UNIVERSITA' COMMERCIALE "LUIGI BOCCONI" – MILANO

Facolta' di Economia

Dottorato di Ricerca in Statistica

XIX Ciclo

Robust Martingale Estimating Functions for Discretely Observed Diffusion Processes

Coordinatore:

Ch.mo Prof. Pietro Muliere

Tesi di:

Davide La Vecchia

MAT. 935585

UNIVERSITA' COMMERCIALE "LUIGI BOCCONI"
ISTITUTO DI METODI QUANTITATIVI

The thesis "**Robust Martingale Estimating Functions for Discretely Observed Diffusion Processes**" by **Davide La Vecchia** **MAT. 935585** is recommended for acceptance by the members of the delegated committee, as stated by the enclosed reports, in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: January 2007

Research Supervisor: **Michael Sørensen**
Fabio Trojani
External Examiners: **Francesco Corielli**

UNIVERSITA' COMMERCIALE "LUIGI BOCCONI"

Author: **Davide La Vecchia** **MAT. 935585**

Title: **Robust Martingale Estimating Functions for Discretely Observed Diffusion Processes**

Department: **Istituto di Metodi Quantitativi**

Signature of Author

To Priscilla

Table of Contents

Table of Contents ix

Acknowledgements xi

Introduction 3

1 Inference for Diffusion Processes 7

1.1 Continuous-time observations 7

1.2 Discrete-time observations 8

1.2.1 Maximum Likelihood Estimator 8

1.2.2 Approximation of the Likelihood by a Gaussian transition density 9

1.2.3 Indirect Inference and EMM 11

1.2.4 Estimating Functions 14

2 Overview about Robust Statistics 29

2.1 Robust Statistics 29

2.2 The role of the Influence Function 31

2.2.1 The i.i.d. case 31

2.2.2 Dependent data 34

3 Robust Inference for Discretely Observed Diffusions 39

3.1 The General framework 40

3.1.1 Local robustness for contaminated stochastic processes in time series 40

3.1.2 Diffusion processes 44
3.2 Conditionally Unbiased M-estimators 48
3.2.1 M-R-T: robust Pseudo ML for ARCH processes . . 48
3.2.2 Kessler and Sørensen: MEFs for discretely observed diffusions 51
3.3 Theoretical Issues 58
3.4 IF^{cond}: Theorem 1.3 in Künsch for diffusions 60
3.5 Robust MEFs for discretely observed diffusions 65
3.5.1 First proposal: Godambe-Heyde Optimal Robust MEFs 65
3.5.2 Second proposal: Robust version of Kessler and Sørensen MEFs 71
3.6 Algorithm for θ and analytical approximation for $\tau(X_{t_{i-1}}, \theta)$ 81
3.7 Asymptotics 83

4 Examples and Implementation of R-algorithms 85
4.1 Example 1: Contaminated CIR process 86
4.1.1 Mathematical Analysis 86
4.1.2 Monte Carlo Analysis 88
4.2 Example 2: Contaminated trigonometric diffusion 94
4.2.1 Mathematical Analysis 95
4.2.2 Monte Carlo Analysis 101

Conclusion 115

Bibliography 117

Acknowledgements

This PhD. Thesis has been written during 2005-2006 in Switzerland and in Denmark.

I wish to thank my supervisor Fabio Trojani (University of St.Gallen) for his guidance and for the long time that he spent with me, discussing ideas, proofs, problems and new topics of research. His precious comments and his criticism was very helpful for the success of this Thesis. Moreover, I would like to thank him for the unfailing patience that he had in reading this document. I am grateful also to Patrick Gagliardini, that spent with me some cold and rainy afternoon in St.Gallen, talking about the theory of Partial Differential Equations. Furthermore, I wish to thank Massimo Lonardi, Andrea Vedolin and all the other PhD students at SBF, for their folk support.

I am very grateful to my second supervisor, Michael Sørensen (University of Copenhagen). He spent long time talking with me about the theoretical construction of this Thesis and he gave to me several invaluable suggestions. His illuminating discussions about diffusion processes and about the theory of Martingale Estimating Functions was crucial for the realization of this work. In Copenhagen, I would like to thank also Martin Jacobsen, for some useful discussions about Markov processes and about the 2006 World Cup, that our team won!!

I am indebted also with Francesco Corielli ("encyclopedic" Professor at Bocconi University), that invested much of his time to introduce me to the problem of statistical inference for discretely observed stochastic processes. He also made some useful comments about this Thesis, suggesting me a new way to study the problem of robustness for diffusions.

I thank my family, my PhD colleagues in Bocconi (in particular Chiara, Gabriella, Matteo and Petra) and my "non-academic" friend Fabrizio, for their unconditional support and encouragement.

Finally, my warmest thanks go to Marialucia, whose love helps me every day of my life.

Introduction

In this PhD. Thesis we focus on diffusion models. Diffusions are very attractive and widely applied in many scientific areas, since they provide a useful mathematical tool: the stochastic calculus. Let us think, for instance, to the Finance area. A large class of financial models use both diffusion processes and the stochastic calculus machinery, in order to price financial derivatives or to solve stochastic optimization problems. Thus convinced that diffusions are very important and useful for modeling natural and financial phenomena, we will focus our attention on their statistical analysis. In particular, the main goal of this Thesis will be the definition of a specific robust method, to make parametric inference for the drift and the diffusion coefficients of discretely observed diffusion processes. To give a general overview, we start by recalling that, in order to characterize a diffusion process, there are mainly three alternatives (see, for instance Karlin and Taylor, [27]):

I) We can define the infinitesimal coefficients (drift and diffusion) of the process, but we have to introduce some additional constraints, specifying how the process behaves at or near the boundaries;

II) We can use the Strock-Varadhan Martingale Characterization;

III) We can specify a Stochastic Differential Equation (in the following SDE), since a diffusion process is (under some regularity conditions) the (weak or strong) the solution of a SDE.

The third approach is the most common one and throughout this Thesis we will refer to diffusion processes as solutions of specific SDEs. In particular, in our framework, we assume we are given the SDE:

$$\begin{cases} dX(t) = \alpha(X(t);\theta)dt + \sigma(X(t);\theta)dW(t) \\ X(0) = x_0, \end{cases}$$

where $X(t)$ takes values in the State Space $S = (l, r)^1 \in \mathbb{R}^k$, $\theta \in \Theta$ and Θ is an open subset of $\mathbb{R}^p$. We assume that the drift $(\alpha(X(t);\theta))$ and the diffusion $(\sigma(X(t);\theta))$ coefficients are functions depending on an unknown parameter θ, which are assumed to be smooth enough to ensure the existence and the uniqueness of a weak solution. The process $\{X(t)\}_{t\geq 0}$ is regular, time-homogeneous, ergodic and strictly stationary. For the moment we do not introduce any additional assumptions. Later on (see Chapter 3) we will define in detail the structure of the probability space we will work on and we will provide conditions on $\alpha(X(t);\theta)$ and on $\sigma(X(t);\theta)$ that ensure the ergodicity and the stationarity of the process. As far as the inference about θ is concerned, we have to be aware that there exist some problems related to the time-lag between the observations. As a matter of fact, in many applications only discrete-time observations of the continuous-time process are available. Therefore, the inference for θ is not straightforward. Another problem is related to the estimating function that we use to draw inference. As a matter of fact, only for a very limited class of models (including, for instance, the Brownian Motion, the Ornstein-Uhlenbeck and the CIR process) the Likelihood Score Function can be computed explicitly. For these processes we can use the Maximum Likelihood approach to estimate the parameter θ. However, these cases are very special cases. In many models, both the SDE cannot be solved explicitly and the solution of the Fokker-Planck (or of its adjoint, the Kolmogorov Backward) equation is not known in closed form. This feature implies that the discrete-time transition probabilities are unknown in closed form and the Maximum Likelihood approach is not feasible.

In Chapter 1, we analyze several alternative approaches to make a parametric inference for θ. In particular, we focus on the the Martingale Estimating Functions theory, studied in 90's by Bibby, Jacobsen, Kessler and Sørensen.

In Chapter 2, we introduce Robust Statistics, a topic widely studied by Huber and Hampel (in the classical i.i.d. case) and by Künsch, Genton, Martin, Yohai, Ronchetti

[1] We could have $-\infty \leq l \leq r \leq \infty$.

and Trojani (in the non i.i.d. case). The basic idea of Robust Statistics is to determine inference procedures to define estimators having a good performance even under slight violations of the assumptions of the reference model. This is a key point for our inferential problem. Let us think, for instance, to the case of diffusion processes applied in Physics. In almost all real situations we can have some measurement errors and, as a consequence, our sample can contain some "strange" observations. Also in Finance we can observe some "outliers", due for instance to some rumors that drive the market in an unusual way only for a few days. In all these situations (or more generally in presence of a misspecification model or data contamination) classical estimators can imply a biased and/or inefficient inference. As a result, we believe that robustness is a key issue in the statistical inference for discretely observed diffusion processes. Nevertheless so far, there does not exist any specific method to draw a robust parametric inference for a discretely observed diffusion. General estimation methods (like Robust GMM [36], Robust EMM [35] or Robust Indirect Inference [12]) are really time-consuming and computationally demanding, since they are based on some internal Monte Carlo simulations. In this Thesis, we solve some of these problems, defining a new robust inference procedure, specific for discretely observed diffusions. More precisely, in Chapter 3 we show how to construct Robust Martingale Estimating Functions by means of two proposals, that define Robust Conditionally Unbiased M-estimators. In Section 3.6, we provide an Algorithm to implement the second proposal. A distinguishing feature of this algorithm is that it defines a procedure that uses the solution for the Sturm-Liouville problem for the Infinitesimal Generator to avoid time-consuming Monte Carlo simulations. Finally, in Section 3.7, we study the asymptotic properties of our robust M-estimators.

In Chapter 4 we propose an interesting example in order to apply our method. Our Monte Carlo analysis shows that Sørensen's M-estimator can define for contaminated samples a very biased and/or inefficient inference procedure, even in presence of small departures from the reference model. More specifically, we analyze Example 2.2 proposed in [28] by Kessler and Søresen, and we show that very small contaminations can have large changes on the bias of Sørensen's M-estimator. To mitigate these effects and to correct the bias, we apply our second proposal. Finally, Monte Carlo study shows that our Robust M-estimator maintains high efficiency at the reference model.

Chapter 1

Inference for Diffusion Processes

In this chapter we review some methods for drawing parametric inference of diffusion processes, both for the case of continuous-time observations and for the case of discrete-time observations. This description is not exhaustive. For the discrete-time case, we have chosen to present the benchmark Maximum Likelihood case, Ait-Sahalia's method based on a series expansion of the transition density, two methods based on auxiliary models (the Indirect Inference and the EMM) and, finally, the method of Martingale Estimating Functions. In the last pages of the chapter, we recall some important results for Sturm-Liouville problem, a mathematical tool very useful to define estimating functions.

1.1 Continuous-time observations

Assume we can observe the process $\{X(t), 0 \leq t \leq T\}$ continuously and we want to estimate the parameter θ specifying the drift and diffusion of the SDE specified in the Introduction. We start by remarking that the parameters in the diffusion coefficient can be determined (rather than estimated) by using the quadratic variation of the process. Therefore the diffusion coefficient is completely known and we can set $\sigma(x, \theta) = \sigma(x)$. After this first step, we can estimate the remaining part of the parameters that are in the drift, by using the Maximum Likelihood approach. More specifically, the Likelihood

Score Function is given by (see Lipster and Shiryayev, [32], Chapter 7):

$$\exp\left(\int_0^T \frac{\alpha(X(s);\theta)}{\sigma(X(s))}dX(s) - \frac{1}{2}\int_0^T \left(\frac{\alpha(X(s);\theta)}{\sigma(X(s))}\right)^2 ds\right) \tag{1.1.1}$$

We do not provide any mathematical proof for formula (1.1.1). However, from a heuristic and intuitive point of view, we could justify this expression for the Likelihood Score Function by using an Euler approximation scheme for the SDE (). As a matter of fact, if we have $X_{t_1}, X_{t_2}, ..., X_{t_n}$ observations, sampled with a fixed (really) small time lag ($\Delta t := t_i - t_{i-1} > 0$) and if the diffusion coefficient is known, then the Likelihood conditional on $X_0 = x_0$, is

$$L_n(\theta) = \prod_{i=1}^{n} p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t), \tag{1.1.2}$$

where $p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)$ is the Gaussian transition probability density. As a result the last equation ca be written as:

$$L_n(\theta) \propto \exp\left(\sum_{i=1}^{n} \frac{\alpha(X_{t_i-1};\theta)}{\sigma(X_{t_i-1})}(X_{t_i} - X_{t_i-1}) - \frac{\Delta t}{2}\sum_{i=1}^{n}\left(\frac{\alpha(X_{t_i-1};\theta)}{\sigma(X_{t_i-1})}\right)^2\right).$$

We observe that the last expression is the Riemann-Itô approximation of (1.1.1).

1.2 Discrete-time observations

From now on we focus on the case where we have $X_{t_1}, X_{t_2}, ..., X_{t_n}$ observations of the process. For convenience we consider observations obtained at equidistant points in time with fix time-lag $\Delta t := t_i - t_{i-1} > 0$, but all the methods can be easily generalized to the case of time non-equidistant fixed points.

1.2.1 Maximum Likelihood Estimator

Assume we are able to solve in closed form the Fokker Planck Partial Differential Equation (or its adjoint the Kolmogorov backward equation). In this setting the transition probability density $p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)$ is known in analytical form and we can use the

Likelihood Score to estimate the parameter θ. Under weak regularity conditions, the Maximum Likelihood Estimator is efficient, i.e. has the smallest asymptotic variance among all unbiased estimators.

Remark. The Likelihood process has an important property. Indeed, if we define

$$s_n(\theta) = \sum_{i=1}^{n} \partial_\theta \log p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t),$$

then $\{s_n(\theta)\}_{n\in N}$ is a martingale. To see this, we remark that:

$$\begin{aligned} E_\theta[\partial_\theta \log p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)|X_{t_{i-1}}] &= E_\theta\left[\frac{\partial_\theta p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)}{p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)}|X_{t_{i-1}}\right] \\ &= \int_S \frac{\partial_\theta p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)}{p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)} p(X_{t_{i-1}}, dX_{t_i}, \theta, \Delta t) \\ &= \partial_\theta \underbrace{\int_S p(X_{t_{i-1}}, dX_{t_i}, \theta, \Delta t)}_{=1} = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} E_\theta\left[s_n(\theta))|X_{t_{n-1}}\right] &= E_\theta\left[s_{n-1}(\theta))|X_{t_{n-1}}\right] + E_\theta\left[\partial_\theta \log p(X_{t_{n-1}}, X_{t_n}, \theta, \Delta t)|X_{t_{n-1}}\right] \\ &= s_{n-1}(\theta). \end{aligned} \tag{1.2.1}$$

However the closed form for $p(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)$ is available only for a small class of models and, as a result, the Maximum Likelihood approach ca be applied in a small number of cases. Nevertheless, to estimate θ when the transition probability is unknown in closed form, we have several alternatives. In this section we briefly discuss some of them.

1.2.2 Approximation of the Likelihood by a Gaussian transition density

We can use a naive approximation of the Likelihood by performing a crude discretization of the SDE. For instance, we can apply the Euler scheme and assume a conditional Gaussian transition probability. In this case the approximation is two-fold: the true transition probability is not necessarily Gaussian and the conditional moments need not to be those specified by the Euler scheme. As a result the estimated value for θ can be inconsistent (see Florens-Zmirou, [9] for a discussion).

A more recent approach, based on an approximation of the Likelihood,has been given by Aït-Sahalia in [1] and in [2]. Aït-Sahalia uses Hermite polynomials to obtain an approximation of the transition probability of X. To understand this approach the following analogy may be helpful. Assume we are given a standardized sum of random variable (i.i.d.) to which the Central Limit Theorem (CLT) applies. If the sample size (n) is very big, we can approximate n by infinity and use the $N(0,1)$ limiting distribution for the properly standardized transformation of the data. If $n <$, higher order terms of the limiting distribution (for instance the Hermite polynomial in the classical Edgeworth expansion) can be calculated to improve the small sample performance of the approximation.

The basic idea of Aït-Sahalia in [1] and in [2] is to create an analogy between this situation (in i.i.d. setting) and that of approximating the transition density of a diffusion. If we let $1/\Delta t$ play the same role of n in the CLT, is possible to properly standardize the process, then we can find the limiting distribution when $\Delta t \to 0$. In the CLT the standardization is obtained by summation and by dividing by $n^{-\frac{1}{2}}$. In the case of diffusion processes, the standardization involves a transformation of the original process into a new one Z. Aït-Sahalia shows that the transition density of this standardized process is a $N(0,1)$ when $\Delta t = 0$. Moreover Aït-Sahalia shows that, if Δt is not zero, is possible to define an approximation for the transition density, where the leading term is a $N(0,1)$ and the higher order terms are obtained by Hermite polynomials. Let us see this procedure in detail. To this end, we identify four steps:

1. Transform the process X into a new process Y,

$$Y := \gamma(X,\theta) = \int^{X} \frac{1}{\sigma(u;\theta)} du,$$

   satisfying the SDE:

$$dY_t = \mu_y(Y_t;\theta)dt + dW_t$$

   We observe that the new process Y has a unitary diffusion coefficient, therefore it is a standardized version of X.

2. For given $\Delta t > 0$ and y_0, define the pseudo normalized increment of Y as

$$Z := \Delta t^{-\frac{1}{2}}(Y - y_0)$$

Aït Sahalia shows [1] that for fixed Δt, Z is close to a $N(0,1)$ variable and that is possible to define a convergent series-expansion for its density p_Z around a $N(0,1)$. In particular, let $p_Y(\Delta t, y|y_0, \theta)$ denote the conditional density of $Y_{t+\Delta t}|Y_t$, and define the density function of Z

$$p_Z(\Delta t, z|y_0, \theta) = p_Y(\Delta t, \Delta t^{\frac{1}{2}} z + y_0|y_0, \theta). \tag{1.2.2}$$

3. Define a truncated Hermite series expansion of the density of Z around the Standard Normal. Once we have obtained an approximation for $p_Z(\Delta t, z|y_0, \theta)$, is possible to invert (1.2.2) and obtain:

$$p_Y(\Delta t, y|y_0, \theta) = \Delta t^{-\frac{1}{2}} p_Z(\Delta t, \Delta t^{-\frac{1}{2}}(y - y_0)|y_0, \theta)$$

4. Finally is possible to come back to the original problem, by applying the Jacobian formula for the change of density:

$$p_X(\Delta t, x|x_0, \theta) = \sigma(x, \theta)^{-1} p_Y(\Delta t, \gamma(x, \theta)|\gamma(x_0, \theta), \theta).$$

In this way a series approximation (with Hermite polynomials) for p_X is defined.

This procedure is very computational intensive, because the coefficients of the Hermite series-expansion cannot be calculated explicitly: they must be approximated by a numerical (very complex) technique, which involves the approximation of the solutions for the PDE of the Sturm-Liouville problem for the Infinitesimal Generator.

1.2.3 Indirect Inference and EMM

Both the Indirect Inference (Gourieroux, Monfort and Renault [15], in the following GMR) and the Efficient Method of Moments (Gallant and Tauchen, [10]) are based on an "incorrect" criterion, in the sense that the optimization of this criterion does not provide directly an unbiased and efficient estimator for the parameter of interest. In fact, the aim of those methods is to draw a correct inference about θ starting from an incorrect auxiliary criterion.

Indirect Inference. Assume we are given the following dynamic model:

$$X_t = r(X_{t-1}, \theta),$$

with $\theta \in \mathbb{R}^p$ and the function $r : S \times \Theta \to S$ describes a very complicated relationship between the value of the process at the time t and the value at the time $t-1$. The basic idea is three-fold:

Step one. We introduce a simpler parametric model (the auxiliary model), depending on a parameter $\beta \in \mathbb{R}^q$ (with $q \geq p$). The parameter β is estimated by maximizing a criterion

$$\max_{\beta \in \mathbb{R}^q} Q_T(X_{t_1}, ..., X_{t_T}, \beta)$$

and we label $\hat{\beta}_T$ the solution to this problem:

$$\hat{\beta}_T := arg \max_{\beta \in \mathbb{R}^q} Q_T(X_{t_1}, ..., X_{t_T}, \beta).$$

A fundamental assumption is the following

$$\lim_{T \to \infty} Q_T(X_{t_1}, ..., X_{t_T}, \beta) = Q_\infty(\theta_0, \beta), \, a.s.$$

where $Q_\infty(\cdot;\cdot)$ is a deterministic function. This assumption guarantees that the criterion tends asymptotically to a non stochastic limit, depending both on the true parameter value for the initial model (θ_0) and on the auxiliary parameter (β). Under the additional assumption that $Q_\infty(\cdot;\cdot)$ is continuous and has a unique maximum, we have:

$$\beta_0 = arg \max_{\beta \in \mathbb{R}^q} Q_\infty(\theta_0, \beta).$$

Under additional assumptions, GMR show that $\hat{\beta}_T$ is a consistent estimator for β_0. However, β_0 is unknown because it depends on θ_0. That's why they introduce the so called binding function, a one-to-one function defined by

$$b(\theta) := arg \max_{\beta \in \mathbb{R}^q} Q_\infty(\theta, \beta) \tag{1.2.3}$$

and we obtain

$$\beta_0 = b(\theta_0).$$

From the last equation, we see that the binding function links the value of the parameter of interest to the auxiliary parameter. Moreover, if the binding function in (1.2.3) was known, we could deduce from $\hat{\beta}_T$ a consistent estimator for the parameter of interest θ_0, simply by looking for the solution $\hat{\theta}_T$ such that

$$\hat{\beta}_T = b(\hat{\theta}_T).$$

However, the binding function is difficult to compute in closed form and GMR propose a procedure to estimate $b(\cdot)$.

Step two. In the second step GMR provide an estimation procedure for the binding function, which makes use of simulations of the process. For a given value of θ they simulate H paths $\left[\tilde{X}^h_{t_i}(\theta), i = 1, ..., T\right], h = 1, ..., H$ with starting point x_0. For each h-th path, they consider the maximization problem:

$$\hat{\beta}^h_T(\theta) := arg \max_{\beta \in \mathbb{R}^q} Q_T(\tilde{X}^h_{t_i}, ..., \tilde{X}^h_{t_T}, \beta).$$

When $T \to \infty$, this solution tends to the solution of the limit problem:

$$\max_{\beta \in \mathbb{R}^q} Q_\infty(\theta_0, \beta)$$

that is

$$\lim_{T \to \infty} \hat{\beta}^h_T(\theta) = b(\theta)$$

As a result, $\hat{\beta}^h_T(\theta)$, for every θ, is a consistent estimator for the binding function.

Step three. GMR calibrate the parameter θ in the starting model in order to have

$$\frac{1}{H} \sum_{h=1}^{H} \hat{\beta}^h_T(\theta)$$

"close" to $\hat{\beta}_T(\theta)$. To define how $\frac{1}{H} \sum_{h=1}^{H} \hat{\beta}^h_T(\theta)$ is close to $\hat{\beta}_T(\theta)$, GMR use a particular metrics and state the following proposition :

Proposition. *An indirect inference estimator of θ is defined as the solution $\hat{\theta}^H_T(\Omega)$ of the minimum distance problem:*

$$\min_{\theta \in \Theta} \left[\hat{\beta}_T(\theta) - \frac{1}{H} \sum_{h=1}^{H} \hat{\beta}^h_T(\theta)\right]^{\intercal} \hat{\Omega}_T \left[\hat{\beta}_T(\theta) - \frac{1}{H} \sum_{h=1}^{H} \hat{\beta}^h_T(\theta)\right], \tag{1.2.4}$$

where $\hat{\Omega}_T$ *is a positive definite matrix, converging to a deterministic positive matrix* Ω. *Under some additional assumptions,* $\hat{\theta}_T^H(\Omega)$ *is a consistent and asymptotically normal estimator of* θ_0.

We do not go further into detail, but we remark that in (1.2.4) we have to provide an estimate of Ω. Moreover, the Indirect Inference estimator depends on the number of the simulated paths (H). As a result, this estimation method is computationally intensive and very time-consuming.

Efficient Method of Moments. Gallant and Tauchen (see [10], in the following GT) propose an alternative method. The basic ideas and results are the same as in the Indirect Inference procedure, but GT modify the minimum distance problem that defines the estimator for θ_0.

In particular, they assume that the criterion $Q_T(\tilde{x}_{t_1}^h, ..., \tilde{x}_{t_T}^h, \hat{\beta}_T(\theta))$ is differentiable. Therefore, the problem given in (1.2.4) becomes:

$$\hat{\theta}_T^H(\Sigma) := arg\min_{\theta\in\Theta} \left[\frac{\partial Q_T(\tilde{X}_{t_1}^h, ..., \tilde{X}_{t_T}^h, \hat{\beta}_T(\theta))}{\partial \beta^\intercal}\right]^\intercal \Sigma \left[\frac{\partial Q_T(\tilde{X}_{t_1}^h, ..., \tilde{X}_{t_T}^h, \hat{\beta}_T(\theta))}{\partial \beta^\intercal}\right] \qquad (1.2.5)$$

Remark. Indirect Inference and EMM look very similar. Both use an auxiliary model (generally obtained by a discretization of the dynamics of the process in equation (1.2.3)) and both use Monte Carlo simulations to correct the bias of the auxiliary estimator. However, the EMM avoids the time consuming simulations that we have to do in Step Two of the Indirect Inference. As a matter of fact, even if the binding function is unknown, GT use the differentiability of the criterion (compare (1.2.5) and (1.2.4)) and the injectivity of the binding function, in order to perform only one optimization in θ.

1.2.4 Estimating Functions

A very useful tool to draw inference for diffusion processes has been developed in the nineties by the Danish School in Aarhus and in Copenhagen. The main authors of this theory are: Bibby, Jacobsen and Sørensen. Their idea is intuitive. Given that the Likelihood Score Function is unknown in closed form, we can use for the inference procedure some other functions, that behave like the Likelihood and that have similar features.

Basic Definitions

To introduce the Estimating Functions (in the following EFs), we start by the following definition:

Definition 1.2.1. *An EF for θ is a p-dimensional function of the parameter and the data:*

$$G_n(\theta, X_{t_1}, X_{t_2}, ..., X_{t_n}).$$

For notational convenience, from now on we will suppress the dependence from the data and we will write only $G_n(\theta)$. We find an estimator for θ by solving the p equations of the system

$$G_n(\hat{\theta}_n) = 0.$$

This last system can be expressed in a more explicit way by:

$$G_n(\hat{\theta}_n) = \sum_{i=1}^{n-m+1} \psi(X_{t_1}, ..., X_{t_{i+m-1}}; \hat{\theta}_n) = 0, \tag{1.2.6}$$

for some given $m \in \mathbb{N}$, and some function $\psi : \mathbb{R}^m \times \Theta \to \mathbb{R}^p$. Moreover, if

$$E_{\theta_0}\left[G_n(\theta_0)\right] = 0,$$

the EF is unbiased, or Fisher Consistent.

Equation (1.2.4) is a natural requirement which ensures (under some additional regularity conditions, see Bibby, Jacobsen and Sørensen, [3]) the consistency of the M-estimator solution to equation (1.2.6).

In the class of unbiased EFs, we focus on Martingale Estimating Functions (in the following MEFs).

Definition 1.2.2. *An EF $G_n(\theta)$ satisfying*

$$E_{\theta_0}\left[G_n(\theta_0)|X_{t_1}, X_{t_2}, ..., X_{t_{n-1}}\right] = G_{n-1}(\theta_0), \text{ for } n = 1, 2, ...$$

is called a Martingale Estimating Function.

In order to define a MEF for a diffusion process, we can use a set of functions h_j , such that $h_j : \mathbb{R}^2 \times \Theta \to \mathbb{R}^p$ has zero conditional expected value, for $j = 1, 2, ..., k$. More specifically, we can set:

$$\psi(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t) := \sum_{j=1}^{k} h_j(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t),$$

and define:

$$G_n(\theta) = \sum_{i=1}^{n} \psi(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t), \tag{1.2.7}$$

where $\psi : \mathbb{R}^2 \times \Theta \to \mathbb{R}^p$ is such that $E_{\theta_0}\left[\psi(X_{t_{n-1}}, X_{t_n}, \theta_0, \Delta t)|X_{t_{i-1}}\right] = 0$.

Let us check that $G_n(\theta)$ in equation (1.2.7) is a MEF. To this end, we calculate the expected value of the EF at θ_0:

$$\begin{aligned} E_{\theta_0}\left[G_n(\theta_0)|X_{t_1}, X_{t_2}, ..., X_{t_{n-1}}\right] &= G_{n-1}(\theta_0) + E_{\theta_0}\left(\psi(X_{t_{n-1}}, X_{t_n}, \theta_0, \Delta t)|X_{t_{n-1}}\right) \\ &= G_{n-1}(\theta_0), \end{aligned}$$

therefore $G_n(\theta)$ is a martingale. Moreover,

$$E_{\theta_0}\left[\psi(X_{t_{n-1}}, X_{t_n}, \theta_0, \Delta t)|X_{t_1}, X_{t_2}, ..., X_{t_{n-1}}\right] = E_{\theta_0}\left[\psi(X_{t_{n-1}}, X_{t_n}, \theta_0, \Delta t)|X_{t_{n-1}}\right] = 0$$

and the law of iterated expectations, yields

$$E_{\theta_0}\left[\psi(X_{t_{n-1}}, X_{t_n}, \theta_0, \Delta t)\right] = 0.$$

Therefore, it follows that

$$E_{\theta_0}\left[G_n(\theta_0)\right] = 0,$$

and the MEF in equation (1.2.7) is Fisher Consistent.

Equation (1.2.1) shows that the Likelihood Score Function is a particular case of a MEF.

MEFs are useful tools that we can use when we are not able to calculate analytically $s_n(\theta)$ and when we need to replace it by a simpler Estimating Function, sharing the same features of Likelihood. However, when we work with MEFs the main problem is how to construct them. We discuss about this topic in the next paragraph.

Construction of MEFs for diffusions

From equation (1.2.7), we observe that it is possible to build a MEF by using h_j functions separately. Kessler and Sørensen in [28] show that, for efficiency reasons, it is better define a MEF as a linear combination of those functions, using as weights the functions $a_j(X_{t_{i-1}}, \theta)$. In particular, we can define a class of MEFs of the form:

$$\psi(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t) := \sum_{j=1}^{k} a_j(X_{t_{i-1}}, \theta) h_j(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)$$

where, for every j, $a_j(X_{t_{i-1}}, \theta)$ is a weighting function (measurable w.r.t. the σ-algebra generated by the observations up the time t_{i-1}) that determines the weight given to each $h_j(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t)$.[1] Quoting Bibby, Jacobsen and Sørensen in [3], we point out that:

> "The choice of the functions h_j is an art more than a science",

but in the next pages we will give some methods that we can use in order to define a MEF.

