Universität St.Gallen

Infinitesimal Robustness for Diffusions

Davide La Vecchia and Fabio Trojani

April 2008 Discussion Paper no. 2008-09

Department of Economics University of St. Gallen

Editor:	Prof. Jörg Baumberger University of St. Gallen Department of Economics Bodanstr. 1 CH-9000 St. Gallen Phone +41 71 224 22 41 Fax +41 71 224 28 85 Email joerg.baumberger@unisg.ch
Publisher:	Department of Economics University of St. Gallen Varnbüelstrasse 19 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 22 98
Electronic Publication:	http://www.vwa.unisg.ch

Infinitesimal Robustness for Diffusions

Davide La Vecchia and Fabio Trojani

Author's address: Prof. Dr. Fabio Trojani
Swiss Institute of Banking and Finance
Rosenbergstrasse 52
9000 St. Gallen
Phone +41 71 2247074
Fax +41 71 2247088
Email fabio.trojani@unisg.ch
Website www.sbf.unisg.ch

Abstract

We develop infinitesimally robust statistical procedures for general diffusion processes. We first prove existence and uniqueness of the times series influence function of conditionally unbiased M–estimators for ergodic and stationary dffusions, under weak conditions on the (martingale) estimating function used. We then characterize the robustness of M–estimators for diffusions and derive a class of conditionally unbiased optimal robust estimators. To compute these estimators, we propose a general algorithm, which exploits approximation methods for dffusions in the computation of the robust estimating function. Monte Carlo simulation shows a good performance of our robust estimators and an application to the robust estimation of the exchange rate dynamics within a target zone illustrates the methodology in a real–data application.

Keywords

Dffusion processes, Eigenexpansion, Influence Function, Infinitesimal Generator, M–Estimators, Saddle– point Approximation.

JEL Classification

C13, C22, C32

Infinitesimal Robustness for Diffusions

Davide La Vecchia and Fabio Trojani*

Abstract

We develop infinitesimally robust statistical procedures for general diffusion processes. We first prove existence and uniqueness of the times series influence function of conditionally unbiased M–estimators for ergodic and stationary diffusions, under weak conditions on the (martingale) estimating function used. We then characterize the robustness of M–estimators for diffusions and derive a class of conditionally unbiased optimal robust estimators. To compute these estimators, we propose a general algorithm, which exploits approximation methods for diffusions in the computation of the robust estimating function. Monte Carlo simulation shows a good performance of our robust estimators and an application to the robust estimation of the exchange rate dynamics within a target zone illustrates the methodology in a real–data application.

MSC: 34L05, 60G10, 62G35, 62M05

Keywords: Diffusion processes, Eigenexpansion, Influence Function, Infinitesimal Generator, M–Estimators, Saddle– point Approximation.

First Version: September 2006. **This Version:** April 2, 2008.

---

*Davide La Vecchia is at the University of Lugano, Switzerland. Fabio Trojani is at the University of St. Gallen, Switzerland. E–mail adresses: Davide.La.Vecchia@lu.unisi.ch and Fabio.Trojani@unisg.ch. We gratefully acknowledges the financial support of the Swiss National Science Foundation (NCCR FINRISK, ProDoc project "PhD in Economics and Finance" and grants 101312-103781/1 and 100012-105745/1). The usual disclaimer applies.

1 Introduction

This paper develops infinitesimally robust statistical procedures for general strictly stationary and ergodic diffusion processes. We first characterize the infinitesimal robustness of estimators for diffusions by deriving existence and uniqueness of their time series influence function. Second, based on this characterization we define and derive optimal robust estimators for diffusions, in the class of conditionally unbiased M–estimators. Third, we develop a computationally feasible general algorithm, which can be applied to compute efficiently our robust estimators in applications. Fourth, we study by Monte Carlo simulation the performance of our estimator. Finally, we demonstrate its applicability to real–data by estimating the exchange rate dynamics of a target zone model.

The need for robust statistical methodologies is now widely recognized, both in the statistical and the econometric literature. Several important contributions have studied infinitesimally robust, i.e. bounded–influence, estimators and tests for i.i.d. data; See, among others, Hampel [36], Koenker and Bassett [29], Huber [23], Koenker [30], Markatou and Ronchetti [37], Krishnakumar and Ronchetti [33] and Genton and Ronchetti [36]. Recent research has addressed the infinitesimal robustness problem in the more general time series context. Künsch [34] introduces a formal definition of the time series influence function and constructs optimal robust estimators from it for linear autoregressive processes. Ronchetti and Trojani [41], Mancini, Ronchetti and Trojani [36] and Ortelli and Trojani [39] construct infinitesimally robust estimators and tests for strictly stationary time series using M–type estimators with bounded estimating function. In these papers, a well–defined time series influence function is assumed implicitly as given, and no general sufficient condition for its existence and uniqueness is stated. Künsch [34] proves existence and uniqueness of the influence function of linear autoregressive processes when the M–estimating function satisfies a linear growth condition. This result is applicable to our setting only for (Gaussian) Ornstein-Uhlembeck diffusion processes. More generally, we prove existence and uniqueness of the influence function for square integrable estimating functions of strictly stationary and ergodic diffusions.

Diffusion processes are used for the statistical analysis of many important problems arising in different research areas.[1] To our knowledge, however, optimal robust estimation for diffusions has been formally studied so far only in Yoshida [44], who develops a Huber-type estimator for problems in which observations can be collected continuously over time. This assumption restricts the usage of robust methods for many applications like, e.g., those typically encountered in many financial models. Therefore, we investigate the general case in which observations can be collected discretely over time. Moreover, we study robust estimation in the class of conditionally unbiased estimators, since these estimators can exploit more conveniently the information embedded in the conditional transition density of the underlying diffusion process. In contrast to these ones, Yoshida's [44] estimator is an unconditional one defined through the invariant measure of the process. This feature simplifies the analysis of its infinitesimal robustness, which is easily characterized using the standard Influence Function, as in in the i.i.d. context; See, among others, Hampel [19] and Hampel et al. [20]. The characterization of robustness in the class of conditionally unbiased estimators requires a formal treatment of the time series influence function, which allows us to define conditionally

[1] Diffusion processes are very popular models in Finance and other disciplines like Engineering, Physics and Biology; see, e.g., Gallant and Tauchen [15] for an application to term structure modeling, and Bodo et al. [5] or Sobczyk [42] for applications to Hydrology and Physics, respectively. Stephens and Donnely [11] present several examples of inference problems related to molecular Population Genetics. Applications in Biology, Population Dynamics, Protein Kinetics, Genetics, Experimental Psychology and Neuronal Activity are reviewed in Kloeden and Platen [28, Ch.7] and Karlin and Taylor [26].

unbiased robust estimators for a general diffusion process.

A major issue in the development of optimal robust estimators for diffusions is that the discrete-time transition density of these processes is rarely available in closed-form.[2] Therefore, it is generally not possible to obtain robust M–estimators for diffusions by robustifying directly the conditional maximum likelihood score function. In order to obtain optimal robust estimators for the general diffusion setting, we exploit several approximation methods to the discrete–time Maximum Likelihood score of diffusion processes. When the conditional Laplace transform of the process exists, we use saddle–point methods to obtain these approximations, following the methodology derived in Aït-Sahalia and Yu [3]. In the other cases, we robustify martingale estimating functions derived either from an eigenexpansion of the discrete–time density of the process or by a projection of this density on a given set of martingale estimating functions. Kessler and Sørensen [27] and Bibby, Jacobsen and Sørensen [4], among others, show how these approximations can be efficiently computed in a general diffusion context.

We obtain optimal robust estimators for diffusions by deriving the most efficient estimator with bounded influence function in the class of conditionally unbiased estimators. These estimators are obtained as solutions to the self–standardized version of Hampel's optimality problem, analyzed – among others – in Krasker and Welsch [32], Künsch, Stefansky and Carroll [7] and Mancini, Ronchetti and Trojani [36]. We show that given a verifiable condition our robust estimators are admissible with respect to a natural class of robust $M-$estimators. Our robust estimators for diffusions are constructed

[2]Many estimation methods for diffusions are based on computationally intensive procedures that compute the expected (pseudo) Maximum-Likelihood score either by Monte Carlo simulation or via the numerical solution of a Fokker-Planck equation; See, e.g., Gallant and Tauchen [14], Gouriéroux Monfort and Renault [18], Duffie and Singleton [12] and Poulsen [40]. Robust estimators based on these approaches are even more computationally demanding; See, e.g., Ortelli and Trojani [39].

using a set of Huber's weights that downweight the impact of influential observations. To preserve conditional unbiasedness at the model, a conditional location correction is needed, which depends on a set of expectations computed under the process distribution. In time series, these corrections have typically to be computed by Monte Carlo simulation, as for instance for the Robust Efficient Method of Moments in Ortelli and Trojani [39]. Using the properties of diffusion processes, we are able to produce general approximations for these corrections that largely reduce the computation time of our estimators.

