EÖTVÖS LORÁND UNIVERSITY

FACULTY OF SCIENCE

Bat-erdene Egshiglen

Inference for Interest Rate Models Using Milstein's Approximation

MSc Applied Mathematics Thesis

Supervisor:

Dr. György Michaletzky

Department of Probability Theory and Statistics

Budapest, 2022

DECLARATION OF AUTHORSHIP

Name: Bat-erdene Egshiglen
Program: MSc in Applied Mathematics, Faculty of Science, ELTE
NEPTUN ID: GFUL45
Title of the thesis:
Inference for Interest Rate Models Using Milstein's Approximation

I, as the sole author of this work hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. All references and quotations are done in a form that is standard for the subject and any use of the works of any other author, in any form, is properly acknowledged at their point of use.

Budapest, 2022.04.20

Signature

Acknowledgements

This thesis would not have been possible without the guidance, time and patience of my supervisor, Professor György Michaletzky. I would like to express my gratitude to him for always being patient and explaining everything I did not understand.
I also would like to thank my family and friends who supported me.

Contents

Chapter 1

Introduction

1.1 Summary

Interest rate is considered as a feature of the debt instrument and it is one of the most important aspect of the economics. From a long time ago, the connection between the price of an asset and interest rate has been studied and, as a result, stochastic variation was introduced. After the volatility of the interest rate was discovered, it has been treated as a stochastic variable for the importance of better risk management and forecasting.

Usually, estimation of parameters in the diffusion process can be done by Maximum Likelihood estimation or the approximation Euler–Maruyama scheme, which is quite straightforward as it deals with the trajectory instead of the transition directly. However, Euler-type schemes depend on the sampling interval and introduce discretization bias in the estimates, also, the Maximum Likelihood estimators can be complicated or not even available for certain type of models. Overall, for extended general models, where the diffusion is a function of observation, we cannot compute the conditional moments by the repetition of Itô's formula. This is where the Milstein's approximation is going to help us obtain the first four conditional moments of the underlying process, and after this, we can use the combined estimating function approach.

The goal of the thesis is to provide information about the inference of interest rate models by using Milstein's approximation and then use combined estimating functions under general conditions for better results.

This paper is organized as follows. The rest of Chapter 1 introduces some notations and definitions. In Chapter 2, we will present the basic information regarding optimal estimation functions and as well as the recursive estimation regarding state space and general models. After that, we show combined estimating functions with discretely observed diffusion processes on general models in Chapter 3, and lastly, we show the application of this method by some examples. In the appendixes, we summarize some basic information connected to semimartingales (appendix A) and the main idea of Milstein's approximation (appendix B).

1.2 Preliminaries

First, we will introduce some notations and definitions that will be necessary for the rest of the thesis. This section contains definitions and statements from [5], [17] and [18].

We define Ω a sample space with a measure μ and a probability density function $p(y, \theta)$ with respect to μ where $y \in \Omega$ and $\theta \in \Theta \subset R^p$.

Definition 1.2.1 (Estimating function)**.** A function $g : \Omega \times \Theta \rightarrow R^p$ is called an estimating function if $g(., \theta)$ is measurable for any $\theta \in \Theta$ and $g(y, .)$ is continuous in a compact subspace of Θ containing the true parameter θ_0 for any sample $y \in \Omega$.

Given an estimating function g and an observation y, the estimating equation can be defined as

$$g(y, \theta) = 0, \tag{1.1}$$

and the solution $\hat{\theta}$ is an estimate of the parameter θ.

Additionally, if two estimating functions lead to the same solution for any given sample y, we call them equivalent.

Definition 1.2.2 (Unbiased estimating functions)**.** We call an estimating function g unbiased if

$$E_\theta(g(Y, \theta)) = 0, \forall \theta \in \Theta. \tag{1.2}$$

Definition 1.2.3 (Regular estimating function)**.** g is regular if the following conditions are satisfied

- $E_\theta(g(Y, \theta)) = 0$, for all θ;
- $\partial g(y, \theta)/\partial \theta$ exists for all θ;
- the order of integration and differentiation can be interchanged for any bounded measurable function $f(y)$ that is independent of θ;
- $0 < E_\theta(g^2(Y, \theta)) < \infty$;
- $0 < (E_\theta(\partial g(Y, \theta)/\partial \theta))^2 < \infty$.

In the following chapters, we will use the information matrix that provides a way to measure the amount of information that a random variable contains about some parameter θ (such as the true mean) of the random variable's assumed probability distribution.

Definition 1.2.4 (Fisher information)**.** The Fisher information or information matrix is denoted by I and is defined as

$$I = Var(\frac{\partial}{\partial\theta} l(\theta \mid y)), \tag{1.3}$$

where $l(\theta \mid y))$ is the log-likelihood function of θ given observed value of y.

Definition 1.2.5 (Godambe information)**.** Given a regular estimating function g and a single observation Y, the Godambe information is denoted by J_g, and is described as

$$J_g(\theta) = \frac{S_g^2(\theta)}{V_g(\theta)}, \tag{1.4}$$

where S_g is the sensitivity

$$S_g = E_\theta(\frac{\partial g(Y,\theta)}{\partial\theta}) \tag{1.5}$$

and V_g is the variability

$$V_g = Var(g(Y,\theta)). \tag{1.6}$$

Now, let us introduce a theorem that shows the connection between the Fisher and Godambe information.

Theorem 1.2.1 (Godambe Inequality)**.** Given an estimating function g,

$$J_g(\theta) \leq I(\theta), \tag{1.7}$$

where the equality holds if g is equivalent to the score function.

Chapter 2

Optimal Estimating functions

Since we now know what is an estimating function, the next question is how to identify an optimal one. And the answer is related to the Godambe information of the function. The following definition is from [18].

Definition 2.0.1 (Optimal estimating function). A regular estimating function g^* is optimal if

$$J_{g^*}(\theta) \geq J_g(\theta), \tag{2.1}$$

for all g and $\theta \in \Theta$.

2.1 Basics of Estimating functions

In this section, we follow [9], which presents the basic information regarding optimal estimating functions and information matrices.

We should introduce our probability space $(\Omega, \mathcal{A}, \mathcal{P}_\theta)$ and our observed time series $\{y_t, t = 1, \ldots, n\}$, which is a realization of a discrete-time stochastic process. Its distribution depends on a vector parameter θ, which belongs to an open subset Θ of the $p-$dimensional Euclidean space. Additionally, we assume $\mathcal{F}_t^y$ to be the σ-field generated by $\{y_1, \ldots, y_t, t \geq 1\}$ and let $h_t = h_t(y_1, \ldots, y_t, \theta)$, $1 \leq t \leq n$ be specified q-dimensional vectors that are martingales.