When looking at the definition of ψ, a simple type of MEF can be defined for $k = 1$, with

$$h_1(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t) := X_{t_i} - F\left(X_{t_{i-1}}, \theta, \Delta t\right), \tag{1.2.8}$$

where $F\left(x_{t_{i-1}}, \theta, \Delta t\right) := \int_S x_{t_i} p(\Delta t, x_{t_{i-1}}, x_{t_i}, \theta) dx_{t_i}$. In this case we obtain a *linear MEF* of the form

$$G_n(\theta) = \sum_{i=1}^{n} a_1(X_{t_{i-1}}, \theta) h_1(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t).$$

A slightly more complex MEF can be defined for $k = 2$, with h_1 as in (1.2.8) and

$$h_2 := \left(X_{t_i} - F\left(X_{t_{i-1}}, \theta, \Delta t\right)\right)^2 - \upsilon(X_{t_{i-1}}, \theta, \Delta t),$$

where

$$\upsilon(X_{t_{i-1}}, \theta, \Delta t) := Var_\theta(X_{t_i}|X_{t_{i-1}}) = \int_S \left(x_{t_i} - F\left(x_{t_{i-1}}, \theta, \Delta t\right)\right)^2 p(\Delta t, x_{t_{i-1}}, x_{t_i}, \theta) dx_{t_i}.$$

In this case the MEF is called *quadratic MEF* and is given by formula:

[1] In Chapter 3,we discuss in detail how one can select the weights $a_j(X_{t_{i-1}}, \theta)$.

$$G_n(\theta) = \sum_{i=1}^{n} \left[a_1(X_{t_{i-1}}, \theta) h_1(X_{t_{i-1}}, X_{t_i}; \theta, \Delta t) + a_2(X_{t_{i-1}}, \theta) h_2(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t) \right].$$

Another way to build a MEF is by using the Infinitesimal Generator of the process. Let use introduce the differential operator L_θ:

$$L_\theta f(x, \theta) := \alpha(X(t), \theta) \frac{d}{dx} f(x, \theta) + \sigma(X(t), \theta) \frac{d^2}{dx^2} f(x, \theta)$$

for twice continuously differentiable functions on (l, r), $f : \mathbb{R}^k \times \Theta \to \mathbb{R}^k$. Restricting the domain of L_θ to the class of bounded, twice continuously differentiable functions having bounded derivatives, the differential infinitesimal operator L_θ is called the *Infinitesimal Generator* of the process[2].

Now, let be $\phi_j : S \times \Theta \to \mathbb{R}^p$ a twice continuously differentiable function and λ_j a scalar, satisfying

$$L_\theta \phi_j(x, \theta) = -\lambda_j(\theta) \phi_j(x, \theta) \tag{1.2.9}$$

for every $x \in S$. In their formula (2.4), Kessler and Sørensen [28] show that

$$E_\theta \left[\phi_j(X_{t_i}, \theta) | X_{t_{i-1}} \right] = \exp(-\lambda_j(\theta) \Delta t) \phi_j(X_{t_{i-1}}, \theta).$$

Therefore we can define a MEF by taking a linear combination of k functions, based on ϕ_j and λ_j, for $j = 1, 2, ..., k$:

$$\psi(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t) := \sum_{j=1}^{k} a_j(X_{t_{i-1}}, \theta) \left\{ \phi_j(X_{t_i}, \theta) - \exp(-\lambda_j(\theta) \Delta t) \phi_j(X_{t_{i-1}}, \theta) \right\}.$$

Some results for general Sturm-Liouville problems

As far as the function ϕ_j and the scalar λ_j are concerned, we provide some definitions and some technical results for Sturm-Liouville problems.

[2]The Infinitesimal Generator can be defined for a broader class of functions, to this end see, for instance, Kessler and Sørensen [28].

Definition 1.2.3. *A twice continuously differentiable function $\phi_j : S \times \Theta \to \mathbb{R}^p$ is called an eigenfunction for L_θ with eigenvalue $\lambda_j(\theta)$, if it satisfies:*

$$L_\theta \phi_j(x,\theta) = -\lambda_j(\theta)\phi_j(x,\theta)$$

Moreover, an eigenvalue problem of the form

$$\frac{d}{dx}\left[p(x)\frac{d\phi_j}{dx}\right] - q(x)\phi_j + \lambda_j \rho(x)\phi_j = 0 \tag{1.2.10}$$

with $\rho(x)$ and $q(x)$ continuous functions, $p(x)$ continuously differentiable and ϕ_j subject to the boundary conditions:

$$c_1\phi_j(l) - c_2\phi_j^{'}(l) = 0 \tag{1.2.11}$$

$$d_1\phi_j(r) - d_2\phi_j^{'}(r) = 0, \tag{1.2.12}$$

is called (regular) Sturm-Liouville problem (in the following S-L).

The following proposition for S-L problem holds

Proposition 1.2.4. *The eigenfunctions corresponding to different eigenvalues are orthogonal w.r.t. the weight function $\rho(x)$.*

Proof. We provide a sketch of the proof.[3] Let λ_j and λ_i, for $i \neq j$, be two distinct eigenvalues with corresponding eigenfunctions ϕ_j and ϕ_i. For the sake of simplicity we consider the case where $c_2 = d_2 = 0$. Then, the S-L problem is

$$\frac{d}{dx}\left[p(x)\frac{d\phi_j}{dx}\right] - q(x)\phi_j + \lambda_j \rho(x)\phi_j = 0$$

[3]For detail see Weinberger in [40].

s.t. the boundary conditions $\phi_j(l) = \phi_j(r) = 0$. Moreover we have

$$\frac{d}{dx}\left[p(x)\frac{d\phi_i}{dx}\right] - q(x)\phi_i + \lambda_i \rho(x)\phi_i = 0.$$

s.t. the boundary conditions $\phi_i(l) = \phi_i(r) = 0$. Multiplying the first equation by ϕ_i, subtracting ϕ_j times the second one, and integrating over the state space, it follows

$$\int_l^r \left(\left[\phi_i(p\phi_j')' - \phi_j(p\phi_i')'\right] + (\lambda_j - \lambda_i)\rho\phi_i\phi_j\right) dx = 0.$$

Since $\phi_i(p\phi_j')' - \phi_j(p\phi_i')' = \left(\phi_j p\phi_i' - \phi_i p\phi_j'\right)'$ and given that both ϕ_i and ϕ_j vanish at the boundaries, the first term in this equation is zero, and we have

$$\int_l^r (\lambda_j - \lambda_i)\rho\phi_i\phi_j dx = 0.$$

Since $\lambda_j \neq \lambda_i$ by assumption, we conclude $\int_l^r \rho\phi_i\phi_j dx = 0$. ■

In addition, a very useful result is the following:

Proposition 1.2.5. *If the following assumptions hold in (1.2.10), (1.2.11) and (1.2.12):*

1. *The functions $p(x)$ and $\rho(x)$ are positive;*

2. *$q(x)$ is non negative for $l \leq x \leq r$ and*

3. *c_1, c_2, d_1, d_2 are non negative,*

then (up to a multiplicative constant) only a single eigenfunction is associated with each eigenvalue. Moreover all the eigenvalues are non-negative.

Proof. We do not show the uniqueness[4] and we focus our attention only on the second part of the Proposition, concerning the sign of the eigenvalues. We first multiply equation

[4] For a proof, see Edward and Penny in [8, p. 661]

(1.2.10) by the corresponding eigenfunction ϕ_j and we integrate over S:

$$\int_l^r \phi_j \left[(p\phi_j^{'})^{'} - q\phi_j + \lambda_i \rho \phi_j\right] dx = 0.$$

Integration by parts yields:

$$0 = \phi_j(p\phi_j^{'})|_l^r - \int_l^r \left(p\phi_j^{'2} + q\phi_j^2 - \lambda_j \rho \phi_j^2\right) dx.$$

Since $\phi_j(l) = \phi_j(r) = 0$ (we remark that this boundary condition implies $c_1 = d_1 = 1$ and $c_2 = d_2 = 0$ in (1.2.11) and (1.2.12)[5].), from the last expression we obtain

$$\lambda_j = \frac{\int_l^r \left(p\phi_j^{'2} + q\phi_j^2\right) dx}{\int_l^r \rho \phi_j^2 dx}.$$

Given that the numerator of this ratio (called Reyleigh quotient) is a sum of squares and given that ρ is positive by assumption, we have that $\lambda_j \geq 0$. ■

Sturm-Liouville problems for diffusion processes. Now we come back to the diffusion setting and show how one can link the last two Propositions with the stochastic processes theory.

Following Karlin and Taylor [27], we write the Infinitesimal Generator of a diffusion process in the so called *Canonical Representation of the Infinitesimal Generator*:

$$L_\theta \phi_j = \frac{1}{2}\left(\frac{1}{m(x)}\right)\frac{d}{dx}\left(\frac{1}{s(x)}\frac{d\phi_j}{dx}\right),$$

where[6]

$$s(x) := \exp\left(-\int_l^x \frac{2\alpha(y,\theta)}{\sigma^2(y,\theta)}dy\right) \text{ and } m(x) := \frac{1}{\sigma^2(x,\theta)s(x,\theta)}$$

Note that both $s(x)$ and $m(x)$ are positive functions. For a diffusion problem with exit boundaries, the S-L problem is given:

$$\frac{1}{2}\left(\frac{1}{m(x)}\right)\frac{d}{dx}\left(\frac{1}{s(x)}\frac{d\phi_j}{dx}\right) = -\lambda_j \phi_j$$

[5]We analyze the case with those values for the coefficients c_0, d_0, c_1 and d_1, because this is the case that often characterize S-L problems for a big class of diffusion processes with entrance, reflecting or natural boundaries (see the next subsection). For a proof with general values of these coefficients, see Weinberger [40] or Edward and Penny in [8]

[6]In Chapter 3, we analyze these functions in detail.

with boundary conditions $\phi_j(l) = \phi_j(r) = 0$. After rearranging the terms, we can rewrite this Partial Differential Equation as a S-L problem:

$$\frac{d}{dx}\left(\frac{1}{s(x)}\frac{d\phi_j}{dx}\right) = -\lambda_j\phi_j 2m(x) = -\tilde{\lambda}_j\phi_j m(x)$$

where $\tilde{\lambda}_j = 2\lambda_j$. This problem satisfies the assumptions of Propositions 1.2.4 and 1.2.5, it then follows :

- From Proposition 1.2.4, we conclude that two eigenfunctions related to two different eigenvalues are orthogonal functions w.r.t. the weighting function $m(x)$;
- From Proposition 1.2.5, we argue that the eigenvalues are all non negative. Therefore the *spectrum* of the operator L_θ (that is the set of all eigenvalues) is a subset of $[0, \infty)$.

Remark. The solutions to S-L problems are strictly related to the boundary conditions: different values of the coefficients in (1.2.11) and (1.2.12) determine different solutions to the same differential equation (1.2.10). However, Karlin and Taylor [27] show that assumptions of Propositions 1.2.4 and 1.2.5 are satisfied even in presence of reflecting boundaries (characterized by $\phi_j'(l) = \phi_j'(r) = 0$), entrance boundaries and natural boundaries.

Finally, it is important to recall that many diffusion models have a discrete spectrum $\Lambda_\theta := \{\lambda_1, \lambda_2, ...,\}$, such that

$$0 \leq \lambda_1 < \lambda_2 < ... < \lambda_n \uparrow \infty$$

For these setting, Kessler and Sørensen [28] propose to use the first k eigenfunctions in order to determine a MEF, because they show (using the Fourier Series Expansion) that the dependence of the process on the past is determined mainly by small eigenvalues (see Kessler and Sørensen [28], eq. (2.4)). When the spectrum is not discrete, it might still be a good approach to use a probability function on the spectrum[7] in order to select k eigenfunctions with associated eigenvalues $0 \leq \lambda_1 < \lambda_2 < ... < \lambda_k$. After this selection procedure we can define a MEF as in the discrete-spectrum case.

[7] A very similar procedure is used in Ito and McKean, see [24].

We conclude this chapter by an illustrative example that shows how to use the Infinitesimal Generator of the process in order to find the eigenfunctions and the eigenvalues for the CIR process.

Example 1.2.6. *Assume we are given the SDE*

$$dX_t = (\beta + \theta X_t)dt + \sigma\sqrt{X_t}dW_t,$$

with initial condition $X_0 = x_0$. *The State Space is* $S = (0, \infty)$. *To ensure the ergodicity of the diffusion process, we impose the following restrictions on the parameters:* $\beta > 0, \theta < 0$ *and* $\sigma > 0$.

The given SDE describes the dynamic of the CIR process, well known in the financial literature because it is used to model the random behavior of the interest rates. The S-L problem for this process is:

$$L_\theta \phi(x, \theta) = -\lambda(\theta)\phi(x, \theta), \text{ for } x \in S,$$

where we have that $\phi(0, \theta) = 0$ *and* ∞ *is a natural boundary (a natural boundary implies only an integrability condition on the eigenfunction:* $\int_S f(x)\phi^2(x, \theta)dx < \infty$, *see Karlin and Taylor [27, p. 333]).*

Given the drift and the diffusion coefficients, the S-L problem becomes:

$$\underbrace{\frac{1}{2}\sigma^2 x\phi^{''}(x, \theta)}_{A} + \underbrace{(\beta + \theta x)\phi^{'}(x, \theta)}_{B} + \underbrace{\lambda(\theta)\phi(x, \theta)}_{C} = 0 \tag{1.2.13}$$

with boundary conditions $\phi(0, \theta) = 0$, $\int_S f(x)\phi^2(x, \theta)dx < \infty$ *and where* $\phi^{'}(x, \theta) := \frac{\partial}{\partial x}\phi(x, \theta)$ *and* $\phi^{''}(x, \theta) := \frac{\partial^2}{\partial x^2}\phi(x, \theta)$.

To find the solution to this problem we can try to guess a function satisfying the differential equation (1.2.13) or we can try to link our PDE to another PDE, with known solution. We follow the second approach and we observe that the Laguerre polynomials $L_n^{(\alpha)}(\xi)$ *satisfy the Laguerre equation*

$$\xi L_n^{(\alpha)''}(\xi) + (\alpha + 1 - \xi) L_n^{(\alpha)'}(\xi) + n L_n^{(\alpha)}(\xi) = 0, \tag{1.2.14}$$

where $L_n^{(\alpha)'}(\xi) := \frac{\partial}{\partial \xi} L_n^{(\alpha)}(\xi)$, $L_n^{(\alpha)''}(\xi) := \frac{\partial^2}{\partial \xi^2} L_n^{(\alpha)}(\xi)$ *and*

$$L_n^{(\alpha)}(\xi) := \sum_{m=0}^{n} (-1)^m \binom{n+\alpha}{n-m} \frac{\xi^m}{m!}. \tag{1.2.15}$$

We remark that (1.2.13) and (1.2.14) look very similar, but in order to find a one-to-one mapping between the solution to Laguerre PDE and our PDE we need to introduce a change of variable, rewriting the (1.2.13) as a function of ξ*, and perform some additional manipulations. Our goal is to show that the Laguerre polynomials are the solution to the S-L problem for the CIR process.*

We start by multiplying (1.2.14) by $(-\theta)$*, obtaining*

$$\underbrace{-\theta \xi L_n^{(\alpha)''}(\xi)}_{A^L} - \underbrace{\theta(\alpha + 1 - \xi) L_n^{(\alpha)'}(\xi)}_{B^L} - \underbrace{\theta n L_n^{(\alpha)}(\xi)}_{C^L} = 0. \tag{1.2.16}$$

We introduce a very simple guess, based on a linear transformation (that is a change of variable) of x*. Our transformation determines a one-to-one mapping from* $x \to \xi$*. As a result, we calculate the partial derivatives involved in the differential equations by means of the chain rule, so that we obtain*

$$\phi'(\xi(x), \theta) = \frac{\partial \phi(\xi, \theta)}{\partial \xi} \frac{\partial \xi}{\partial x}$$

and for the second derivative

$$\phi^{''}(\xi(x),\theta) = \frac{\partial^2\phi(\xi,\theta)}{\partial\xi^2}\frac{\partial\xi^2}{\partial x^2}.$$

Now we observe that, in order to link (1.2.13) to (1.2.16), we must have:

$$\begin{cases} A = A^L \\ B = B^L \\ C = C^L \end{cases}$$

Step 1: $A = A^L$ *implies*

$$-\theta\xi L_n^{(\alpha)''}(\xi) = \frac{1}{2}\sigma^2 x\phi^{''}(x,\theta).$$

We observe that the right and the left side of the last equation are written in terms of ξ *and* x*, respectively.*

In order to define the change of variable, we have to write the last equation in terms of ξ*, obtaining*

$$-\theta\xi L_n^{(\alpha)''}(\xi) = \frac{1}{2}\sigma^2 x\frac{\partial^2\phi(\xi,\theta)}{\partial\xi^2}\frac{\partial\xi^2}{\partial x^2} = \frac{1}{2}\sigma^2 x\frac{\partial^2\phi(\xi,\theta)}{\partial\xi^2}\left(\frac{\partial\xi}{\partial x}\right)^2, \tag{1.2.17}$$

where $x := x(\xi)$ *and it means that we are considering* x *as a function of* ξ*. The (1.2.17) holds if*

$$-2\theta\xi\sigma^{-2}\frac{1}{x} = \left(\frac{\partial\xi}{\partial x}\right)^2.$$

Taking the squared root of the last equation[8]*, we obtain*

$$\left(-2\theta\xi\sigma^{-2}\frac{1}{x}\right)^{1/2} = \frac{\partial\xi}{\partial x},$$

[8] *When we calculate the squared root we exclude the roots that contradict the conditions imposed for the ergodicity of the process.*

that is an ordinary differential equation that we can solve by the method of separation of variables.

In particular

$$\left(-2\theta\sigma^{-2}\frac{1}{x}\right)^{1/2} dx = \xi^{-1/2} d\xi$$

then

$$\zeta_1 + \left(-2\theta\sigma^{-2}\right)^{1/2} x^{1/2} = \xi^{1/2}, \tag{1.2.18}$$

where ζ_1 is an additive constant. We remark that if we want to write the solution of (1.2.13) in terms of $L_n^{(\alpha)}(\xi(x))$, the boundary condition $\phi(0,\theta) = 0$ gives us an additional information. As a matter of fact, looking at Laguerre polynomial formula in equation (1.2.15), we observe that if $L_n^{(\alpha)}(\xi(0))$ must be equal to zero in order to respect the boundary condition in 0. As a result, we have that $x = 0$ implies $\xi = 0$, that is $\zeta_1 = 0$. Then, from (1.2.18), we obtain:

$$x = \xi\left(-2\theta\sigma^{-2}\right)^{-1}. \tag{1.2.19}$$

***Step 2**: $B = B^L$ implies*

$$(-\theta(\alpha+1) + \theta\xi) L_n^{(\alpha)'}(\xi) = (\beta + \theta x)\phi'(x,\theta).$$

As in Step 1, we write the left member of the last equality in terms of ξ and we have

$$\begin{aligned}
(-\theta(\alpha+1) + \theta\xi) &= (\beta + \theta x(\xi))\frac{\partial\xi}{\partial x} \\
&= \left(\beta + \theta\xi\left(-2\theta\sigma^{-2}\right)^{-1}\right)\frac{\partial\xi}{\partial x} \\
&= \left(\beta + \theta\xi\left(-2\theta\sigma^{-2}\right)^{-1}\right)\left(-2\theta\sigma^{-2}\right),
\end{aligned}$$

implying

$$-\theta(\alpha+1) = \beta\left(-2\theta\sigma^{-2}\right)$$

and then

$$\alpha = 2\beta\sigma^{-2} - 1.$$

Step 3: $C = C^L$ *implies*

$$-\theta n L_n^{(\alpha)}(\xi) = \lambda(\theta)\phi(x,\theta)$$

and by the change of variable in equation (1.2.19), we obtain

$$\lambda_n(\theta) = -\theta n.$$

for $n \in \mathbb{N}$.

Conclusion. *Finally, we can conclude that the spectrum for the S-L problem in (1.2.13) is the set*

$$\Lambda_\theta = \{-n\theta, n = 0, 1, ...\}$$

with corresponding eigenfunctions:

$$\phi(x,\theta) = L_n^{(\alpha)}(-2x\theta\sigma^{-2}),$$

where α *is given in (1.2.6) and* $L_n^{(\alpha)}(\cdot)$ *is the Laguerre polynomial of* n*-th order. We observe that for the CIR process the spectrum is discrete, since the eigenvalues are indexed by the set of natural number.*

Chapter 2

Overview about Robust Statistics

In this chapter we give a short introduction to the theory of Robust Statistics and we provide an overview about the role of the Influence Curve.[1] For details, we refer to book of Hampel et al. [17].

2.1 Robust Statistics

In 1974, Hampel [16] has underlined, for the first time, the central role of the Influence Function in Robust Statistics. Essentially, he has studied the derivative of an estimator, viewed as a statistical functional and he has proposed a way how this derivative can be used to analyze some local robustness properties. In particular, Hampel has shown that the first order Von Mises' expansion is very useful to understand the behavior of an estimator in presence of misspecification models, and/or in presence of contaminations in the data. More precisely, Hampel has remarked that it may be the case (and often turns out to be the case) that classical optimal inference procedures behave quite poorly already under slight violations of the strict model assumptions. This is a key point in Robust Statistics, and Hampel [16] has shown that, in this context, the study of the

[1] Following Hampel et al. in [17], throughout the Thesis we will use the word Influence Function, instead of Influence Curve.

Influence Function is important for two reasons. On one hand, the Influence Function can be a useful tool to deepen our understanding of some properties of classical estimators. On the other hand, the Influence Curve is necessary to derive new inference procedures, having pre-specified robustness features.

To make it clear, we highlight that Robust Statistics focuses on a neighborhood of a parametric model (see for instance, Hampel at al., [17]). More specifically, given the assumptions that define a classical parametric setting, a robust inference procedure determines a probability neighborhood of the reference parametric model. This neighborhood contains the distributions that are close[2] to reference model's distribution and it defines a broader class of models, having small departures from reference model's assumptions.

In Robust Statistics, the Influence Function is the mathematical tool that we need in order to understand the behavior of an estimators in this neighborhood. Moreover, working with the Influence Function, is possible to construct new estimators that have a good performance for all the distributions in the probability neighborhood.

We will study in detail this topic in Chapter 3, when we will provide a mathematical definition of the probability neighborhood. In this brief introduction, we only do some introductory remarks.

First, we point out that the robust approach is different from classical parametric approach in Statistics. As a matter of fact, by defining a probability neighborhood of a reference distribution, we weaken the strict assumptions given by the reference model. In particular, Robust Statistics assume that the data can have small departures from the reference distribution and robust estimators work for a broader class of models.

Nevertheless, we point out that the aforementioned robust approach is different also from Non-Parametric Statistics. As a matter of fact, in robust inference procedures, we do not work with the whole class of probability distributions, but we fix a reference model. In a robust analysis, we allow the data to have small departures from this reference model and we study the performance of the estimator restricting our attention to a specified class of models, close to a fixed point in the space of probability measures.

Finally, we provide some qualitative remarks.

The main feature of robust estimators is that they have a better performance than

[2]In distributional sense, i.e. typically by means of Kolmogorov distance, see [17].

classical estimators, when the data are contaminated and/or in case of misspecification.

Heuristically, we could define the robust estimators as a sort of “hedging” (or an insurance) for the uncertainty that characterizes the real world. The cost for this better robustness properties (that is the cost of the hedging) is a loss of efficiency at the reference model. As a matter of fact, it is well known that, at the reference model, robust estimators have a larger variance than classical estimators. This feature implies a well-known trade-off between robustness and efficiency (see, for instance, Hampel et al [17]).

2.2 The role of the Influence Function

2.2.1 The i.i.d. case

In this subsection, we introduce the basic definition of Influence Function and we provide a heuristic interpretation of its role. For an exhaustive analysis in the i.i.d. case, we refer to Hampel [16], Huber [23] and Hampel et al. [17]. In Section 2.2.2, we introduce some basic ideas for the case of dependent data. Nevertheless we study in detail the definition of Influence Function for the non standard i.i.d. framework, in Chapter 3.

Let us start by the fundamental:

Definition 2.2.1. *Let Ω be a complete separable metric space. Let $T : \mathcal{D} \rightarrow \mathbb{R}^p$ be a vector-valued mapping from (a subset of) the space of probability measures ($\mathcal{D}$) on Ω into the p-dimensional Euclidean space ($\mathbb{R}^p$). Let F be in the domain of T. Finally, let δ_ω denote the atomic probability measure concentrated in any given $\omega \in \Omega$. The vector-valued Influence Function of T at F is defined pointwise by*

$$IF(\omega, T, F) := \lim_{\varepsilon \downarrow 0} \left(\frac{T\left[(1-\varepsilon)F + \varepsilon\delta_\omega\right] - T[F]}{\varepsilon} \right),$$

if this limit exists for every $\omega \in \Omega$.

From now on, we assume for the sake of simplicity that we are working on random variables (x) on the Real Line.

Using Von Mises' expansion, Hampel [16] shows that the asymptotic bias for a probability measure G close to F can be written as:

$$\begin{aligned} T(G) &\approx T(F) + \int_{\mathbb{R}} IF(x, T, F) d(G - F)(x) \\ &= T(F) + \int_{\mathbb{R}} IF(x, T, F) dG(x), . \end{aligned} \tag{2.2.1}$$

For a Fisher Consistent estimator we have that

$$\int_{\mathbb{R}} IF(x, T, F) dF(x) = 0. \tag{2.2.2}$$

Roughly speaking, the Influence Function gives the infinitesimal standardized effect on the asymptotic bias of a contamination in the reference model (F).[3] This is a very important feature of the Influence Function and it is useful in order to determine the asymptotic properties of the estimator at the reference model. For instance, Hampel has shown that the asymptotic variance can be expressed as:

$$V(T, F) = \int_{\mathbb{R}} IF(x, T, F) IF^{\top}(x, T, F) dF(x). \tag{2.2.3}$$

Moreover, the first order Von Mises' expansion is a useful tool also to study the asymptotic distribution of the estimator (see for instance, Hampel et al. [17]). In Chapter 3 we will analyze in detail the properties of the Influence Function, in this introductory chapter we highlight only three fundamental ideas:

- The asymptotic bias is the key tool that we need in order to study the performance of the estimators in presence of contaminations (and/or mispecification);
- Hampel has shown that the asymptotic bias can be approximated by the first order Von Mises' expansion;
- Given a statistical functional $T(\cdot)$, is possible to calculate the first order Von Mises' expansion by using the Influence Function (see equation (2.2.1)).

[3] If we contaminate F with a contaminating distribution that coincides with the reference model (as we have done in equation (2.2.2)), the influence of this kind of contamination on the asymptotic bias of the estimator must be zero.

Quoting Hampel [16], we interpret the Influence Curve as the functional derivative of $T(\cdot)$, where the derivative:

> "can be described by the infinite set of partial derivatives in the directions of the point masses, i.e., along mixtures of the form $(1-\varepsilon)F + \varepsilon\delta_x$, where δ_x si the point mass 1 at x".

As a result, we conclude by this fundamental remark: the Influence Function is the leading term in Von Mises' series expansion and it provides an approximation for the asymptotic bias of the estimator. Furthermore, the Influence Function can be applied to calculate the asymptotic variance of estimators as in equation (2.2.3).

From functional analysis we know that there exist several alternatives to calculate a functional derivative and to write down the first term of Von Mises' expansion, but here we do not provide any further detail about this topic (see, for instance, Yoshida in [42]). In this introductory chapter we only recall that, in Robust Statistics, the functional derivative of $T(\cdot)$ is calculated by using a generalized definition of derivative. In particular Fréchet, Gateaux or Hadammard differentiability play an important role in the study of Von Mises' expansion (e.g., see Huber [23] or Clarke [5]). By definition, a robust estimator has a bounded Influence Curve, that is it has a bounded Fréchet derivative. This implies that small departures from the reference model have a small bounded impact on the asymptotic bias. Roughly speaking this means that the outliers due to contaminations have a bounded negative impact on the performance of the estimator.

Example 2.2.2. *For illustrative purposes, we notice that in the i.i.d. case there are several famous examples of robust estimators. These estimators can be derived directly from a robustification procedure of their classical counterparts or they can be defined ex novo, in order to have some specific robust properties. Let us think, for instance, to the Trimmed mean, to the Median or to the Winsored mean estimators. The common characteristic of those estimators is that they imply a bounded impact of outliers on their asymptotic bias because they have a bounded Gateaux derivative (see Hampel*

[16]). For instance, assume that we are given a random variable x, having Cumulative Density Function F with continuous unimodal density f, symmetric around zero[4]*. Assume that in order to estimate the location of the distribution, we use an estimator based on the Median. In this framework, we obtain that the Influence Function is: $IF(x) = sign(x)/2f(0)$, where $sign(x)$ expresses the signum of the random variable x and it assumes only two values (± 1). We remark that the estimator has a bounded Influence Function, so it has a bounded asymptotic bias. As a result, this estimator performs well even in presence of some outliers that are far away from the bulk of the observations.*

2.2.2 Dependent data

The main part of Robust Statistics has been based on the hypothesis of i.i.d. data and only in the last 20 years, Statisticians and Econometricians have focused on the non i.i.d. case. Some important results in this direction are due to Künsch [30]. In his seminal paper he has shown that the Influence Function is a very important tool, also in the case of dependent data.

According to Künsch's theorems and following Hampel's theoretical approach, the importance of the Influence Function is two-fold:

- It can be applied to understand the behavior of the estimator in presence of local departures from the reference model;
- It is a key instrument to define robust estimators.

We discuss in detail Künsch's theory in Chapter 3. In this subsection we only give a general overview, quoting the main results in Robust Statistics for dependent data.

[4] *This example can be found in Hampel [16, p. 385].*

We start by observing that, during the last decade, Statisticians have developed and implemented robust estimators in a time series setting mainly thanks to Künsch's theorems. As a matter of fact, Künsch has defined the fundamental mathematical tool that can be used to construct robust inference procedures for stochastic processes with a dependence structure between the observations. For instance, the widely applied robust version of GMM [36] and the robust version of EMM [35] are both based on Künsch definition of Influence Function. As far as diffusion processes are concerned, one of the most applied tools is the Robust Indirect Inference, recently developed by Genton and Ronchetti [12]. Also this statistical procedure is based on Künsch results.