We analyze by Monte Carlo simulation the performance of our robust estimator in two estimation problems, which consider a trigonometric and a Jacobi diffusion, respectively, and find that it implies a quite favorable tradeoff between efficiency and robustness. Finally, we estimate in a real-data example the parameters of a Jacobi diffusion modeling the exchange rate dynamics in a target zone. Here, we find that both the mean reversion and volatility parameter estimates of classical procedures are highly sensitive to influential data points, which can be linked ex-post, using our robust procedure, to abnormal market events realized before and during periods of turbulent financial markets.

Section 2 introduces our general diffusion setup, defines the infinitesimal robustness problem in this setting, and proves existence and uniqueness of the influence function for estimators of diffusion processes. It also summarizes several approaches to obtain martingale estimating functions for these processes. In Section 3, we first define our robust conditionally unbiased estimators for diffusions and derive their optimality properties. In a second step we describe the algorithm to compute it and approximation procedures that can be applied to reduce the computation time. The Monte Carlo simulation study and the real-data application are presented in Section 4. Section 5 summarizes and concludes. All proofs are in the Appendix.

2 Infinitesimal Robustness for Diffusions

As a statistical model for discrete–time observations $\{X_0, \ldots, X_T\}$, we consider a diffusion process $\mathcal{X} := \{X(t) : t \geq 0\}$ taking values in the state space $\mathcal{S} := (l, r)$, with $-\infty \leq l < r \leq \infty$, and defined on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$. $\mathcal{X}$ satisfies the stochastic differential equation:

$$dX(t) = \alpha(X(t), \theta)dt + \sigma(X(t), \theta)dW(t) \ , \ X(0) = x_0 \ , \tag{1}$$

where $\theta \in \Theta \subset \mathbb{R}^p$, $W := \{W(t) : t \geq 0\}$ is a standard Brownian motion, and for known functions $\alpha(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ that satisfy regularity conditions detailed below. If the statistical model implied by diffusion process (1) is correctly specified, one can draw inference about parameter θ by assuming that $X_i := X(i\Delta)$ for all $i = 0, 1, \ldots, n$ and some fixed $\Delta > 0$. The goal of infinitesimal robustness is to construct accurate estimators of θ that are robust, in an appropriate sense, with respect to moderate deviations from parametric diffusion process (1).

2.1 General Setting

Given the diffusion process (1), the scale measure $S(\theta)$ and the speed measure $M(\theta)$ of this process are absolutely continuous measures on $\mathcal{S}$, having densities

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

and

$$m(x, \theta) = \frac{1}{\sigma^2(x, \theta)s(x, \theta)} \ ; \ x \in \mathcal{S} \ ,$$

respectively, with respect to Lebesgue measure. $s(x, \theta)$ and $m(x, \theta)$ are standard tools that are useful to study the stationarity and the ergodicity of diffusion processes; see, e.g., Genon-Catalot, Jeantheau and Laredo [16]. The following general assumption is needed for our theory.

Assumption 1 **(I.)** *For all* $x \in \mathcal{S}$*,* $\alpha(x, \cdot)$ *is continuously differentiable,* $\sigma(x, \cdot)$ *is twice continuously differentiable and* $\sigma(x, \cdot) > 0$*. Moreover,* α *and* σ *satisfy the growth conditions:*

$$|\alpha(x, \theta)| \leq C(\theta) (1 + \|x\|) \;\; ; \;\; \sigma^2(x, \theta) \leq C(\theta) \left(1 + \|x\|^2\right) \tag{2}$$

for some function $C(\theta) > 0$ *and all* $x \in \mathcal{S}$*.* **(II.)** *For every* $\theta \in \Theta$ *and* $x^{\#} \in \mathcal{S}$*:*

$$\int_{x^{\#}}^{r} s(x, \theta) dx \;=\; \int_{l}^{x^{\#}} s(x, \theta) dx = \infty \;, \tag{3}$$

$$A(\theta) \;=\; \int_{l}^{r} m(x, \theta) dx < \infty \;. \tag{4}$$

(III.) *For every* $\theta \in \Theta$*, the probability distribution* $\mu(\cdot, \theta)$ *of* $X(0)$ *has density defined by*

$$\mu(dx, \theta) := \frac{m(x, \theta)}{A(\theta)} dx \tag{5}$$

with respect to Lebesgue measure. **(IV.)** $\sigma(x, \theta) m(x, \theta) \to 0$ *as* $x \downarrow l$ *and* $x \uparrow r$*.* **(V.)** *Let* $\rho(x, \theta) = \partial_x \sigma(x, \theta) - 2\alpha(x, \theta)/\sigma(x, \theta)$*. The limits* $\lim_{x \downarrow l} 1/\rho(x, \theta)$ *and* $\lim_{x \uparrow r} 1/\rho(x, \theta)$ *are both finite for all* $\theta \in \Theta$*.*

Conditions (I.) and (II.) in Assumption 1 ensure existence and uniqueness of the solution of (1) together with its recurrence. Condition (III.) is needed for the strict stationarity of $\mathcal{X}$. Under this condition, $\mu(\theta)$ is the stationary invariant distribution of $X(t)$. Conditions (IV.) and (V.) ensure the ergodicity of both $\mathcal{X}$ and the discrete-time Markov chain $\{X(i\Delta) : i \in \mathbb{N}\}$.

Our main focus is on conditionally unbiased M–estimators, which are defined by estimating functions that satisfy the martingale difference property.

Definition 2 **(I.)** *A Martingale Estimating Function (MEF) for diffusion process* $\mathcal{X}$ *is a function* $\psi : \mathbb{R}^2 \times \Theta \to \mathbb{R}^p$ *such that:*

$$E_\theta[\psi(X_i, X_{i-1}; \theta) | X_{i-1}] = 0 \tag{6}$$

for all $\theta \in \Theta$. **(II.)** *A conditionally unbiased estimator* $\hat{\theta} := \{\hat{\theta}_T : T \in \mathbb{N}\}$ *is a sequence of solutions of the implicit equations:*

$$\sum_{i=1}^{T} \psi(X_i, X_{i-1}; \hat{\theta}_T) = 0 \ ; \ T \in \mathbb{N} \ . \tag{7}$$

By the law of iterated expectations, a conditionally unbiased estimator that is consistent for unknown parameter $\theta_0 \in \Theta$ is asymptotically the unique solution of the population moment equation, when diffusion process (1) is correctly specified:

$$E_{P_{\theta_0}}[\psi(X_i, X_{i-1}; \theta_0)] = 0 \ , \tag{8}$$

where $E_{P_\theta}[\cdot]$ denotes expectation with respect to the marginal probability distribution P_θ of (X_i, X_{i-1}), defined for any Borel set B by:

$$P_\theta(B) = \int_B p_\theta(x_i|x_{i-1}) dx_i \mu(dx_{i-1}, \theta) \ , \tag{9}$$

with the conditional density function $p_\theta(x_i|x_{i-1})$ of the diffusion process.

It follows that the asymptotic statistical functional implied by a conditionally unbiased $M-$estimator in our diffusion setting takes the form:

$$\hat{\theta}(F) = \theta \quad \Leftrightarrow \quad E_F[\psi(X_i, X_{i-1}; \hat{\theta})] = 0 \ , \tag{10}$$

where F is the two–dimensional marginal distribution of a strictly stationary process. In the sequel, we denote by $\mathcal{M}$ the family of all two–dimensional marginal distributions of a stationary process and consider conditionally unbiased M–estimator functionals, i.e., functionals $\hat{\theta}(\cdot) : dom(\hat{\theta}) \subset \mathcal{M} \to \mathbb{R}^p$ of the form (10) that satisfy condition (6) with respect to the underlying process probability P. $\mathcal{P} := \{P_\theta : \theta \in \Theta\}$ is the statistical model implied by diffusion process (1).