Now, we consider the class of zero mean and square integrable p-dimensional martingale estimating functions:

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

where a_{t-1} are $p \times q$ matrices depending on $y_1, \ldots, y_{t-1}$, $1 \leq t \leq n$.

Moreover, we also assume the followings about our estimating function $g_n(\theta)$ for every $n \geq 1$:

- $g_n(\theta)$ are almost surely differentiable with respect to θ,
- $E\left(\frac{\partial g_n(\theta)}{\partial \theta} \mid \mathcal{F}_{n-1}^y\right)$ is non-singular for all θ,
- $E\left(g_n(\theta) g_n(\theta)' \mid \mathcal{F}_{n-1}^y\right)$ is non-singular and positive definite for all θ.

In the class $\mathcal{M}$, our optimal estimating function

$$\begin{aligned} g_n^*(\theta) &= \textstyle\sum_{t=1}^n a_{t-1}^* h_t \\ &= \textstyle\sum_{t=1}^n \left(E\left(\frac{\partial h_t(\theta)}{\partial \theta} \mid \mathcal{F}_{t-1}^y\right)\right)' \left(E\left(h_t h_t' \mid \mathcal{F}_{t-1}^y\right)\right)^{-1} h_t \end{aligned} \tag{2.2}$$

will maximize the information matrix

$$\begin{aligned} I_g(\theta) = &\left(\sum_{t=1}^{n} a_{t-1} E\left(\frac{\partial h_t(\theta)}{\partial \theta} \mid \mathcal{F}_{t-1}^y \right) \right)' \\ &\left(\sum_{t=1}^{n} E\left((a_{t-1} h_t)(a_{t-1} h_t)' \mid \mathcal{F}_{t-1}^y \right) \right)^{-1} \left(\sum_{t=1}^{n} a_{t-1} E\left(\frac{\partial h_t(\theta)}{\partial \theta} \mid \mathcal{F}_{t-1}^y \right) \right) \end{aligned} \tag{2.3}$$

and reduces the corresponding optimal information to

$$E\left(g_n^*(\theta)g_n^*(\theta)' \mid \mathcal{F}_{n-1}^y\right). \tag{2.4}$$

These will be used in Section 3.2, 3.3 and as well in the Section 3.4, where we apply our knowledge in an example.

2.2 Recursive estimation

In this section, we will focus on the recursive estimation for continuous time models using Milstein's approximation, following paper [10]. Recursive estimation uses the parameter at time t to estimate the parameter in $t+1$ and some adjustment based on the observation at time $t+1$. It helps us to break the problem to smaller problems and obtain a mathematical model of the system in real time.

2.2.1 State space models

Before jumping into the optimal recursive estimate functions for general models, we will see a discrete-time state space model of an observed process $\{y_t\}$ and a state process $\{\theta_t\}$.

$$y_{t+1} = A\theta_t + az_{t+1} + b(z_{t+1}^2 - 1) \tag{2.5}$$

$$\theta_{t+1} = B\theta_t + c\eta_{t+1} + d(\eta_{t+1}^2 - 1), \tag{2.6}$$

where A, B, a, b, c and d are positive constants, moreover, possibly measurable with respect to the σ-field $\mathcal{F}_t^y$. Additionally, $\{z_{t_i}\}$, $\{\eta_{t_i}\}$ are two independent standard Gaussian sequences of i.i.d random variables with

$\text{Corr}(z_t, \eta_t) = \rho$. Now, we introduce a lemma that will be useful to prove next theorem.

Lemma 2.2.1. Assume $Z_1 \sim N(0,1)$ and $Z_2 \sim N(0,1)$ with $\text{Corr}(Z_1, Z_2)=\rho$. Then $\text{Corr}(Z_1^2, Z_2^2)=\rho^2$.

Theorem 2.2.2. Given the previous state space model and the class of all estimators of the form :

$$\hat{\theta}_{t+1} = B\hat{\theta}_t + \hat{G}_t(y_{t+1} - A\hat{\theta}_t), \tag{2.7}$$

we will have $\hat{G}_t$

$$\hat{G}_t = \frac{AB\gamma_t + \rho(ac+2)}{A^2\gamma_t + a^2 + 2b^2} \tag{2.8}$$

minimizing the mean-square error

$$\gamma_{t+1} = E\Big((\theta_{t+1} - \hat{\theta}_{t+1})^2 \mid \mathcal{F}_t^y\Big). \tag{2.9}$$

Moreover, the mean-square error is given as

$$\gamma_{t+1} = (B - A\hat{G}_t)^2\gamma_t + c^2 + 2d^2 + \hat{G}_t^2(a^2 + 2b^2) - 2\rho\hat{G}_t(ac + 2\rho bd). \tag{2.10}$$

Proof. Firstly, we calculate the difference $\theta_{t+1} - \hat{\theta}_{t+1}$

$$\begin{aligned}
\theta_{t+1} - \hat{\theta}_{t+1} &= B\theta_t + c\eta_{t+1} + d(\eta_{t+1}^2 - 1) - (B\hat{\theta}_t + \hat{G}_t(y_{t+1} - A\hat{\theta}_t)) \\
&= B(\theta_t - \hat{\theta}_t) + c\eta_{t+1} + d(\eta_{t+1}^2 - 1) \\
&- G_t(A\theta_t + az_{t+1} + b(z_{t+1}^2 - 1) - A\hat{\theta}_t) \\
&= (B - AG_t)(\theta_t - \hat{\theta}_t) + c\eta_{t+1} + d(\eta_{t+1}^2 - 1) - aG_t z_{t+1} - bG_t(z_{t+1}^2 - 1).
\end{aligned} \tag{2.11}$$

If we take the square of the above expression and calculate the expected value, we can use the previous lemma 2.2.1 to see that the conditional mean-square error at $t+1$ is given by

$$\gamma_{t+1} = (B - AG_t)^2\gamma_t + c^2 + 2d^2 + G_t^2(a^2 + 2b^2) - 2\rho G_t(ac + 2\rho bd). \tag{2.12}$$

Now, we differentiate it with respect to G_t and assume the first derivative is zero, we get

$$-2A(B - AG_t)\gamma + 2G_t(a^2 + 2b^2) - 2\rho(ac + 2\rho bd) = 0$$

Solving for G_t, we obtain

$$\hat{G}_t = \frac{2AB\gamma_t + \rho(ac + 2\rho bd)}{2A^2\gamma_t + a^2 + 2b^2}.$$