As in the i.i.d. case, the starting point of Robust Statistics for time series is the remark that classical estimators (for instance EMM, GMM or Pseudo Maximum Likelihood) typically have an unbounded Influence Function. This feature implies that small departures from the reference model can determine dramatic changes in the performance of the estimators. In particular, the asymptotic bias can become very large. For instance, if we use a (Pseudo) Maximum Likelihood score function, we may observe that a small contamination in the data implies strongly biased (with a big variance) estimated values (see, for instance, Mancini, Ronchetti e Trojani's, [33]).

Robust Statistics proposes a solution to this problem. The main idea of the robust versions of classical inference procedures in a time series setting is to define estimators that have a bounded Influence Function. Thanks to this property, robust estimators have a small bias in presence of contaminations and/or misspecification model. For instance, Robust GMM, Robust EMM or Robust Indirect Inference provide stable estimated values, having a small asymptotic bias in presence of contaminations.

It is worth to notice that these inference procedures for time series are very general and can be applied also in case of diffusion processes. Nevertheless, we observe that RGMM, REMM or Robust Indirect Inference do not take into account the specific features of diffusions and because of their generality, these methods are inefficient and very slow. For instance, both RGMM or REMM need to perform internal Monte Carlo simulations, which are a very time consuming and computationally demanding task.

At the time of writing, the only specific results for diffusion processes are due to Nakahiro Yoshida [41]. In his pioneering paper, he follows Künsch's approach and he defines robust M-estimators for diffusion processes. However, Yoshida's M-estimators can not be applied in practice. As a matter of fact, the main drawbacks of Yoshida's approach are:

- The assumption that the Statistician has access to a continuum of observations;
- The assumption that the Statistician defines an estimating function using the Likelihood Score Function.

The first assumption is too strong, because in most applications we have only discrete time observations of the continuous time process. Even if we assume that we can observe the process in continuous time, the second assumption still implies a problem. Yoshida's Estimating Function is based on a robust version of the Likelihood Score Function, that in our setting is unknown in closed form. [5] Moreover, Yoshida works with models obtained for small departures from the invariant measure. This is a strong simplification, since this robust estimation procedure does not take into account the memory structure of the process.

Yoshida is aware of these problems and he leaves an open question in his conclusions:

> " Theoretically, diffusion coefficients can be calculated without error. But practically, it may be necessary to estimate them. It would be an interesting problem to investigate the effect of discretization of continuous observations on estimations of diffusion coefficients and to seek the robust procedures".

We achieve this goal in the next Chapter. In our development, we follow Künsch's theory and we define a particular version of the Influence Function for M-estimators: the Conditional Influence Function. We show how to use this mathematical tool, in order to define Robust Conditionally Unbiased M-estimators for discretely observed diffusion processes.

In this Chapter, we do not give any further detail, but we remark that the main difference between our approach and Yoshida's theory is related to two important features. More precisely, in Chapter 3, we define two robust inference procedures, having two main properties. First, our procedure is based on Martingale Estimating Functions, that take into account the sampling frequency (the lag Δt between the observations).

[5] See the Introduction and Chapter 1, for a discussion.

In this way, we remove the first Yoshida's unrealistic assumption. Second, we will define Conditionally Unbiased M-estimators that focus on the Markovian structure of diffusion processes. In particular, our inference procedures is based on estimating functions related to the specific memory structure of the stochastic process we are working on. In this way we remove second Yoshida's assumption as well.

Chapter 3

Robust Inference for Discretely Observed Diffusions

This is the main chapter of the Thesis. In the the first section, we introduce Künsch's theory of M-estimators for time series. In Subsection 3.1.2, we define our specific framework for diffusion process.

Subsection 3.2.1 presents a brief overview of Mancini, Ronchetti, Trojani's Conditionally unbiased M-Estimators (related to Künsch's theory) for ARCH processes. In subsection 3.2.2, we review Kessler and Sørensen's Martingale Estimating Functions theory for discretely observed diffusions. Both Kessler and Sørensen's theory and Mancini, Ronchetti, Trojani's theory define Conditionally Unbiased M-estimators that could be applied to define a robust inference procedure for discretely observed diffusion processes. However, Kessler and Sørensen's MEFs define M-estimators that are not robust. Moreover, also Mancini, Ronchetti, Trojani's approach can not be directly applied in our framework. In Section 3.3, we explain which theoretical problems have to be solved to define Robust Martingale Estimating Functions for discretely observed diffusions. In Section 3.4 and 3.5 we solve these problems. In particular, we extend to our setting of diffusion processes Künsch's Theorem 1.3 and we state a Proposition showing the existence and the uniqueness of the main mathematical tool of our development: the

Conditional Influence Function. In Section 3.5, we provide two proposals, defining optimal Robust Martingale Estimating Functions. In section 3.6, we give an algorithm to calculate our robust M-estimators for the second proposal. Finally in Section 3.7, we analyze the asymptotic properties ($\sqrt{n}$-consistency and asymptotic distribution) of our robust M-estimators.

3.1 The General framework

3.1.1 Local robustness for contaminated stochastic processes in time series

Let $\mathcal{X} = \{X_t\}_{t\geq 0}$ be a real valued, stationary and ergodic stochastic process, defined on a complete probability space $(\Omega, \mathcal{F}, \gamma)$. Following Künsch [30], we focus on M-estimators defined as the implicit solutions of equations:

$$\sum_{j=1}^{n-m+1} \psi(X_j, ..., X_{j+m-1}; \hat{\theta}_n) = 0 \tag{3.1.1}$$

for some given $m \in \mathbb{N}$ and a function $\psi : \mathbb{R}^m \times \Theta \to \mathbb{R}^p$, with $\Theta \subset \mathbb{R}^p$.

Define the m-dimensional marginal distribution γ^m induced by $\mathbf{X}_m := (X_1, X_2, ..., X_m)$. As in Künsch [30], we can generalize the class of estimators in equation (3.1.1) as follows.

Let $x = (x_1, x_2, ..., x_n)$ be n observations generated by the process $\mathcal{X}$,

1. The "empirical m-dimensional marginal distribution" $\rho(x, n)^m$ of $\mathcal{X}$, with $(m < n)$ is defined by:

$$\rho(x, n)^m = \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i,...,x_{i+m-1})}, \text{ with } x_i = x_{i-n} \text{ for } i > n,$$

   where $\delta_{(x_i,...,x_{i+m-1})}$ is the Dirac measure in $(x_i, ..., x_{i+m-1}) \in \mathbb{R}^m$. We use $x_i = x_{i-n}$ as a technical device so that $\rho(x, n)^m \in \mathcal{M}^m_{stat}$, where

$$\mathcal{M}^m_{stat} = \{m\text{-dimensional marginals of ergodic strictly stationary processes}\};$$

2. We consider estimators that can be written as functionals

$$T : dom(T) \subset \mathcal{M}^m_{stat} \to \Theta,$$

evaluated in $\rho(x,n)^m$, for m independent of n:

$$\hat{\theta}_n(x_1, x_2, ..., x_n) = T(\rho(x,n)^m). \quad (3.1.2)$$

M-estimators belong to this class, if we define

$$T(\gamma^m) = \theta \text{ iff } \int \psi(x_1, ..., x_m; \theta) d\gamma^m(x_1, ..., x_m) = 0.$$

Remark 3.1.1. *The value m of process coordinates in ψ depends on the process "historical" structure. For Markov processes, of course, $m = 2$.*

Let $\wp = \{\mu_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$ be some parametric model for γ. Since we are concerned about stochastic processes with a "m memory structure", we can focus on the m-dimensional marginal distributions. In order to define robust estimators, we assume that γ^m belongs to a non-parametric neighborhood of $\mu^m_{\theta_0}$. $\mu^m_{\theta_0}$ will be called in the sequel the "reference model". In particular we assume that $\gamma^m \in U^\eta\left(\mu^m_{\theta_0}\right)$, where

$$U^\eta\left(\mu^m_{\theta_0}\right) := \left\{\mu^m_\eta = (1-\eta)\,\mu^m_{\theta_0} + \eta\nu^m\right\}, \eta \le \varepsilon, \varepsilon \in [0,1], \nu^m \in \mathcal{M}^m_{stat} \quad (3.1.3)$$

The neighborhood defined in equation (3.1.3) is a mathematical way to describe local perturbations of the reference model. Our goal is to construct estimators that are "resistant" to local deviations of γ^m from the given m-dimensional marginal distribution $\mu^m_{\theta_0}$.

Remark 3.1.2. *Consider the space of probability measures on $(\Omega, \Im)$, where Ω is a complete and separable space and $\Im$ is the Borel σ-algebra. The "weak-topology" is the weakest topology such that, for every bounded and continuous function ψ, the functional*

$$\mu \to \int \psi d\mu$$

from the space of probability measures into $\mathbb{R}^p$, is continuous. In our Markovian setting, we are focusing on functionals defined on $\mathcal{M}^m_{stat}$. Denoting by $d_K(\cdot;\cdot)$ the Kolmogorov distance, we obtain that for every $\nu^m \in \mathcal{M}^m_{stat}$ such that $\mu^m_\eta \in U^\eta\left(\mu^m_{\theta_0}\right)$, it follows $d_K(\mu^m_\eta;\mu^m_{\theta_0}) \leq \eta$. Following Huber [23], this implies

$$d_L(\mu^m_\eta;\mu^m_{\theta_0}) \leq d_K(\mu^m_\eta;\mu^m_{\theta_0}) \leq \eta,$$

where $d_L(\cdot;\cdot)$ is the Lèvy distance, which metrizes the weak-topology in the space of probability measures over $\mathbb{R}^m$. Therefore, the reference model $\mu^m_{\theta_0}$ can be interpreted as an "approximation" for the true Data Generating Process (DGP): in the sense that γ^m is near to $\mu^m_{\theta_0}$, in distribution.

Under a contamination model with marginal μ^m_η, the asymptotic bias of a Fisher consistent estimator is:

$$b(\eta) := T(\mu^m_\eta) - \theta_0 = T(\mu^m_\eta) - T(\mu^m_{\theta_0}).$$

Its derivative is

$$b^{'}(\eta) = \lim_{\eta\to 0}\left(\frac{T(\mu^m_\eta) - T(\mu^m_{\theta_0})}{\eta}\right).$$

By Von Mises expansion (see [30] and [33]), the following expression for the linearized (asymptotic) bias holds

$$b(\eta) = \eta\, b^{'}(\eta)\Big|_{\eta=0} + o\left(d_*\left(\mu^m_\eta, \mu^m_{\theta_0}\right)\right), \tag{3.1.4}$$

where $d_*\left(\mu^m_\eta, \mu^m_{\theta_0}\right)$, denotes the Lèvy or Kolmogorov distance.

In addition, if we approximate the arc $\mu^m_\eta \in \mathcal{M}^m_{stat}$ by a Taylor expansion around $\mu^m_{\theta_0}$, we obtain:

$$\mu^m_\eta = \mu^m_{\theta_0} + c\eta(\nu^m - \mu^m_{\theta_0}) + o(\eta) \tag{3.1.5}$$

where

$$c := \lim_{\eta\to 0}\left(\frac{\Pr\{\text{at least one outlier in a patch of length } m\}}{\eta}\right)$$

and $\nu^m \in \mathcal{M}^m_{stat}$ depends both on the distribution of the outliers and $\mu^m_{\theta_0}$.

Remark 3.1.3. *Since $\mathcal{M}^m_{stat}$ is not a linear space, some care is needed when we use the linearization (3.1.5). To this end, Künsch [30] observes that when we have outliers in long patches, it is possible to approximate the arc μ^m_η, for any $\nu^m \in \mathcal{M}^m_{stat}$, by equation (3.1.5), with $c \equiv 1$.*

Using equation (3.1.5), the linearized (asymptotic) bias of the functional T is (see Künsch [30]):

$$\begin{aligned} b'(\eta) &= T'(\theta_0, \nu^m) = \lim_{\eta \to 0} \frac{1}{\eta} \left(T(\mu^m_\eta) - T(\mu^m_{\theta_0}) \right) \qquad (3.1.6) \\ &= \lim_{\eta \to 0} \frac{1}{\eta} \left(T(\mu^m_{\theta_0} + \eta(\nu^m - \mu^m_{\theta_0})) - T(\mu^m_{\theta_0}) \right) \end{aligned}$$

with $\nu^m \in \mathcal{M}^m_{stat}$. Following Künsch ([30]), we define the Influence Function (IF) of T in θ_0, any function $IF_T(\mathbf{X}_m, \theta_0) : \mathbb{R}^m \times \Theta \to \mathbb{R}^p$ such that, for every $\nu^m \in \mathcal{M}^m_{stat}$, it follows:

$$T^{'}(\theta_0, \nu^m) = \int IF_T(\mathbf{x}, \theta_0) \nu^m(d\mathbf{x}), \qquad (3.1.7)$$

where $\mathbf{x} := (x_1, ..., x_m)$. By definition, the IF satisfies:

$$\int IF_T(\mathbf{x}, \theta_0) \mu^m_{\theta_0}(d\mathbf{x}) = 0 \qquad (3.1.8)$$

and equation (3.1.4) becomes, using (3.1.5):

$$b(\eta) = \eta \int IF_T(\mathbf{x}, \theta_0) \left. \frac{\partial \mu^m_\eta(d\mathbf{x})}{\partial \eta} \right|_{\eta=0} + o\left(d_* \left(\mu^m_\eta, \mu^m_{\theta_0} \right) \right).$$

Künsch ([30], page 848) remarks that the last expression for the asymptotic bias holds more generally than in the contamination models (3.1.5). Indeed, if a contamination model is such that

$$\lim_{\eta \to 0} \frac{\mu^m_\eta - \mu^m_{\theta_0}}{\eta}$$

converges weakly to a finite signed measure $\dot{\mu}^m$, then it follows:

$$b^{'}(\eta) = \int IF_T(\mathbf{x}, \theta_0) \dot{\mu}^m(d\mathbf{x}).$$

Künsch (Theorem 1.2, [30]) shows that the Influence Function is an equivalence class of kernels, each satisfying (3.1.7). Given a kernel $IF_T(\mathbf{x}, \theta_0)$, any other version of the IF is of the form:

$$IF_T(\mathbf{X}_m, \theta_0) + g(X_1, X_2, ..., X_{m-1}, \theta_0) - g(X_2, ..., X_m, \theta_0),$$

where $g : \mathbb{R}^{m-1} \times \Theta \rightarrow \mathbb{R}^p$ is an arbitrary integrable function such that:

$$\int [g(x_1, x_2, ..., x_{m-1}, \theta_0) - g(x_2, ..., x_m, \theta_0)]\, d\nu^m(\mathbf{x}) = 0, for\ all \nu^m \in \mathcal{M}^m_{stat}.$$

Among the different kernel functions, Künsch selects a particular version of the IF, the Conditional Influence Function:

Definition 3.1.4. *The Conditional IF (in the following IF^{cond}) any version of the IF which satisfies:*

$$\int IF^{cond}_T(\mathbf{x}, \theta_0) d\mu_{\theta_0}(x_m | x_1, x_2, ..., x_{m-1}) = 0 \tag{3.1.9}$$

IF^{cond} has a natural heuristic interpretation. It gives the influence of the observation x_m, given past observations $x_1, x_2, ..., x_{m-1}$. The conditional expectation of IF^{cond} for the clean process is zero. Hence IF^{cond} satisfies equation (3.1.8). Finally, Theorem 1.3 in Künsch [30] shows that IF^{cond} exists and is unique, when $\mu^m_{\theta_0}$ is the m-dimensional marginal distribution of a linear autoregressive process.

3.1.2 Diffusion processes

In this section, we define the reference model in our setting of diffusion processes. Consider a SDE:

$$\begin{cases} dX(t) = \alpha(X(t); \theta)dt + \sigma(X(t); \theta)dW(t) \\ X(0) = x_0. \end{cases} \tag{3.1.10}$$

We assume that $\alpha(X(t);\theta)$ and $\sigma(X(t);\theta)$ satisfy the conditions for the existence of a weak solution to the SDE (3.1.10)[1]. The solution process $\{X_t\}_{t\geq 0}$ takes values in the state space $S = (l,r)^2 \subset \mathbb{R}^k$, $\theta \in \Theta$ and Θ is open set of $\mathbb{R}^p$.

We now introduce some tools that we will be useful in our future development. We assume that there exists the Scale Function (or Scale Measure) $S(x,\theta)$, an absolutely continuous measure w.r.t. Lebesgue measure [3], with density

$$s(x,\theta) := \exp\left(-\int_0^x \frac{2\alpha(y,\theta)}{\sigma^2(y,\theta)}dy\right).$$

Furthermore, we assume the existence of the Speed Measure $M(x,\theta)$, an absolutely continuous measure w.r.t. Lebesgue measure, whose density is:

$$m(x,\theta) = \frac{1}{\sigma^2(x,\theta)s(x,\theta)}.$$

The following standard conditions on $s(x,\theta)$ and $m(x,\theta)$ are assumed in order to ensure the existence of an invariant measure for the diffusion process:

Condition 3.1.5. *If for every $\theta \in \Theta$ and $x \in S$*

$$\int_{x^\#}^r s(x)dx = \int_l^{x^\#} s(x)dx = \infty$$

and

$$\int_l^r \left[s(x,\theta)\sigma^2(x,\theta)\right]^{-1} dx = A(\theta) < \infty,$$

then $\{X(t)\}_{t\geq 0}$ is strictly stationary, with invariant measure having density

$$\mu(x;\theta) = \left[A(\theta)s(x,\theta)\sigma^2(x,\theta)\right]^{-1} = m(x,\theta)A(\theta)^{-1},$$

We focus on regular and time-homogenous diffusion processes sampled at discrete points in time, with fix lag between the observations:

$$\Delta t := t_i - t_{i-1} > 0.$$

[1]We will provide more detail about the existence and uniqueness of a weak solution in the first statement of Condition 3.1.6.

[2]We could have $-\infty \leq l \leq r \leq \infty$.

[3]We can generalize this sentence, using any dominating measure, see for instance Sørensen in [38]

We indicate the "discrete time" observation of the process at time t_i by X_{t_i}. Without loss of generality we set $t_0 = 0$.

Let $C(S)$ denote the class of bounded and continuous functions on S, such that $\lim_{x\to l} f(x)$ and $\lim_{x\to r} f(x)$ exist. $C(S)$ is a linear subspace of $L_2(\mu(x,\theta))$ (see for instance Yosida in [42]). We denote by

$$P(\Delta t, y, x, \theta) := \Pr\left\{X_{t_i} \le y | X_{t_{i-1}} = x\right\}$$

the transition distribution function of X_{t_i} given $X_{t_{i-1}} = x$. We assume that $P(\Delta t, y, x, \theta)$ is differentiable over S, with derivative:

$$p(\Delta t, y, x, \theta) := \frac{\partial P(\Delta t, y, x, \theta)}{\partial y}.$$

We define a family $\{T_t\}_{t\ge 0}$ of conditional expectation operators over $C(S)$ by:

$$T_t f(x) := E_x[f(X_t)] := E[f(X_t)|X_0 = x] \tag{3.1.11}$$

for $t > 0$ and $x \in S$.

The conditional expectation in the last equation is given by the formula:

$$E_x[f(X_t)] := \int_S f(y)p(t, y, x, \theta)dy \tag{3.1.12}$$

and the operator is a contraction semi-group in our specified Hilbert space, with the L_2-norm[4].

Additional conditions on the drift and diffusion coefficients are needed in order to ensure the ergodicity. We use results from Jacobsen, Bibby and Sørensen [3] and Genon-Catalot, Jeantheau and Laredo [11] to obtain this. At the same time, these conditions provide useful information about the spectrum of the operator defined in (3.1.12).

Condition 3.1.6. *I) The function $\alpha(x;\theta)$ is continuously differentiable w.r.t. x and $\sigma(x;\theta)$ is twice continuously differentiable w.r.t. x, $\sigma(x;\theta) > 0$ for all $x \in S$ and there exists a constant K_θ such that*

$$|\alpha(x;\theta)| \le K_\theta(1 + |x|)$$

[4] See for instance, Itô and McKean in [24]

and

$$\sigma^2(x;\theta) \leq K_\theta \left(1 + x^2\right),$$

for all $x \in S$.

II) $\sigma(x;\theta)\mu(x;\theta) \to 0$ *as* $x \downarrow l$ *and* $x \uparrow r$.

III)$\frac{1}{\rho(x,\theta)}$ *has a finite limit as* $x \downarrow l$ *and* $x \uparrow r$, *where*

$$\rho\left(x,\theta\right) := \partial_x \sigma(x;\theta) - 2\frac{\alpha(x;\theta)}{\sigma(x;\theta)}.$$

Genon-Catalot, Jeantheau and Laredo [11] show that Condition 3.1.6 implies that the process $\{X_t\}_{t\geq 0}$ is time reversible. Moreover both the continuous process $\{X_t\}_{t\geq 0}$ and the Markov chain $(X_{t_i})_{i\in N}$ are ergodic. Furthermore, is possible to show (see Genon-Catalot, Jeantheau and Laredo [11] for a proof) that under Condition 3.1.6, the two following statements are equivalent:

1. The Markov chain $(X_{t_i})_{i\in\mathbb{N}}$ is ergodic;
2. 1 is a simple eigenvalue of the conditional expectation operator

$$T_{t_i} := E[f(X_{t_i})|X_{t_{i-1}} = x]$$

(this means that the space

$$\{h \in L_2\left(\mu(x,\theta)\right) : T_{t_i}h = h\}$$

is the one-dimensional space spanned by the constant functions).

The last statement implies that from the ergodicity follows that the spectrum (Λ) of the Infinitesimal Generator of the process contains an eigenvalue strictly larger than zero:

$$\lambda_1 := \inf\left\{\lambda \in \Lambda | \lambda > 0\right\}.$$

The gap between zero and λ_1 is called the "spectral gap". The second statement in Theorem 2.3 in Genon-Catalot, Jeantheau and Laredo [11] implies that for any $h \in L_2\left(\mu(x,\theta)\right)$, such that $\int_S h(x)\mu(x,\theta)dx = 0$ we have:

$$\|T_{t_i}h\|_2^\mu \leq \exp(-\lambda_1\Delta t)\,\|h\|_2^\mu,$$

where $\|\cdot\|_2^\mu$ is the L_2-norm. In this case the conditional expectation operator is a strong contraction (see Hansen and Scheinkeman, [19]) and the resolvent $(I - T_{t_i})^{-1}$ is well defined:

$$(I - T_{t_i})^{-1}h(x) = \sum_{k=0}^{\infty} T_{t_i}^k h(x).$$

The series converges in $L_2(\mu(\theta, x))$ (see, for instance, Bibby, Jacobsen and Sørensen [3]). Finally we remark that, under Condition 3.1.6, every stationary scalar diffusion is time reversible (e.g. see Hansen and Scheinkman [19], page 776 and Genon-Catalot et al. [11]) and its Infinitesimal Generator is self-adjoint. Moreover, in our development we will work with diffusion processes having a discrete spectrum. It is worth to notice that, when the spectrum of the Infinitesimal Generator is discrete, then there exists a spectral gap and a spectral representation for the conditional expectation operator T_{t_i} (see Hansen and Scheinkman, [20], page 22).

3.2 Conditionally Unbiased M-estimators

In order to introduce our robust inference procedure for discretely sampled diffusions, we first briefly review Mancini-Ronchetti-Trojani's (in the following M-R-T) approach for robust conditionally unbiased M-estimators and Kessler and Sørensen's Martingale Estimating Functions theory.

3.2.1 M-R-T: robust Pseudo ML for ARCH processes

M-R-T ([33]) define an inference procedure to construct Conditionally Unbiased Robust M-estimators . In the specific case of a Markov process, they define an estimator such that:

1. $\hat{\theta}_n$ is the solution of the equation[5]:

$$\sum_{i=1}^{n} \psi\left(X_{t_i}, X_{t_{i-1}}; \hat{\theta}_n\right) = 0,$$

[5]This is equation (3.1.1), with $m = 2$.

2. Estimating function ψ is such that

$$E_{\theta_0}\left[\psi\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right) | \mathcal{F}_{t_{i-1}}\right] = 0 \tag{3.2.1}$$

where $\mathcal{F}_{t_{i-1}}$is the augmentation of $\mathcal{G}_{t_{i-1}} = \sigma\left(X_{t_0}, X_{t_1}, ..., X_{t_{n-1}}\right)$, that is the smallest σ-algebra generated by the observations;

3. Estimator $\hat{\theta}_n$ has a bounded Conditional IF.

The most common approach to derive a bounded IF^{cond} estimators is to impose a bound on its "self standardized sensitivity", defined by:

$$\gamma(\psi) := \sup_{\left(X_{t_i}, X_{t_{i-1}}\right)\in(S\times S)} \left\| V\left(\psi; \theta_0\right)^{-1/2} IF^{cond}_{\psi}\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right)\right\| \tag{3.2.2}$$

where

$$V\left(\psi; \theta_0\right) := E_{\theta_0}\left[IF^{cond}_{\psi}\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right) IF^{cond}_{\psi}\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right)^{\intercal}\right].$$

A non robust estimator is such that $\gamma(\psi) = \infty$ and a robust one has $\gamma(\psi) \leq b < \infty$, for some positive constant $b \geq \sqrt{p}$.

M-R-T [33] define a robust estimator for θ_0 that has some optimality properties, in the class of conditionally unbiased M-estimators. They study (G)ARCH processes in a Pseudo Maximum Likelihood framework and they propose optimal robust M-estimators based on estimating functions with the following structure:

$$\psi_b\left(X_{t_i}, X_{t_{i-1}}; \theta\right) := A(\theta)\psi^{bif}\left(X_{t_i}, X_{t_{i-1}}; \theta\right)$$

$$\psi^{bif}\left(X_{t_i}, X_{t_{i-1}}; \theta\right) := \left(s\left(X_{t_i}, X_{t_{i-1}}, \theta\right) - \tau\left(X_{t_{i-1}}, \theta\right)\right)\omega\left(X_{t_i}, X_{t_{i-1}}, \theta\right), \tag{3.2.3}$$

where

$$\omega\left(X_{t_i}, X_{t_{i-1}}, \theta\right) := \min\left(1, b\left\|A(\theta)\left(s\left(X_{t_i}, X_{t_{i-1}}, \theta\right) - \tau\left(X_{t_{i-1}}, \theta\right)\right)\right\|^{-1}\right)$$

and $s\left(X_{t_i}, X_{t_{i-1}}, \theta\right)$ is the Pseudo ML score function for the given parametric model.

The non singular matrix $A(\theta)$, with dimension $p\times p$, and the $\mathcal{F}_{t_{i-1}}$- measurable vector $\tau\left(X_{t_{i-1}},\theta\right)\in\mathbb{R}^p$ are determined by solving the implicit equations:

$$E_{\theta_0}\left[\psi_b\left(X_{t_i},X_{t_{i-1}},\theta_0\right)\psi_b\left(X_{t_i},X_{t_{i-1}},\theta_0\right)^\intercal\right]=I \tag{3.2.4}$$

and

$$E_{\theta_0}\left[\psi_b\left(X_{t_i},X_{t_{i-1}},\theta_0\right)|\mathcal{F}_{t_{i-1}}\right]=0. \tag{3.2.5}$$

Notice that ψ_{bif} is a truncated version of Pseudo Likelihood Score Function and that ψ_{bif} and ψ_b have zero conditional expected value at the reference model. By usual computations (implicit function theorem) for a contamination in the direction $\delta_{\left(X_{t_i},X_{t_{i-1}}\right)}-\mu_{\theta_0}$we get:

$$IF^{cond}\left(X_{t_i},X_{t_{i-1}},\theta_0\right)=-D(\psi_b;\theta_0)^{-1}\psi_b\left(X_{t_i},X_{t_{i-1}},\theta_0\right),$$

where

$$D(\psi_b;\theta_0):=E_{\theta_0}\left[\frac{\partial\psi_b\left(X_{t_i},X_{t_{i-1}},\theta_0\right)}{\partial\theta}\right].$$

Künsch's Theorem 3.1 ensures the existence and the uniqueness of IF^{cond}. Moreover, M-R-T remark that the IF^{cond} is, by definition, the Gateaux derivative of the given statistical functional related to the M-estimators satisfying (3.1.1).

Finally, in their setting, M-R-T show that the truncated Pseudo Likelihood score function satisfies the following optimality criterion:

Proposition 3.2.1. *If for a given constant*[6] $b\geq\sqrt{p}$ *equations (3.2.4) and (3.2.5) have solutions* $A(\theta_0)$ *and* $\tau\left(X_{t_{i-1}},\theta_0\right)$, *respectively, then* ψ_{bif} *minimizes the criterion given by the*

$$tr\left(V\left(\psi,\theta_0\right)V\left(\psi_{bif},\theta_0\right)^{-1}\right),$$

among all ψ *satisfying equation (3.2.1) and such that:*

$$\sup_{\left(X_{t_i},X_{t_{i-1}}\right)\in S^2}\left(IF_\psi^{cond}\left(X_{t_i},X_{t_{i-1}},\theta_0\right)^\intercal V\left(\psi_{bif},\theta_0\right)^{-1}IF_\psi^{cond}\left(X_{t_i},X_{t_{i-1}},\theta_0\right)\right)\leq b^2$$

[6] *The constraint* $b\geq\sqrt{p}$ *is due to Hampel et al. [17], who showed that equations (3.2.4) and (3.2.5) do not have any solution if* $b<\sqrt{p}$.

Therefore, the following corollary holds:

Corollary 3.2.2. *If there exists a conditionally unbiased, efficient score function ψ_{opt} satisfying $\gamma(\psi) \leq b < \infty$, then ψ_{opt} is equivalent to ψ_{bif} whenever the latter is defined.*

Under standard conditions (see Heritier, Ronchetti [21] and Clarke,[5] and [6]), the optimal robust estimator is $\sqrt{n}$-consistent for θ_0 and asymptotically normally distributed at the reference model.

3.2.2 Kessler and Sørensen: MEFs for discretely observed diffusions

We observe that M-R-T's approach is based on the Pseudo Maximum Likelihood theory. We here introduce Kessler and Sørensen's Martingale Estimating Functions (in the following MEFs) that define Conditionally Unbiased M-estimators for discretely sampled diffusion processes.