2.2 Infinitesimal Robustness

Infinitesimal robustness studies robust statistical procedures that are robust with respect to local deviations from a given parametric model. Let $P_{\epsilon,\nu} := \epsilon P_{\theta_0} + (1-\epsilon)\nu$, with $\nu \in \mathcal{M}$, be a generic $\epsilon-$contamination of the parametric diffusion probability P_{θ_0}, where $\epsilon \leq \eta$ for fixed $0 < \eta < 1$. By $\mathcal{U}_\eta(P_{\theta_0})$ we denote the local neighborhood of all such contaminations. First order infinitesimal robustness studies estimators with uniformly bounded linearized asymptotic bias B, defined by:

$$B(\epsilon, \nu) = \epsilon \frac{\partial \hat{\theta}(P_{\epsilon,\nu})}{\partial \epsilon}\Big|_{\epsilon=0} \tag{11}$$

for all $\nu \in \mathcal{M}$ such that this derivative exists. An influence function is just a kernel representing this bias.

Definition 3 **(I.)** *An Influence Function (IF) for conditionally unbiased estimator $\hat{\theta}(\cdot)$ of diffusion process (1) is any function $IF : \mathcal{S}^2 \times \Theta \to \mathbb{R}^p$ such that:*

$$\frac{\partial \hat{\theta}(P_{\epsilon,\nu})}{\partial \epsilon}\Big|_{\epsilon=0} = \int_{\mathcal{S}^2} IF(x_1, x_0; \theta_0)\nu(dx_0, dx_1) \tag{12}$$

for all $\nu \in \mathcal{M}$. **(II.)** *The Conditional Influence Function is any influence function $IF^c : \mathcal{S}^2 \times \Theta \to \mathbb{R}^p$ such that:*

$$E_{\theta_0}\left[IF^c(X_i, X_{i-1}; \theta_0)|X_{i-1}\right] = 0 \ . \tag{13}$$

If an IF exists, it can be used to describe the linearized asymptotic bias of estimator functional $\hat{\theta}(\cdot)$ over neighborhood $\mathcal{U}_\eta(P_{\theta_0})$. In general, several versions of the IF exist, which build a class of equivalent kernels satisfying condition (12); Given a kernel $IF(x_1, x_2; \theta_0)$ any other version of the IF is of the form:

$$IF(x_2, x_1; \theta_0) + g(x_1; \theta_0) - g(x_2; \theta_0) \ , \tag{14}$$

where $g : \mathcal{S} \times \Theta \to \mathbb{R}^p$ is an arbitrary function such that for all $\nu \in \mathcal{M}$:

$$\int_{\mathcal{S}^2} (g(x_1;\theta_0) - g(x_2;\theta_0))\nu(dx_2, dx_1) = 0 \ . \tag{15}$$

Künsch [34, Theorem 1.3] proves existence an uniqueness of the conditional IF for linear autoregressive processes under a linear growth condition on the estimating function used. He then uses the conditional IF to define and construct optimal robust estimators for autoregressive processes. These robust estimators imply a bounded asymptotic bias under contamination because their conditional IF is bounded. To construct robust estimators for diffusion processes, we first prove appropriate existence and uniqueness results for the conditional IF of these processes. Unfortunately, Künsch [34] result applies to our diffusion context only for the very specific setting of a (Gaussian) Ornstein Uhlembeck diffusion, which has a linear autoregressive process as its exact time–discretization.

Proposition 4 *Let Assumption 1 be satisfied.* **(I.)** *If function $f : \mathcal{S}^2 \to \mathbb{R}^p$ is $P_{\theta_0}-$square integrable and such that:*

$$E_{\theta_0}[f(X_i, X_{i-1})] = 0 \tag{16}$$

then there exist a $\mu(\theta_0)-$square integrable function $g : \mathcal{S} \to \mathbb{R}^2$, which is unique up to an additive constant, such that:

$$E_{\theta_0}[f(X_i, X_{i-1}) + g(X_{i-1}) - g(X_i)] = 0 \ . \tag{17}$$

(II.) *If, in addition:*

$$E_{\theta_0}\left(f(X_i, X_{i-1})\,|X_{i-1}\right) = 0,$$

then function g is constant.

Remark 5 *Proposition 4 extends Theorem 1.3 in Künsch [34] to our diffusion setting. It holds for all diffusion processes satisfying Assumption 1. This* *class includes the strictly stationary and ergodic Ornstein Uhlembeck process. Moreover, it follows from the proof of Proposition 4 that these results apply more generally to any discrete–time strictly stationary and ergodic Markov process.*

The main implication of Proposition 4 is that there can exit at most one P_{θ_0}–square integrable IF having zero conditional expectation under the diffusion process distribution. Therefore, the conditional IF of conditionally unbiased M–estimators for such processes is unique in the class of P_{θ_0}–square integrable IF and can be characterized easily:

Corollary 6 *Let $\hat{\theta}$ be a conditionally unbiased M−estimator for diffusion process (1), defined by a P_{θ_0}–square integrable estimating function $\psi : \mathcal{S}^2 \times \Theta \to \mathbb{R}^p$. Then the conditional IF of $\hat{\theta}$ is uniquely given by:*

$$IF_{\psi}^{c}\left(x_i, x_{i-1}; \theta_0\right) = -D(\psi, \theta_0)^{-1} \psi\left(x_i, x_{i-1}; \theta_0\right), \tag{18}$$

where $D(\psi, \theta_0) := E_{\theta_0}\left(\nabla_{\theta'} \psi(X_i, X_{i-1}; \theta_0)\right)$.

An important consequence of Corollary 6 is that infinitesimally robust conditionally unbiased M–estimators for diffusions are those with a bounded estimating function, because their linearized asymptotic bias under contamination, which is uniquely determined by the conditional IF, is bounded. It follows that any other conditionally unbiased M-estimator with unbounded P_{θ_0}–square integrable estimating function implies an unbounded bias under contamination over neighborhood $\mathcal{U}_\eta(P_{\theta_0})$. This characterization is the starting point to define and construct optimal robust estimators for diffusion processes.

2.3 Martingale Estimating Functions for Diffusions

Several procedures to obtain efficient martingale estimating functions for discretely-observed diffusions are known in the literature. Most of them imply M–estimators that are not robust.

2.3.1 Exact and Approximate Maximum Likelihood

Maximum Likelihood: If the discrete time conditional transition density $p_\theta(x_i|x_{i-1})$ of the diffusion process is known in closed–form, conditionally unbiased Maximum Likelihood M-estimators for θ_0 are readily available, using the estimating function:

$$\psi_{ML}(X_{t_i}, X_{t_{i-1}}; \theta) = \nabla_{\theta'} \ln p_\theta(X_i|X_{i-1}) \ ; \ \theta \in \Theta \ . \tag{19}$$

Unfortunately, for diffusion processes $p_\theta(x_i|x_{i-1})$ is rarely known in closed–form. Moreover, when it is known explicitly it is most of the times unbounded, which implies the non–robustness of associated Maximum Likelihood estimator.

Example 7 *Consider the well–known CIR [8] diffusion process:*

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

This process satisfies Assumption 1 for the parameter choice $\beta, \alpha, \sigma > 0$ and its discrete–time conditional transition density is known in closed–form. However, the conditional IF of this estimator is unbounded because ψ_{ML} is unbounded. The top left Panel of Figure 1 illustrates this feature by plotting $\psi_{ML}(X_i, X_{i-1}; \theta)$ for the case where only parameter β is estimated[3]*. This function diverges to infinity as either X_i or X_{i-1} increases.*

Approximate Maximum Likelihood by saddle–point methods: If the discrete time conditional transition density is not known explicitly, one possibility is to approximate it accurately using saddle–point methods. The necessary assumption is the existence of the cumulant generating function $K_\theta(\varrho|X_{i-1}) :=$

[3]In Figures 1-6, we plot the absolute value of the corresponding estimating functions.

