned}$$