The main feature of MEFs theory is the definition of Godambe-Heyde [22] optimal Conditionally Unbiased M-estimators. Kessler and Sørensen's optimal M-estimators are $\sqrt{n}$-consistent, optimal, have an asymptotically Normal distribution (see Kessler and Sørensen, [22]) and they do not involve the (Pseudo) Likelihood Score. Thanks to this feature, Kessler and Sørensen MEFs are a useful tool that we can use in order to define Robust Conditionally Unbiased M-estimators for discretely observed diffusion processes.

In the next subsection, we briefly review Kessler and Sørensen's theory and we explain in what sense the optimal MEF provides an approximation of the true Likelihood score function.

The general case

Consider k functions

$$h_j : \mathbb{R}^i \times \Theta \to \mathbb{R}^p, \tag{3.2.6}$$

indexed by $j = 1, 2, ...k$, such that

$$E_{\theta_0}\left[h_j(X_{t_1}, X_{t_2}, ..., X_{t_{i-1}}, X_{t_i}, \theta_0)|\mathcal{F}_{t_{i-1}}\right] = 0. \tag{3.2.7}$$

These functions define relationships between the present observation X_{t_i}, its past history $\mathbf{X}_{t_{i-1}} := (X_{t_1}, X_{t_2}, ..., X_{t_{i-1}})$ and the value of θ_0.

Let $a\left(\mathbf{X}_{t_{i-1}}, \theta\right)$ be a function from $\mathbb{R}^{i-1} \times \Theta$ into the set of $p \times k$-matrices that is differentiable in θ. It then follows that

$$G_n(\theta) = \sum_{i=1}^{n} a\left(\mathbf{X}_{t_{i-1}}, \theta\right) h\left(\mathbf{X}_{t_i}, \theta\right) \tag{3.2.8}$$

defines a p-dimensional MEF. Here $h := (h_1, h_2, ..., h_k)^\top$ is a column vector in $\mathbb{R}^k$.

Consider the class

$$\mathcal{G}_n := \{G_n : G_n \textit{ is a MEF of the form in (3.2.8)}\}.$$

Bibby, Jacobsen and Sørensen [3] show that there exists in $\mathcal{G}_n$ an optimal MEF (i.e. the one with the lowest asymptotic variance of the corresponding M-estimator or, equivalently, with the highest Godambe-Hyde Information matrix). This MEF is of the form:

$$G_n(\theta) = \sum_{i=1}^{n} a^*\left(\mathbf{X}_{t_{i-1}}, \theta\right) h\left(\mathbf{X}_{t_i}, \theta\right) = \sum_{i=1}^{n} \psi_*^{KS}\left(\mathbf{X}_{t_i}, \theta\right),$$

where

$$a^*\left(\mathbf{X}_{t_{i-1}}, \theta\right) := -E_\theta\left(\partial_{\theta^\intercal} h\left(X_{t_1}, ..., X_{t_{i-1}}, X_{t_i}, \theta\right)|\mathcal{F}_{i-1}\right)^\intercal V_h\left(X_{t_1}, ..., X_{t_{i-1}}, \theta\right)^{-1} \tag{3.2.9}$$

and

$$V_h\left(X_{t_1}, ..., X_{t_{i-1}}, \theta\right)^{-1} := E_\theta\left(h\left(X_{t_1}, ..., X_{t_{i-1}}, X_{t_i}, \theta\right) h\left(X_{t_1}, ..., X_{t_{i-1}}, X_{t_i}, \theta\right)^\intercal |\mathcal{F}_{t_{i-1}}\right)^\intercal.$$

The function $a^*\left(\mathbf{X}_{t_{i-1}}, \theta\right)$ gives the optimal weights as in Bibby, Jacobsen and Sørensen [3] and the vector ψ_*^{KS} belongs to $\mathbb{R}^p$.

Remark 3.2.3. *Keeping in mind the Markovian structure of diffusion processes, is possible to define a set of functions depending only from $X_{t_{i-1}}$ and X_{t_i}:*

$$h_j : \mathbb{R}^2 \times \Theta \to \mathbb{R}^p,$$

where $j = 1, 2, ..., k$ and having zero conditional expected value. Then we set

$$G_n^*(\theta) = \sum_{i=1}^{n} a^* \left(X_{t_{i-1}}, \theta \right) h \left(X_{t_{i-1}}, X_{t_i}, \theta \right) = \sum_{i=1}^{n} \psi_*^{KS} \left(X_{t_{i-1}}, X_{t_i}, \theta \right),$$

where a^* is a $p \times k$-matrix such that

$$a^* \left(X_{t_{i-1}}, \theta \right) := -E_\theta \left(\partial_\theta^\intercal h \left(X_{t_{i-1}}, X_{t_i}, \theta \right) | \mathcal{F}_{i-1} \right)^\intercal V_h \left(X_{t_{i-1}}, \theta \right)^{-1}$$

and $h := (h_1, h_2, ..., h_k)^\top$ is a $k \times 1$-vector.

G_n^* defines Conditionally Unbiased M-estimator, having the same structure as M-R-T's estimators in Section 3.2.1.

Let $p(y, \theta | \mathbf{x}_{t_{i-1}})$ be the conditional density of X_i given its history:

$$\mathbf{x}_{t_{i-1}} = (x_{t_1}, ..., x_{t_{i-1}}).$$

The Likelihood function based on n observations $(x_1, x_2, ..., x_n)$ is:

$$L_n(\theta) = \prod_{i=1}^{n} p(x_{t_i}, \theta | \mathbf{x}_{t_{i-1}}).$$

The Likelihood score is

$$s_n(\theta) = \sum_{i=1}^{n} \partial_\theta \ln p(x_{t_i}, \theta | \mathbf{x}_{t_{i-1}}).$$

For given i, $\mathbf{x}_{t_{i-1}}$ and θ, consider now the L_2-space $\Gamma \left(\mathbf{x}_{t_{i-1}}, \theta \right)$ of square-integrable functions $f : \mathbb{R} \to \mathbb{R}$ such that:

$$\int_S f(y)^2 p(y, \theta | \mathbf{x}_{t_{i-1}}) dy < \infty.$$

We equip $\Gamma \left(\mathbf{x}_{t_{i-1}}, \theta \right)$ with the inner product:

$$\langle f, g \rangle := \int_S f(y) g(y) p(y, \theta | \mathbf{x}_{t_{i-1}}) dy$$

and denote by $\Upsilon\left(\mathbf{x}_{t_{i-1}},\theta\right)$ the k-dimensional subspace of $\Gamma\left(\mathbf{x}_{t_{i-1}},\theta\right)$ spanned by the set of functions

$$\left\{h_j\left(\cdot,\mathbf{x}_{t_{i-1}},\theta\right):j=1,2,...k\right\}.$$

We assume $\partial_{\theta_s}\ln p(y,\theta|\mathbf{x}_{t_{i-1}})\in\Gamma\left(\mathbf{x}_{t_{i-1}},\theta\right)$ for every $s=1,2,...,p$. Denote by z_{s_*} the orthogonal projection of $\partial_{\theta_s}\ln p(y,\theta|\mathbf{x}_{t_{i-1}})$ onto $\Upsilon\left(\mathbf{x}_{t_{i-1}},\theta\right)$ and define $z_*:=(z_{1_*},..,z_{p_*})\in\mathbb{R}^p$. Kessler and Sørensen [28] show that:

$$\psi_*^{KS}(y,\mathbf{x}_{t_{i-1}},\theta)=a^*\left(\mathbf{x}_{t_{i-1}};\theta\right)h\left(y,\mathbf{x}_{t_{i-1}};\theta\right)=z_*(y,\mathbf{x}_{t_{i-1}},\theta).$$

The optimal MEF ψ_* is then defined by the weights in a^* and it satisfies the orthogonality condition:

$$\left\langle\partial_{\theta_s}\ln p\left(y,\theta|\mathbf{x}_{t_{i-1}}\right)-\psi_{*_s}^{KS}\left(y,\mathbf{x}_{t_{i-1}},\theta\right),h_j\left(y,\mathbf{x}_{t_{i-1}},\theta\right)\right\rangle=0 \tag{3.2.10}$$

for all $s=1,2,...,p$, $j=1,2,...k$ and for h_j in $\Upsilon\left(\mathbf{x}_{t_{i-1}},\theta\right)$.

In other words, the optimal MEF is the projection of $\partial_\theta\ln p\left(y,\theta|\mathbf{x}_{t_{i-1}}\right)$ onto the sub-space of L_2 spanned by the h_j functions in $\Upsilon\left(\mathbf{x}_{t_{i-1}},\theta\right)$. Therefore, as noted by Bibby, Jacobsen and and Sørensen [3], the optimal MEF is an approximation in L_2 of the Likelihood score function. If the functions h_j are chosen such that for $k\to\infty$ the $\Upsilon\left(\mathbf{x}_{i-1},\theta\right)$ tends to $\Gamma\left(\mathbf{x}_{i-1},\theta\right)$, then, the optimal MEF approaches (for $k\to\infty$) the true Likelihood score function.

Sturm-Liouville problem for the Infinitesimal Generator

In Chapter 1, we have shown that is possible to define the system of h_j functions in (3.2.6) using the Infinitesimal Generator or by the conditional moments of the process. Kessler and Sørensen [28], focus on eigenfunctions and eigenvalues that are solutions to S-L problem with discrete spectrum.[7].

In equation (1.2.9) (see Chapter 1), we have defined $(\lambda_j(\theta);\phi_j(\theta,x))$ as the pair eigenvalue-eigenfunction for the differential operator L_θ. Proposition 3.1 in Kessler and Sørensen [28] shows that the weights $\alpha_j^*(X_{t_{i-1}},\theta)$ as in (3.2.9) defines the Godambe-Heyde optimal

[7] Their approach can be generalized to the case with continuous spectrum, by introducing a specific selection procedure of eigenvalues and eigenfunctions (see Kessler and Sørensen [28])

MEF:

$$G_n^*(\theta) = \sum_{i=1}^{n} \psi_*^{KS}\left(X_{t_i}, X_{t_{i-1}}, \theta, \Delta t\right),$$

with

$$\psi_*^{KS}\left(X_{t_i}, X_{t_{i-1}}, \theta, \Delta t\right) = \sum_{j=1}^{k} \alpha_j^*(X_{t_{i-1}}, \theta)\left\{\phi_j(X_{t_i}, \theta) - \exp[-\lambda_j(\theta)\Delta t]\phi_j(X_{t_{i-1}}, \theta)\right\}.$$

This formula is strictly related to the conditional moments approach in Chapter 1. Indeed, by the Fourier Generalized Series Expansion it follows that:

$$E_\theta\left(\phi_j(X_{t_i}, \theta)\right) = \exp\left[-\lambda_j(\theta)\Delta t\right]\phi_j(X_{t_{i-1}}, \theta). \tag{3.2.11}$$

Therefore we can define a MEF using the system of martingale differences given by:

$$\left\{h_j = \phi_j(X_{t_i}, \theta) - \exp\left[-\lambda_j(\theta)\Delta t\right]\phi_j(X_{t_{i-1}}, \theta)\right], \text{ for } j = 1, ..., k.$$

Moreover, we know that ψ_*^{KS} is the projection of the Likelihood score function $s := \frac{\partial_\theta p(\Delta t, y, x, \theta)}{p(\Delta, y, x, \theta)}$ onto the (closed) sub-space of L_2 generated by the functions h_j. Therefore, the optimal MEF satisfies the conditional moment restrictions:

$$E_\theta\left[\psi\left(s - \psi_*^{KS}\right)\middle| X_{t_{i-1}} = x\right] = 0,$$

for all functions ψ in the given subspace of L_2.

To conclude this section about the review of MEFs theory, we provide a first example of the above constructions. First, we show how we can use the first two conditional moments of the CIR diffusion process to define a Godambe-Heyde optimal M-estimator for the parameters in the diffusion coefficient. Then, we explain how to define an optimal MEF using the eigenfunctions-eigenvalues of the Infinitesimal Generator[8].

Example 3.2.4. *(see Bibby, Jacobsen and Sørensen, [3]). Consider the process solution of the SDE:*

$$dX_t = -\beta(X_t - \alpha)dt + \sigma\sqrt{X_t}dW_t.$$

[8] We remark that for the CIR-process the transition density is known in closed form. We have decided to present this example just to illustrate Kessler and Sørensen's theory.

This is the SDE of the CIR-process, well known in the financial literature. The state space is the positive real line and the process is ergodic for $\beta > 0$, $\alpha > 0$ and $\sigma > 0$. Suppose that σ is known and that we want draw inference about $\theta = (\beta, \alpha) \in \mathbb{R}^2$ from discrete-time observations.

We can define an optimal MEF for the parameter θ, by using the first conditional moment of the process:

$$\phi(x, y, \theta) = y - F(y, x, \theta, \Delta t)$$

where $y = X_{t_i}, x = X_{t_{i-1}}$ and

$$F(y, x, \theta, \Delta t) = E[y|x] = y - x\exp(-\beta\Delta t) - \alpha(1 - \exp(-\beta\Delta t)).$$

The optimal linear estimating function is:

$$\begin{aligned} G_n(\theta) &= \sum_{i=1}^{n} \frac{\partial_\theta F(X_{t_i}, X_{t_{i-1}}, \theta, \Delta t)}{v(X_{t_{i-1}}, \theta, \sigma)} [X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta t) - \alpha(1 - \exp(-\beta\Delta t))] \\ &= \sum_{i=1}^{n} \tilde{\psi}\left(X_{t_i}, X_{t_{i-1}}, \theta, \Delta t\right), \end{aligned}$$

where

$$v(X_{t_{i-1}} = x, \theta, \sigma) = Var(X_{t_i}|X_{t_{i-1}} = x) = \sigma^2 x.$$

As far as the definition of a MEF is concerned, we observe that, generally, the first and the second conditional moments are unknown in closed form and, quite often, we need to compute them by Monte Carlo simulations. Moreover, we remark that, to define an optimal MEF, we need to calculate the derivative w.r.t. θ of the first conditional moment. Those computational aspects can determine some technical problems in the inference procedure.

A possible approach to simplify the calculations, is to use an approximation for $F(X_{t_i} = y, X_{t_{i-1}} = x, \theta, \Delta)$ and for $v(X_{t_{i-1}} = x, \theta, \sigma)$, by formulas in Bibby, Jacobsen and Sørensen [3][9].

For illustration purposes, we apply here those formulas and we obtain:

$$F(y, x, \theta, \Delta t) = x - \Delta t \beta (x - \alpha) + \frac{1}{2}\left[\Delta t^2 \beta^2 (x - \alpha)\right] + O(\Delta t^3)$$

and for the conditional variance:

$$v(X_{t_{i-1}} = x, \theta, \sigma) = \Delta t \sigma^2 x + \Delta t^2 \left\{ -\frac{\beta}{2}(x - \alpha)\sigma^2 - \sigma^2 x \beta \right\} + O(\Delta t^3).$$

For Δt small, we then neglect the terms in Δt^2 and Δt^3 and we obtain

$$\partial_\theta F(X_{t_i} = y, X_{t_{i-1}} = x, \theta, \Delta t) \simeq \Delta t \partial_\theta \left(-\beta(X_t - \alpha)\right)$$

and

$$v(X_{t_{i-1}} = x, \theta, \sigma) \simeq \Delta t \sigma^2 x.$$

The approximately optimal linear estimating function is:

$$G_n(\theta) = \sum_{i=1}^{n} \frac{\partial_\theta((-\beta(X_t - \alpha)))}{v(x, \theta, \sigma)} \phi\left(X_{t_i}, X_{t_{i-1}}, \theta, \Delta t\right),$$

that is

$$G_n(\theta) = \begin{pmatrix} \sum_{i=1}^{n} \frac{\beta}{\sigma^2 X_{t_{i-1}}}[X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta) - \alpha(1 - \exp(-\beta\Delta))] \\ \sum_{i=1}^{n} \frac{\alpha - X_{t_{i-1}}}{\sigma^2}[X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta) - \alpha(1 - \exp(-\beta\Delta))] \end{pmatrix}.$$

Multiplying $G_n(\theta)$ by the invertible matrix

$$B(\theta, \sigma) = \begin{pmatrix} \frac{\sigma^2}{\beta} & 0 \\ \frac{\alpha\sigma^2}{\beta} & -\sigma^2 \end{pmatrix},$$

[9] *See formulas (5.14) and (5.15), in [3].*

we obtain an estimator $\hat{\theta}_n$ for θ by solving the estimating equations:

$$\begin{cases} \sum_{i=1}^{n} \frac{1}{X_{t_{i-1}}}[X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta t) - \alpha(1 - \exp(-\beta\Delta t))] = 0 \\ \sum_{i=1}^{n}[X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta t) - \alpha(1 - \exp(-\beta\Delta t))] = 0. \end{cases} \tag{3.2.12}$$

We can also use an alternative construction and define a MEF by solving the PDE involved in the S-L problem for the Infinitesimal Generator. In the next example, we explain how to use the eigenfunction and the eigenvalue solutions of the S-L problem in order to construct a MEF for the drift coefficient of the CIR-process.

Example 3.2.5. *(see Kessler and Sørensen [28]). In Chapter 1, we have solved the Sturm-Liouville problem for the Infinitesimal Generator of the CIR-process and we have found that the eigenfunctions are the Laguerre Polynomials ($L_n^{(v)}(x)$) of n-order. Therefore, we can define a MEF for the parameter $\theta = (\beta, \alpha) \in \mathbb{R}^2$ using the martingale difference:*

$$\psi(y, x, \theta, \Delta t) := L_1^{(v)}(2\beta\sigma^{-2}y) - \exp(\lambda_1(\beta,\alpha)\Delta t)L_1^{(v)}(2\beta\sigma^{-2}x),$$

where $v := 2\alpha\beta\sigma^{-2} - 1$ and $\lambda_1(\beta, \alpha)$ is the first eigenvalue. Calculating the first derivative of $\psi(y, x, \theta, \Delta t)$ wrt β and wrt α, we obtain the optimal weights and the resulting optimal estimating function.

3.3 Theoretical Issues

The goal of this Thesis is to define Robust Martingale Estimating Functions for discretely sampled diffusions. At a first glance, we observe that, in our setting, we can apply the seminal paper of Florens-Zmirou [9] and we can define a Gaussian PML score function. Therefore, according to M-R-T [33], is possible to find a robust M-estimator for a discretely sampled diffusion, by implementing a robustification procedure of the

Gaussian PML score. Applying this inference procedure, we can argue the optimality of our M-estimator by a direct application of Proposition 3.2.1 and Corollary 3.2.2(see M-R-T, [33]) and we can conclude that the robustified version of the Gaussian PML score defines efficient M-estimators for discretely sampled diffusion processes.

The drawback of this approach is the consistency. It is well known that Florens-Zmirou estimating functions can determine inconsistent estimates, unless some unrealistic assumptions about the sampling frequency are provided. Therefore we conclude that M-R-T's approach is not directly applicable in our setting.

An alternative solution to our inferential problem can be Kessler and Sørenen's approach. As a matter of fact, it is worth to notice that Kessler and Sørensen's theory defines some estimating functions that provide consistent estimates, whatever be Δt. Their procedure is useful to draw a parametric inference about the drift and the diffusion coefficients of discretely sampled diffusion processes, without the assumption of a Gaussian transition probability density and without the unrealistic assumption of a shrinking sampling interval, as in [9].

Nevertheless, we notice that Kessler and Sørenen's MEFs are often unbounded. Consider, for instance, the estimating functions in (3.2.12). In this case $\tilde{\psi}$ diverges toward infinity, when we are close to the boundaries.

From Künsch [30], we know that (one version of) the conditional $IF_{\tilde{\psi}}$, for a direction of contamination $(\delta_{(Z_{t_i},Z_{t_{i-1}})} - \mu^2_{\theta_0})$, is

$$IF_{\tilde{\psi}}(Z_{t_i}, Z_{t_{i-1}}, \theta_0, \Delta t) := \frac{\tilde{\psi}\left(Z_{t_i}, Z_{t_{i-1}}; \theta_0, \Delta t\right)}{-E_{\theta_0}\left[\frac{\partial}{\partial\theta_0}\tilde{\psi}\left(X_{t_i}, X_{t_{i-1}}; \theta_0, \Delta t\right)\right]}.$$

Therefore

$$IF_{\tilde{\psi}}(Z_{t_i}, Z_{t_{i-1}}, \theta_0, \Delta t) \propto \tilde{\psi}\left(Z_{t_i}, Z_{t_{i-1}}; \theta_0, \Delta t\right),$$

up to a multiplicative constant depending on θ_0. Since $\tilde{\psi}$ is unbounded, according to our considerations in Section 3.1.1, we notice that the asymptotic bias of the estimator obtained from $\tilde{\psi}$ could be very large. The last remark leads to a direct consideration involving the robustness of Kessler and Sørensen's inference procedure. Generally, Kessler and Søresen's MEFs define Conditionally Unbiased M-estimators that are not robust in Huber and Hampel's sense. This implies that a little perturbation in the reference model can determine dramatic changes in the performance of their M-estimators, with an asymptotic bias that can be arbitrarily large. Therefore, as we have conclude for M-R-T's theory, also Kessler and Sørensen's approach is not directly applicable for our purposes.

According to this discussion, the goal of our research is the definition of Robust Conditionally Unbiased M-estimators, that do not involve the (Pseudo)Likelihood score and that behave well (in terms of asymptotic bias and with a small cost in terms of asymptotic variance) even in presence of contaminations. We observe that, to achieve this goal, we first need to modify Künsch's Theorem 1.3 (that is specific for AR(p) processes) in order to derive the existence and the uniqueness of IF^{cond} in our setting of diffusion processes. Then, we can use IF^{cond} to define Conditionally Unbiased, $\sqrt{n}$-consistent, optimal Robust M-estimators.

3.4 IF^{cond}: Theorem 1.3 in Künsch for diffusions

In this section we will define a Proposition, that ensures the existence and the uniqueness of the IF^{cond} for robust MEFs. We start by observing that Künsch's Theorem 1.3 in [30] is specific for linear autoregressive processes. As far as diffusion processes are concerned, we observe that Künsch's Theorem ensures the existence and uniqueness of the IF^{cond} for those diffusions, whose solution to the SDE (3.1.10) is an AR-process (for instance the Ornstein-Uhlenbeck process). For this class of processes, we can define Robust Conditionally Unbiased M-estimators by a direct application of Künsch's theory. However, only a small number of diffusion processes have a solution that is an AR-process. To define Robust MEFs for a broader class of diffusions, we prove the following Proposition, which extends Künsch's arguments to the processes satisfying Conditions 3.1.5 and 3.1.6.

Proposition 3.4.1. *Assume we are given a diffusion process with drift and diffusion coefficient satisfying the Conditions 3.1.5 and 3.1.6. Let the joint stationary distribution of* $(X_{t_i}, X_{t_{i-1}})$ *have a density* $P_2^\theta(x_{t_i}, x_{t_{i-1}}, \Delta t)$*:*

$$P_2^\theta(x_{t_i}, x_{t_{i-1}}, \Delta t) := \mu(x_{t_{i-1}}, \theta) p(\Delta t, x_{t_i}, x_{t_{i-1}}, \theta),$$

where $p(\Delta t, x_{t_i}, x_{t_{i-1}}, \theta)$ *is the conditional density of* X_{t_i} *given* $X_{t_{i-1}} = x_{t_{i-1}}$*. Then, for every* $f \in C(S \times S)$ *such that*

$$\int_{S\times S} f(x_{t_i}, x_{t_{i-1}}) P_2^{\theta}(x_{t_i}, x_{t_{i-1}}, \Delta t) d(x_{t_i} \cdot x_{t_{i-1}}) = 0, \tag{3.4.1}$$

there exists a function $g \in C(S)$ *satisfying, for all* $x_{t_{i-1}}$*:*

$$\int_S \left(f(x_{t_i}, x_{t_{i-1}}) + g(x_{t_{i-1}}) - g(x_{t_i})\right) p(\Delta t, x_{t_i}, x_{t_{i-1}}, \theta) dx_{t_i} = 0. \tag{3.4.2}$$

The function g *is unique up to an additive constant.*

Proof. **Uniqueness of** g**.** Assume there exist functions g_1 and g_2 satisfying equation (3.4.2). We show that $g := g_1 - g_2$ is constant. To this end, it is sufficient to prove that $f \equiv 0$ in equation (3.4.2) implies that g is constant. For $f \equiv 0$, equation (3.4.2) becomes:

$$\int_S \left(g(x_{t_{i-1}}) - g(x_{t_i})\right) p(\Delta t, x_{t_i}, x_{t_{i-1}}, \theta) dx_{t_i} = 0. \tag{3.4.3}$$

From this equation, we can argue that $(g(X_{t_i}))_{i\in N}$ is a martingale process[10].

Iterating equation (3.4.3), we obtain:

$$g(X_{t_1}) = E\left\{ \frac{1}{n} \sum_{i=1}^{n} g(X_{t_i}) \middle| X_{t_1} \right\} \tag{3.4.4}$$

Indeed, from equation (3.4.3), we have for $i \geq 2$:

$$g(X_{t_{i-1}}) = E\left(g(X_{t_i}) | X_{t_{i-1}}\right)$$

[10]In realty, the martingale property follows under the weaker assumption:

$$\int_S f(x_{t_i}, x_{t_{i-1}}) p(\Delta t, x_{t_i}, x_{t_{i-1}}, \theta) dx_{t_i} = 0.$$

For $i = 2$ this gives

$$g(X_{t_1}) = E\left(g(X_{t_2})\,|X_{t_1}\right)$$

and for each $g(X_{t_i})$ with $i > 2$, taking conditional expectation given X_{t_1}, we have :

$$\begin{aligned} E\left(g(X_{t_i})\,|X_{t_1}\right) &= E\left(E(g(X_{t_i})|X_{t_{i-1}})\,|X_{t_1}\right) \\ &= E\left(E(g(X_{t_{i-1}})|X_{t_{i-2}})\,|X_{t_1}\right) = ... = g(X_{t_1}). \end{aligned}$$

Now we apply the Ergodic Theorem and the Dominated Convergence Theorem to the right-hand side of (3.4.4). It the follows

$$\begin{aligned} g(X_{t_1}) &= \lim_{n\to\infty} E\left\{ \frac{1}{n}\sum_{i=1}^{n} g(X_{t_i}) \middle| X_{t_1} \right\} = E\left\{ \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g(X_{t_i}) \middle| X_{t_1} \right\} \\ &= E\left\{E\left[g(X_{t_i})\right]|X_{t_1}\right\} = E\left[g(X_{t_1})\right] \text{ a.s.} \end{aligned}$$

Therefore, g is constant. This shows the uniqueness up to an additive constant.

Existence of g. We show that for any f satisfying equation (3.4.1), there exists $g \in C(S)$ satisfying equation (3.4.2). Let

$$J = \left\{ h \in C(S) : \int_S h(x)\mu(x,\theta)dx = 0 \right\}.$$

For simplicity of notation, we rewrite integral equation (3.4.2) as:

$$g - T_{t_i} g = (I - T_{t_i})g = -T_{t_i} f. \tag{3.4.5}$$

Applying iterated expectations, we observe that $\tilde{f} := -T_{t_i} f \in J$. From now on, we consider every conditional expectation operator T_{t_i} as an operator

$$T_{t_i} : J \to J$$

restricted to the class J and that maps J in J. The operator $(I - T_{t_i})$ is the inverse of the resolvent of the semi-group T_{t_i}Given the Condition (3.1.6) (ergodicity), we know that there exists a spectral gap between the eigenvalues of the Infinitesimal Generator. Then we know that the resolvent is well defined, therefore for each function $\tilde{f} \in J$ there exists

$$g := (I - T_{t_i})^{-1} \tilde{f} \in J,$$

satisfying eq. (3.4.5). ■

In the Thesis, we use the last Proposition to ensure the existence and the uniqueness of the IF^{cond} for Robust Conditionally Unbiased M-estimators for discretely observed diffusion processes. To this end, we remark[11] that the IF^{cond} has some desirable properties. First, for Conditionally Unbiased M-estimators, IF^{cond} can be easily computed by the limit

$$IF_T^{cond}(X_{t_i}, X_{t_{i-1}}, \theta_0) = \lim_{\eta \to 0} \frac{T((1-\eta)\,\mu^2_{\theta_0} + \eta \delta_{(x_{t_i}, x_{t_{i-1}})}) - T(\mu^2_{\theta_0})}{\eta},$$

where $\mu^2_{\theta_0}$ is the joint measure of $(X_{t_i}, X_{t_{i-1}})$ and $\delta_{(x_{t_i}, x_{t_{i-1}})}$ is the point mass at $(x_{t_i}, x_{t_{i-1}})$. This implies that for MEFs of the form

$$G_n(\theta) = \sum_{i=1}^{n} \psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right),$$

with $\psi : \mathbb{R}^2 \times \Theta \to \mathbb{R}^p$ bounded and continuous function and such that

$$E_{\theta_0}\left[\psi(X_{t_{i-1}}, X_{t_i}, \theta_0) | \mathcal{F}_{t_{i-1}}\right] = 0, \tag{3.4.6}$$

the IF_ψ^{cond} is

$$IF_\psi^{cond}\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right) = -D(\psi, \theta_0)^{-1} \psi\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right), \tag{3.4.7}$$

where

$$D(\psi, \theta_0) := E_{\theta_0}\left[\frac{\partial \psi\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right)}{\partial \theta}\right].$$

[11] For details, see M-R-T, [33].

Proposition 3.4.1 guarantees the existence and the uniqueness of the IF_{ψ}^{cond}, for this class of MEFs . Therefore (3.4.7) is the unique representation. Another desirable properties of IF^{cond} is related to the expression of the asymptotic variance.

To make it clear, let us introduce the shift-operator $S : \mathbb{R}^2 \to \mathbb{R}^2$ and its iterations

$$S^g\left(X_{t_l}, X_{t_{l-1}}\right) = \left(X_{t_{l+g}}, X_{t_{l-1+g}}\right)$$

for $g = 0, 1, 2....$ The asymptotic variance-covariance matrix of an M-estimator defined by ψ is [12]

$$\begin{aligned} V(\psi, \theta_0) &= E_{\theta_0}\left[IF_{\psi}(X_{t_i}, X_{t_{i-1}}, \theta_0) IF_{\psi}(X_{t_i}, X_{t_{i-1}}, \theta_0)^{\intercal}\right] \\ &\quad + \sum_{g=1}^{\infty} E_{\theta_0}\left[IF_{\psi}(X_{t_i}, X_{t_{i-1}}, \theta_0) IF_{\psi}(S^g(X_{t_i}, X_{t_{i-1}}), \theta_0)^{\intercal}\right. \\ &\quad \left. + IF_{\psi}(S^g\left(X_{t_i}, X_{t_{i-1}}\right), \theta_0) IF_{\psi}(X_{t_i}, X_{t_{i-1}}, \theta_0)^{\intercal}\right]. \end{aligned} \tag{3.4.8}$$

This expression is independent of the particular choice of IF. This means that formula (3.4.8) holds for every element belonging to the class of kernels satisfying (3.1.8). In particular, for Conditionally Unbiased M-estimators, (3.4.8) becomes

$$V(\psi, \theta_0) = E_{\theta_0}\left[IF_{\psi}^{cond}(X_{t_i}, X_{t_{i-1}}, \theta_0) IF_{\psi}^{cond}(X_{t_i}, X_{t_{i-1}}, \theta_0)^{\intercal}\right]. \tag{3.4.9}$$

To obtain the last equation, we can proceed as follow.