$\log E_\theta[\exp(\varrho X_i)|X_{i-1}]$ of the process. When $K_\theta(\varrho|X_{i-1})$ is known explicitly, the saddle–point approximation method can be applied directly; see, e.g., Daniels [9] . When $K_\theta(\varrho|X_{i-1})$ is not known explicitly, an additional approximation with respect to the time step Δ can be applied; see Aït-Sahalia and Yu [3]. Let $\hat{\varrho}$ be the solution of the following saddle–point equation:

$$\frac{\partial K_\theta(\hat{\varrho}|X_{i-1})}{\partial \varrho} = X_i. \tag{20}$$

The leading term in the saddle-point approximation of $p_\theta(X_i|X_{i-1})$ is:

$$p_\theta^{(0)}(X_i|X_{i-1}) = \exp(K_\theta(\hat{\varrho}|X_{i-1}) - \hat{\varrho}X_i)\left(2\pi\frac{\partial^2 K_\theta(\hat{\varrho}|X_{i-1})}{\partial \varrho^2}\right)^{-1/2}. \tag{21}$$

This approximation is accurate for many purposes. Moreover, it can be easily improved by considering higher order approximations. The estimating function for the implied approximate Maximum Likelihood estimator is:

$$\psi_{SP}(X_i, X_{i-1}, \theta) = \nabla_{\theta'} \ln p_\theta^{(0)}(X_i|X_{i-1}). \tag{22}$$

As for Maximum–Likelihood estimators, (the absolute value of) this estimating function is most of the time unbounded, as is illustrated in the top right Panel of Figure 1. When $K_\theta(\varrho|X_{i-1})$ is not known explicitly, it can be approximated with respect to the time step Δ using the structure of the infinitesimal generator of the diffusion process. Precisely, let $f : \mathcal{S} \to \mathbb{R}$ be a smooth function of class at least C^2. It then follows (see also Bibby, Jacobsen and Sørensen [4]):

$$E_\theta\left(f\left(X_i\right)|X_{i-1}\right) = \sum_{i=0}^{\varsigma} \frac{\Delta^i}{i!}\mathcal{L}_\theta^i f(X_{i-1}) + O\left(\Delta^{\varsigma+1}\right), \tag{23}$$

where

$$\mathcal{L}_\theta f(x_0) := \alpha(x_0,\theta)\frac{df(x_0)}{dx} + \sigma(x_0,\theta)\frac{d^2 f(x_0)}{dx^2} \tag{24}$$

is the infinitesimal generator of the diffusion process applied to function f.[4] In this way, one obtains the approximate cumulant generating function (see Aït-Sahalia and Yu [3, p. 521]):

$$\begin{aligned} K_\theta(\varrho|X_{i-1}) &= \log E_\theta(\exp(\varrho X_i)|X_{i-1}) \\ &= \log\left(\sum_{i=0}^{\varsigma}\frac{\Delta^i}{i!}\mathcal{L}_\theta^i e^{\rho\cdot}(X_{i-1})\right) + O\left(\Delta^{\varsigma+1}\right) \\ &= K_\theta^{(\varsigma)}(\varrho|X_{i-1}) + O\left(\Delta^{\varsigma+1}\right) \end{aligned}$$

Using this result, another approximation to $p_\theta(X_i|X_{i-1})$ follows by inserting $K_\theta^{(\varsigma)}(\varrho|X_{i-1})$ in the saddle–point equation (26) and the saddlepoint approximation (27):

$$p_\theta^{(0,\varsigma)}(X_i|X_{i-1}) = \exp(K_\theta^{(\varsigma)}(\hat{\varrho}|X_{i-1}) - \hat{\varrho}X_i)\left(2\pi\frac{\partial^2 K_\theta^{(\varsigma)}(\hat{\varrho}|X_{i-1})}{\partial\varrho^2}\right)^{-1/2}, \quad (25)$$

with $\hat{\rho}$ solving:

$$\frac{\partial K_\theta^{(\varsigma)}(\hat{\varrho}|X_{i-1})}{\partial\varrho} = X_i. \quad (26)$$

The corresponding martingale estimating function is:

$$\psi_{SP,\varsigma}(X_i, X_{i-1}, \theta) = \nabla_{\theta'}\ln p_\theta^{(0,\varsigma)}(X_i|X_{i-1}) \ . \quad (27)$$

However, also this estimating function is in most cases unbounded, as illustrated in the middle left Panel of Figure 1 for the choice $\varsigma = 1$.

2.3.2 Martingale Estimating Functions

A more general method to construct efficient martingale estimating functions can be applied also when the cumulant generating function of the process is

[4]Note that the computation of higher order terms in this approximation can be easily achieved, which implies good approximations for realistic time steps Δ arising in applications.

not known. The basic idea to obtain efficiency here is to project the unknown Maximum Likelihood score on a given space of available martingale estimating functions. This space can be generated, e.g., by estimating functions defined through some conditional moments of the diffusion process.

Closed-form conditional moments: Consider the family of martingale differences $\Gamma_k(\theta) := \Gamma_k(X_i, X_{i-1}; \theta) := X_i^k - E_\theta[X_i^k | X_{i-1}]$, $k = 1, \ldots, K \geq p$, and let:

$$\Gamma(\theta) := \Gamma(X_i, X_{i-1}; \theta) := (\Gamma_1(X_i, X_{i-1}; \theta), \ldots, \Gamma_K(X_i, X_{i-1}; \theta))' . \quad (28)$$

Using Godambe-Heyde theory, Bibby, Jacobsen and Sørensen [4] show that across estimating functions of the form $\nu(X_{i-1}; \theta)\Gamma(X_i, X_{i-1}; \theta)$, where $\nu(X_{i-1}; \theta)$ takes values in $\mathbb{R}^{p \times K}$, the $\mathcal{A}$–optimal function has a weight ν^* given by:

$$\nu^*(\theta) := \nu^*(X_{i-1}; \theta) = -E_\theta[\nabla_{\theta'} \Gamma(\theta) | X_{i-1}] E_\theta[\Gamma(\theta)\Gamma(\theta)' | X_{i-1}]^{-1} .$$

In this way, one obtains the optimal martingale estimating function:

$$\psi_{BJS}(X_i, X_{i-1}; \theta) = \nu^*(X_{i-1}; \theta)\Gamma(X_i, X_{i-1}; \theta) \quad (29)$$

that is nearest, in L_2-norm, to the unknown Maximum Likelihood score.[5] However, also these estimating functions are in most cases unbounded. The middle right Panel of Figure 1 illustrates this point for the case where $K = 1$ using the closed form expression for the first conditional moment of the CIR process, i.e., $E[X_i | X_{i-1}] = X_{i-1} \exp(-\beta\Delta) + \alpha(1 - \exp(-\beta\Delta))$.

Approximate conditional moments: When closed–form conditional moments are not available, it is possible to approximate them by applying an expansion in the time step Δ with formula (23) applied to function $f(X_i) = X_i^k$, for $k = 1, \ldots, K$. E.g., for $\varsigma = 1$ the approximation of the first moment of the CIR process reads: $E[X_i | X_{i-1}] = X_{i-1} + \Delta\beta(\alpha - X_{i-1}) + O(\Delta^2)$ and the

[5] Precisely, the L_2-norm here is measured with respect to the conditional distribution of X_i given X_{i-1} under probability measure P_θ.

approximation to function $\Gamma_1(X_i, X_{i-1}; \theta)$ readily follows. As highlighted in the bottom left Panel of Figure 1, also this estimating function is unbounded.
Sturm-Liouville problem: When the eigenvalues and eigenfunctions of the conditional expectation operator $T_\theta f(X_{i-1}) := E_\theta[f(X_i)|X_{i-1}]$ are known, we can define for $k = 1, \ldots, K$ the martingale differences $\Gamma_k(X_i, X_{i-1}; \theta) := \phi_k(X_i; \theta) - \mu_k(\theta)\phi_k(X_{i-1}; \theta)$, where $\phi_k(\cdot; \theta)$ is the k–th eigenvector of operator T_θ for eigenvalue $\mu_k(\theta)$. Equivalently, $\phi_k(\cdot : \theta)$ is eigenvector of the infinitesimal generator $\mathcal{L}_\theta$ in equation (24) for eigenvalue $\lambda_k(\theta) := -(\log \mu_k(\theta))/\Delta$, i.e., it is a solution of the Sturm-Liouville (S–L) problem (see also Karlin and Taylor [26]). Given the so–defined process $\Gamma(\theta) = [\Gamma_1(\theta), \ldots, \Gamma_K(\theta)]'$, Godambe-Heyde theory can be applied again to obtain the optimal estimating function across functions of the form $\nu(\theta)\Gamma(\theta)$. The solution of S–L problem is known for some processes, like Brownian motion, the Ornstein Uhlembeck and the Cox-Ingersoll-Ross process, or the Jacobi diffusion. Moreover, the solution of the S–L problem is also known for all diffusions that are obtained as invertible C^2 transformation of these processes; see, e.g., Kessler and Sørensen [27, p. 306]. However, most optimal estimating functions derived from solutions of S–L problems are also unbounded. For instance, the eigenfunctions of the Cox-Ingersoll-Ross process are the Laguerre polynomials (see Kessler and Sorensen [27]).[6] Therefore, the implied M–estimators are not infinitesimally robust.