□

Corollary 2.2.2.1. Let the state space model be of the form

$$y_{t+1} = A\theta_t + z_{t+1}, \quad \theta_{t+1} = B\theta_t + \eta_{t+1},$$

where $\{z_t\}$ and $\{\eta_t\}$ are two sequences of i.i.d random variables with mean zero and variance σ_z^2 and σ_η^2 respectively. In the class of estimates of the form:

$$\hat{\theta}_{t+1} = B\hat{\theta}_{t+1} + \hat{G}_t(y_{t+1} - A\hat{\theta}_t), \tag{2.13}$$

the G_t, which minimizes the mean-square error

$$\gamma_t = E\bigg((\theta_t - \hat{\theta}_t) \mid \mathcal{F}_t^y\bigg),$$

is given by

$$\hat{G}_t = \frac{BA\gamma_t}{A^2\gamma_t + \sigma_z^2}.$$

In addition, the mean-square error is given as

$$\gamma_{t+1} = (B - \hat{G}_t A)^2\gamma_t + \sigma_\eta^2 + \hat{G}_t^2\sigma_z^2.$$

Proof. It can be proven by setting $a = \sigma_z$, $b = \sigma_e ta$, $c = d = 0$, and $\rho = 0$ and using the previous theorem. □

2.2.2 General models

In the continuous setting, we consider the following general state space model

$$\partial y_t = A(y_t)\theta_t dt + \alpha(y_t) dW_1(t), \tag{2.14}$$

$$\partial \theta_t = B(y_t)\theta_t dt + \beta(y_t, \theta_t) dW_2(t) \tag{2.15}$$

where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions. After applying the Milstein's approximation considering discretisation in small intervals of time, then we obtain the following non-Gaussian discrete state-space model

$$y_{t_{i+1}} - y_{t_i} = A(y_{t_i})\theta_{t_i} h + \alpha(y_{t_i})\sqrt{h} z_{t_{i+1}} + \frac{h}{2}\alpha(y_{t_i})\dot{\alpha}_y(y_{t_i})(z^2_{t_{i+1}} - 1), \tag{2.16}$$

$$\theta_{t_{i+1}} = (1 + B(y_{t_i})h)\theta_{t_i} + \beta(y_{t_i})\sqrt{h}\eta_{t_{i+1}} + \frac{h}{2}\beta(y_{t_i}, \theta_{t_i})\dot{\beta}_\theta(y_{t_i}, \eta_{t_i})(\eta^2_{t_{i+1}} - 1), \tag{2.17}$$

where $\dot{\alpha}_y = \frac{\partial \alpha}{\partial y}$, $\dot{\beta}_y = \frac{\partial \beta}{\partial \theta}$.

We can relate this to our discrete model (2.5)-(2.6), and use the previous knowledge to calculate the recursive estimator and mean-square error.

Now since we covered both general models and state space models, let us see how we apply this theory to an interest rate model.

Example 2.2.1 (CIR Model)**.** Consider the Cox–Ingersoll–Ross model for the observed process y_t, given that

$$dy_t = k(\theta_t - y_t)dt + \sigma\sqrt{y_t} dW_1(t)$$

and the state process θ_t follows a diffusion process of the form

$$d\theta = B(y_t)\theta_t dt + \beta(y_t, \theta_t) dW_2(t),$$

with $E(dW_1(t), dW_2(t)) = 0$.

Applying the Milstein's approximation will lead to

$$y_{t_{i+1}} - y_{t_i} + ky_{t_i} = k\theta_{t_i}h + \sigma\sqrt{y_{t_i}h}z_{t_{i+1}} + \frac{1}{4}\sigma^2 h(z^2_{t_{i+1}} - 1),$$

and

$$\theta_{t_{i+1}} = \Big(1 + B(y_{t_i})h\Big)\theta_{t_i} + \beta(y_{t_i})\sqrt{h}\eta_{t_{i+1}} + \frac{h}{2}\beta(y_{t_i}, \theta_{t_i})\dot{\beta}_\sigma(y_{t_i}, \theta_{t_i})(\eta^2_{t_{i+1}} - 1).$$

We would like to relate these to (2.5) and (2.6) by letting $y_{t+1} = y_{t_{i+1}} - y_{t_i} + ky_{t_i}$, $\theta_{t+1} = \theta_{t_{i+1}}$, $z_{t+1} = z_{t_{i+1}}$, $\eta_{t+1} = \eta_{t_{i+1}}$, and $\rho = 0$. Additionally, we also say $A = kh$, $a = \sigma\sqrt{y_{t_i}h}$, $b = \frac{1}{4}\sigma^2 h$, $B = 1 + B(y_{t_i})h$, $c = \beta(y_{t_i})\sqrt{h}$, and $d = \frac{h}{2}\beta(y_{t_i}, \theta_{t_i})\beta(\widehat{y_{t_i}, \theta_{t_i}})$. Now, from Theorem 2.2.2 we can get the recursive estimator

$$\hat{\theta}_{t+1} = \Big(1 + B(y_{t_i})h\Big)\hat{\theta}_t + \hat{G}_t\Big(y_{t+1} - kh\hat{\theta}_t\Big),$$

where

$$\hat{G}_t = \frac{k\Big(1 + B(y_t)h\Big)\gamma_t}{k^2 h\gamma_t + \sigma^2\Big(y_t + \frac{1}{8}\sigma^2\Big)},$$

and the mean-square error is given as

$$\gamma_{t+1} = \Big(1+B(y_{t_i})h-kh\hat{G}_t\Big)^2\gamma_t+\beta^2(y_t)h+\frac{h^2}{2}\beta^2(y_t, \hat{\theta}_t)\dot{\beta}^2_t(y_t, \hat{\theta}_t)+\sigma^2 h\hat{G}^2_t\Big(y_t+\frac{1}{8}\sigma^2 h\Big).$$

As we can see, we found the form of the recursive estimator and the mean square error, which is the mean of the squared difference between the actual value and the estimated value.

2.3 Optimal estimation for semimartingales

This section is based on [14], which extends the theory of parametric estimation for discrete-time stochastic processes [3] to the continuous-time case. Firstly, we introduce the optimality criterion and later we will mention the recursive case.

2.3.1 Optimality criterion

Assume we have a complete probability space $(\Omega, \mathcal{A}, \mathcal{P})$ for each $P \in \mathcal{P}$ with a family $\mathcal{F} = \{F_t, t \geq 0\}$ of σ-algebras. We also denote the space of right-continuous functions $x = (x_t, t \geq 0)$ having limits on the left by D. We will use $X = (X_t, F_t)$ to denote an $\{F_t\}$-adapted process (X_t) with trajectories in the space D, with initial assumption $X_0 = 0$. Additionally, we denote by $M(F, P), M_{\text{loc}}(F, P), M^2_{\text{loc}}(F, P)$ classes of uniformly integrable, local and locally square-integrable martingales $X = (X_t, F_t)$ respectively. Next, we have the class of random processes $V = (V_t, F_t)$ which have locally bounded variation P-a.s, denoted by $V_{\text{loc}}(F, P)$.

Since we assume the process $X = (X_t, F_t)$ is a semimartingale for each P, then we can represent it as

$$X_t = V_{t,\theta} + H_{t,\theta}, \tag{2.18}$$

where $V = (V_t, F_t) \in V_{\text{loc}}(F, P)$ and $H = (H_t, F_t) \in M_{\text{loc}}(F, P)$.