Since ψ satisfies the equation (3.4.6), we can argue that, for every $i \neq u$, the functions $\psi\left(X_{t_u}, X_{t_{u-1}}, \theta_0\right)$ and $\psi\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right)$ are uncorrelated.

Consider for instance, the case where $u > i + 1$. Straight calculations lead to:

$$\begin{aligned} & Cov_{\theta_0}(\psi\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right); \psi\left(X_{t_u}, X_{t_{u-1}}, \theta_0\right) \\ = \; & E_{\theta_0}\left[\psi\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right) \psi\left(X_{t_u}, X_{t_{u-1}}, \theta_0\right)^{\intercal}\right] \\ = \; & E_{\theta_0}\left\{E_{\theta_0}\left[\psi\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right) \psi\left(X_{t_u}, X_{t_{u-1}}, \theta_0\right)^{\intercal} | \mathcal{F}_{t_{u-1}}\right]\right\} \\ = \; & 0. \end{aligned}$$

Then we have

$$Cov_{\theta_0}(IF_{\psi}^{cond}\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right); IF_{\psi}^{cond}\left(X_{t_u}, X_{t_{u-1}}, \theta_0\right)) = 0.$$

Using the last formula in equation (3.4.9), equation (3.4.9) follows.

[12]See Künsch, [30]

3.5 Robust MEFs for discretely observed diffusions

To define Robust MEFs, we provide two proposals. The main difference between the first and the second proposal is in the path that we follow to achieve robustness. In the first proposal (see subsection 3.5.1), we robustify a given set of unbounded martingale differences and then define the Godambe-Heyde optimal MEF.

In the second proposal (see subsection 3.5.2), we apply an "opposite" construction. We define the optimal linear combination of a given set of unbounded martingale differences and then robustify it in order to define a bounded (and then robust) MEF.

3.5.1 First proposal: Godambe-Heyde Optimal Robust MEFs

Our first proposal is based on the Godambe-Heyde theory for MEFs and defines robust estimating functions by finding the optimal (i.e. with the minimum variance) linear combination of some bounded martingale differences in a given sub-space of $L_2(p(\Delta t, y, x, \theta))$. Assume for simplicity that $\theta \in \Theta \subset \mathbb{R}$ (the multidimensional case shows the same calculations, but the notation is more complicated). Let, for $j = 1, 2, ...k$,

$$\chi_j : \mathbb{R}^2 \times \Theta \to \mathbb{R}$$

be a function such that:

$$E_\theta \left[\chi_j \left(X_{t_i}, X_{t_{i-1}}, \theta \right) | \mathcal{F}_{t_{i-1}} \right] = 0 \tag{3.5.1}$$

We assume that every function χ_j is a bounded and continuous function, with finite variance and having a bounded sensitivity:

$$\sup_{\left(X_{t_i}, X_{t_{i-1}}\right) \in S \times S} \left| \chi_j \left(X_{t_i}, X_{t_{i-1}}; \theta \right) \right| \leq b_j(\theta). \tag{3.5.2}$$

Define the k-dimensional vector

$$\Gamma \left(X_{t_i}, X_{t_{i-1}}; \theta \right) := \left(\chi_1 \left(X_{t_i}, X_{t_{i-1}}, \theta \right), \chi_2 \left(X_{t_i}, X_{t_{i-1}}, \theta \right), ..., \chi_k \left(X_{t_i}, X_{t_{i-1}}, \theta \right) \right)^\intercal . \tag{3.5.3}$$

Now we consider the class of MEFs:

$$\mathcal{G}_n := \left\{ G_n(\theta) = \sum_{i=1}^{n} \psi \left(X_{t_i}, X_{t_{i-1}}; \theta \right) \right\},$$

where

$$\psi\left(X_{t_i}, X_{t_{i-1}};\theta\right) = \alpha(\theta)\Gamma\left(X_{t_i}, X_{t_{i-1}},\theta\right) \tag{3.5.4}$$

and $\alpha(\theta)$ is a $(1 \times k)$-dimensional vector of weights (differentiable in θ).

In order to define the $\Gamma\left(X_{t_i}, X_{t_{i-1}},\theta\right)$ vector, we can start, for instance, using the eigenfunctions and the eigenvalues of the Infinitesimal Generator (see Kessler and Sørensen, [28]). Typically, the S-L eigenfunctions are unbounded.[13] Therefore, to satisfy equation (3.5.2), we first robustify every function χ_j by using an approach close to that one of M-R-T [33].

To clarify this procedure, we split our construction of a robust MEF into two steps:

1. We solve the S-L problem, calculating the eigenfunctions $\phi_j(x,\theta)$ and the eigenvalues $\lambda_j(\theta)$, for $j = 1, ..., k$. We then define a martingale difference $\varkappa_j(X_{t_i}, X_{t_{i-1}},\theta)$ as follows

$$\kappa_j(X_{t_i}, X_{t_{i-1}},\theta) := \phi_j(X_{t_i},\theta) - \exp[-\lambda_j(\theta)\Delta t]\phi_j(X_{t_{i-1}},\theta); \tag{3.5.5}$$

2. We robustify every function κ_j by following M-R-T 's approach [14]:

$$\chi_j\left(X_{t_i}, X_{t_{i-1}},\theta\right) := \omega_j(X_{t_i}, X_{t_{i-1}},\theta)(\kappa_j(X_{t_i}, X_{t_{i-1}},\theta) - \tau_j(X_{t_{i-1}},\theta)),$$

where

$$\omega_j(X_{t_i}; X_{t_{i-1}},\theta) := \min\left(1; b_j \left\|(\kappa_j\left(X_{t_i}, X_{t_{i-1}},\theta\right) - \tau_j\left(X_{t_{i-1}},\theta\right))\right\|^{-1}\right).$$

and $\tau_j\left(X_{t_{i-1}},\theta\right)$ is such that equation (3.5.1) holds.

To define an optimal MEF we state the following

Proposition 3.5.1. *Let be*

$$\psi_*\left(X_{t_i}, X_{t_{i-1}},\theta\right) = \alpha^*(\theta)\Gamma\left(X_{t_i}, X_{t_{i-1}},\theta\right),$$

[13] This argument remains valid if we define the functions in Γ by using the conditional moments of the process, because, often, conditional moments define polynomial functions.

[14] We provide more details on the role of the functions $\omega_j(X_{t_i}, X_{t_{i-1}},\theta)$ and $\tau_j(X_{t_{i-1}},\theta)$ in the next section.

where the weights

$$\alpha^*(\theta) = -E_\theta \left[\nabla_\theta \Gamma\right]^\top E_\theta \left(\Gamma \Gamma^\intercal\right)^{-1}.$$

Then ψ_* *can be used to construct a MEF*

$$G_n^*(\theta) = \sum_{i=1}^{n} \psi_* \left(X_{t_i}, X_{t_{i-1}}, \theta\right) \tag{3.5.6}$$

in $\mathcal{G}_n$*, defining an M-estimator having the following properties:*

I - It is unbiased and Fisher Consistent;

II - It is F*-optimal in the class* $\mathcal{G}_n$ *and*

III - It has a unique bounded IF^{cond}.

Proof. Our problem can be defined as in Künsch [30]. In particular we want to find a $\psi_* \left(X_{t_i}, X_{t_{i-1}}, \theta\right)$, satisfying the equation (3.5.1) and such that it is the solution to the following minimization problem (that we label (KP1)):

$$\min_\psi E\left(\psi\psi\right)$$

among all martingale differences $\psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right)$, of the form (3.5.4) and such that

$$E_\theta \left[\psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right) | \mathcal{F}_{t_{i-1}}\right] = 0, \tag{3.5.7}$$

$$E_\theta \left[\psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right) s\left(X_{t_i}, X_{t_{i-1}}, \theta\right)\right] = -E_\theta[\partial_\theta \psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right)], \tag{3.5.8}$$

where $s\left(X_{t_i}, X_{t_{i-1}}, \theta\right)$ is the Maximum Likelihood score function, and

$$\sup_{\left(X_{t_i}, X_{t_{i-1}}\right) \in (S \times S)} \left|\psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right)\right| \le c(\theta). \tag{3.5.9}$$

We observe that every $\psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right)$ is a linear combination of bounded martingale differences χ_j[15]. By construction ψ defines an unbiased and Fisher Consistent M-estimator. That shows property I.

To solve our problem (KP1) it is enough to find a function of the form (3.5.4), having the smallest variance in the reference class $\mathcal{G}_n$. To achieve this goal, we use Godambe and Heyde theory. In particular, setting the weights

$$\alpha^*(\theta) := -E_\theta\left[\nabla_\theta \Gamma\right]^\top E_\theta\left(\Gamma\Gamma^\intercal\right)^{-1}, \tag{3.5.10}$$

we obtain the linear combination

$$\psi_*\left(X_{t_i}, X_{t_{i-1}}, \theta\right) := \alpha^*(\theta)\Gamma\left(X_{t_i}, X_{t_{i-1}}, \theta\right). \tag{3.5.11}$$

From Theorem 2.1 in Heyde [22] we know that for a convex (closed under addition) class $\mathcal{G}_n$, $G_n^*(\theta)$ is F-optimal, if

$$\left(E_\theta\left(\nabla_\theta G_n(\theta)\right)\right)^{-1} E_\theta\left(G_n^*(\theta) G_n(\theta)\right) = const$$

for every $G_n(\theta) \in \mathcal{G}_n$.

We observe that our class $\mathcal{G}_n$ is convex. Moreover,

$$\begin{aligned}
& \left(E_\theta\left(\nabla_\theta G_n(\theta)\right)\right)^{-1} E_\theta\left(G_n^*(\theta) G_n(\theta)\right) \\
= & \left(E_\theta\left(\nabla_\theta \alpha(\theta)\Gamma(\theta)\right)\right)^{-1} E_\theta\left(\sum_{i=1}^n \psi_*\left(X_{t_i}, X_{t_{i-1}}, \theta\right) \sum_{i=1}^n \psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right)\right) \\
= & \left(E_\theta\left(\alpha(\theta)\nabla_\theta \Gamma(\theta)\right)\right)^{-1} E_\theta\left(\sum_{i=1}^n \psi_*\left(X_{t_i}, X_{t_{i-1}}, \theta\right) \psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right)\right) \\
= & \left(E_\theta\left(\alpha(\theta)\nabla_\theta \Gamma(\theta)\right)\right)^{-1} \alpha^*(\theta) E_\theta\left(\sum_{i=1}^n \Gamma\left(X_{t_i}, X_{t_{i-1}}, \theta\right) \Gamma\left(X_{t_i}, X_{t_{i-1}}, \theta\right)^\top \alpha(\theta)^\intercal\right) \\
= & \; const.
\end{aligned} \tag{3.5.12}$$

[15]Every admissible ψ satisfies (3.5.7), (3.5.8) and (3.5.9).

is satisfied for every $G_n(\theta) \in \mathcal{G}_n$. Equation (3.5.12) follows from (3.5.1)) and the last equality follows from the definition of the optimal weights in equation (3.5.10). Therefore we conclude that $\psi_* \left(X_{t_i}, X_{t_{i-1}}, \theta\right)$ defines an F-optimal MEF in $\mathcal{G}_n$. That shows property II.

Finally, from equation (3.5.1), we observe that every $G_n(\theta) \in \mathcal{G}_n$ is a MEF. By construction, $\psi_* \left(X_{t_i}, X_{t_{i-1}}, \theta\right)$ is a bounded and continuous function, having zero conditional expected value. Therefore, our optimal MEF satisfies the assumptions in Proposition 3.4.1. This implies that for $\psi_* \left(X_{t_i}, X_{t_{i-1}}, \theta\right)$ there exists a unique bounded $IF^{cond}_{\psi_*}$, showing property III. ■

We conclude this subsection about Godambe-Heyde's optimality, with two remarks.

Remark 3.5.2. *The bound $c(\theta)$ on $IF^{cond}_{\psi_*}$ it is proportional to the single bounds introduced for every $\chi_j \left(X_{t_i}, X_{t_{i-1}}, \theta\right)$:*

$$\begin{aligned}
& \sup_{\left(X_{t_i}, X_{t_{i-1}}\right) \in (S \times S)} \left|\psi_* \left(X_{t_i}, X_{t_{i-1}}, \theta\right)\right| \\
= & \sup_{\left(X_{t_i}, X_{t_{i-1}}\right) \in (S \times S)} \left|\sum_{j=1}^{k} \alpha_j^*(\theta) \chi_j \left(X_{t_i}, X_{t_{i-1}}, \theta\right)\right| \\
\leq & \sum_{j=1}^{k} \left|\alpha_j^*(\theta)\right| \sup_{\left(X_{t_i}, X_{t_{i-1}}\right) \in (S \times S)} \left|\chi_j \left(X_{t_i}, X_{t_{i-1}}, \theta\right)\right| \\
\leq & \sum_{j=1}^{k} \left|\alpha_j^*(\theta)\right| b_j(\theta) =: c(\theta).
\end{aligned}$$

The bound on the optimal robust MEF depends on the selected parameterization and this feature can cause some problems with the converge of the algorithms used to compute the M-estimator defined by our first proposal (see Künsch [30], page 851 for a related discussion). This technical problem could make our first proposal very difficult to implement.

Remark 3.5.3. *The F-optimal weights $\alpha^*(\theta)$ as in (3.5.4) do not depend on observations. Let us introduce a class of weights depending on the past observation. We define*

$$\mathcal{G}_n := \left\{ G_n(\theta) = \sum_{i=1}^{n} \alpha(X_{t_{i-1}}, \theta)\Gamma\left(X_{t_i}, X_{t_{i-1}}, \theta\right) \right\}$$

with $\Gamma\left(X_{t_i}, X_{t_{i-1}}, \theta\right)$ as in (3.5.3). Heyde [22] shows that, setting

$$\alpha^*(X_{t_{i-1}}, \theta) := -E_\theta\left[\nabla_\theta \Gamma | \mathcal{F}_{t_{i-1}}\right]^\intercal E_\theta\left(\Gamma\Gamma^\intercal | \mathcal{F}_{t_{i-1}}\right)^{-1}$$

and

$$\psi^*_{A-optim}\left(X_{t_i}, X_{t_{i-1}}, \theta\right) = \alpha^*(X_{t_{i-1}}, \theta)\Gamma\left(X_{t_i}, X_{t_{i-1}}, \theta\right) \tag{3.5.13}$$

is possible to define an A-optimal MEF in the class $\mathcal{G}_n$ (see [22], Chapter II). According to Godambe and Heyde's theory, the difference, in terms of asymptotic variance, between F-optimal and A-optimal MEFs is very small. However, from a "robust" point of view, we note that the A-optimal weights depend on $X_{t_{i-1}}$ and are typically unbounded. As a result, that implies that $\psi^*_{A-optim}\left(X_{t_i}, X_{t_{i-1}}, \theta\right)$ in equation (3.5.13) is generally a linear combination of unbounded functions. Therefore A-optimal MEFs define quite often Conditionally Unbiased M-estimators that are not robust. In contrast, our F-optimal MEFs are linear combinations of bounded martingale differences, therefore they always define robust M-estimators.

In the Section 3.5.2 we show how to define Robust, Conditionally Unbiased M-estimator, using A-optimal MEFs.

3.5.2 Second proposal: Robust version of Kessler and Sørensen MEFs

In this subsection we introduce our second proposal, that defines an inference procedure closely related to M-R-T's approach [33]. The main idea is to construct a robust version of Kessler and Sørensen A-optimal MEFs. To achieve our goal we proceed as follows:

Step 1. We define the system

$$\Xi := \{\varkappa_j : j = 1, ..., k\} \tag{3.5.14}$$

of functions having zero conditional expected values, with finite variance (w.r.t. to $P_2^\theta(y, x)$) and having form as in equation (3.5.5). Denote by C_k the k-dimensional subspace of L_2 spanned by the functions in Ξ . We denote by ψ_*^{KS} the projection of the Likelihood score function onto C_k.

Step 2. We propose an estimating function φ_b, with bounded Conditional Influence Function, satisfying the following equations:

$$E_\theta \left[\varphi_b \left(X_{t_i}, X_{t_{i-1}}, \theta\right) \varphi_b \left(X_{t_i}, X_{t_{i-1}}, \theta\right)^\intercal\right] = I, \ \theta \in \Theta \tag{3.5.15}$$

and

$$E_\theta \left[\varphi_b \left(X_{t_i}, X_{t_{i-1}}, \theta\right) | \mathcal{F}_{t_{i-1}}\right] = 0, \ \theta \in \Theta. \tag{3.5.16}$$

Step 3. We define a robust M-estimator $\hat{\theta}_{n_{rob}}$ as the implicit solution of the equations:

$$G_n(\hat{\theta}_{n_{rob}}) = \sum_{i=1}^{n} \varphi_b(X_{t_i}, X_{t_{i-1}}, \hat{\theta}_{n_{rob}}) = 0. \tag{3.5.17}$$

To define the MEFs in our second proposal we set

$$\varphi_b \left(X_{t_i}, X_{t_{i-1}}, \theta\right) := A(\theta)\psi_S \left(X_{t_i}, X_{t_{i-1}}, \theta\right), \tag{3.5.18}$$

where

$$\psi_S := \left(\psi_*^{KS} \left(X_{t_i}, X_{t_{i-1}}, \theta\right) - \tau \left(X_{t_{i-1}}, \theta\right)\right) \omega \left(X_{t_i}, X_{t_{i-1}}, \theta\right) \tag{3.5.19}$$

with ψ_*^{KS} representing Kessler and Sørensen optimal MEF, and

$$\omega \left(X_{t_i}, X_{t_{i-1}}, \theta\right) := \min \left(1, b \left\|A(\theta) \left(\psi_*^{KS} \left(X_{t_i}, X_{t_{i-1}}, \theta\right) - \tau \left(X_{t_{i-1}}, \theta\right)\right)\right\|^{-1}\right).$$

Both φ_b and its unscaled version ψ_S are column vectors in $\mathbb{R}^p$.

The non singular matrix $A(\theta)$, with dimension $p \times p$, and the $\mathcal{F}_{t_{i-1}}$-measurable vector $\tau\left(X_{t_{i-1}};\theta\right) \in \mathbb{R}^p$ are determined, respectively, by solving the implicit equations (3.5.15) and (3.5.16). Moreover, by construction, φ_b satisfies the robustness constraint: $\gamma(\varphi_b) < \infty$.

Differently from the first proposal, φ_b has a bounded self-standardized sensitivity and its bound does not depend on the selected parameterization. As a result, our second proposal it is simpler to implement than the first proposal, because it solves the technical problem discussed in Remark 3.5.2.

Furthermore, by usual formulas, it follows:

$$IF^{cond}_{\varphi_b}\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right) = -D(\varphi_b, \theta_0)^{-1}\varphi_b\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right). \qquad (3.5.20)$$

Since φ_b is a bounded and continuous function, the $IF^{cond}_{\varphi_b}$ is unique and it defines a martingale difference process (see Proposition 3.4.1). Hence, equation (3.5.20) is the only admissible representation.

Finally, by equation (3.5.15), the Asymptotic Variance of $\hat{\theta}_{n_{rob}}$ is:

$$\begin{aligned} V(\varphi_b, \theta_0) &= D(\varphi_b, \theta_0)^{-1} E_{\theta_0}\left[\varphi_b\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right)\varphi_b\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right)^{\intercal}\right]\left[D(\varphi_b, \theta_0)^{-1}\right]^{\intercal} \\ &= D(\varphi_b, \theta_0)^{-1}\left[D(\varphi_b, \theta_0)^{-1}\right]^{\intercal}. \end{aligned}$$

Optimality

In order to study the optimality of φ_b we define a further linear subspace F of $L_2\left(p(\Delta t, y, x, \theta)\right)$.

Every $f \in F$ has zero conditional expectation

$$E_{\theta_0}\left[f\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right)|\mathcal{F}_{t_{i-1}}\right] = 0 \qquad (3.5.21)$$

and satisfies the orthogonality condition

$$E_{\theta_0}\left[f\left(\underbrace{s - \psi^{KS}_*}_{\varepsilon}\right)\middle|\mathcal{F}_{t_{i-1}}\right] = 0, \qquad (3.5.22)$$

Equation (3.5.22) is a natural requirement in our setting, because it implies that no function in F can be used to improve the error in the projection of the Likelihood score function onto C_k.

Since $C_k \subseteq F$, equation (3.5.22) is an analogous for functions in F to the moment condition (3.2.10), holding for functions in C_k. Assume, furthermore, that every function in F is continuously differentiable in θ (almost everywhere). We show that there does not exist a robust function f belonging to F, having smaller (asymptotic) variance than ψ_S.

We label the asymptotic variance of the MEF ψ by $V(\psi, \theta)$ and we show the following

Proposition 3.5.4. *Let $b \geq \sqrt{p}$. Assume, that equations (3.5.15) and (3.5.16) have solutions $A(\theta_0)$ and $\tau\left(X_{t_{i-1}}; \theta_0\right)$, respectively, and assume $\psi_S \in F$. If there exists a $f_{opt} \in F$, such that $V_{f_{opt}} - V_{\psi_S} \leq 0$ and that $\gamma(f_{opt}) \leq b^2 < \infty$, then $f_{opt} = \psi_S$ (up to a multiplicative constant).*

Proof. Let $\psi_S \in F$.

Without loss of generality, we consider $f \in F$ written in "canonical form", so that $f = IF_f^{cond}$ (see [16]). The martingale difference property of f implies that

$$E_{\theta_0}\left[f\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right)\right] = 0.$$

Differentiating both sides of this equation wrt θ, we obtain:

$$E_{\theta_0}\left[\frac{\partial}{\partial \theta_0} f\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right)\right] = -E_{\theta_0}\left[f\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right) s^{\intercal}(X_{t_i}; X_{t_{i-1}}; \theta_0)\right]$$

where $f(X_{t_i}; X_{t_{i-1}}; \theta_0)$ and $s(X_{t_i}; X_{t_{i-1}}; \theta_0)$ are column vectors in $\mathbb{R}^p$ and $s(X_{t_i}; X_{t_{i-1}}; \theta_0)$ is the Likelihood score function.Since f is in canonical form, it follows:

$$E_{\theta_0}\left[f\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right) s^{\intercal}(X_{t_i}; X_{t_{i-1}}; \theta_0)\right] = I \qquad (3.5.23)$$

and

$$-D(f; \theta_0) := -E_{\theta_0}\left[\frac{\partial}{\partial \theta_0} f\left(X_{t_i}, X_{t_{i-1}}; \theta_0\right)\right] = I$$

Assume that there exists a function $f_{opt} \in F$ such that, $V_{f_{opt}} \leq V_{\psi_S}$. It then follows:

$$(V_{f_{opt}} V_{\psi_S}^{-1}) \leq (V_{\psi_S} V_{\psi_S}^{-1})$$

and

$$tr(V_{f_{opt}} V_{\psi_S}^{-1}) \leq tr(V_{\psi_S} V_{\psi_S}^{-1}) = p.$$

We compute the function f_{opt} by solving the following optimization problem:

$$\begin{cases} \min_{f \in F} tr\left(V_f V_{\psi_S}^{-1}\right) \\ s.t.\ \gamma(f) \leq b. \end{cases} \tag{P2*}$$

For the sake of brevity, we use the following notation: $s = s(X_{t_i}, X_{t_{i-1}}, \theta_0)$, $f = f\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right)$, $D_{\psi_S} = D(\psi_S, \theta_0)$, $\tau = \tau\left(X_{t_{i-1}}, \theta_0\right)$ and $V_{\psi_S} = V(\psi_S, \theta_0)$.

The estimating function ψ_*^{KS} can be written as:

$$\psi_*^{KS} = s - \varepsilon, \tag{3.5.24}$$

where random vector ε satisfies the orthogonality condition (3.2.10) and condition (3.5.22).

Consider the expression

$$\begin{aligned} & E_{\theta_0}\left\{\left[D_{\psi_S}^{-1}(\overbrace{s-\varepsilon}^{\psi_*^{KS}}-\tau)-f\right]\left[D_{\psi_S}^{-1}(\overbrace{s-\varepsilon}^{\psi_*^{KS}}-\tau)-f\right]^{\intercal}\right\} \\ = \; & E_{\theta_0}\left\{D_{\psi_S}^{-1}(s-\tau-\varepsilon)(s-\tau-\varepsilon)^{\intercal}\left(D_{\psi_S}^{-1}\right)^{\intercal}\right. \\ & \left. -D_{\psi_S}^{-1}(s-\tau-\varepsilon)f^{\intercal} - f\left(D_{\psi_S}^{-1}(s-\tau-\varepsilon)\right)^{\intercal} + ff^{\intercal}\right\} \\ = \; & \delta - D_{\psi_S}^{-1}E_{\theta_0}\left\{(s-\tau-\varepsilon)f^{\intercal}\right\} - E_{\theta_0}\left\{f(s-\tau-\varepsilon)^{\intercal}\right\}\left(D_{\psi_S}^{-1}\right)^{\intercal} + E_{\theta_0}\left\{ff^{\intercal}\right\} \end{aligned}$$

where

$$\delta := D_{\psi_S}^{-1}E_{\theta_0}\left\{(s-\tau-\varepsilon)(s-\tau-\varepsilon)^{\intercal}\right\}\left(D_{\psi_S}^{-1}\right)^{\intercal}.$$

Taking into account equations (3.5.21) and (3.5.22), it follows:

$$\begin{aligned} & E_{\theta_0}\left\{\left[D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)-f\right]\left[D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)-f\right]^\intercal\right\} \\ = \; & \delta - D_{\psi_S}^{-1} - \left(D_{\psi_S}^{-1}\right)^\intercal + E_{\theta_0}\left\{ff^\intercal\right\}. \end{aligned}$$

As a result minimizing $tr(V_f V_{\psi_S}^{-1})$ over F is equivalent to solving

$$\min_{f\in F} tr\left(E_{\theta_0}\left\{\left[D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)-f\right]\left[D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)-f\right]^\intercal\right\}V_{\psi_S}^{-1}\right)$$

that is, under the additional robustness constraint,

$$\begin{cases} \min_{f\in F}\left(E_{\theta_0}\left\{\left[D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)-f\right]^\intercal V_{\psi_S}^{-1}\left[D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)-f\right]\right\}\right) \\ \qquad s.t.\ \gamma(f)\le b<\infty \end{cases} \tag{P2}$$

For $f \in F$, we define $\phi(X_{t_i}, X_{t_{i-1}}) := V_{\psi_S}^{-1/2} f(X_{t_i}, X_{t_{i-1}}, \theta_0)$ and we rewrite problem (P2) in the form

$$\begin{cases} \min_{\phi} E_{\theta_0}\left[\left\|\phi(X_{t_i}, X_{t_{i-1}}) - V_{\psi_S}^{-1/2} D_{\psi_S}^{-1}(\psi_*^{KS}(X_{t_i}, X_{t_{i-1}}, \theta_0) - \tau\left(X_{t_{i-1}}, \theta_0\right))\right\|^2\right] \\ \qquad s.t.\ \gamma(\phi)\le b<\infty. \end{cases}$$

The target function in the expectation of the last optimization problem is quadratic. As a result, we can solve the minimization problem by minimizing (wrt ϕ) pointwise the Lagrangian function

$$\Lambda = \left\|\phi - V_{\psi_S}^{-1/2} D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)\right\|^2 - \lambda\left(\|\phi\| - b\right).$$

where λ is the Lagrange multiplier for the constraint. The solution is

$$\begin{aligned} \phi &= V_{\psi_S}^{-1/2} D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)\min\left(1; b\left\|V_{\psi_S}^{-1/2} D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)\right\|^{-1}\right) \\ &= V_{\psi_S}^{-1/2} D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)\min\left(1; b\left\|D_{\psi_S}^{-1}(\psi_*^{KS}-\tau)\right\|_{V_{\psi_S}^{-1}}^{-1}\right). \end{aligned}$$

After pre-multiplying this last expression by $V_{\psi_S}^{-1/2\intercal}$, we obtain:

$$f = D_{\psi_S}^{-1}\psi_S.$$

Define the following subclasses of F:

$$T_1 := \left\{ f : \sup_{\left(X_{t_i}, X_{t_{i-1}}\right)} f^\intercal(X_{t_i}; X_{t_{i-1}}; \theta_0) V_f^{-1} f(X_{t_i}; X_{t_{i-1}}; \theta_0) \leq b^2 \right\}$$

and

$$T_2 := \left\{ f : \sup_{\left(X_{t_i}, X_{t_{i-1}}\right)} f^\intercal(X_{t_i}; X_{t_{i-1}}; \theta_0) V_{\psi_S}^{-1} f(X_{t_i}; X_{t_{i-1}}; \theta_0) \leq b^2 \right\}.$$

If there exists $f_{opt} \in T_1$, it is such that $V_{f_{opt}} \leq V_f$ for all $f \in T_1$. By definition, $D_{\psi_S}^{-1}\psi_S$ $\in T_1$. Moreover,

$$f_{opt}^\intercal V_{\psi_S}^{-1} f_{opt}^\intercal \leq f_{opt}^\intercal V_{f_{opt}}^{-1} f_{opt} \leq b^2.$$

Then $f_{opt} \in T_2$. Therefore $f_{opt} \in T_1 \cap T_2$. However, $V_{f_{opt}} < V_{\psi_S}$ implies $tr(V_{f_{opt}} V_{\psi_S}^{-1}) < tr(V_{\psi_S} V_{\psi_S}^{-1})$, in contradiction to the fact ψ_S is the solution to problem (P2*). ■

We observe that if the bounding constant $b \rightarrow= \infty$, then

$$\omega\left(X_{t_i}, X_{t_{i-1}}, \theta_0\right) \equiv 1 \text{ and } \tau\left(X_{t_{i-1}}, \theta_0\right) \equiv 0. \tag{3.5.25}$$

and $\psi_S \equiv \psi_*^{KS}$. We remark that in this limit case, ψ_*^{KS} is optimal in Godambe-Heyde's sense and there is no function in F having a smaller variance than ψ_*^{KS}. To obtain a proof of the last sentence, we observe that Theorem 4.3, in [3] shows the A-optimality of ψ_*^{KS}. Moreover, using (3.5.25) and moving along the same lines of our proof for Proposition 3.5.4, we can show that there does not exist a function $f_{opt} \in F$, such that $V_{fopt} < V_{\psi_*^{KS}}$. In this sense, we conclude that asymptotic variance of ψ_*^{KS} represents the (limit) lower bound that the asymptotic variance for functions in F achieves, when $b \rightarrow \infty$.