3 Optimal Robust Conditionally Unbiased Estimators

In this section, we derive optimal robust M–estimators based on bounded martingale estimating functions, i.e., bounded conditional IF. As is common

[6]See also the bottom right Panel of Figure 1.

in the infinitesimal robustness literature, we robustify unbounded martingale estimating functions by introducing a set of Huber weights that downweight accordingly observations that are excessively influential.

3.1 Estimators with Bounded Self–Standardized Sensitivity

Let $\psi_\star(X_i, X_{i-1}; \theta)$ be an unbounded martingale estimating function for diffusion process (1). We consider standardized robust $M-$estimators, in Künsch [34] terminology.[7] Our robust estimator $\overline{\theta}$ is the M-estimator defined by the estimating function:

$$\begin{aligned} \psi_r(X_i, X_{i-1}; \theta) &:= A(\theta)\psi_b(X_i, X_{i-1}; \theta) \\ &:= A(\theta)(\psi_\star(X_i, X_{i-1}; \theta) - \tau(X_{i-1}; \theta))\omega(X_i, X_{i-1}; \theta) \end{aligned} \quad (30)$$

where for given $b \geq \sqrt{p}$ the Huber weight ω is defined by:

$$\omega(X_i, X_{i-1}; \theta) = \min\left(1, \frac{b}{\|A(\theta)(\psi_\star(X_i, X_{i-1}; \theta) - \tau(X_{i-1}, \theta))\|}\right) \quad (31)$$

with matrix $A(\theta) \in \mathbb{R}^{p\times p}$ and $\mathcal{F}_{i-1}$–measurable $p-$dimensional random vector $\tau(X_{i-1}; \theta)$ solving the implicit equations:

$$E_\theta[\psi_r(X_i, X_{i-1}; \theta)\psi_r(X_i, X_{i-1}; \theta)'] = I_p \ , \quad (32)$$

$$E_\theta[\psi_r(X_i, X_{i-1}; \theta)|X_{i-1}] = 0 \ . \quad (33)$$

M–estimator $\overline{\theta}$ defined by estimating function ψ_r is conditionally unbiased, because by construction $E_\theta[\psi_r(X_i, X_{i-1}; \theta)|X_{i-1}] = 0$, from the definition of the conditional location vector $\tau(X_{i-1}; \theta)$ in equation (33). Moreover, $\overline{\theta}$ has

[7] Following Huber [23], it is possible to consider also unstandardized optimal robust estimators. For brevity, we focus on standardized estimators here, even if our optimality results can be adapted to consider unstandardized estimators.

a self–standardized sensitivity, $\Upsilon_{\psi_r}(\theta)$, bounded by b^2; see also Hampel et al. [20]:

$$\begin{aligned}\Upsilon_{\psi_r}(\theta) &:= \sup_{(X_{i-1},X_i)\in\mathcal{S}^2} \Upsilon_{\psi_r}(X_i, X_{i-1};\theta)\\ &:= \sup_{(X_{i-1},X_i)\in\mathcal{S}^2} |IF^c_{\psi_r}(X_i, X_{i-1};\theta)' V^{-1}_{\psi_r}(\theta) IF^c_{\psi_r}(X_i, X_{i-1};\theta)|\\ &= \sup_{(X_{i-1},X_i)\in\mathcal{S}^2} |\psi_r(X_i, X_{i-1};\theta)'\psi_r(X_i, X_{i-1};\theta)| \leq b^2 \ ,\end{aligned}$$

using the formula for the covariance matrix V_{ψ_r} of M–estimator $\overline{\theta}$, the expression for the conditional IF in Corollary 6, and equation (32). Under standard conditions, estimator $\overline{\theta}$ is asymptotically normally distributed with covariance matrix $V_{\psi_r}(\theta_0) = [D(\psi_r, \theta_0)D(\psi_r, \theta_0)']^{-1}$.

Given a choice for $\psi_\star$, the robust estimating function ψ_b solves an optimization problem that implies admissibility of robust estimator $\overline{\theta}$ in an appropriate class of estimators. Let $D_\star(\psi, \theta) := E_\theta[\psi\psi_\star']$ and $V_\star(\psi, \theta) = D_\star(\psi, \theta)^{-1} W_\psi(\theta) D_\star(\psi, \theta)'^{-1}$, where $W_\psi(\theta) := E_\theta(\psi\psi')$. Next Proposition gives the optimality result for estimating function ψ_b.

Proposition 8 *If for given $b \geq \sqrt{b}$ equations (32) and (33) have solution $A(\theta_0)$ and $\tau(X_{i-1};\theta_0)$ then ψ_b minimizes tr $[V_\star(\psi, \theta_0)V_\star(\psi_b;\theta_0)^{-1}]$ among all martingale estimating functions ψ such that*

$$\sup_{(X_i,X_{i-1})\in\mathcal{S}^2} \psi(X_i, X_{i-1};\theta_0)' V_\star(\psi_b, \theta_0)^{-1} \psi(X_i, X_{i-1};\theta_0) \leq b^2 \ . \tag{34}$$

ψ_b is unique, up to multiplication by a constant matrix. **(II.)** *Assume that $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$. If there exists a (strongly efficient) martingale estimating function ϕ such that $E_{\theta_0}(\phi\psi_\star') = E_{\theta_0}[\nabla_{\theta'}\phi]$, $\Upsilon_\phi(\theta_0) \leq b^2$ and $V_\phi(\theta_0) - V_{\psi_b}(\theta_0)$ is negative definite, then ϕ is equivalent to ψ_b whenever the latter is defined.*

Statement (II.) of the proposition states that if estimating function ψ_b satisfies $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$ then there cannot exist another robust martingale estimating function ϕ more efficient than ψ_b and such that $E_{\theta_0}[\phi\psi_\star'] =$

$E_{\theta_0}[\nabla_{\theta'}\phi]$. Since $E_{\theta_0}[\nabla_{\theta'}\phi] = E_{\theta_0}[\phi(\nabla_{\theta'}\log p_\theta)']$ these estimating functions are orthogonal to the difference between $\psi_\star$ and the Maximum Likelihood score $\nabla_\theta \log p_\theta$. Note that condition $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$ depends only on the known estimating function $\psi_\star$. Therefore, it can be verified in applications.

Corollary 9 *(I.) Assume that $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$. Then, there cannot exist a robust martingale estimating function ϕ strictly more efficient than ψ_b and such that both $\Upsilon_\phi(\theta_0) \leq b^2$ and $E_{\theta_0}[\phi(\psi_\star - \nabla_{\theta'}\log p_\theta)'] = 0$ hold. (II.) Let $\psi_\star(X_i, X_{i-1}; \theta) = \nabla_{\theta'}\log p_\theta(X_i, X_{i-1}; \theta)$ be the Maximum Likelihood score. Then, $V_\star(\psi_b, \theta_0) = V_{\psi_b}(\theta_0)$ and there cannot exist a robust martingale estimating function ϕ strictly more efficient than ψ_b and such that $\Upsilon_\phi(\theta_0) \leq b^2$.*

Statement (II.) in Corollary 9 is the version of Corollary 1.1 in Stefansky, Carroll and Ruppert [6] for our time series diffusion setting. It applies when the discrete–time conditional Maximum Likelihood score function of the process is explicitly known, a situation that arises rarely. In this case, the robust estimator $\overline{\theta}$ is admissible in the class of conditionally unbiased robust estimators with $\Upsilon_\psi(\theta_0) \leq b^2$. More generally, admissibility of our robust estimator follows, provided that $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$, with respect to the smaller class of robust conditionally unbiased estimating functions orthogonal to $\psi_\star - \nabla_{\theta'}\log p_\theta$.