Remark. If V is predictable, then this representation (2.18) is called the Doob-Meyer decomposition.

Now, we consider a parameter θ to be a function of P, and let $G = (G_{t,\theta}, F_t)$ be a family of processes indexed by θ such that $E(G_{t,\theta}) = 0$ for each t with respect to P. Godambe's optimality criterion in [3] states that in the class of unbiased estimating functions an estimating function is optimum if it minimizes

$$E(G_{t,\Theta}^2) \Big/ \left(E(\frac{\partial G_t}{\partial \theta}) \right)^2 \tag{2.19}$$

for each t. In this section, we specialize in the functions of the form

$$G_{t,\theta} = \int_0^t a_{s,\theta}\, dH_{s,\theta}, \tag{2.20}$$

where $H = (H_{t,\theta}, F_t) \in M^2_{\text{loc}}(F, P)$ and $I_{s,\theta} = \partial H_{s,\theta}/\partial\theta$ exists.
Also, we assume the following: $(a_{s,\theta}, F_s)$ is predictable and differentiable with respect θ, $\int_0^t (\partial a_s/\partial\theta)\, dH_{s,\theta}$, $\int_0^t a_{s,\theta}\, dH_{s,\theta}$ exist and lastly, $\partial G_{t,\theta}/\partial\theta$ has non-zero expectation and can be expressed as

$$\int_0^t (\partial a_{s,\theta}/\partial\theta)\, dI_{s,\theta} \int_0^t a_{s,\theta}\, dI_{s,\theta}. \tag{2.21}$$

Using the properties of the stochastic integral with respect to martingales, we get

$$E(G_{t,\Theta}^2) = E(\int_0^t a_{s,\theta}^2\, d\langle H\rangle_{s,\theta}), \tag{2.22}$$

and

$$E(\frac{\partial G_{t,\theta}}{\partial\theta}) = E(\int_0^t a_{s,\theta}\, dI_{s,\theta}). \tag{2.23}$$

A sufficient criterion for the optimal process $G^*_{t,\theta} = \int_0^t a^*_{s,\theta}\, dH_{s,\theta}$ to exists is that

$$E\left(G_{t,\Theta} G^*_{t,\theta} / E(\frac{\partial G_{t,\theta}}{\partial\theta} \right) \tag{2.24}$$

is the same for all G. After using the properties of stochastic integrals, we can write it in the form

$$E\left(\int_0^t a_{s,\theta}a^*_{s,\theta}\,d\langle H\rangle_{s\theta})\Big/E(\int_0^t a_{s,\theta}\,dI_{s,\theta}\right); \tag{2.25}$$

therefore, $a^*_{s,\theta}$ can be represented as $-dI_{s\theta}/d\langle H\rangle_{s,\theta}$ and it will be optimal. This extends a result of [4] for discrete time processes.

2.3.2 Recursive optimal estimating functions

Following [14], we view the estimation of θ in a dynamic sense, as a special case of filtering, where the goal is to compute the estimate for some time-varying parameter from an observation and we will examine the properties of the general solutions of the recursive algorithms. In our case, θ is a constant over time, which leads to the filtering corresponding to a continual updating of the estimate $\hat{\theta}_t$ as more data comes available.

Now, consider the following semimartingale model:

$$X_t = \theta R_t + H_{t,\theta}, \tag{2.26}$$

where R_t is absolutely continuous with respect to $d\langle H\rangle_{t,\theta}$.

Now rewriting this as

$$dX_t = \theta dR_t + dH_{t,\theta}. \tag{2.27}$$

The optimal estimate $\hat{\theta}$ will satisfy the equation

$$\int_0^t a^*_{s,\hat{\theta}}(dX_s - \hat{\theta}_t dR_s) = 0. \tag{2.28}$$

Assuming $d\hat{X}_t = \hat{\theta}_t dR_t$ and $a^*_{s,\theta}$ is independent of θ, then by differentiating (2.28) we get

$$d\hat{\theta}_t = N_t a^*_t(dX_t - d\hat{X}_t), \tag{2.29}$$

where $N_t^{-1} = \int_0^t a_s^* dR_s$ and

$$dN_t = -N_t^2 a_t^{*2} d\langle H \rangle_t. \tag{2.30}$$

Equation (2.29) and (2.30) gives us the recursive procedure for estimating θ. The general solution gives the estimate

$$\theta_t^0 = \frac{\theta_0 N_0^{-1} + \int_0^t a_s^* dX_s}{N_0^{-1} + \int_0^t a_s^* dR_s}. \tag{2.31}$$

In this general solution, N_0^{-1} will correspond to the prior variance of θ_t^0, and as it goes to infinity, θ_t^0 becomes the estimate $\hat{\theta}_t$.

Now let us see a specific example.

Considering the Aalen's model (see [1]), we have

$$dX_t = \theta J(t, X)dt + dM_{t,\theta}.$$

Without going into details, for the counting process $\theta J(t, X)dt = \langle M \rangle_{t,\theta}$, the estimating function

$$G_{t,\theta}^* = \int_0^t a_{s,\theta}^* dM_{s,\theta}$$

will be optimal for

$$a_{s,\theta}^* = \frac{J(s, X)ds}{\langle M \rangle_s}.$$

If $a_{s,\theta}^* = 1/\theta$ and $dR_s = J(s, X)ds$, then equation (2.29) will be

$$d\hat{\theta}_t = \hat{\theta}_t (dX_t - d\hat{X}_t)/X_t = N_t (dX_t - d\hat{X}_t),$$

where $N_t^{-1} = \int_0^t J(s, X)ds$. Then

$$\dot{N}_t = -J(t, X)N_t^2,$$

and the general solution is

$$\theta_t^0 = \frac{\theta_0 N_0^{-1} + \int_0^t dX_s}{N_0^{-1} + \int_0^t J(s,X)ds}, \tag{2.32}$$

where as N_0 converges to infinity, θ_t^0 becomes the estimate $\hat{\theta}_t$.

Also this solution provides the filtering algorithm of Van Shuppen for counting processes with Aalen's model.

Chapter 3

Combined estimating functions

From now on we will now consider combined estimating functions to get a better estimate.

3.1 Optimal estimating function combinations

This section follows the paper [15]. Let us consider a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, on which θ is a real valued random variable. We also let g_1 and g_2 be unbiased estimating functions with finite positive variance and the expectation of $\frac{\partial g_1}{\partial \theta}$ and $\frac{\partial g_2}{\partial \theta}$ are also finite, with $\mathrm{E}(\frac{\partial g_1}{\partial \theta}) \neq 0$.