If $\varphi_b \notin F$, Proposition 3.5.4 does not apply. Nevertheless, we observe that, also in this case, φ_b remains a natural choice, since it is a truncation of the Conditionally Unbiased M-estimators based on Likelihood score, studied by M-R-T. To understand this key feature, we observe that[16] ψ_*^{KS} is an approximation of the Likelihood score function s to the k-th order in $L_2(p(\Delta t, y, x, \theta))$.

In particular, if we consider the space C_k spanned by the functions in (3.5.14), the space

$$B_\infty := \cup_{k=1}^{\infty} C_k$$

is dense in $L_2(p(\Delta t, y, x, \theta))$. This implies that, if $k \to \infty$,

$$\psi_*^{KS} \to s,$$

leading to a vanishing projection error.

For $k \to \infty$, we obtain a Conditionally Unbiased robust M-estimator using the ML score function:

$$s_{rob} = (s - \tau) \min\left(1; b \left\| s - \tau \right\|^{-1}\right).$$

This estimating function defines the same Conditionally Unbiased M-estimator studied by M-R-T in [33] (see equation (3.2.3)). Moreover, from Corollary 3.2.2 (see M-R-T for a proof, [33]), it follows that there does not exist any function satisfying equations (3.2.1), the robustness constraint and having (asymptotic) variance smaller than s_{rob}. As a result, we conclude that, for $k \to \infty$, s_{rob} is the limit benchmark in the class of Estimating Functions having zero conditional expected value and having self-standardized sensitivity bounded by b. Furthermore, we observe that the smaller (bigger) is b, the bigger (smaller) is the loss of efficiency of s_{rob} wrt the s.

According to last considerations, φ_b must be considered the natural candidate to use, when we replace the ML score function (unknown in closed form), by its approximation. In particular, we notice that φ_b defines a robust version of Kessler and Sørensen's approximation of the ML score to the k-th order.[17]

Computational remarks: the matrix $A(\theta)$ and the $\tau(X_{t_{i-1}}, \theta)$

In equation (3.5.18) , ψ_S is multiplied by the matrix $A(\theta)$. This is a technical device, that standardizes the function φ_b so that b^2 is the upper bound on $\gamma(\varphi_b)$. From equation

[16] See Kessler and Sørensen [28] for a discussion.

[17] We highlight that if we have "high frequency data" (i.e. Δt is very small), the efficiency of ψ_*^{KS} is very close to the one of the ML score, even if we use a small number of functions in Ξ. As a matter of fact, for Δt very small, the transition density is close to the Gaussian and we can set $k = 2$.

(3.5.15), it follows:

$$\begin{aligned} b^2 &\geq \left\| V(\varphi_b, \theta_0)^{-1/2} IF_{\varphi_b}^{cond} \right\|^2 \\ &= IF_{\varphi_b}^{cond\intercal} V(\varphi_b, \theta_0)^{-1} IF_{\varphi_b}^{cond} \\ &= \left\| \varphi_b \right\|^2 . \end{aligned}$$

The τ-vector in equation (3.5.19) has a simple heuristic interpretation. To see this, we start from ψ_*^{KS}. To find an estimating function with bounded IF, we apply Huber's weights in their original from

$$\varpi_H \left(X_{t_i}, X_{t_{i-1}}, \theta \right) := \min \left(1; \frac{b}{\left\| \psi_*^{KS} \left(X_{t_i}, X_{t_{i-1}}, \theta \right) \right\|} \right).$$

Now, define

$$\psi_*^b \left(X_{t_i}, X_{t_{i-1}}, \theta \right) = \psi_*^{KS} \left(X_{t_i}, X_{t_{i-1}}, \theta \right) \varpi_H \left(X_{t_i}, X_{t_{i-1}}, \theta \right).$$

By construction $\psi_*^b \left(X_{t_i}, X_{t_{i-1}}, \theta \right)$ has a bounded IF. However, we can not use $\psi_*^b \left(X_{t_i}, X_{t_{i-1}}; \theta \right)$ to define a Conditionally Unbiased M-estimator, because condition in (3.2.5) could not be satisfied. To preserve the Fisher Consistency, each truncated estimating function $\psi_*^b \left(X_{t_i}, X_{t_{i-1}}, \theta \right)$ should be shifted by a specific τ-vector[18]. In particular, for a general $\psi \left(X_{t_i}, X_{t_{i-1}}, \theta \right)$ the shifting τ-vector is given by the equation:

$$\tau(X_{t_{i-1}}, \theta) = \frac{E_\theta \left(\omega \left(X_{t_i}, X_{t_{i-1}}, \theta \right) \psi \left(X_{t_i}, X_{t_{i-1}}, \theta \right) \middle| \mathcal{F}_{t_{i-1}} \right)}{E_\theta \left(\omega \left(X_{t_i}, X_{t_{i-1}}, \theta \right) \middle| \mathcal{F}_{t_{i-1}} \right)}. \tag{3.5.26}$$

In the general time series setting, the expectations in (3.5.26) are not known in closed form. This technical problem makes robust estimation procedures based on M-estimators very time consuming. In our diffusion setting, we propose in the sequel a way to circumvent this issue, when the S-L problem for the Infinitesimal Generator can be solved.

Eigenfunction-eigenvalue expansion: the spectral decomposition. Following Karlin and Taylor [27], we study a regular diffusion process with a compact state space

[18] As M-R-T- remark, the existence of such a vector is guaranteed by the continuity of the mapping

$$\tau \left(X_{t_{i-1}}, \theta \right) \mapsto \left(\psi_* - \tau \left(X_{t_{i-1}}, \theta \right) \right) \omega \left(X_{t_i}, X_{t_{i-1}}, \theta \right)$$

and by the mean value theorem (see M-R-T [33], page 632).

$[l, r]$, where $\{l\}$ and $\{r\}$ are exit boundaries.[19] For a bounded and continuous function $f : \mathbb{R} \to \mathbb{R}$ we study the function

$$u(t, x, \theta) := E_\theta \left(f(X_t) \,|x \right),$$

which satisfies the backward PDE

$$u_t(x, \theta) = \frac{1}{2} \left(\frac{1}{m(x)} \right) \frac{d}{dx} \left(\frac{1}{s(x)} \frac{du}{dx} \right) = L_\theta u(x, \theta), \tag{3.5.27}$$

with initial condition $u(0, x, \theta) = f(x)$ and boundary conditions $u(t, l, \theta) = u(t, r, \theta) = 0$. L_θ is the Infinitesimal Generator in canonical representation. By the method of separation of variables, the solution to (3.5.27) is of the form

$$u(t, x, \theta) = c(t, \theta)\phi(x, \theta). \tag{3.5.28}$$

Then, the PDE (3.5.27) implies

$$\frac{c'(t)}{c(t, \theta)} = \frac{L_\theta \phi(x, \theta)}{\phi(x, \theta)}.$$

The right-hand member does not depend on x, while the left-hand member does not depend on t. This is possible only if both the sides of the equation are equal to a constant depending on θ:

$$\frac{c'(t, \theta)}{c(t, \theta)} = -\lambda(\theta) = \frac{L_\theta \phi(x, \theta)}{\phi(x, \theta)}.$$

Therefore

$$c'(t, \theta) = -\lambda(\theta) c(t, \theta) \tag{3.5.29}$$

and

$$L_\theta \phi(x, \theta) = -\lambda(\theta)\phi(x, \theta),$$

where $\phi(x, \theta)$ satisfies the boundary conditions $\phi(l, \theta) = \phi(r, \theta) = 0$.

The last equation define a S-L problem having solutions defined by eigenfunctions $\{\phi_n(x, \theta)\}_{n=0,1,\ldots}$ and eigenvalues $\{\lambda_n(\theta)\}_{n=0,1,\ldots}$ We remark that from (3.5.29), it follows that

$$c_n(t, \theta) = c_n(\theta) \exp(-\lambda_n(\theta)t).$$

[19] We can relax this assumption and we can work in a general environment. To this end, see Karlin and Taylor ([27], pages 330-331-332)

From (3.5.28), the solution of the PDE (3.5.27) are for $n \in N$:

$$c_n(\theta)\exp(-\lambda_n(\theta)t)\phi_n(x,\theta)$$

Using linearity of the PDE (3.5.27) and the principle of superposition, we write

$$u(t,x,\theta) = \sum_{n=0}^{\infty} c_n(\theta)\exp(-\lambda_n(\theta)t)\phi_n(x,\theta) = \sum_{n=0}^{\infty} c_n(\theta,t)\phi_n(x,\theta). \qquad (3.5.30)$$

This is a Fourier Generalized Series Expansion, with coefficients $\{c_n(t,\theta)\}_{n=0,1,\ldots}$ which have to be determined to guarantee the initial condition $u(0,x,\theta) = f(x)$. From the classical theory of Fourier Series Expansion, it follows that

$$c_n(\theta,t) = \frac{\int_l^r u(t,x,\theta)\phi_n(x,\theta)m(x,\theta)dx}{\int_l^r \phi_n^2(x,\theta)m(x,\theta)dx}.$$

Now, taking into account that the sequence of $\{c_n(\theta,t)\}_{n=0,1,\ldots}$ is subject to the initial (temporal) condition at $t = 0$, we have

$$c_n(0,\theta) = \frac{\int_l^r f(y)\phi_n(y,\theta)m(y,\theta)dy}{\int_l^r \phi_n^2(y,\theta)m(y,\theta)dy},$$

then

$$c_n(\theta,t) = \frac{\int_l^r f(y)\phi_n(y,\theta)m(y,\theta)dy}{\int_l^r \phi_n^2(y,\theta)m(y,\theta)dy}\exp(-\lambda_n(\theta)t) = c_n(0,\theta)\exp(-\lambda_n(\theta)t), \qquad (3.5.31)$$

that satisfies equation (3.5.29).

We can use the above procedure to approximate the transition probability of the underling diffusion process. Indeed, given $l \le a < x < b \le r$, let us define:

$$f(y) := \begin{cases} 1 & y \in (a,b) \\ 0 & elsewhere \end{cases}$$

After some manipulations (see for instance Karlin and Taylor, [27], page 332), we can write

$$\begin{aligned} p(\Delta t,x,y,\theta) &= \sum_{n=0}^{\infty} \frac{\exp(-\lambda_n(\theta)\Delta t)\phi_n(x,\theta)\phi_n(y,\theta)m(y,\theta)}{\int_l^r \phi_n^2(y,\theta)m(y,\theta)dy} \\ &= \sum_{n=0}^{\infty} \exp(-\lambda_n(\theta)\Delta t)\phi_n(x,\theta)\phi_n(y,\theta)m(y,\theta)\pi_n(\theta), \qquad (3.5.32) \end{aligned}$$

where $\pi_n(\theta) := \frac{1}{\int_l^r \phi_n^2(y,\theta)m(y,\theta)dy}$.

Using a finite order series approximation for the conditional expected value of a bounded and continuous function based on (3.5.30), we can provide accurate analytical approximations for $\tau(X_{t_{i-1}};\theta)$ in equation (3.5.26).

In the following section we give more details about this procedure, providing an algorithm that applies this formula and that can be used to calculate our second proposal for a robust M-estimator.

3.6 Algorithm for θ and analytical approximation for $\tau(X_{t_{i-1}},\theta)$

To compute our robust estimator, given a number k of martingale differences, we use the following iterative algorithm. We split the algorithm in four different steps:

1. We set an initial value $\theta^{(0)}$, and define $\tau^{(0)} = \tau\left(X_{t_{i-1}},\theta^{(0)}\right) = 0$. We calculate the stating value A^0 by equation (3.5.15) and by using the unbounded estimating function in Kessler and Sørensen [28]:
$$A^{(0)}A^{(0)\intercal} = \left(\frac{1}{n}\sum_{i=1}^{n}\psi_*^{KS}\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)\psi_*^{KS}\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)^\intercal\right)^{-1};$$

2. We calculate $\tau^{(1)}$as:
$$\tau^{(1)}(X_{t_{i-1}},\theta^{(0)}) = \frac{E_{\theta^{(0)}}\left(\omega\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)\psi_*^{KS}\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)\middle|\mathcal{F}_{t_{i-1}}\right)}{E_{\theta^{(0)}}\left(\omega\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)\middle|\mathcal{F}_{t_{i-1}}\right)}$$
and $A^{(1)}$ (see equation (3.5.15)) as:
$$A^{(1)}A^{(1)\intercal} = \left(\frac{1}{n}\sum_{i=1}^{n}\psi_*^{KS}\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)\psi_*^{KS}\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)^\intercal\omega\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)\right)^{-1}$$
where
$$\omega\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right) = \min\left(1;\frac{b}{\left\|A^{(0)}(\psi_*^{KS}\left(X_{t_i},X_{t_{i-1}},\theta^{(0)}\right)-\tau^{(0)}\left(X_{t_{i-1}},\theta^{(0)}\right))\right\|}\right).$$

Using the spectral representation (3.5.30), we have a way to approximate analytically $\tau^{(1)}(X_{t_{i-1}}, \theta^{(0)})$. By the Fourier Generalized Series Expansion we obtain for the denumerator[20] in equation (3.5.26):

$$E_{\theta^{(0)}}\left(\omega\left(X_{t_i}, X_{t_{i-1}}, \theta^{(0)}\right)\left|\mathcal{F}_{t_{i-1}}\right.\right) = \sum_{z=0}^{\infty} c_z(\theta^{(0)}; 0)\exp(-\lambda_z(\theta^{(0)})\Delta t)\phi_z(\theta^{(0)}, X_{t_{i-1}})$$

where

$$c_z(\theta^{(0)}, 0) := \frac{\int_l^r \omega\left(x_{t_i}, x_{t_{i-1}}, \theta^{(0)}\right)\phi_z(\theta^{(0)}, x_{t_i})m(\theta^{(0)}, x_{t_i})dx_{t_i}}{\int_l^r \phi_z^2(\theta^{(0)}, x_{t_i})m(\theta^{(0)}, x_{t_i})dx_{t_i}}.$$

To define the approximation, we truncate the sum up to a selected value q, leading to the Fourier Generalized Series:

$$E_{\theta^{(0)}}\left(\omega\left(X_{t_i}, X_{t_{i-1}}, \theta^{(0)}\right)\left|\mathcal{F}_{t_{i-1}}\right.\right) \approx \sum_{z=0}^{q} c_z(\theta^{(0)}, 0)\exp(-\lambda_z(\theta^{(0)})\Delta t)\phi_z(\theta^{(0)}, X_{t_{i-1}});$$

3. Given $A^{(1)}$ and $\tau^{(1)}$, we update the parameter $\theta^{(1)}$, by solving the implicit equation

$$\sum_{i=1}^{n} A^{(1)}\left(\psi_*^{KS}\left(X_{t_i}, X_{t_{i-1}}, \theta^{(0)}\right) - \tau^{(1)}\right)\omega^{(01)}\left(X_{t_i}, X_{t_{i-1}}, \theta^{(0)}\right) = 0,$$

where

$$\omega^{(01)}\left(X_{t_i}, X_{t_{i-1}}, \theta^{(0)}\right) = \min\left(1, \frac{b}{\left\|A^{(1)}(\psi_*\left(X_{t_i}, X_{t_{i-1}}, \theta^{(0)}\right) - \tau^{(1)})\right\|}\right);$$

4. We go back to Step 2 and replace $A^{(0)}$ by $A^{(1)}$, $\tau^{(0)}$ by $\tau^{(1)}$ and $\theta^{(0)}$ by its new value $\theta^{(1)}$. Then we iterate the routine just until there is the convergence of the sequences $\left\{\theta^{(j)}\right\}_{j=1,..}$, $\left\{A^{(j)}\right\}_{j=1,..}$ and $\left\{\tau^{(j)}\right\}_{j=1,..}$.

To define the truncated Fourier Generalized Series, we need to fix the number q of eigenvalues and eigenfunctions. There is no golden rule to determine the optimal number of terms and we have to keep in mind that the higher is q, the higher is the reliability of the approximation, but the bigger is the computational burden of the algorithm. In the next Chapter, we provide an example and we implement one possible strategy to select the value of q.

[20]The calculations for the numerator are identical.

3.7 Asymptotics

Our robust MEFs (both the first and the second proposal) define Robust, Conditionally Unbiased M-estimators. Therefore consistency of our estimators at the reference model, can be shown by standard arguments.

In particular, under Condition 3.1.5 and under some suitable regularity conditions on the Bounded Estimating Function ψ (see for instance Kessler and Sørensen, [28], page 309), a solution $\hat{\theta}_n$ to equation (3.5.6) (or equivalently to equation (3.5.17)) exists with a μ_{θ_0}-probability tending to 1, and $\hat{\theta}_n$ is consistent for θ_0.

As far as the asymptotic distribution at the reference model is concerned, we apply the Martingale Central Limit Theorem as in Sørensen, [38] to obtain:

$$\sqrt{n}\left(T[\rho(x,n)^2] - T[\mu_2^0]\right) = \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N\left(0, V_{\theta_0}\left(\psi, \theta_0\right)\right).$$

Observing carefully the asymptotic Variance-Covariance matrix, we do a useful remark.

By standard formulas (see equation (3.4.9)) we have:

$$V_{\theta_0}\left(\psi, \theta_0\right) = E_{\theta_0}(IF_\psi^{cond} IF_\psi^{cond^\intercal}) = D_\psi^{-1}(\theta_0) E_{\theta_0}(\psi\psi^\intercal) D_\psi^{-\intercal}(\theta_0).$$

The last member defines the inverse of the Godambe Information Matrix (see Bibby, Jacobsen and Sørensen, [3]). Therefore we can conclude that the asymptotic Variance-Covariance matrix of an M-estimator is the inverse of Godambe Information Matrix for ψ, that is nothing else but the inverse of $E_{\theta_0}\left(IF_\psi IF_\psi^\intercal\right)$.

Chapter 4

Examples and Implementation of R-algorithms

In this final Chapter we illustrate some examples, that provide some numerical values to the mathematical and theoretical considerations that we have done in Chapter 3. Example I analyzes the performance of Sørensen MEFs for the CIR process. We show how we can use the IF^{cond} to study the behavior of the M-estimator defined by Kessler and Sørensen. We remark that in this case the Likelihood score is known in analytical closed form and we could define Conditionally Unbiased M-estimators by the classical ML approach. Nevertheless, we present this example for demonstrative purposes and in order to show how is possible to apply the mathematical tools defined in Chapter 3. As a matter of fact, in the first part of the example, we provide the formulas defining Sørensen's M-estimator for the parameters in the drift of the CIR process introduced in Chapter 3 (see Example 3.2.4, for the analytical calculations). Moreover we show how to derive the Conditional Influence Function of Sørensen's M-estimator and we analyze the behavior of the IF^{cond}, in order to study (from a theoretical point of view) the robust features of the M-estimator. In the second part of the example, we introduce a Monte Carlo analysis. We explain how to solve some computational aspects and we show how to simulate a contaminated sample for the CIR process, using a contamination scheme called “replacement model”. Our Monte Carlo analysis provides some numerical values for the bias that Sørensen’s M-estimator can have under contaminations. In Example II, we study a trigonometric transformation of the Jacobi diffusion. This example analyzes the robust features of MEFs defined in Example 2.2, in Kessler and Sørensen [28]. We split the example in two parts. In the first part we show the calculations that define a Conditionally Unbiased M-estimator by using the solutions to the Sturm-Liouville problem for the Infinitesimal Generator. Following Kessler and Sørensen ([28], page 306), we define two MEFs: the Godambe-Heyde optimal MEF (ψ_*) and the approximately optimal MEF ($\tilde{\psi}$). Moreover we calculate the Conditional Influence Function of the M-estimators implied by ψ_* and by $\tilde{\psi}$ and we investigate their robust properties. In particular, we show how to use the IF^{cond} to analyze the different asymptotic bias of the two M-estimators. Our mathematical analysis highlights some robust features of ψ_*. In the second part, we show the results of a Monte Carlo analysis. In particular, we show that, even if Sørensen’s optimal MEF has a bounded IF^{cond}, it is “not enough robust” and a Monte Carlo analysis gives a numerical value of the bias that Sørensen’s M-estimator can have in presence of contamination. We reduce this bias by defining a Robust MEF. To this end, we apply our second proposal. In a specific paragraph, we show how to use the spectral representation for the Infinitesimal Generator, in order to approximate analytically the $\tau(X_{t_{i-1}}, \theta)$, in equation (3.5.26). The Monte Carlo analysis shows the better behavior of our Robust Conditionally Unbiased M-estimator, in presence of contamination.

4.1 Example 1: Contaminated CIR process

4.1.1 Mathematical Analysis

From Chapter 3, we know that the Martingale Estimating Function for the parameter α in the CIR process is:

$$\sum_{i=1}^{n} \frac{1}{X_{t_{i-1}}}[X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta t) - \alpha(1 - \exp(-\beta\Delta t))] = 0. \tag{4.1.1}$$

So the first component of the IF^{cond} in this setting is proportional to:

$$\frac{1}{X_{t_{i-1}}}[X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta t) - \alpha(1 - \exp(-\beta\Delta t))]. \qquad (4.1.2)$$

In Figure1, we present the graph of this function for the parameter choice $\alpha = 0.0305$, $\beta = 1$ and $\Delta t = 0.2$. From Figure 1, we notice that the function (considered as a function of (y, x)) is unbounded: in particular if $x = X_{t_{i-1}}$ takes small values and/or $y = X_{t_i}$ takes large values, the function in (4.1.2) diverges to infinity.

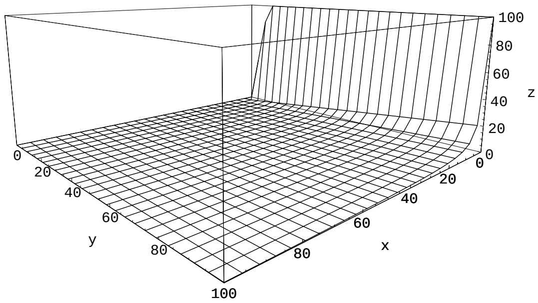

Fig.1: Estimating function for α

The Martingale Estimating Function for β is:

$$\sum_{i=1}^{n}[X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta t) - \alpha(1 - \exp(-\beta\Delta t))] = 0. \qquad (4.1.3)$$

So the second component of the IF^{cond} is proportional to

$$[X_{t_i} - X_{t_{i-1}} \exp(-\beta\Delta t) - \alpha(1 - \exp(-\beta\Delta t))]. \qquad (4.1.4)$$

In Figure 2, we present the graph of function (4.1.4) for the same parameter choice as before. Also this function is unbounded in (y, x). More specifically if $x = X_{t_{i-1}}$ takes large values and/or $y = X_{t_i}$ takes large values, the function in (4.1.4) diverges to infinity.

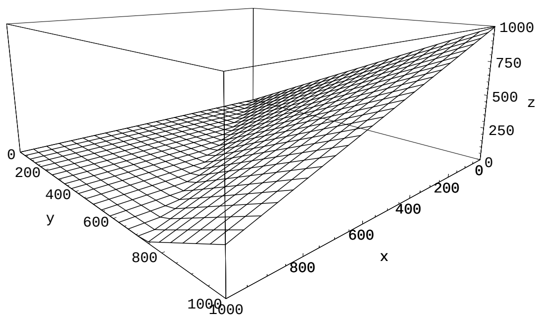

Fig.2 Estimating function for β

According to our considerations about Von Mises' expansion, the asymptotic bias of an M-estimator implied by the estimating functions in formulas (4.1.1) and (4.1.3) is unbounded. As a consequence the inference for $\theta = (\alpha, \beta)$ could be very affected by data contaminations.

Assume, for instance, that the contaminating distribution stresses the points (y, x) in the regions where the function (4.1.2) or (4.1.4) (or both of them) are unbounded. If this is the case, the gross error sensitivity diverges toward infinity and the optimal estimator, obtained by solving (4.1.1) and (4.1.3), is not robust in Huber and Hampel's sense. As a result, a small departure from the underlying distribution could determine dramatic changes in the bias of the estimator. Let us quantify this bias by a Monte Carlo simulation.

4.1.2 Monte Carlo Analysis

Technical Issues for Monte Carlo Analysis

In this subsection we briefly discuss some technical issues. In particular, we present a simulation scheme that generates a contaminated sample, having a specified frequency of outliers (i.e. a specified probability of having a contaminated value in our sample). Let us see this procedure in detail.

In our simulation study, the Clean process for the CIR is defined by the SDE

$$dX_t = -\beta(X_t - \alpha)dt + \sigma\sqrt{X_t}dW_t.$$

To obtain one trajectory for the reference model, we use the Milstein scheme with this parameters:

δ	Δt	α	β	σ
0.1	0.2	0.00305	1	0.01

Table for CIR

and starting value $X_0 = x_0 = 0.03$. Every trajectory for the discrete-sampled process contains 1000 observations.

To generate the Contaminated process, we follow Ortelli-Trojani's approach [35] and we use a replacement model, as in Martin-Yohai [34].

To implement this scheme, we use a Data Generating Process for the Contaminated process $\{Y(t)\}_{t\in T}$ given by:

$$Y(t) := (1 - H^{\eta}(t))X_t + H^{\eta}(t)\xi, \tag{4.1.5}$$

where X_t is the reference process (CIR), ξ is a constant, $H^{\eta}(t)$ is an iid process 0-1 valued and it is such that $\Pr(H^{\eta}(t) = 1) = \eta$.

According to Martin and Yohai, we can interpret this structure in the following way: the contaminated process (Y) is equal to the clean (reference) process X, when there is not any contamination. Otherwise, when there is a contamination, Y is equals to ξ. Roughly speaking, on average there is a η%-number of outliers that perturbs the theoretical reference process.

As far as this contamination model is concerned, we do an important remark that is related to Künsch assumptions discussed in Chapter 3. If we label the contaminated distribution by μ^2_{η} and by $\mu^2_{\theta_0}$ the reference distribution, it can be shown (see Ortelli's PhD thesis, pages 65-66) that

$$\lim_{\eta\to 0} \frac{\mu^2_{\eta} - \mu^2_{\theta_0}}{\eta}$$

converges weakly to a signed measure ($\dot{\mu}^2$). As a consequence, according to Künsch [30], we can express the asymptotic bias by formulas given in Chapter 3. Furthermore, we observe that, in equation (4.1.5), we have defined $H^{\eta}(t)$ as an iid process. This is only one possible way to model the contamination of the reference model and we have selected this specification of $H^{\eta}(t)$ only for practical purposes. From a theoretical point of view, is possible to define $H^{\eta}(t)$ as a stationary and ergodic 0-1 valued process with any dependence structure. For instance, we can use a $H^{\eta}(t)$ with strong memory and define outliers in long patches (see Hampel et al., [17], for a discussion).

After having generated a trajectory for the reference model, we draw random observations from a Bernoulli distribution (having probability of success η) and we replace some of the simulated values by the outliers ξ. In this way, we simulate the Contaminated process.

Here below, in Figure 3 and Figure 4, we present one simulated trajectory for the Clean and for the Contaminated process, respectively. We have chosen the following parameter values for the contaminated process:

- $\xi = 0.0338$ and
- $\eta = 0.02$ = contamination probability.

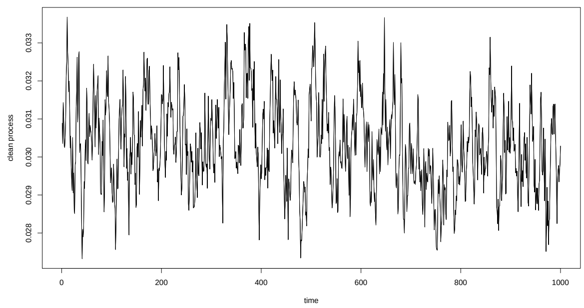

Fig.3 Clean CIR process: $\alpha = 0.0305$, $\beta = 1$, $\sigma = 0.01$. $\Delta t = 0.2$

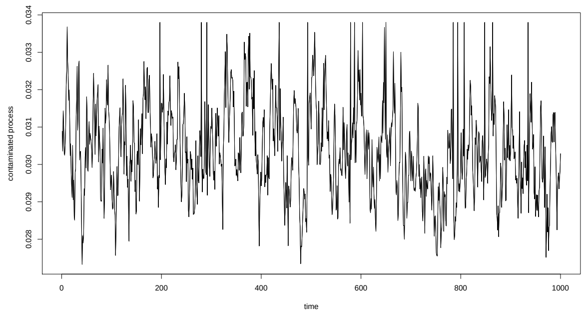

Fig.4 Contaminated CIR process:$\eta = 0.02$ and $\xi = 0.0338$

We higlight that, by construction, the contaminated sample has some extreme points. Nevertheless, we notice that the selected contamination stresses only the probability of the extreme values and it does not introduce any "new and unsual" outlier, very far away from bulk of the observations.

To make it clear, look carefully at Figure 3 and at Figure 4. It is easy to notice that the selected contamination stresses only the frequency of some borderline points, that we can already observe also under the reference model.

Monte Carlo Analysis: Output

In order to study the behavior of Sørensen's M-estimator for θ in presence of contaminations, we have written an algorithm in R, implementing the estimation procedure, given by solving (4.1.1) and (4.1.3). We have applied this algorithm (200 times) both to the Clean process and to the Contaminated process and we have observed the performance of the M-estimator. In Table 1 and in Table 2, we present the output:

for α

Clean Process		Contaminated Process	
Mean	SD	Mean	SD
0.03049	1.5e-08	0.03056	1.5e-08

Table 1

and for β

Clean Process		Contaminated Process	
Mean	SD	Mean	SD
1.07	0.013	1.78	0.076

Table 2

Looking at the first column of Table 1 and Table 2, we remark that Sørensen's M-estimator applied to the Clean process is very precise and it defines an M-estimator having a very small bias and a small variance.