Remark 10 *When $\psi_\star$ is an efficient martingale estimating function in the sense of Bibby, Jacobsen and Sørensen [4], it is the orthogonal projection of the Maximum Likelihood score on a closed k–dimensional subspace $S_k(\theta_0)$ of martingale estimating functions, with respect to the L_2–scalar product under the P_{θ_0}–conditional distribution of X_i given X_{i-1}. Therefore, $V_\star(\psi_\star, \theta_0) = V_{\psi_\star}(\theta_0)$. Moreover, ψ_b converges to $\psi_\star$ in L_2–norm as $b \to \infty$, using Lebesque Theorem. Therefore, for b sufficiently large we can heuristically expect ψ_b to* *be admissible, up to a small approximation error, in the class of robust martingale estimating functions with $\Upsilon_\psi(\theta_0) \le b^2$ and orthogonal to the projection error $\psi_\star - \nabla_{\theta'} \log p_\theta$.*

Remark 11 *When state space $\mathcal{S}$ is bounded, the projection space S_k defining efficient martingale estimating function $\psi_\star$ in Bibby, Jacobsen and Sørensen [4] can be generated by the first k eigenfunctions $\{\phi_1(X_i), \ldots, \phi_k(X_i)\}$ of the generator of the diffusion process. It is well–known that in this case $\lim_{k\to\infty} S_k$ is dense in the space of L_2–estimating functions. Therefore, $\psi_\star$ converges to $\nabla_{\theta'} \log p_\theta$ in L_2–norm as $k \to \infty$, and $V_\star(\psi_b, \theta) \to V_{\psi_b}(\theta)$. It follows that for k sufficiently large we can heuristically expect ψ_b to be admissible, up to a small approximation error, in the class of martingale estimating functions with $\Upsilon_\psi(\theta) \le b^2$. When $\mathcal{S}$ is unbounded, additional conditions are needed to ensure this feature.*[8]

3.2 Computation of $\tau(X_{i-1}, \theta_0)$

In the definition of our robust estimator $\overline{\theta}$, the auxiliary $\mathcal{F}_{i-1}-$measurable random vector $\tau(X_{i-1}, \theta_0)$ solves the implicit equation (33), which is key to ensure conditional unbiasedness of the robust estimator at the parametric diffusion model. Solving equation (33) for τ, we obtain:

$$\tau(X_{i-1}; \theta_0) = \frac{E_{\theta_0}[\psi_\star(X_i, X_{i-1}; \theta_0)\omega(X_i, X_{i-1}; \theta_0)|X_{i-1}]}{E_{\theta_0}[\omega(X_i, X_{i-1}; \theta_0)|X_{i-1}]} \tag{35}$$

Note that since the Huber weight $\omega(X_i, X_{i-1}; \theta_0)$ depends on τ itself, this is a fixed–point equation for $\tau(X_{i-1}; \theta_0)$. Moreover, this equation depends on the ratio of two conditional expectations with respect to the discrete–time transition density of the underlying diffusion process. This density is most of the times unknown in closed–form. One possibility is to compute these

[8] See Kessler and Sørensen [27, p. 310] for a more detailed discussion.

expectations by Monte Carlo simulation. However, this approach can make the computation of the robust estimating function a very time consuming task already for simple models. Therefore, we propose different ways to circumvent this problem, which make strong use of the intrinsic properties of diffusion processes.

3.2.1 Computation by Means of Eigenexpansions

When the spectrum $\{\lambda_n(\theta_0) : n \in \mathbb{N}\}$ of the generator of the diffusion process is discrete, it is possible to produce a good approximation of $\tau(X_i, X_{i-1}; \theta_0)$ by means of two eigenexpansions for the numerator and the denominator in equation (35). Since $\psi_\star\omega$ and ω are bounded continuous functions, their conditional expectation $u(\Delta, X_{i-1}, \theta_0) := E_{\theta_0}[f(X_i)|X_{i-1}]$, with $f(X_i) := \omega(X_i, X_{i-1}; \theta_0)$ and $f(X_i) := \psi_{\star k}(X_i, X_{i-1}; \theta_0)\omega(X_i, X_{i-1}; \theta_0)$, with subscript $k = 1, \ldots p$ indicating the corresponding component of $\psi_\star$, solves a Kolmogorov backward differential equation with initial condition $u(0, X_{i-1}, \theta_0) = f(X_{i-1})$ and boundary conditions $u(\Delta, l, \theta_0) = u(\Delta, r, \theta_0) = 0$; see, e.g., Karlin and Taylor [26, p. 330–333]. Given the eigenfunctions $\{\phi_n(\cdot; \theta_0) : n \in \mathbb{N}\}$, solutions of the S–L problem (see again Section 2.3.2), u can be written as:

$$\begin{aligned} E_{\theta_0}[f(X_i)|X_{i-1}] &= u(\Delta, X_{i-1}, \theta_0) \\ &= \sum_{n=0}^{\infty} c_{f,n}(\theta_0) \exp(-\lambda_n(\theta_0)\Delta)\phi_n(X_{i-1}, \theta_0) \qquad (36) \end{aligned}$$

where convergence is in the space $L_2(\mu(\theta_0))$ and with Fourier coefficients given by:

$$c_{f,n}(\theta) = \frac{\int_l^r f(x_i)\phi_n(x_i, \theta)m(x_i, \theta)dx_i}{\int_l^r \phi_n^2(x_i, \theta)m(x_i, \theta)dx_i} \ ; \ f = \omega, \psi_{\star k}\omega \ ,$$

for $k = 1, \ldots, p$. These coefficients have usually to be computed numerically, but their computation is typically fast. To define an approximation for (35), we truncate after $q > 0$ terms the series (36) and replace these approximations in the numerator and the denominator of equation (35). Since the coefficients $c_{n,f}(\theta_0)$ are weighted by a weight that decreases exponentially with n, the convergence of these approximations can be often achieved for practical purposes with a relatively moderate number of terms. In our Monte Carlo simulations of Section 4, we find that $q = 5$ terms already produced quite accurate results.

3.2.2 Computation by Means of Saddlepoint Methods

Approximations using the cumulant generating function of $X_i|X_{i-1}$. When the conditional cumulant generating function of the diffusion process exists, it is possible to exploit saddlepoint methods to produce a good approximation of $\tau(X_i, X_{i-1}; \theta_0)$. Let, e.g., $p_{\theta_0}^{(0)}(X_i|X_{i-1})$ be the leading term (21) in the saddlepoint approximation. An accurate approximation for $\tau(X_{i-1}; \theta_0)$ can be obtained by computing the two integrals in the numerator and the denominator of (35) with respect to the saddlepoint approximation $p_{\theta_0}^{(0)}$:

$$\tau^{(0)}(X_{i-1}; \theta_0) = \frac{\int_l^r \omega(x_i, X_{i-1}; \theta_0)\psi_\star(x_i, X_{i-1}; \theta_0) p_{\theta_0}^{(0)}(x_i|X_{i-1}) dx_i}{\int_l^r \omega(x_i, X_{i-1}; \theta_0) p_{\theta_0}^{(0)}(x_i|X_{i-1}) dx_i} \tag{37}$$

This expression has to be evaluated numerically, but its computation is typically fast and can be computed with standard methods when the dimension p of the parameter space is not too large. Higher order approximations of τ are also possible, using higher order saddlepoint approximations of p_{θ_0}, essentially for the same numerical computational costs as $\tau^{(0)}$. Moreover, if the cumulant generating function of the process is not know explicitly, we can use an additional expansion in the time step Δ as in (25), and obtain the approximation (see again Aït Sahalia and Yu [3]):

$$\tau^{(0,\varsigma)}(X_{i-1}; \theta_0) = \frac{\int_l^r \omega(x_i, X_{i-1}; \theta_0)\psi_\star(x_i, X_{i-1}; \theta_0) p_{\theta_0}^{(0,\varsigma)}(x_i|X_{i-1}) dx_i}{\int_l^r \omega(x_i, X_{i-1}; \theta_0) p_{\theta_0}^{(0,\varsigma)}(x_i|X_{i-1}) dx_i} \tag{38}$$

In applications where the time step Δ is small, a moderate order ς of the expansion often produces good results. Depending on the dimension of θ and the specific form of $\psi_\star$ some further simplifications might arise in the computation of the robust estimating function.[9] Alternatively, expressions (37) and (38) might be further approximated by analytical methods using, e.g., Lugannani and Rice–type of formulas (see e.g. [25]). In our Monte Carlo simulations of Section 4, we find that a saddle–point approximation based on the leading term $p_\theta^{(0)}$ produced quite accurate results.