Theorem 3.1.1. In the class of all unbiased estimating functions

$$g = g_1 + cg_2, \tag{3.1}$$

we have

- g^* minimizes $Var(g)$:

$$g^* = g_1 + C^* g_2, \tag{3.2}$$

where

$$C^* = -\frac{Cov(g_1, g_2)}{Var(g_2)}, \tag{3.3}$$

- g^0 minimizes $Var(g)/(E(\frac{\partial g}{\partial \theta}))^2$:

$$g^0 = g_1 + C^0 g_2, \tag{3.4}$$

where

$$C^0 = \frac{E(\frac{\partial g_2}{\partial \theta})Var(g_1) - E((\frac{\partial g_1}{\partial \theta})Cov(g_1, g_2)}{E(\frac{\partial g_1}{\partial \theta})Var(g_2) - E((\frac{\partial g_2}{\partial \theta})Cov(g_1, g_2)}. \tag{3.5}$$

Remark. If $E(\frac{\partial g_2}{\partial \theta}) = 0$, then these two are equivalent, in which case $C^* = C^0 = -Cov(g_1, g_2)/Var(g_2)$.

Remark. In case when g_1^0 and g_2^0 are orthogonal and information unbiased estimating functions, then $C^0 = 1$ and the optimal combined estimating function is

$$g^0 = g_1^0 + g_2^0. \tag{3.6}$$

3.1.1 Application to State Space models

Following [15], which extend the result on optimal combination on estimating functions for discrete time stochastic processes [7] to discrete time state space models and to continuous time counting process models, we will consider the following state space model in discrete time:

$$\theta_{t+1} = a(t)\theta_t + c(t) + b(t)u_{t+1}$$
$$\xi_{t+1} = A(t)\theta_{t+1} + B(t)v_{t+1},$$

where $\{\theta_t\}$ is an unobserved sequences of random variables, $\{\xi_t\}$ is an observed sequence of variables and $\{u_t\}, \{v_t\}$ are independent sequences of independent variables with variance σ_u^2, σ_v^2 respectively. Also, the functions

$a(t), b(t), c(t), A(t)$ and $B(t)$ are $\mathcal{A}_t^\xi$ measurable. Let $\tilde{\theta}_t = E(\theta_t \mid \mathcal{A}_t^\xi)$,and $\gamma_t = Var(\theta_t \mid \mathcal{A}_t^\xi)$.

We consider the combinations of

$$g_1 = \theta_{t+1} - a(t)\tilde{\theta}_t - c(t) \tag{3.7}$$

$$g_2 = \xi_{t+1} - A(t)a(t)\tilde{\theta}_t - A(t)c(t). \tag{3.8}$$

We should note that θ_{t+1} serves the same role as the θ we mentioned above. By (3.2), the optimal combination is

$$\theta_{t+1} - a(t)\tilde{\theta}_t - c(t) - \frac{Cov(g_1, g_2 \mid \mathcal{A}_t^\xi)}{Var(g_2 \mid \mathcal{A}_t^\xi)}(\xi_{t+1} - A(t)a(t)\tilde{\theta}_t - A(t)c(t)). \tag{3.9}$$

3.1.2 Application to Continuous time models

Similarly as the previous subsection, we follow [15], showing how the combination of estimating function suggests a prescription for updating a filter in continuous time. As in Section 2.3.1, we have a probability space with right continuous filtration and here we also have cadlag processes $(\theta_t, t \geq 0)$(unobserved state process) and $(\eta_t, t \geq 0)$(observed state process) such that

$$d\theta_t = \alpha(t)\theta_{t-}dt + \sigma(t)dt + dM_t, \tag{3.10}$$

and

$$d\eta_t = h_t dt + dV_t, \tag{3.11}$$

where $(M_t, t \geq 0)$ is a square integrable martingale with respect to the filtration generated by $(\theta_s, t \leq s)$ and $(V_t, t \geq 0)$ is a square integrable martingale with respect to $(G_t, t \geq 0)$, where G_t is generated by $(\theta_t, t \geq 0; \eta_s, s \leq t)$. Additionally, α and σ have continuous paths and measurable with respect to F_{t-}^{η}, which is a σ-field generated by $\eta_s, s \leq t$. Lastly, $(h_t, t \geq 0)$ is a G_t-predictable and independent of the two cadlag processes.

To find the combined estimating functions, we will think of $d\theta_t$ as $\theta_{t+dt} - \theta_{t-}$ and let us say the differential $\widehat{d\tilde{\theta}_t}$ is an estimate of

$$d\theta_t + \theta_{t-} - \breve{\theta}_t, \tag{3.12}$$

where $\tilde{\theta}_t$ stands for $E(\theta_t \mid \mathcal{F}_t^{\eta})$ and $\breve{\theta}_t$ is $E(\theta_{t-} \mid \mathcal{F}_{t-}^{\eta})$.

Equation 3.12 is a part of the first estimating component

$$g_1 = d\theta_t - \alpha(t)\breve{\theta}_{t-}dt - \sigma(t)dt + \theta_{t-} - \breve{\theta}_{t-} \tag{3.13}$$

and the second component could be

$$g_2 = d\eta_t - \breve{h}_{t-}dt. \tag{3.14}$$

The optimal combination will be

$$d\theta_t - \alpha(t)\breve{\theta}_{t-}dt - \sigma(t)dt + \theta_{t-} - \breve{\theta}_{t-} - \frac{C_t}{W_t}(d\eta_t - \breve{h}_{t-}dt), \tag{3.15}$$

where C_t stands for the covariance

$$Cov(g_1, g_2 \mid \mathcal{F}_{t-}^{\eta}) = E((dM_t + \theta_{t-} - \breve{\theta}_{t-})(dV_t + h_t dt - \breve{h}_{t-}dt) \mid \mathcal{F}_{t-}^{\eta}))$$

and W_t is the variance

$$Var(g_2 \mid \mathcal{F}_{t-}^{\eta} = E(d\langle V\rangle_t \mid \mathcal{F}_{t-}^{\eta}))$$

assuming $Var(h_t dt \mid \mathcal{F}_{\sqcup-}^{\eta})$ is negligible. By setting (3.15) to be zero we get a solution

$$\widehat{d\tilde{\theta}_t} = \alpha(t)\breve{\theta}_{t-}dt + \sigma(t)dt + \frac{C_t}{W_t}(d\eta_t - \breve{h}_{t-}dt).$$

The last equation, by a few little modification is essentially the standard filtering equation. Thus, it gives a justification through estimating functions for stochastic differential equation satisfied by the usual filter $\breve{\theta}_t = E(\theta_t \mid \mathcal{F}_t^{\eta})$.