A totally different conclusion follows from the analysis of the performance under contamination. The second column of Table 1 and Table 2 shows that the inference in presence of contaminated sampled is biased. This is the main effect of the unbounded IF^{cond}, which causes the bad performance of the M-estimator in presence of contaminated observations. As far as the effects of the contamination are concerned, it is interesting to notice that we are simulating a CIR process with a small diffusion coefficient and a long-run mean $\alpha = 0.0305$. Our contamination scheme stresses some points above the long run mean. We can interpret this kind of contamination as an exogenous factor, that artificially changes the dynamic of the process. In particular, any ξ influences strongly the mean reversion speed. For instance, after every replacement of one observation, the process needs a stronger mean reversion in order to converge towards α. Our Monte Carlo simulation provides numerical values for these effects of the contamination. As a matter of fact, we observe that the sample mean of the M-estimator for α has a shift toward the contaminating values. Its sample standard deviation remains quite stable. Looking at Table 2, we can notice the strong effect of the contamination on the distribution of Sørensen's M-estimator for β. As a matter of fact, we observe that, even if we are in presence of contamination with small probability, the mean and the standard deviation of the M-estimator change dramatically. The bias on the mean is about 78% and the SD rises of about 5.6 times. As a result, we can argue that under the given contamination scheme, the M-estimator for β is more instable than the M-estimator of α. To give a graphical overview of the effects of the contamination, we present in Figure 9 and Figure 10 the (fitted) kernel density of the estimated values, under the Clean and under the Contaminated process. In Figure 5, the continuous line represents the kernel density of Sørensen M-estimator for α at the reference model. The dotted line, represents the kernel density of the M-estimator, calculated for contaminated samples.

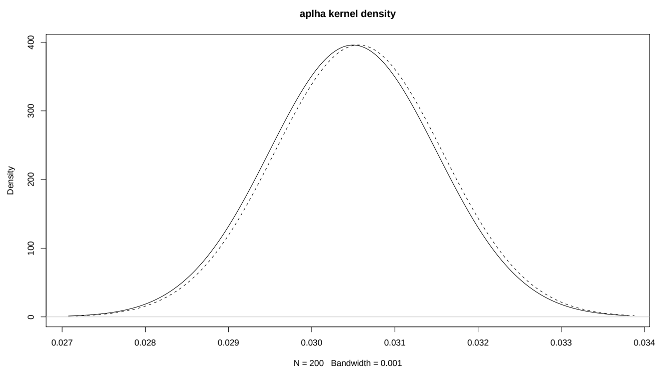

Fig.5 Kernel Density of Sørensen's M-estimator for α: reference (continuous line) and contaminated (dotted line) model

We observe that, even if the two distributions are very similar, the contamination determines up-ward biased estimates for the long-run mean.

The effect of the contaminations is larger for the M-estimator for the speed of mean-reversion. In particular, in Figure 6, we may notice the strong impact of the contamination on the values of the M-estimator for β. We remark that the fitted distribution of the M-estimator in presence of contaminations is very biased. Furthermore (look at the shape of the dotted curve) it is clear that the M-estimator calculated under contamination has a larger standard deviation, wrt the standard deviation of the M-estimator without contamination.

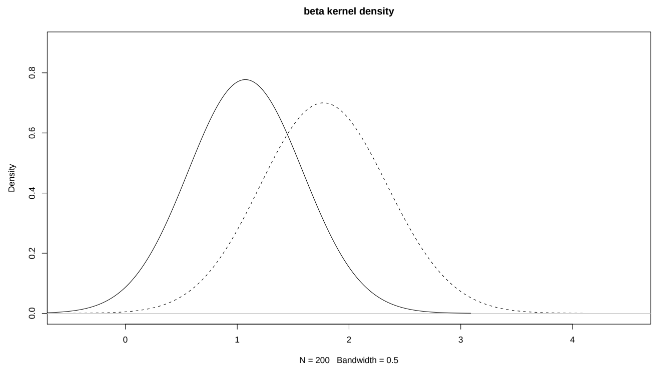

Fig.6 Kernel density of Sørensen's M-estimator for β: under the reference model (continuous line) and under contamination (dotted line)

Conclusion: Sørensen's MEF for the parameters in the drift coefficient of the CIR process defines Conditionally Unbiased M-estimator that are really precise under the reference model. However this M-estimator can have a strong bias under slight violations of model assumptions, even with a small contamination probability and with contaminating values ξ very close to the bulk of the observations.

To study the behavior of Sørensen's M-estimator in different settings, we have performed further Monte Carlo simulations. In particular, we have changed some parameters both for the Clean and for the Contaminated process. For instance we have changed the long-run mean or the strength of mean revertion. Moreover we have changed the impact of contamination, setting different values of η and different values of ξ. In all these simulation studies, we have observed a very good performance of Sørensen's M-estimator at the reference model. Nevertheless, in presence of contaminations, Sørensen's M-estimator has shown a bad performance, both in terms of its bias and in terms of its standard deviation.

4.2 Example 2: Contaminated trigonometric diffusion

We now consider a second example, given by a "trigonometric transformation" of the Jacobi diffusion. In Example 1, we have defined a MEF using the first conditional moment of the CIR-process. Example 2 is interesting, because, following Kessler and Sørensen [28], we can define an A-optimal[1] MEF, using the solutions to the Sturm-Liouvlille problem for the Infinitesimal Generator. Monte Carlo studies show that the optimal MEF defines a Conditionally Unbiased M-estimator, having a small degree of robustness. Therefore we need to define a new Robust Conditionally Unbiased M-estimator. To this end, we implement our second proposal. In particular, given that we know the eigenfunctions and the eigenvalues of the Infinitesimal Generator, we use the Generalized Fourier Series Expansion, introduced in Chapter 3 (see formulas (3.5.30)), to calculate an analytical approximation of $\tau(X_{t_{i-1}}, \theta)$, in equation (3.5.26).

As far as this topic is concerned, Example 2 highlights one of the main contributions of the Thesis. As a matter of fact, thanks to this analytical approximation for the solution of the parabolic PDE defining the conditional expected value of a function of the process (see Chapter 3), our second proposal defines Robust Conditionally Unbiased M-estimators that are faster than Robust GMM, Robust EMM or Robust Indirect Inference estimators. In fact, our inference procedure avoids time-consuming internal Monte Carlo simulations, a very computational demanding task that the other methods need to approximate $\tau(X_{t_{i-1}}, \theta)$.

Example 2 is organized in two subsections:

1. **Section 4.2.1**: A mathematical part, where we calculate the mathematical tools that we use for the inference procedure and where we provide a theoretical analysis of the IF^{cond}.

2. **Section 4.2.2**: A Monte Carlo Analysis, with a subsection where we specify how we can use in practice the Generalized Fourier Series Expansion.

[1]We will use Godambe-Heyde optimality in a specific class of MEFs

4.2.1 Mathematical Analysis

The Jacobi diffusion is defined by the SDE

$$dX(t) = -\beta\left[X(t) - (m + \gamma z)\right]dt + \sigma\sqrt{z^2 - (X(t) - m)^2}dW(t), \tag{4.2.1}$$

where $m \in \mathbb{R}, \beta > 0, \gamma \in (-1, 1)$ and W is a standard Wiener process. The calculations of the speed and the scale measure show that this process is ergodic and stationary on $(m-z, m+z)$ if and only if $\beta(1-\gamma) \geq \sigma^2$ and $\beta(1+\gamma) \geq \sigma^2$ (see [31], for calcualations).

This process is mean-reverting, that is, it reverts to the mean $(m+\gamma z)$ with a mean-revertion speed β. When the trajectory is near to the values $m + z$ or $m - z$, the drift drives the process away from the boundaries and the diffusion coefficient goes to zero.

Now consider the process:

$$dX(t) = -\rho\frac{\sin\left(\frac{\pi}{2}\left(X(t) - m\right)/z\right) - \varphi}{\cos\left(\frac{\pi}{2}\left(X(t) - m\right)/z\right)}dt + \sigma dW(t), \tag{4.2.2}$$

where $m \in \mathbb{R}, \rho > 0, \varphi \in (-1, 1), \sigma > 0$ and W is a standard Wiener process.

This process can be interpreted as a "trigonometric transformation" of the solution to the SDE (4.2.1). To show the link between the SDE (4.2.2) and the SDE (4.2.1), we can apply Itô's formula to (4.2.1) with $m = 0$ and $z = 1$, that is

$$dX(t) = -\beta\left[X(t) - \gamma\right]dt + \sigma\sqrt{1 - X(t)^2}dW(t)$$

and define $Y(t) = \arcsin(X(t)) = \sin^{-1}(X(t))$

$$\begin{aligned} y_t &= 0, \\ y_x &= \frac{1}{\sqrt{1-x^2}} = \frac{1}{\cos(y)}, \\ y_{xx} &= (1-x^2)^{-\frac{3}{2}}x, \end{aligned}$$

then

$$\begin{aligned} dY(t) &= \left[-\beta X(t) + \beta\gamma + \frac{1}{2}\sigma^2 X(t)\right]\frac{dt}{\cos(Y(t))} + \sigma dW(t) \\ &= \left[-\beta\sin(Y(t)) + \beta\gamma + \frac{1}{2}\sigma^2\sin(Y(t))\right]\frac{dt}{\cos(Y(t))} + \sigma dW(t) \\ &= -\rho\frac{\sin\left(Y(t)\right) - \varphi}{\cos\left(Y(t)\right)}dt + \sigma dW(t), \end{aligned} \tag{4.2.3}$$

where $\rho := \beta - \frac{1}{2}\sigma^2$, $\varphi := \frac{\beta\gamma}{\beta-\frac{1}{2}\sigma^2}$.

This SDE is a particular version of (4.2.2), with $m = 0$ and $z = \frac{\pi}{2}$. Larsen and Sørensen [31] show that this diffusion process is ergodic and stationary over $(-\frac{\pi}{2}, \frac{\pi}{2})$ iff $\rho \geq \frac{\sigma^2}{2}$ and $(-1 + \frac{1}{2\rho} \leq \varphi \leq 1 - \frac{1}{2\rho})$.

For this model, the $j - th$ eigenfunction of the Infinitesimal Generator is:

$$\phi_j(x, \rho, \sigma, \varphi) = P_j^{\left(\rho(1-\varphi)\sigma^{-2}, \rho(1+\varphi)\sigma^{-2}\right)}(\sin(x)),$$

where $P_j^{\left(\rho(1-\varphi)\sigma^{-2}, \rho(1+\varphi)\sigma^{-2}\right)}(\sin(x))$ is the Jacobi polynomial of order j, with the corresponding eigenvalue:

$$\lambda_j(\rho, \sigma, \varphi) := j\left(\rho + \frac{\sigma^2}{2}j\right).$$

Assume now we are given the parameters $\gamma = \frac{1}{2}$, $\sigma = 1$ and assume that we want to draw inference about the unknown parameter β. This means that we are working on the process that is the solution of the following SDE:

$$dY(t) = \left(\frac{1}{2} - \beta\right)\frac{\sin(Y(t)) - \frac{\beta}{1-2\beta}}{\cos(Y(t))}dt + dW(t).$$

In addition, we assume that the process is discretely sampled at fixed intervals of length Δt.

Let be

$$\psi\left(X_{t_i}, X_{t_{i-1}}, \theta, \Delta t\right) = \sum_{j=1}^{k} \alpha_j(X_{t_{i-1}}, \theta)\left\{\phi_j(X_{t_i}, \theta) - \exp[-\lambda_j(\theta)\Delta t]\phi_j(X_{t_{i-1}}, \theta)\right\}.$$

and define the class of MEFs:

$$\mathcal{G}_n := \left\{G_n(\theta) = \sum_{i=1}^{n} \psi\left(X_{t_i}, X_{t_{i-1}}, \theta\right)\right\}.$$

In order to find an optimal estimator for β, we can define the MEF $G_n(\theta)$ with the smallest variance in $\mathcal{G}_n$. To this end we can use equations (3.3)-(3.6), in Kessler and Sørensen [28]. In particular, for the first eigenfunction and the first eigenvalue, we obtain

$$E_\theta\left[P_1^{(\frac{\beta}{2}-1, \frac{3\beta}{2}-1)}(\sin(X_{t_i}))\,\middle|X_{t_{i-1}} = x\right] = \exp(-\beta\Delta t)P_1^{\left(\frac{1}{2}(\beta-2), -1+\frac{3\beta}{2}\right)}(\sin(x)),$$

then

$$E_\theta\left[\sin(X_{t_i})\,\middle|X_{t_{i-1}} = x\right] = \exp(-\beta\Delta t)\left(\frac{1}{2}(-1 + \exp(\beta\Delta t)\right) + \sin(x)).$$

In a similar way (but with more complicated calculations), we can then compute from the second eigenvalue and from the second eigenfunction, the conditional expected value $E_\theta\left[\sin^2(X_{t_i})\,\middle|X_{t_{i-1}}=x\right]$.

Godambe-Heyde's optimal MEF in $\mathcal{G}_n$ is

$$G_n^*(\beta)=\sum_{1=i}^{n}\psi_*\left(X_{t_i},X_{t_{i-1}},\beta,\Delta t\right),$$

where

$$\psi_*\left(X_{t_i},X_{t_{i-1}},\beta,\Delta t\right)=\frac{A(X_{t_i},X_{t_{i-1}},\beta,\Delta t)}{B(X_{t_i},X_{t_{i-1}},\beta,\Delta t)},$$

with

$$\begin{aligned}A(X_{t_i},X_{t_{i-1}},\beta,\Delta t)\quad:\quad&=2e^{\Delta t}\left(\frac{1}{2}+\beta\right)(1+\beta)\\&\times\left(-\frac{1}{2}(1+e^{\beta\Delta t})+\frac{1}{2}\beta(1+\Delta t)+(1+\beta\,(-1+\Delta t))\sin(X_{t_{i-1}})\right)\\&\times\left(-\frac{1}{2}+e^{\beta\Delta t}\left(\frac{1}{2}+1\right)+\sin(X_{t_{i-1}})+\sin(X_{t_i})\right)\end{aligned}$$

and

$$\begin{aligned}B(X_{t_i},X_{t_{i-1}},\beta,\Delta t)\quad:\quad&=e^{\Delta t(1+\beta)}\left(-\frac{1}{2}-\beta\right)+\\&-\frac{3}{4}e^{\Delta t(1+2\beta)}\left(1+\beta\Delta t\right)-\frac{1}{2}\left(-\frac{89}{20}+\beta\right)\left(\frac{9}{20}+\beta\right)+\\&+\frac{1}{2}e^{\Delta t}\left(\frac{1}{2}+\beta\right)(1+\beta)+\\&+\left(-2e^{\Delta t}\left(\frac{1}{2}+\beta\right)(1+\beta)+(e^{(1+\beta)\Delta t}+\beta)\,(1+2\beta)\right)\sin(X_{t_{i-1}})\\&+2\left(-1+e^{\Delta t}\right)\left(\frac{1}{2}+\beta\right)(1+\beta)\sin^2(X_{t_{i-1}}).\end{aligned}$$

In Figure 7, we show the graph of ψ_*, for the particular parameterization $\beta=2$, $\Delta t=0.2$ and where we label $X_{t_i}:=y$ and $X_{t_{i-1}}:=x$

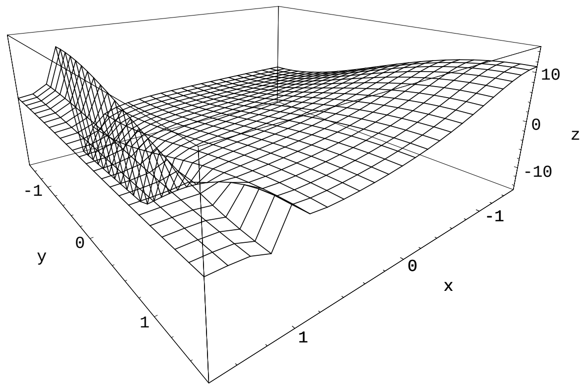

Fig.7 Estimating function $\psi_*(y, x, \beta, \Delta t)$, with $\beta = 2$ and $\Delta t = 0.2$)

This function is bounded over the state space and then its self standardized sensitivity is bounded. According to our considerations about the Von Mises expansion for the asymptotic bias of an M-estimator (see equation (3.1.4), in Chapter 3), this feature should imply that the M-estimator defined by ψ_* should be, somehow, robust.

This is an interesting property of this example, which is mainly related to the specific form of the eigenfunctions. To analyse in more details this feature of our trigonometric transformations of the Jacobi diffusion, we select a specific parametrization of the SDE in equation (4.2.3). To this end, we simplify the expression of the drift in the SDE to obtain more manageble calculations of the optimal MEF. Therefore, we will work on the diffusion process solution to the SDE[2]

$$\begin{cases} dX(t) = -\theta tanX(t)dt + dW(t) \\ \qquad X(0) = x_0 = 0. \end{cases}$$

We first determine the eigenvalues and eigenfunctions that are necessary to define the optimal MEF for this process (with the simplified drift).

The Sturm-Liouville problem for the Infinitesimal Generator is

$$L_\theta \phi(x, \theta) = -\lambda \phi(x, \theta) = -\theta \tan x \phi^{'}(x, \theta) + \frac{1}{2}\phi^{''}(x, \theta),$$

with spectrum

$$\left\{ i\left(\theta + \frac{i}{2}\right),\ i = 0, 1, 2.. \right\},$$

[2]This is the trigonometric diffusion studied by Kessler and Sorensen in Example 2.2., in [28], page 306.

and eigenfunctions

$$\phi_i(x,\theta) = Geg_i^{(\theta)}(\sin(x)),$$

where $Geg_i^{(\theta)}(x)$ are the Gegenbauer polynomials.

Given that we are working on the drift and we are interested in the estimation of the parameter determining the strength of the mean reversion ($\theta \in \mathbb{R}^+\backslash[0,\frac{1}{2})$), the optimal MEF can be defined by using only $Geg_1^{(\theta)}(\sin(x))$ and it is:

$$G_n^*(\theta) = \sum_{i=1}^{n} \frac{\sin(X_{t_i-1})\left[\sin(X_{t_i}) - \exp(-\theta\Delta t - \frac{1}{2}\Delta t)\sin(X_{t_{i-1}})\right]}{\frac{1}{2(1+\theta)}\left(\exp(2\left(1+\theta\right)\Delta t) - 1\right) - \left(\exp(\Delta t). - 1\right)\sin^2(X_{t_{i-1}})}. \qquad (4.2.4)$$

In Figure 8 we present the graph of the function

$$\psi_*(X_{t_{i-1}}, X_{t_i}, \theta, \Delta t) = \frac{\sin(X_{t_i-1})\left[\sin(X_{t_i}) - \exp(-\theta\Delta t - \frac{1}{2}\Delta t)\sin(X_{t_{i-1}})\right]}{\frac{1}{2(1+\theta)}\left(\exp(2\left(1+\theta\right)\Delta t) - 1\right) - \left(\exp(\Delta t) - 1\right)\sin^2(X_{t_{i-1}})}$$

for the parameter choice $\theta = 2$ and $\Delta t = 0.2$. As usual we set $x = X_{t_{i-1}}$ and $y = X_{t_i}$.

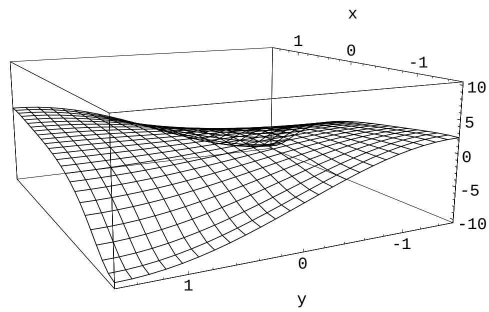

Fig.8 Estimating function ψ_*, with $\theta = 2$ and $\Delta t = 0.2$

Figure 8 shows that this function is bounded over the state space. Figure 9, furthermore, presents the graph of the self-standardized sensitivity.

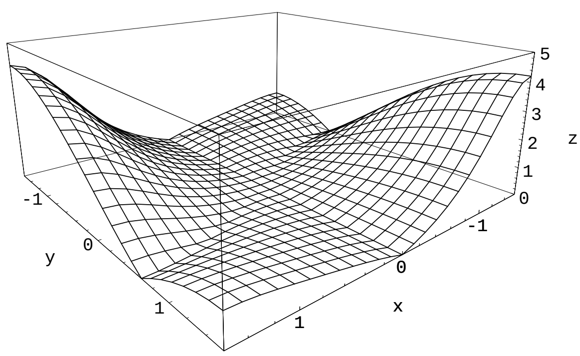

Fig.9 Function $\left|V_{\psi^*}^{-1/2}(\theta_0) IF_{\psi^*}^{cond}(y = X_{t_i}, x = X_{t_{i-1}}, \theta_0)\right|$

Since ψ_* is bounded, it defines a Conditionally Unbiased M-estimator for θ, having a bounded self-standardized sensitivity (see Figure 9). Therefore this MEF defines a Robust M-estimator, that should give a good performance, even in presence of contaminations.

An interesting question is to understand how much is robust the M-estimator implied by ψ_*. This is an important question, that highlights the fundamental role of the bounding constant. As a matter of fact, to different values of the bound on the self-standardized sensitivity, correspond different degrees of robustness[3]. In particular, the higher is the bounding constant, the lower is the degree of robustness of the M-estimator. Nevertheless, there is a trade-off between robustness and efficiency. From Robust Statistics theory we know that the smaller is bounding constant, the larger is the loss of efficiency of the M-estimator.

In order to understand the degree of robustness of the M-estimator implied by ψ_* and in order to have a numerical value of its asymptotic bias, we performe a Monte Carlo analysis. Before starting this analysis, we do a further important consideration. In their paper for *Bernoulli*, Kessler and Sørensen suggest to use a Taylor expansion for the function at the denominator of $G_n^*(\theta)$. More specifically, for Δt small, they prose an approximately optimal MEF

$$\begin{aligned}\tilde{G}_n(\theta) &= \sum_{1=i}^{n} \frac{\sin(X_{t_{i-1}})\left[\sin(X_{t_i}) - \exp(-\theta\Delta t - \frac{1}{2}\Delta t)\sin(X_{t_{i-1}})\right]}{\cos^2(X_{t_{i-1}})} \\ &= \sum_{1=i}^{n} \tilde{\psi}(X_{t_i}, X_{t_{i-1}}, \theta, \Delta t) \qquad (4.2.5)\end{aligned}$$

According to this approximation, Kessler and Sørensen observe that $G_n^*(\theta)$ and $\tilde{G}_n(\theta)$ are equivalent, when we have a small sampling interval. However those two MEFs are not equivalent. In Figure 10, we present the graph of the approximately optimal estimating function, for the parameter choice $\theta = 2$ and $\Delta t = 0.2$. As usual we set $x = X_{t_{i-1}}$ and $y = X_{t_i}$.

[3]By definition (see Chapter 3), a not robust estimator has a bounding constant diverging to infinity.

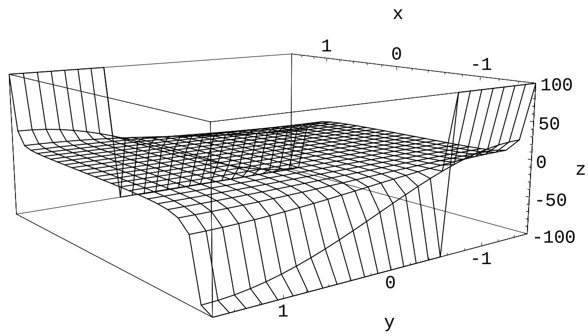

Fig.10 Estimating function $\tilde{\psi}$, for $\theta = 2$ and $\Delta t = 0.2$

Near to the boundaries, the function $\tilde{\psi}$ diverges to infinity. As a consequence, the asymptotic bias of the M-estimator implied by $\tilde{\psi}$, is unbounded.

Therefore, we conclude that, even if $G_n^*(\theta)$ and $\tilde{G}_n(\theta)$ are equivalent (for Δt small) in "classical sense", they have completely different robust implications.

We test those theoretical considerations in some Monte Carlo simulations.

4.2.2 Monte Carlo Analysis

The Monte Carlo analysis provides a useful tool in order to study the behavior of the M-estimator, implied by the MEFs in equation (4.2.4) and in equation (4.2.5).

In the Section 2.2.1, we do a direct comparison between the performaces of the M-estimators defined by ψ_* and by $\tilde{\psi}$. In particular, using the replacement model to simulate contaminated trajectories, we obtain a numerical value for the bias under contamination that both the M-estimators can have. Moreover, we show that, under the same contamination, the approximately optimal MEF defines an M-estimator that can have a bigger bias than the bias of the M-estimator defined by ψ_*. This different performance is due to the different degree of robustness owned by the two M-estimators. As a matter of fact, the M-estimator defined by ψ_* has a stronger degree of robustness than the M-estimator defined by $\tilde{\psi}$ (that, we recall, is not robust, since it has an unbounded IF^{cond}, see Figure 10).

Nevertheless, our Monte Carlo analysis in presence of contaminated samples shows that ψ_* is not enough robust, because it defines an M-estimator that can have a large bias. Therefore, if we want an M-estimator, having a small bias in presence of contaminations, we need to implement our robustification procedure. To this end, in the Section 2.2.2, we implement our Second Proposal and we define a Robust Conditionally Unbiased M-estimator, that reduces the asymptotic bias in presence of contamination. In order to apply our algorithm defined in Section 3.6, we show how we can use the spectral decomposition to approximate analytically the τ.

Optimal MEF vs Approximately Optimal MEF: robust features

In this subsection we study the performances of the M-estimators defined by ψ_* and by $\tilde{\psi}$. We start our analysis, by a simulation of the reference model in SDE (4.2.4). We use the Milstein scheme, while the contaminated process is simulated by the replacement model (see Example 1).

In our Monte Carlo analysis, we use the following parameters:

δ	Δt	θ	η	ξ
0.001	0.2	2	1.2	0.01

Table 3

where δ is the discretization step for the Milstein scheme. Every trajectory used for the inference procedure contains 2250 observations.

Figure 11 and Figure 12 present one simulated trajectory, for the Clean and for the Contaminated process, respectively.

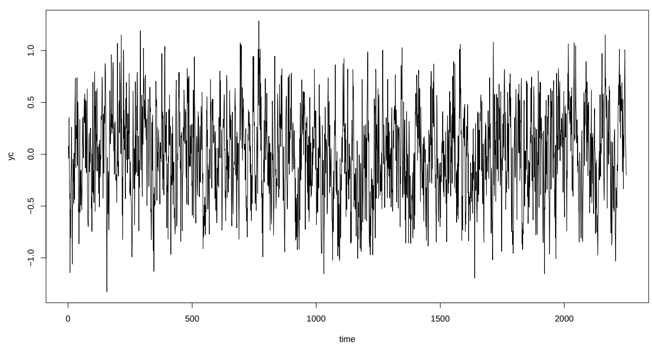

Fig.11 Clean trigonometric process, $\theta = 2$ and $\Delta t = 0.2$

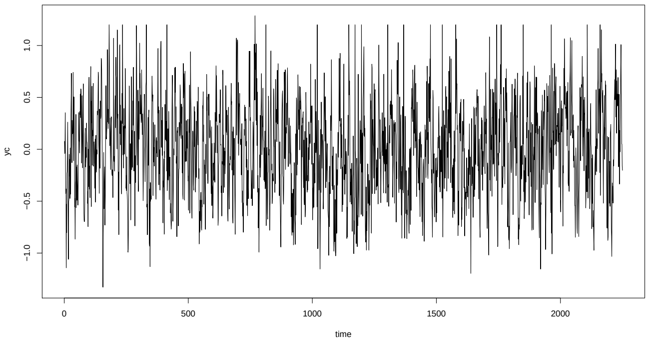

Fig.12 Contaminated trigonometric process:$\eta = 0.01$ and $\xi = 1.2$

We observe that the contamination probability η is very small and the contaminating values ξ are very close to the bulck of the observations. As a consequence, the trajectory for the the Clean and for the Contaminated process look very similar. In this simulation study, we use some positive contaminating values that drive the process in a non natural way and that determine some changes in the speed of the mean reversion.

We label by θ_n^* the M-estimator defined as the implicit solution to

$$G_n^*(\theta_n^*) = \sum_1^k \psi_*(X_{t_i}, X_{t_{i-1}}, \theta_n^*) = 0,$$

and e label by $\tilde{\theta}_n$ the M-estimater obtained from

$$\tilde{G}_n(\tilde{\theta}_n) = \sum_1^k \tilde{\psi}(X_{t_i}, X_{t_{i-1}}, \tilde{\theta}_n) = 0.$$

In order to study the behavior of the M-estimators $\tilde{\theta}_n$ and θ_n^*, we run our R-algorithm 100 times. Table 4 and Table 5 give the estimated values of the strength of mean reversion, under the Clean and the Contaminated Process:

1. For θ_n^*

Clean Process		Contaminated Process	
Mean	SD	Mean	SD
2.00	0.132	2.42	0.181

Table4

2. And for $\tilde{\theta}_n$

Clean Process		Contaminated Process	
Mean	SD	Mean	SD
2.01	0.170	2.97	0.350

Table5

Look at the first column of Table 4 and of Table 5. We observe that at the reference model, both ψ_* and $\tilde{\psi}$ define Conditionally Unbiased M-estimators that show a very similar performance. Both the M-estimators are very precise and have a small SD.

As far as the performance under contamination is concerned, this Monte Carlo analysis gives a numerical value to the theoretical considerations that we have done in Section 4.2.1. As a matter of fact, looking at Figure 8 and at Figure 10, we have concluded that ψ_* has a higher degree of robustness than $\tilde{\psi}$.

The results in the second column of Table 4 and Table 5, quantify the different behavior of the M-estimators, under contamination. We may notice that the bias of $\tilde{\theta}_n$ is larger than the bias of θ_n^*. In particular, the selected contamination determines a bias of 21% for θ_n^*, while it determines a bias of 48.5% for $\tilde{\theta}_n$. More than double!

In Figure 13, we present the kernel density for θ_n^* (left graph) and for $\tilde{\theta}_n$ (right graph) both under the Clean (continuous line) and the Contaminated (dotted line) processes.