Approximations using the cumulant generating function of $\psi_\star|X_{i-1}$. When the conditional cumulant generating function

$$K_{\star\theta}(\varrho|X_{i-1}) := \log E_\theta[\exp(\varrho'\psi_\star(X_i, X_{i-1};\theta))|X_{i-1}]$$

of $\psi_\star$ exists, we can compute $\tau(X_{i-1};\theta)$ by using directly a multivariate saddlepoint approximation for the conditional density $p_{\star\theta}(Y_i(\theta)|X_{i-1})$ of $Y_i(\theta) := \psi_\star(X_i, X_{i-1};\theta)$ given X_{i-1}. The leading term in this approximation is:

$$p_{\star\theta}^{(0)}(Y_i|X_{i-1}) = \exp(K_{\star\theta}(\hat{\varrho}) - \hat{\varrho}'Y_i)(2\pi)^{-p/2}\det\left(\frac{\partial^2 K_{\star\theta}(\hat{\varrho}|X_{i-1})}{\partial\varrho\partial\varrho'}\right)^{-1/2}, \quad (39)$$

with the saddlepoint $\hat{\varrho}$ solution of the saddlepoint equation:

$$Y_i = \frac{\partial K_{\star\theta}(\hat{\varrho}|X_{i-1})}{\partial\varrho}. \quad (40)$$

Let $\omega(Y_i) := \min\left(1, b/|(Y_i-\tau)'A'A(Y_i-\tau)|^{1/2}\right)$, given invertible matrix $A \in \mathbb{R}^{2p}$ and vector $\tau \in \mathbb{R}^p$. Note that ω is a symmetric function of $A(Y_i - \tau)$. Formula (35) can be then written as:

$$\begin{aligned}\tau(X_{i-1};\theta_0) &= \frac{\int_{\mathbb{R}^p} y_i\omega(y_i)p_{\star\theta_0}(y_i|X_{i-1})dy_i}{\int_{\mathbb{R}^p}\omega(y_i)p_{\star\theta_0}(y_i|X_{i-1})dy_i}\\ &= \frac{\int_{\|A(y_i-\tau)\|>b} y_i(\omega(y_i)-1)p_{\star\theta_0}(y_i|X_{i-1})dy_i}{1+\int_{\|A(y_i-\tau)\|>b}(\omega(y_i)-1)p_{\star\theta_0}(y_i|X_{i-1})dy_i}, \quad (41)\end{aligned}$$

[9] E.g., in the case of the Gaussian Ornstein Uhlembeck process, it is well known that $\tau(X_{i-1};\theta_0) = 0$ when $\psi_\star(X_i, X_{i-1};\theta)$ is the discrete–time Maximum–Likelihood score function; see, for instance Künsch [34].

using the conditional unbiasedness of $\psi_\star$. This expression depends on the ratio of two $p-$dimensional tail integrals of Y_i. Therefore, starting from the saddlepoint approximation (39), a simpler analytical expression for $\tau(X_{i-1};\theta_0)$ can be obtained by using Laplace method to approximate these integrals.

3.2.3 Computation by Means of Small–Time Expansions

If $\psi_\star$ is of class C^2 with respect to X_i, it is possible to approximate the expression for $\tau(X_{i-1};\theta_0)$ in (35) by means of an expansion of the form (23). Some care is needed here, because function ω is not of class C^2 everyhere with respect to X_i. However, we can always approximate ω arbitrarily well by a smooth function $\tilde{\omega}$, say, of class at least C^2. In this case, it is possible to apply the small Δ approximation (23) to $f(X_i) := \tilde{\omega}(X_i, X_{i-1};\theta_0)$ and $f(X_i) := \psi_{\star k}(X_i, X_{i-1};\theta_0)\tilde{\omega}(X_i, X_{i-1};\theta_0)$, respectively, where $k = 1, \ldots, p$. The approximate expression for (35) takes the form:

$$\tilde{\tau}^{(\varsigma)}(X_{i-1};\theta_0) = \frac{\sum_{i=0}^{\varsigma} \frac{\Delta^i}{i!} \mathcal{L}_\theta^i \psi_\star \tilde{\omega}(X_i, X_{i-1};\theta_0)}{\sum_{i=0}^{\varsigma} \frac{\Delta^i}{i!} \mathcal{L}_\theta^i \tilde{\omega}(X_i, X_{i-1};\theta_0)} + O\left(\Delta^{\varsigma+1}\right) \tag{42}$$

where $\mathcal{L}_\theta \psi_\star \tilde{\omega}(X_i, X_{i-1};\theta_0)$ is defined as the $p-$dimensional vector with components $\mathcal{L}_\theta \psi_{\star k} \tilde{\omega}(X_i, X_{i-1};\theta_0)$, $k = 1, \ldots, p$.

3.3 Algorithm

To compute our robust estimator, an iterative algorithm is applied, because the Huber weight $\omega(X_i, X_{i-1};\theta_0)$, the matrix $A(\theta_0)$ and the random vector $\tau(X_{i-1};\theta_0)$ depend on the estimator itself in a nonlinear way. Given a constant $b \geq \sqrt{p}$ (see, e.g., Hampel et al. (1986), p. 228), robust estimator $\overline{\theta}$ is computed by the following algorithm.

1. Set initial values $\theta^{(0)}$, $\tau_i^{(0)} := \tau^{(0)}(X_{i-1}) := 0$ and $A^{(0)}$, by solving

equation (32) for the given (unbounded) estimating function $\psi_\star$:

$$\left(A^{(0)}A^{(0)\prime}\right)^{-1} = \frac{1}{n}\sum_{i=1}^{n}\psi_{\star i}^{(0)}\psi_{\star i}^{(0)\prime}$$

where $\psi_{\star i}^{(0)} := \psi_\star\left(X_i, X_{i-1};\theta^{(0)}\right)$. Moreover, set:

$$\omega_i^{(0)} := \min\left(1; \frac{b}{\left\|A^{(0)}\left(\psi_{\star i}^{(0)} - \tau_i^{(0)}\right)\right\|}\right). \tag{43}$$

2. Calculate $\tau_i^{(1)} := \tau^{(1)}(X_{i-1})$ as:

$$\tau_i^{(1)} = \frac{E_{\theta^{(0)}}\left[\omega_i^{(0)}\psi_{\star i}^{(0)}|X_{i-1}\right]}{E_{\theta^{(0)}}\left[\omega_i^{(0)}\middle| X_{i-1}\right]} \tag{44}$$

and $A^{(1)}$, using equation (32):

$$\left(A^{(1)}A^{(1)\prime}\right)^{-1} = \frac{1}{n}\sum_{i=1}^{n}\left[\omega_i^{(0)}\left(\psi_{\star i}^{(0)} - \tau_i^{(0)}\right)\left(\psi_{\star i}^{(0)} - \tau_i^{(0)}\right)^{\prime}\omega_i^{(0)}\right]$$

3. Given $\tau^{(1)}$ and $A^{(1)}$, compute parameter $\theta^{(1)}$ as the solution of the implicit equation:

$$0 = \sum_{i=1}^{n} A^{(1)}\left(\psi\left(X_i, X_{i-1};\theta^{(1)}\right) - \tau_i^{(1)}\right)\min\left(1; \frac{b}{\left\|A^{(1)}\left(\psi_{\star i}^{(0)} - \tau_i^{(1)}\right)\right\|}\right) \tag{45}$$

Given $\theta^{(1)}$, set $\psi_{\star i}^{(1)} := \psi_\star(X_i, X_{i-1};\theta^{(1)})$ and:

$$\omega_i^{(1)} := \min\left(1; \frac{b}{\left\|A^{(1)}\left(\psi_{\star i}^{(1)} - \tau_i^{(1)}\right)\right\|}\right). \tag{46}$$

4. Go back to Step 2 and replace $\omega_i^{(0)}$ by $\omega_i^{(1)}$, $\psi_{\star i}^{(0)}$ by $\psi_{\star i}^{(1)}$ and $\tau_i^{(0)}$ by $\tau_i^{(1)}$. Then iterate Steps 2. and 3. until convergence of the sequences $\{\theta^{(j)}\}$, $\{A^{(j)}\}$ and $\{\tau^{(j)}\}$.

In Step 2. of the algorithm, we obtain $\tau_i^{(j)}$ by computing the two conditional expectations in the numerator and the denominator of equation (44). In the time series setting, these expectations are typically unknown in closed–form. One possibility is to compute them by Monte Carlo simulation. However, this procedure can make the robust estimation of diffusion processes very time consuming already for simple models. Therefore, it is convenient to circumvent this problem by exploiting the approximation procedures introduced in Section 3.2 for computing $\tau(X_{i-1}; \theta_0)$ in equation (35). Note that in Step 3. of the algorithm the solution of equation (45) is found by holding vector $\tau_i^{(j)}$ fixed. In this way, Step 3. can be maintained computationally not too demanding also for models in which $\tau_i^{(j)}$ has to be computed with some numerical integration procedure.

4 Monte Carlo Simulation and Empirical Application

We first study by Monte Carlo simulation the performance of our robust estimator in two applications to the estimation of a trigonometric and a Jacobi diffusion process; see also Kessler and Sørensen [27]. In a second step, we use the Jacobi diffusion to estimate a model for the exchange rate dynamics in a target zone.