3.2 Discretely observed diffusion

This section follows closely [9], where we deal with discrete time results on combining estimating functions and obtain closed form expression for the gain in information. We assume the continuous time process $\{y_t\}$ is recorded discretely at time points $h.2h, \ldots$, where h is the discrete interval of observations. Let $\mathcal{F}^y_{(t-1)h}$ be the σ-field generated by $y_{1h}, y_{2h}, \ldots, y_{(t-1)h}$, now we consider the observable discrete time process $\{y_{th}, t = 1, 2 \ldots\}$ with the following conditional moments

$$\mu_t(\theta) = E(y_{th} \mid \mathcal{F}^y_{(t-1)h}), \tag{3.16}$$

$$\sigma^2_t(\theta) = Var(y_{th} \mid \mathcal{F}^y_{(t-1)h}), \tag{3.17}$$

$$\gamma_t(\theta) = E\big((y_{th} - \mu_t(\theta))^3 \mid \mathcal{F}^y_{(t-1)h}\big), \tag{3.18}$$

$$\kappa_t(\theta) = E\big((y_{th} - \mu_t(\theta))^4 \mid \mathcal{F}^y_{(t-1)h}\big). \tag{3.19}$$

In order to estimate θ based on the observations, we consider two classes of martingale differences $\{m_t(\theta) = y_{th} - \mu_t(\theta), t = 1, \ldots, n\}$ and $\{M_t(\theta) = m^2_t(\theta) - \sigma^2_t(\theta), t = 1, \ldots, n\}$, where the quadratic variation and covariation of these two are

$$\langle m \rangle_t = \sigma^2_t(\beta), \tag{3.20}$$

$$\langle M \rangle_t = \kappa_t(\beta) - \sigma^4_t(\beta), \tag{3.21}$$

and

$$\langle m, M \rangle_t = \gamma_t(\beta) \tag{3.22}$$

respectively.

The optimal estimating functions based on the martingale differences are given by

$$g_m^* = -\sum_{t=1}^{n} \frac{\partial \mu_t}{\partial \theta} \frac{m_t}{\langle m \rangle_t}, \tag{3.23}$$

$$g_M^* = -\sum_{t=1}^{n} \frac{\partial \sigma_t^2}{\partial \theta} \frac{M_t}{\langle M \rangle_t}, \tag{3.24}$$

moreover, we have the corresponding information

$$I_{g_m}^* = -\sum_{t=1}^{n} \frac{\partial \mu_t}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \frac{1}{\langle m \rangle_t}, \tag{3.25}$$

$$I_{g_M}^* = -\sum_{t=1}^{n} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{1}{\langle M \rangle_t}. \tag{3.26}$$

Theorem 3.2.1. Given a discretely observed process, in the class of all combined estimating functions of the form

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

- The optimal estimating function is $g_c^* = \sum_{t=1}^{n} (a_{t-1}^* m_t + b_{t-1}^* M_t)$, where

$$a_{t-1}^* = \left(1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t}\right)^{-1} \left(-\frac{\partial \mu_t}{\partial \theta} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t}\right) \tag{3.27}$$

and

$$b_{t-1}^* = \left(1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t}\right)^{-1} \left(\frac{\partial \mu_t}{\partial \theta} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t} - \frac{\partial \sigma_t^2}{\partial \theta} \frac{1}{\langle m \rangle_t}\right). \tag{3.28}$$

- The information $I_{g_C^*}(\theta)$ is given by

$$I_{g_C^*}(\theta) = \sum_{t=1}^{n} \left(1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t \langle M \rangle_t}\right)^{-1} \left[\frac{\partial \mu_t}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{1}{\langle M \rangle_t} + \left(\frac{\partial \mu_t}{\partial \theta} \frac{\partial_t^2}{\partial \theta'} + \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \mu_t}{\partial \theta'} \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t}\right)\right]; \tag{3.29}$$

- The information gain is given by

$$I_{g_C^*}(\theta) - I_{g_m^*}(\theta) = \textstyle\sum_{t=1}^{n} \left(1 - \frac{\langle m,M\rangle_t^2}{\langle m\rangle_t \langle M\rangle_t}\right)^{-1}$$
$$\left[\frac{\partial\mu_t}{\partial\theta}\frac{\partial\mu_t}{\partial\theta'}\frac{\langle m,M\rangle_t^2}{\langle m\rangle_t} + \frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\frac{1}{\langle m\rangle_t} + \left(\frac{\partial\mu_t}{\partial\theta}\frac{\partial_t^2}{\partial\theta'} + \frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\mu_t}{\partial\theta'}\frac{\langle m,M\rangle_t}{\langle m\rangle_t\langle M\rangle_t}\right)\right]. \tag{3.30}$$

Before going into how it works with specific models, it would be wise to see what is the case with general models.

3.3 General models

In this section we will use Milstein's approximation to obtain the first four conditional moments and construct the optimal estimating functions as in paper [9].

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

$$dy_t = a(\alpha, y_t)dt + b(\beta, y_t)dW_t, \tag{3.31}$$

where a is the drift, b is the diffusion functions, and W_t is the standard Brownian motion.

For extended general models, the diffusion is a function of the observation y_t, hence closed form of expressions of the conditional distribution, as well as closed form expression for the conditional moments cannot be easily obtained by repeating the Itô's formula. That is why we use the Milstein's approximation to do so.

After applying Milstein's approximation to (3.31), we get

$$
\begin{aligned}
y_{th} =& y_{(t-1)h} + a(\alpha, y_{(t-1)h})dt + b(\beta, y_{(t-1)h})\sqrt{h}z_t \\
&+ \tfrac{1}{2} b(\beta, y_{(t-1)h}) b_y(\beta, y_{(t-1)h})(z_t^2 - 1)h,
\end{aligned}
\tag{3.32}
$$

where $b_y = \frac{\partial b}{\partial y}$ and $z \sim N(0, 1)$, i.i.d.

With (3.32), we can compute the first four conditional moments of y_{th} given $y_{(t-1)h}$:

$$\mu_t(\alpha) = y_{(t-1)h} + a(\alpha, y_{(t-1)h})h, \tag{3.33}$$

$$\sigma_t^2(\beta) = b^2(\beta, y_{(t-1)h})h + \frac{1}{2} b^2(\beta, y_{(t-1)h}) b_y^2(\beta, y_{(t-1)h})h^2, \tag{3.34}$$

$$\gamma_t(\beta) = 3b^3(\beta, y_{(t-1)h}) b_y(\beta, y_{(t-1)h})h^2 + b^3(\beta, y_{(t-1)h}) b_y^3(\beta, y_{(t-1)h})h^3, \tag{3.35}$$

$$
\begin{aligned}
\kappa_t(\beta) = & 15b^4(\beta, y_{(t-1)h}) b_y^2(\beta, y_{(t-1)h})h^3 \\
&+ \tfrac{15}{4} b^4(\beta, y_{(t-1)h}) b_y^4(\beta, y_{(t-1)h})h^4 + 3b^4(\beta, y_{(t-1)h})h^2.
\end{aligned}
\tag{3.36}
$$