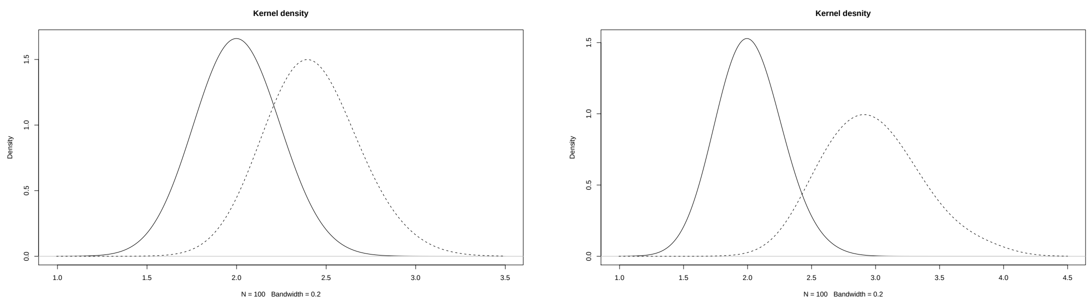

Fig.13: Kernel density for θ_n^* (left) and for $\tilde{\theta}_n$ (right). Continuous line: reference model,

Dotted line: contaminated model

The last Figure presents a graphical overview of the different effects that the same contamination has on the M-estimators. We may notice that the contamination determines for both the M-estimators a shift in the mean and an increase in the standard deviation. Nevertheless, $\tilde{\theta}_n$ has a stronger shift in the mean and a bigger rise in the standard deviation than θ_n^*.

This example has two main implications:

- It shows that a small contamination can determine dramatic changes in the performance of $\tilde{\theta}_n$ and θ_n^*. Nevertheless, the optimal MEF ψ_* defines a more stable M-estimator than the approximated one.
- It highlights that, even if θ_n^* has a bounded IF^{cond}, it can have a bad performance in presence of contamination.

Therefore we conclude that if we want to define a Conditionally Unbaised M-estimator that is stable under contamination and that has a good performance even under small departures from the reference model, we need to improve the degree of robustness of ψ_*. To this end, we define a robustifyied version of ψ_* by applying our second proposal.

Second Proposal for the Trigonometric diffusion

In the next example we study the effects on the performance of θ_n^*, determined by a very small contamination. This Monte Carlo analysis is useful to show that Kessler and Sørensen optimal MEF ψ_* defines an M-estimator that can be biased, even if we have very small departures from the reference model. In particular, we show that the M-estimator defined by ψ_* can have a bad performance, even if we contaminate the reference model by a small probability (η) and by some values ξ very close to the bulk of the observations. This is a serious inferential problem, because in almost every real situations, we have a sample that contains some "slightly strange" values that do not follow exactly the reference model. Think, for instance, to the case of small measurement errors in Physics or to the case of strange observations in Finance, due to some rumors that, for a few days, move the market in an unusual way. This Monte Carlo study will show that, in those situations, ψ_* can determine a biased inference.

This is the reason why we apply our second proposal: our robust inference procedure provides a very useful tool to correct the bias in this kind of situations.

In the Monte Carlo analysis we use the following parameters:

δ	Δt	θ	η	ξ
0.001	0.2	2	1	0.005

Table 6

where δ is the discretization step for the Milstein scheme. Any trajectory used for the inference procedure contains 2250 observations.

In Figure 14 and Figure 15, we present one simulated trajectory for the Clean and for the Contaminated Process, respectively.

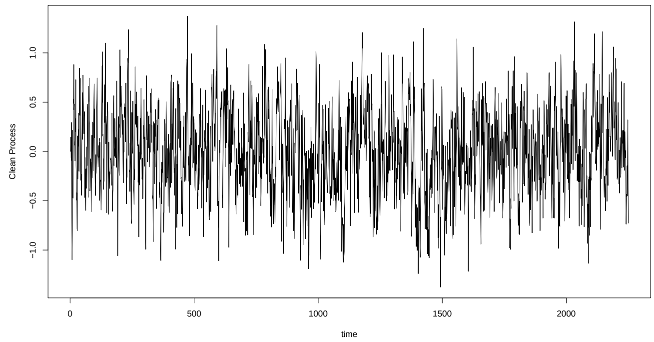

Fig.14 Clean trigonometric Process, $\theta = 2$ and $\Delta t = 0.2$

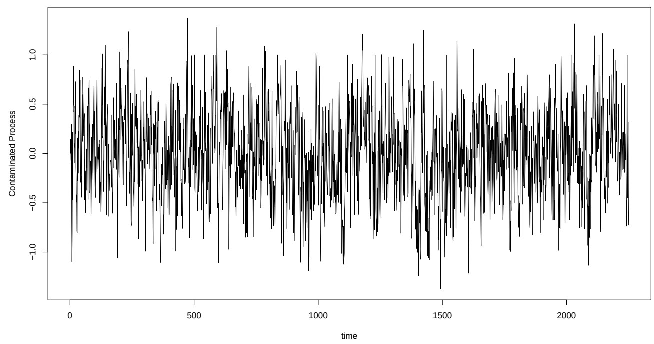

Fig.15 Contaminated trigonometric Process: $\eta = 0.005$ and $\xi = 1$

Figure 14 and Figure 15 show that the trajectories for the reference model and for the contaminated model look very similar. Nevertheless, out Monte Carlo analysis highlights that this small contamination can have strong effect on the bias of Sørensen's M-estimator.

Before starting our analysis and applying our second proposal, we remark a key point, necessary for the definition of our Robust Conditionally Unbiased M-estimator.

Key Point: Analytical Approximation for τ If we want to implement our robust inference procedure in a R-algorithm, there is a technical problem to solve. As a matter of fact, in Chapter 3 we pointed out that, to define a Robust MEF, we need to compute a τ-vector, necessary to preserve the Fisher Consistency.

Classical Robust methods (see for instance RGMM, [36] or REMM,[35]) calculate τ by some internal Monte Carlo simulations. This feature makes robust algorithms very slow, time consuming and computational demanding. However in the setting of Example 2 we can avoid these Monte Carlo simulations. As a matter of fact, we know the analytical solution to the S-L problem for the Infinitesimal Generator (see Section 4.2.1). Therefore we can avoid the time-consuming internal simulations and we can define an analytical approximation for τ.

As a matter of fact, from Section 3.6, we know that is possible to use a Generalized Fourier Series Expansion, in order to approximate the conditional expected values in the equation for $\tau(X_{t_{i-1}},\theta)$ (see equation (3.5.26)).

We recall that (see equation (3.5.32) in Section 3.5):

$$\begin{aligned} p(\Delta t,x,y,\theta) &= \sum_{n=0}^{\infty} \frac{\exp(-\lambda_n(\theta)\Delta t)\phi_n(x,\theta)\phi_n(y,\theta)m(y,\theta)}{\int_l^r \phi_n^2(y,\theta)m(y,\theta)dy} \\ &= \sum_{n=0}^{\infty} \exp(-\lambda_n(\theta)\Delta t)\phi_n(x,\theta)\phi_n(y,\theta)m(y,\theta)\pi_n(\theta), \end{aligned}$$

where $\pi_n(\theta) := \frac{1}{\int_l^r \phi_n^2(y,\theta)m(y,\theta)dy}$.

We observe that, using a finite order series for the transition density, we can provide accurate analytical approximations for $\tau(X_{t_{i-1}},\theta)$.

To this end let us set

$$\begin{aligned} p(\Delta t,x,y,\theta) &\approx \sum_{n=0}^{q} \frac{\exp(-\lambda_n(\theta)\Delta t)\phi_n(x,\theta)\phi_n(y,\theta)m(y,\theta)}{\int_l^r \phi_n^2(y,\theta)m(y,\theta)dy} \\ &= \sum_{n=0}^{q} \exp(-\lambda_n(\theta)\Delta t)\phi_n(x,\theta)\phi_n(y,\theta)m(y,\theta)\pi_n(\theta), \end{aligned} \tag{4.2.6}$$

with $q < \infty$. We can use this approximated transition probability in order to obtain a numerical value for the conditional expected values in the numerator and in the denominator of $\tau(X_{t_{i-1}},\theta)$.

A crucial point in this procedure is the choice of the number (q) of eigenfunctions-eigenvalues that we will use in the spectral series representation (4.2.6). From a practical point of view, the main problem is due to the definition of a criterion that we can use in order to select the number of terms that we will take into account in the series expansion.

In section 3.6, we have observed that there is not any golden rule to select this value. Furthermore, we have highlighted that there exists a trade-off between the accuracy of the approximation (related to the number q) and the computational burden of the algorithm. In this paragraph we provide a practical approach to select the number q of eigenvalues-eigenfunctions.

In a preliminary mathematical study we have analyzed the behavior of the series expansion in (4.2.6) for some known transition probability densities. For instance, we have studied the case of the OU-process (having a Gaussian transition density) and the case of the CIR-process (whose transition density is related to the the First Kind Modified Bessel Function). In this analysis, we have noticed that the series expansion in (4.2.6) behaves very well in the central region of the state space for whatever degree of the expansion, but the goodness of the fit in the tails of the distribution depends on q. In particular, we have noticed that for small values of q (implying a low approximation order), the approximated transition probability density takes negative values in the tails. As a consequence, to select the q value in formula (4.2.6) for the diffusion process of our example, we have plotted the approximated transition probability functions, obtained from (4.2.6), for different q-order expansions. We have calculated these approximation in $X_{t_i-1} = \xi = 1$, since ξ is the contamination value that we artificially introduce to stress the reference distribution in an extreme point, near to right tail. To select the value of q, we have chosen the order of series expansion that provides a well behaving approximation in both the tails.

To make it clear and to give a graphical overview of this selection procedure, in Figure 16 and in Figure 17, we present the graph of the approximated transition probability density that we obtain, taking into account two different numbers of eigenvalues: $q = 1$ and $q = 5$. In particular, Figure 15 shows the approximated transition probability at $X_{t_i-1} = \xi = 1$, obtained by using only the first term of the series ($q = 1$).

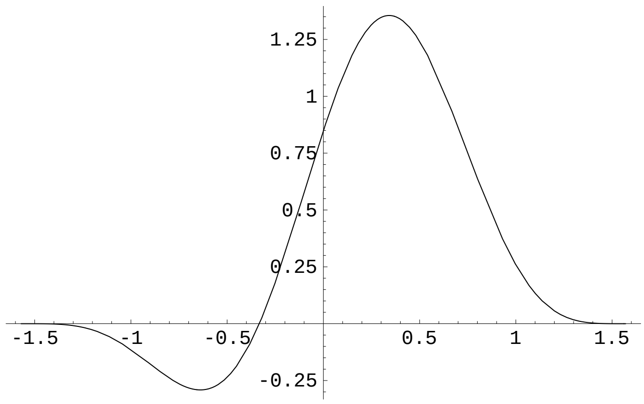

Fig.16 Approximated $p(X_{t_i-1} = \xi, X_{t_i} = 1, \theta = 2, \Delta t = 0.2$, with $q = 1)$

Figure 16, shows that setting $q = 1$ gives a rough approximation. As a matter of fact, we observe that the left tail is very irregular and the function takes negative values. In order to define a well behaving series expansion, we have increased the number of eigenvalues-eigenfunctions. In this selection procedure, we have noticed that the first approximation defining a function that at $X_{t_i-1} = \xi$ takes only positive values is the series with five terms ($q = 5$).

In Figure 17, we present the graph of the approximated transition probability obtained using the first five eigenfunctions-eigenvalues: $q = 5$.

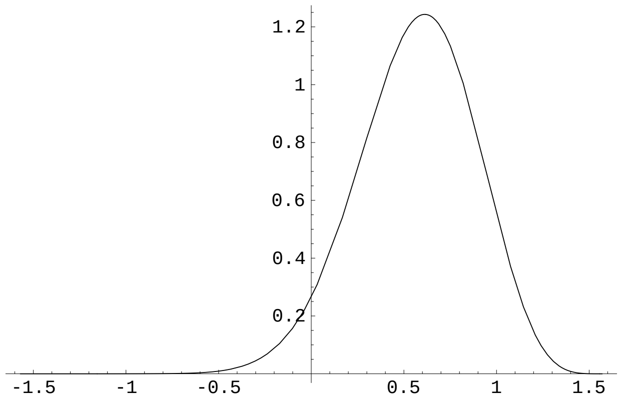

Fig.17 Approximated $p(X_{t_i-1} = \xi, X_{t_i} = 1, \theta = 2, \Delta t = 0.2$, with $q = 5)$

Figure 17 shows that the function obtained with $q = 5$ has a good behavior in both the tails.

To conclude our selection procedure, we have asked ourself what happens if we augment the number of eigenfunctions and eigenvalues. As far as this topic is concerned, we remark that (see equation (4.2.6)) the weights of the terms related to the eigenvaules of high order decrease exponentially. As a matter of fact if we calculate the weights related to eigenvalues of higher order than 5, we may notice that they are very small. As a consequence, if we add new terms, related to the sixth or seventh (or more) eigenvalue (that is if we use $q = 6, 7...$), we obtain only a small improvement in the accuracy of the approximation. Therefore we have decided to set $q = 5$, in formula (4.2.6).

Monte Carlo Analysis: the Output In this final paragraph, we present the output of a Monte Carlo Analysis. To test the behavior of our Robust MEF we do a direct comparison between the performances of the Sørensen's M-estimator implied by ψ_* and our Robust M-estimator, defined by the second proposal. According to the discussion in the last paragraph, in our R-algorithm, we approximate the conditional expected values in τ, by a spectral decomposition of the probability transition density as in (4.2.6), setting $q = 5$.

The bounding constant for the robust Conditionally Unbiased M-estimator was set $b = 3$.[4] We have observed that this choice of b provides a strong reduction in the bias of the M-estimator and implies a small efficiency loss. As a matter of fact, at the reference model, our robust M-estimator has a slightly bigger SD than Sørensen's M-estimator. If we calculate the ratio between the SD of our robust M-estimator and the SD of Sørensen's M-estimator, we obtain a value of 96.15% and we observe that the our robust M-estimator determines a loss of efficiency of about 3.85%.

In Table 7, we present the performances of the two M-estimators of θ at the Clean and at the Contaminated model. The first and the second column give the performance of Sørensen's M-estimator, respectively, at the reference model and under contamination. In the third and the fourth column we provide the performance of our robust M-estimator, at the reference model and under contamination.

[4] At the time being, there does not exist a formal criterion for choosing b in our specific setting. Nevertheless, in future developments, we could follow M-R-T's approach ([33], page 638) and we could define a bounding constant based on the theory of Robust Tests. In particular, M-R-T [33] defined this theory in an ARCH setting and in some future researchs we could try to define an analogous theory for diffusion processes observed in discrete points in time. We consider this topic as a possible field for future researchs.

	Sør_Cln_Pr	Sør_Cnt_Pr	Rob_Cln_Pr	Rob_Cnt_Pr
Mean	2.00	2.16	1.98	2.05
Median	1.99	2.16	1.97	2.04
SD	0.125	0.142	0.130	0.134
Mean_bias%	0%	8%	1%	2.5%
Median_bias%	0.5%	8%	1.5%	2%
MSE	0.016	0.047	0.017	0.020

Table 7

The performance of Sørensen's M-estimator at the reference model is very good. Nevertheless, we notice that, under contamination, Sørensen's M-estimator gives upward biased estimates. The bias (in percentage) on the mean (and/or on the median) is about 8%. Moreover, as we have already noticed for the other Monte Carlo analysis (see, for instance, Example 1), under contamination, Sørensen's M-estimator has a bigger SD. This bad performance, both in terms of bias and in terms of SD, is the effect of the small degree of robustness owned by Sørensen's M-estimator.

Column 3 and 4 of Table 7 present the good performance of our robust M-estimator. First we observe that at the reference model our robust M-estimator has a slightly larger SD than Sørensen's M-estimator. This implies that the second proposal defines, for this example, a Robust, Conditionally Unbiased M-estimator, having a small loss of efficiency. Furthermore, we notice that our robust M-estimator corrects the bias under contamination, providing a bias-reduction of about 4 times (both for the median and for the mean).

In Figure 18 we show a graphical comparison of the boxplots for the distributions of Sørensen's and our robust M-estimator, both at the reference model and under contamination.

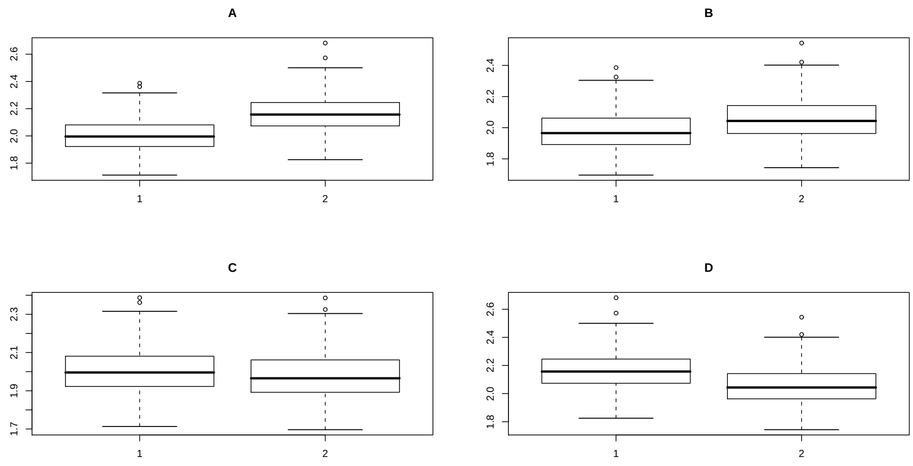

Fig.18 A: Sør Est. for the Clean (1) and for the Cont.(2) process, B: Robust Est. for the Clean (1) and for the Cont.(2) process, C: Sør Est. (1) and Robust Est (2) for the Clean process, D: Sør Est.(1) and Robust Est(2) for the Contamianted process

This figure presents the different behaviors of the two M-estimators. We highlight that our robust M-estimator corrects the effect due to the contamination, having a smaller shift in the mean and a smaller rise in the SD, than Søresen's M-estimator.

To observe the good performance of our M-estimator, look at Figure 18 C and Figure 18 D, that present a direct comparison of the distributions of the two M-estimators. Figure 18 C presents the spread between the distributions of Sørensen's and our Robust M-estimator, under the reference model: the two distributions are very similar! Figure 18 D, highlights the spread between the distributions of Sørensen's and our Robust M-estimator in presence of contamination: the two distribution are different! Our robust M-estimator is closer to the true value $\theta = 2$ and has a smaller dispersion.

Furthermore, from Figure 18 A, it is evident the impact of the contamination on the distribution of Sørensen's M-estimator: there is a shift in mean and a rise in the SD. Figure 18 B, shows that the same contamination has a smaller impact on on the distribution of our robust M-estimator: there is a small shift in mean and a very small rise in the SD.

Finally, in Figure 19 we provide a general overview and we present a graph containing the boxplots for the distributions of Sørensen's M-estimator (both under the reference model (1) and under contamination (2)) and of our robust M-estimator (both under the reference model (3) and under contamination (4)).

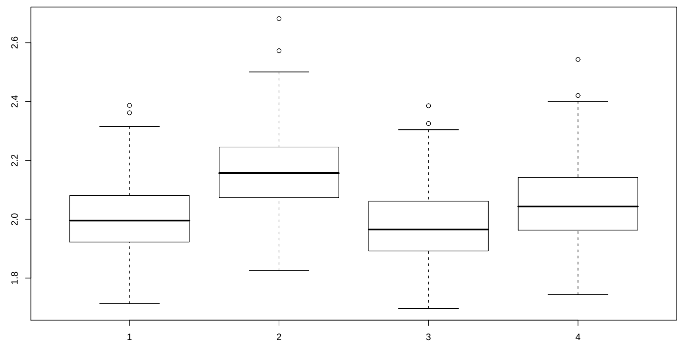

Fig.19 :(1) Sørensen estimator at the reference model; (2) Sørensen estimator under contamination; (3) Robust estimator at the reference model; (4) Robust estimator under contamination;

Looking at those graphs, we conclude that our second proposal defines a very useful tool to correct the bias that Sørensen M-estimator has, even in presence of small departures from the reference model.

Finally, as far as the asymptotic distribution at the reference model is concerned, from Chapter 3 (see section 3.7) we know that our second proposal defines a Conditionally Unbiased M-estimator having a Gaussian Asymptotic Distribution. Moreover, form Kessler and Sørensen [28], we know that ψ_* has a Gaussian Asymptotic Distribution, as well.

In Figure 20, we present a comparison between the quantiles of a Gaussian distribution and the quantiles of the distribution both of Sørensen's (left graph) M-estimator and of our robust (right graph) M-estimator.

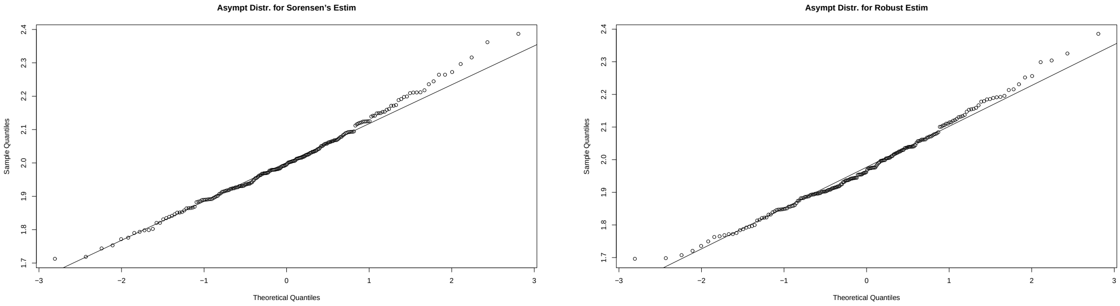

Fig. 20: QQ-plots for the finite sample distributions of the two M-estimators

Figure 20 confirms the good approximation of the distribution of both the M-estimators by means of the Gaussian distribution.

Conclusion

This PhD. Thesis defines Robust Martingale Estimating Functions for discretely observed diffusion processes. We start our analysis from the remark that in many diffusion models the solution of the Fokker-Planck equation is not known in analytical closed form. Moreover only discrete-time observations are available. This implies that the discrete-time transition probabilities are unknown in analytical closed form, therefore the classical Maximum Likelihood approach is usually not tractable. Furthermore, we notice that in Robust Statistics there does not exist a specific inference procedure for diffusion processes. More precisely, in Chapter 2, we highlight that the robust methods applied in time series setting (RGMM, REMM or Robust Indirect Inference) are too generic and they do not take into account the specific features of the diffusion processes. Because of that, these methods define very time-consuming inference procedures, based on internal Monte Carlo simulations.

So far, the only specific results in a diffusion setting are due to Nakairo Yoshida [41] (1988). Yoshida's pioneering job is an important theoretical research, but it does not define estimators that can be applied in practice. As a matter of fact, Yoshida proposes a robustified version of the Likelihood score and defines a contamination of the invariant measure. Moreover he assumes that the Statistician is given a continuum of observations. In Chapter 2, we highlight that these assumptions have some important drawbacks. Therefore, in the Thesis we propose a new class of feasible Robust M-estimators, specific for discretely-observed diffusion processes and that do not involve the unknown Likelihood score. In particular, in Chapter 3, we define our framework and we clarify (following the Künsch's theory in [30]) the concept of local robustness, applied in our development.

We first suggest that a possible solution to define robust M-estimators could be to use the theory of Martingale Estimating Functions (MEFs), defined by the Danish School (represented by Bibby, Jacobsen, Kessler and Sørensen). Nevertheless, we highlight that Søresen's MEFs generally define Conditionally Unbiased M-estimators that are not robust. An alternative solution could be represented by Mancini, Ronchetti and Trojani's research [33], that following Künsch's theory, defined Robust Conditionally Unbiased M-estimators. Nevertheless their approach is based on the (Pseudo) Likelihood score function and this feature determines some drawbacks (see Chapter 3).

In Section 3.4 we solve our inferential problem and we determine two proposals, that define Robust Conditionally Unbiased M-estimators. Both our first and our second proposal take into account the special features of diffusion processes and they define M-estimators that, under standard assumptions, are $sqrt(n)$-consistent, have an asymptotically normal distribution and are optimal in L_2 (in some specific sense). Nevertheless we remark that if we define an algorithm for our first proposal, we can meet some technical problems. In contrast, the second proposal does not have this kind of problems. As a consequence, in Section 3.6, we write an algorithm that implements the second proposal. As far as this topic is concerned we remark, that one of the main contributions of the Thesis, is the definition of an inference procedure that, using the solutions of Sturm-Liouville problem for the Infinitesimal Generator, approximates analytically some conditional expected values. When the eigenfunctions and the eigenvalues are known in analytical closed form, our second proposal avoids internal Monte Carlo simulations, a computational demanding task that typically characterizes other robust inferential procedures in time series.

Finally, in the last chapter, we illustrate the theoretical construction discussed in Chapter 3. Monte Carlo simulations shown the negative impact that small departures from the reference model can have on the performance of Sørensen's M-estimators. In Section 4.2.2 we show that our second proposal defines a Conditionally Unbiased M-estimator that reduces this impact, maintaining high efficiency at the reference model.

Bibliography

[1] Aït-Sahalia, Y.: "Nonparametric pricing of interest rate derivative securitires", *Econometrica*, 1996, 64, 527-560

[2] Aït-Sahalia, Y.: "Maximum likelihood estimation of discretely sampled diffusions: a closed-form approach", *Econometrica*, 2002 , 70, 223-262.

[3] Bibby, B.M. and Sørensen, M.: "Martingale estimating function for discretely observed diffusion processes". *Bernoulli*, 1995, 1, 17-39.

[3] Bibby, B.M, Jacobsen, M. and Sørensen, M.: *"Estimating Functions for Discretely Sampled Diffusion-Type Models"*, for Ait-Sahalia's Handbook, 2004.

[4] Boos, D.D. and Serfling, R.J.: "A note on differentials and the CLT and LIL for statistical functions, with applications to M-estimates". *Annal. Statist.*, 1980 , 8, 618-624.

[5] Clarke, B.R.: "Uniqueness and Fréchet differentiability of functionals solutions to maximum likelihood type equations". *Annal. Statist.*, 1983 , 11, 1198-1205.

[6] Clarke, B.R.: "Nonsmooth analysis and Fréchet differentiability of M-functionals". *Probab. Th. Rel. Fileds.*, 73, 197-209.

[7] Clarke, F.R. *"Optimization and nonsmooth analysis"*, New york, Wiley, 1983.

[8] C.H. Edward jr, D.E. Penney "Elementary Differential Equations with Boundary Value Problens", III ed, Prentice Hall, New Jersey, 1993.

[9] Florens-Zmirou, D.: "Approximate discrete-time schemes for statistics in diffusion processes", *Statistics*, 1989, 20, 547-557

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

[11] Genon-Catalot,V, Jantheau, T and Laredo, C.: "Stochastic Volatility Models as Hidden Markov Models annd Statistical Applications", *Bernoulli*, 2000, 6, 1051-1079

[12] Genton,M. and Ronchetti, E.: "Robust Indirect Inference", *JASA*, 2003, 98, 67-76

[15] Gourieroux, C., Monfort, A., Renault E.: "Indirect Inference", "Journal of Applied Econometrics", 1993, 8, S85-S118

[16] Hampel, F.R.: "The Influence Curve and Its Role in Robust Estimation". *JASA*, 1974, 69, 383-393

[17] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., *Robust Statistics: The Approach based on Influence Functions*, John Wiley, New York, 1986

[18] Hansen, L.P, "Large Sample Properties of GMM estimators". *Econometrica*, 1982, 50, 1029-1054

[19] Hansen,L.P and Scheinkman, J.A., "Back to the future: Generating moment implication for continuous time markov processes". *Econometrica*, 1995, 63 (4), 767-804

[20] Hansen,L.P, Scheinkman, J.A., Touzi N.: ""Spectral methods for identifyng scalar diffusions". *Journal of Econometrics*, 1998, 86 (1), 1-32",

[21] Heritier, S. and Ronchetti E.: "Robust bounded-influence tests in general parametric models", *JASA*, 1994, 897-904

[22] Heyde, C. C., *"Quasi-Likelihood and Its Application"*, 1997, Springer-Verlag, New York

[23] Huber, P.J., *Robust Statistics*, New York,1981, Wiley

[24] Itô,K. and McKean, H.P., *Diffusion Processes and their Sample Paths*, New York,1974, Springer-Verlag.

[25] Jacobsen, M., “Discretely Observed Diffusions: Classes of Estimating Functions and Small Δ-optimality”. *Scandivian Journal of Statisitcs*, 2001, 28, 123-150

[26] Karlin S. and Taylor, H.M., “*A First Course in Stochastic Procesess*”, 1981, Orlando, FL: Academic Press

[27] Karlin S. and Taylor, H.M., “*A Second Course in Stochastic Procesess*”, 1981, Orlando, FL: Academic Press

[28] Kessler, M. and Sørensen, M., “Estimating functions based on eigenfunctions for a discretely observed diffusion process”, *Bernoulli*, 1999, 5, 299-314

[29] Kloeden, P.E. and Platen, E.,“*Numerical Solution of Stochastic Differential Equations*”, 1999, Springer-Verlang, New York

[30] Künsch, H.: “Infinitesimal Robustness for Autoregressive Processes”, *The Annals of Statistics*, 1984, 12, 843-863

[31] Larsen, K.S. and Sørensen, M. : “A diffusion model for exchange rates in a target zone”. Preprint No.6, Department of Applied Mathematics and Statistics, University of Copenhagen, 2003.

[32] Lipster, R.S. and Shiryayev, A. N.,“*Statistics of Random Processes*”, Vol. I, 1977, Springer-Verlang, New York

[33] Mancini, C., Ronchetti, E. and Trojani, F.: “Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models”, *JASA*, 2005, 628-641

[34] Martin, R.D. and Yohai, V.J.: “Influence function for Time Series”. *The Annals of Statistics*, 1986, 14, 781-818

[35] Ortelli, C. and Trojani, F.: “Robust Efficient Method of Moments”, *Journal of Econometrics* , 2005, 128, 69-97

[36] Ronchetti, E. and Trojani, F.: “Robust Inference With GMM Estimators”, *Journal of Econometrics* , 2001, 101, 37-69

[37] Sørensen H., *"Inference for Diffusion Processes and Stochastic Volatily Models"*, PhD. Thesis, Department of Theoretical Statistics, University of Copenhagen, 2000

[38] Sørensen, M.: "On Asymptotics of Estimating Functions", *Brazilian Journal of Probability and Statistics*, 1999, 13:111-136

[39] Stefanski, L.A., Carroll, R.J. and Ruppert, D.: "Optimally Bounded Score Functions for GLM with Applications to Logistic Regression ", *Biometrika*, 1986, 73, 413-425

[40] H.F. Weinberger "A First Course in Partial Differential Equations", 1965, Xerox

[41] Yoshida, N.: "Robust M-Estimators in diffusion processes," *Ann. Inst. Statist. Math.*, 1988, 4, 799-820.

[42] Yosida, K.: "*Functional Analysis*", Springer-Verlag, 1968, New York.