Then based on the observations, we can see the martingale differences $m_t = y_{th} - \mu_t(\alpha)$ and $M_t = m_t^2 - \sigma_t^2(\beta)$ as we did in Section 2.1. In this case, we can also compute the quadratic variation and covariation as we did in (3.21)

$$
\begin{aligned}
\langle M \rangle_t = & \kappa_t(\beta) - \sigma_t^4(\beta) = \\
&= 14b^4(\beta, y_{(t-1)h}) b_y^2(\beta, y_{(t-1)h})h^3 \\
&+ \tfrac{7}{2} b^4(\beta, y_{(t-1)h}) b_y^4(\beta, y_{(t-1)h})h^4 + 2b^4(\beta, y_{(t-1)h})h^2,
\end{aligned}
\tag{3.37}
$$

$$\langle m \rangle_t = \sigma_t^2(\beta), \langle m, M \rangle_t = \gamma_t(\beta). \tag{3.38}$$

Now using from (3.23)- (3.26), we can compute the optimal estimating functions and corresponding information:

$$g_m^*(\alpha) = -\sum_{t=1}^{n} \frac{\partial \mu_t}{\partial \alpha} \frac{m_t}{\langle m \rangle_t}, \tag{3.39}$$

$$g_M^*(\beta) = -\sum_{t=1}^{n} \frac{\partial \sigma_t^2}{\partial \beta} \frac{M_t}{\langle M \rangle_t}, \tag{3.40}$$

moreover, we have the corresponding information:

$$I_{g_m}^*(\alpha) = -\sum_{t=1}^{n} \frac{\partial \mu_t}{\partial \alpha} \frac{\partial \mu_t}{\partial \alpha'} \frac{1}{\langle m \rangle_t}, \tag{3.41}$$

$$I_{g_M}^*(\beta) = -\sum_{t=1}^{n} \frac{\partial \sigma_t^2}{\partial \beta} \frac{\partial \sigma_t^2}{\partial \beta'} \frac{1}{\langle M \rangle_t}. \tag{3.42}$$

3.4 Application

Finally, let us see how we obtain the optimal estimating function and information matrix based on the previous theorems. (We are using paper [9].)

Example 3.4.1 (NLD process)**.** Consider the following Nonlinear Drift (NLD) diffusion process for modeling interest rates

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

where $\beta_1, \beta_2 > 0$, $0 < \alpha_1 < \beta_2/2$ and $\alpha_2 > \beta_1/2$. To relate this to our general model, we have $a(\alpha, y) = (\alpha_1 + \alpha_2/y)$, $B(\alpha, y) = \sqrt{\beta_1 + \beta_2 y}$ and $b_y = (\alpha, y) = 1/2(\beta_1 + \beta_2 y)^{-1/2}$. The Milstein's approximation will give us the following:

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

One can find the first four conditional moments and then the estimating function and information based on m_t is:

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

$$I_m^*(\alpha) = \begin{pmatrix} h^2 \sum_{t=1}^n \frac{1}{\sigma_t^2(\beta)} & h^2 \sum_{t=1}^n \frac{1}{y_{(t-1)h}\sigma_t^2(\beta)} \\ h^2 \sum_{t=1}^n \frac{1}{y_{(t-1)h}\sigma_t^2(\beta)} & h^2 \sum_{t=1}^n \frac{1}{y_{(t-1)h}^2\sigma_t^2(\beta)} \end{pmatrix}.$$

Also, the estimating function and information based on M_t are given as

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

$$I_m^*(\alpha) = \begin{pmatrix} h^2 \sum_{t=1}^n \frac{1}{\langle M \rangle_t} & h^2 \sum_{t=1}^n \frac{(y_{(t-1)h}+1/4\beta_2 h)}{\langle M \rangle_t} \\ h^2 \sum_{t=1}^n \frac{(y_{(t-1)h}+1/4\beta_2 h)}{\langle M \rangle_t} & h^2 \sum_{t=1}^n \frac{(y_{(t-1)h}+1/4\beta_2 h)^2}{\langle M \rangle_t}. \end{pmatrix}$$

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:

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

where

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

with

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

and

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

Finally, the information for the combined estimating function is

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

Now, let us see another example.

Example 3.4.2. In this example, we will attempt to use trinomial trees for the purpose of simulating an interest rate model and Milstein's approximation. Let's consider the modified version of the CIR model, for the purpose of better convergence we will take the square root of

$$dy_t = k(\theta_t - y_t)dt + \sigma\sqrt{y_t}dW_t.$$

We will have $x_t = \sqrt{y_t}$,

$$dx_t = (-\frac{k}{2}x_t + \frac{1}{2}\frac{k\theta}{x_t} - \frac{1}{8}\frac{\sigma^2}{x_t}) + \frac{\sigma}{2}dW_t.$$

Trinomial trees are mostly used for option pricing and it also can be used to provide a discrete-time and discrete-space Markov approximation for x (see [13] for more details). We denote the process value on node (i, j) by $y(i, j)$. Hence, from a node $y(i, j)$ we have three possible steps: to go up, to stay, or to go down, with respective probabilities. For the theoretical part, we used [2].

Using the probability structure of the trinomial tree, we have generated a trajectory (orange), which then we compare with the Milstein's approximation (blue) of the CIR model.

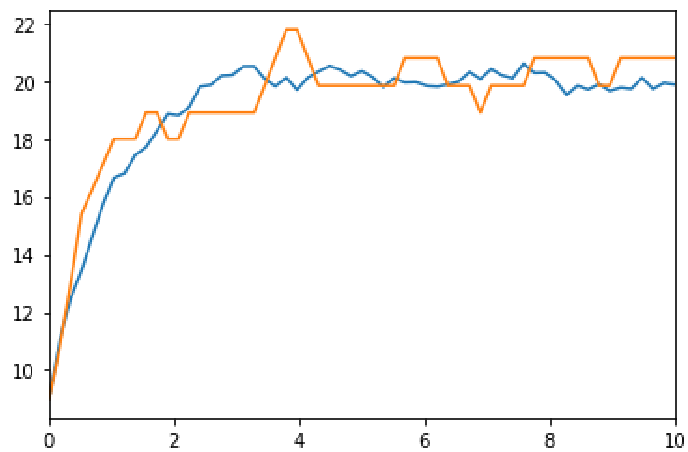

Figure 1: Milstein's approximation and the trajectory of the trinomial tree of the CIR model

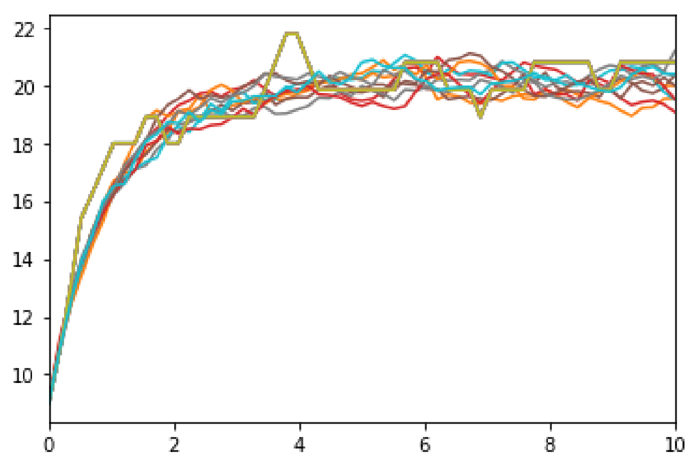

Figure 2: Multiple trajectories of the Milstein's approximation of the CIR model

As we can see, the approximation is very close to our trajectory of the trinomial tree.

Chapter 4

Conclusion

As we have seen in the examples, Milstein's method works well when it comes to some interest rate models. Additionally, we saw how we can use Milstein's method to obtain the first four conditional moments of the diffusion and then constructing an optimal estimating function. In the end, we can use the martingale differences and obtain information gain.

Appendix A

Semimartingales

A semimartingale is a stochastic process which can be decomposed as a sum of local martingales and a finite variation process. It has a significant role in financial application, with being able to represent arbitrage-free asset prices and interest rates.

Definition A.0.1. Consider a filtered probability space and X a real valued process on it. We say X is a semimartingale if it can be decomposed as

$$X_t = M_t + F_t,$$

where M_t is a local martingale and F_t is an adapted process of locally bounded variation.

The continuous case is also quite similar.

Definition A.0.2. X is a continuous semimartingale if it can be uniquely decomposed as

$$X_t = M_t + F_t,$$

where M_t is a continuous local martingale and F_t is a continuous finite variation process starting at zero.

To just name a few semimartingales: Brownian motion, Itô process, Levy process etc.

An important property to note is that semimartingales form the largest class of processes for which Itô integral can be defined.

Property 1. A linear combination of semimartingales is a semimartingale too.

Property 2. For every semimartingale, the quadratic variation exists.

One thing to note is that in this paper we use a special case of semimartinagles, that can be written with respect to the Brownian motion. The reason for that is, we use Milstein's approximation, that is obtained as a result of application of stochastic Taylor expansion, or more easily, by Itô's formula.

Appendix B

Milstein's approximation

This appendix is based on [8] and [12].

The simplest way to approximate a stochastic differential equation may be the Euler–Maruyama Method, but it may not the most accurate in some cases. Hence, we can use Milstein's approximation which adds a second-order "correction" term, that is derived from the Taylor expansion and Itô's lemma. Considering the homogeneous scalar stochastic differential equation

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t,$$

where a, b are the drift and diffusion coefficients, both smooth, and W_t is a Brownian motion, the Milstein's method yields the following form:

$$X_{n+1} - X_n = a(t, X_n)\Delta t + b(t, X_n)\Delta W_n + 0.5b(t, X_n)b'(t, X_n)((\Delta W_n)^2 - \Delta t).$$

Now, to show why this method is useful we can check its convergence. Without going to the specifics, we can see that it has strong and weak order of convergence of $\beta = 1$. For the definition and more details, see [12]. Specifically, in [8], we can see how the Milstein approximation is closer to the exact solution, when compared to the Euler–Maruyama method, for which it has a strong order of convergence of $\beta = 1/2$, while having weak order of convergence of $\beta = 1$.

Finally, the Milstein approximation is better when we deal with scalar diffusions, since it is easier to implement. However, when it comes to multi-dimensional cases it becomes drastically more challenging.

Bibliography

[1] Aalen O. O.,*Non parametric inference for a family of counting processes*, Ann. Statist. 6, 1978.

[2] Damiano Brigo, Fabio Mercurio,*Interest Rate Models Theory and Practice*, With Smile, Inflation and Credit (Springer Finance) 2nd Edition 2013.

[3] Godambe V.P.,*An Optimum Property of Regular Maximum Likelihood Estimation*, Ann. Math. Statist. 31(4), 1960.

[4] Godambe V.P.,*The foundations of finite sample estimation in stochastic processes*, Biometrika 72, 1985.

[5] Godambe V.P.,*Estimating Functions*, Oxford Statistical Science, 1991.

[6] Godambe V.P.,*Estimating Functions: A Synthesis of Least Squares and Maximum Likelihood Methods*, Lecture Notes-Monograph Series, vol. 32, Institute of Mathematical Statistics, pp. 5–15, 1997.

[7] Godambe V.P.,*Linear Bayes and optimal estimation*, Ann. Inst. Statist. Math. Vol.51 , 1999.

[8] Hautahi Kingi, *Numerical SDE Simulation - Euler vs Milstein Methods*, 2019.
`https://hautahi.com/sde_simulation`

[9] Koulis T. and Thavaneswaran A., *Inference for Interest Rate Models Using Milstein's Approximation*, Journal of Mathematical Finance, Vol. 3 No. 1, pp. 110-118, 2013.

[10] Koulis T., Paseka A. and Thavaneswaran A., *Recursive Estimation for Continuous Time Stochastic Volatility Models Using the Milstein Approximation*, Journal of Mathematical Finance, Vol. 3 No. 3, pp. 357-365, 2013.

[11] López-Pérez, Alejandra, Manuel Febrero-Bande and Wencesalo González-Manteiga., *Parametric Estimation of Diffusion Processes: A Review and Comparative Study*, 2021.

[12] Martin Haugh, *Simulating Stochastic Differential Equations*, Monte-Carlo Simulation, Columbia University, 2017.
`http://www.columbia.edu/~mh2078/MonteCarlo/MCS_SDEs.pdf`

[13] Markus Leippold and Zvi Wiener, *On Trinomial Trees for One-Factor Short Rate Models*, 2003.

[14] Thavaneswaran A. and Thompson M., *Optimal estimation for semi-martingales*, Journal of Applied Probability, Vol. 23 Issue 2, pp. 409 - 417, June 1986.

[15] Thompson M. E. and Thavaneswaran A., *Filtering via estimating functions*, Applied Mathematics Letters 12, 1999, 61-67.

[16] Tse Y. K., *Stochastic Models of Interest Rates in Economics, Finance and Actuarial Science*, Keynote Address, Modsim 95. International Congress on Modelling and Simulation, Newcastle, New South Wales, Research Collection School Of Economics, Australia 1995.

[17] Sachin Date, *An Intuitive Look At Fisher Information*, 2021. `https://towardsdatascience.com/an-intuitive-look-at-fisher-information-2720c40867d8`

[18] Qian Zhou, *Information Matrices in Estimating Function Approach: Tests for Model Misspecification and Model Selection*, Thesis, Canada, 2009.