4.1 Monte Carlo Setting

The first model we consider is a trigonometric diffusion satisfying the stochastic differential equation:

$$dX(t) = -\theta \tan X(t)dt + dW(t) \ ; \ X(0) = 0 \ . \tag{47}$$

This process satisfies Assumption 1 and has a bounded state space $\mathcal{S} = (-\pi/2, \pi/2)$. The transition density of this process is not known explicitly. However, the solution for its S–L problem is known. Therefore, we can use the eigenfunctions of the generator of this process to obtain an efficient martingale estimating function. The $\mathcal{A}$–optimal martingale estimating function in Kessler and Sørensen [27] for estimating parameter θ is given by:

$$\psi_\star(X_i, X_{i-1}; \theta) = \frac{\sin(X_{i-1})\left[\sin(X_i) - \exp(-\theta\Delta - \Delta/2)\sin(X_{i-1})\right]}{\frac{1}{2(1+\theta)}\left(\exp(2\,(1+\theta)\,\Delta) - 1\right) - \left(\exp(\Delta) - 1\right)\sin^2(X_{i-1})}. \quad (48)$$

This estimating function can be also approximated by a simpler expression, using a Taylor expansion in the time step Δ for the denominator:

$$\tilde{\psi}_\star(X_i, X_{i-1}; \theta) = \frac{\sin(X_{i-1})\left[\sin(X_i) - \exp(-\theta\Delta - \Delta/2)\sin(X_{i-1})\right]}{\cos^2(X_{i-1})}. \quad (49)$$

For illustration purposes, we plot these two estimating functions in the left and right Panels of Figure 2. Interestingly, even if for small Δ these functions should have very similar efficiency properties, they have different robustness implications, since the first function is bounded but the second is not. This feature implies the non–robustness of the M–estimator based on the approximate estimating function $\tilde{\psi}_\star$. However, even if the estimating function $\psi_\star$ is bounded over the support $\mathcal{S}$, its absolute value can grow quite fast in some regions of the state space, e.g., as $X_i \to \pi/2$ and $X_{i-1} \to -\pi/2$. This feature can imply an excessive sensitivity of the corresponding estimator to potential influential points. Therefore, the robustification of the estimating function $\psi_\star$ by means of the robust estimator in Section 3 can prove useful in applications.

The second model we consider is a Jacobi Diffusion, satisfying the stochastic differential equation:

$$dX(t) = -\beta(X(t) - m)dt + \sigma\sqrt{Z^2 - (X(t) - m)^2}dW(t). \quad (50)$$

where parameter m represents the log of the central tendency of $X(t)$, assumed known, and Z is the maximal deviation from the central tendency, also assumed known. This process is mean–reverting and heteroscedastic; We estimate the parameter vector $\theta := (\beta, \sigma)$, including the mean-reversion speed and the volatility coefficient of the process, respectively.

Since the first two conditional moments of the Jacobi diffusion are known in closed form, a conditionally unbiased estimator can be defined using a quadratic MEF. The left and right Panels of Figure 3 illustrate the form of each component of the implied estimating function. They are both bounded on $\mathcal{S} = (-Z, Z)$, but can grow quite fast over some regions of their domain, e.g., close to the boundaries. If follows that a robustification of the classical M-estimator implied by quadratic MEF can prove useful also in this second diffusion setting.

4.2 Monte Carlo Results

We first simulate discrete-time trajectories of process (47) for a sample size $T = 2000$, a Monte Carlo size 2000 and parameters $\theta = 2$ and $\Delta = 0.2$. We also simulate contaminated trajectories using the simple replacement model:

$$Y(t) = H_t^\eta X(t) + (1 - H_t^\eta)\xi \ , \tag{51}$$

where $X(t)$ is the clean diffusion process (47), H_t^η is a binary 0/1–random variable with $\eta := P(H_t^\eta = 1) = 0.005$, $\xi = 1$ is the value of the contaminated observation $Y(t)$ and Y is the observed process.[10]

Table 1 summarizes the results of our first Monte Carlo exercise. The comparison is between the M–estimator based on estimating function (48) and the robust version of it implied by a bounding constant $b = 3$. To reduce the computation time of random vector τ in the robust estimating function

[10] See also Ortelli and Trojani [39] for a similar Monte Carlo exercise in a different setting.

we apply the approximation methods introduced in Section 3.2 using an eigenexpansion of order $q = 5$.[11]

The performance of classical and robust estimators under the clean process is quite similar (see column 2 and 4 of Table 1), even if the classical estimator has a slightly lower Mean Square Error (MSE). Under the replacement model, the classical estimator generates both a large bias and a high MSE (see column 3 of Table 1), despite the low contamination probability η. The MSE and the bias of the robust estimator under the replacement model, given in row 5 of Table 1, are less than half those of the classical estimator.

In the second Monte Carlo exercise, we simulate discrete-time trajectories of the Jacobi diffusion (50), both using the clean process and replacement model (51). Monte Carlo parameters are $\delta = 0.1$, $\beta = 0.05$, $\sigma = 0.1$, $\eta = 0.005$, $\xi = 1.2$, $Z = 2.25$, $m = 0$ and $\Delta = 1$. The sample size is $T = 300$. This choice implies a very small likelihood of a model contamination and a size of the contamination well inside the support of the given Jacobi diffusion. Figure 4 illustrates these features by presenting a random trajectory generated from the clean process and the replacement model (51), in the top and bottom panel, respectively. To reduce the computation time of the robust estimator in our simulations, we compute the random vector $\tau(X_{i-1}; \theta)$ using a saddle–point approximation with the leading term $p_\theta^{(0)}$, as described in equation (37).

To summarize briefly the results of our Monte Carlo simulation, we present in Figure 5 and 6 the boxplots of the point estimates for β and σ, respectively, under the clean process and the replacement model. As in the previous exercise, even a low likelihood of a model contamination implies a sizable positive bias and a high mean square error for the classical estimator of both model

[11]In our simulations, we found that truncating the Fourier series expansion (36) after five terms already provides a very accurate approximation for random vector τ in equation (35).

parameters (see colums 1 and 2) relative to our robust estimator (columns 3 and 4).

4.3 Real–Data Application

To test empirically our robust estimation procedure, we model the exchange rates in a target zone of the European Monetary System (EMS) for the period January 1991 to July 1993 using the Jacobi diffusion (50).[12] We estimate the mean reversion and volatility parameters using classical and robust estimators. In the EMS, each currency had official fluctuation bands around a central parity, fixed by bilateral agreements of the Central Banks. In equation (50), this is modeled by parameter Z, that fixes the maximal deviation from central parity (which until September 1993 was 2.25%), and parameter m, which represents the log of the central parity.

We estimate the model using weekly data from January 1991 to July 1993, collected from Datastream, for the Dutch Guilder (Dfl), the French Franc (Ffr) and the Danish Kruna (DK) versus the Deutsche Mark (DM).[13] Table 2 presents estimation results for the classical and robust M–estimator. The main difference between the results are a lower volatility estimated with the robust method for all currencies, and a lower mean reversion for the French Franc, similarly to the results obtained in the Monte Carlo simulation of the last section. To understand these differences, it is instructive to study the Huber weights, plotted in Figure 7, that have been estimated by the robust M–estimator for each observation in our sample. The French Franc and the Danish Kruna data have a common period, from June '92 to December '92, in which low estimated Huber weights cluster. Moreover, all currencies highlight a second common subset of observations, towards the end of our sample, in

[12] See also Werker [10] and Larsen and Sørensen [35] for two related applications.

[13] We define $X_t = \log(S_t/\mu)$, where S_t is the spot exchange return at date t and μ is the given central parity.

which low estimated Huber weights arise. Interestingly, the period from June '92 to December '92 coincides with a sequence of particularly stressing events for the EMS. In June '92, the Danish referendum rejected the Maastrich Treaty and this event initiated a period of strong instability for all currencies in the EMS: From August '92 to September '92 the French Franc, the Italian Lira and the British Pound were subject to repeated speculative attacks that determined the exit of the Italian Lira and the British Pound from the EMS in September '92. The last subset of observations in our dataset is also linked to a second period of serious stress in the EMS. In August '93, just a few weeks after our last data point, the Monetary Policy Makers decided to widen the fluctuation bands from $\pm 2.25\%$ to $\pm 15\%$. This event is likely to have been at least partly anticipated by the financial markets, leading to a structural instability in the data that is detected by the Huber weights estimated with our robust procedure.

Figure 8 highlights the consequences of these particular events for the stability properties of the classical estimator, using a simple sensitivity analysis for the Danish Krona data. We perturb observation X_{89} in the interval $(0, 1.5)$, which includes the actual value 1.052 of this observation in the data.[14] This perturbation of one out of 134 observations is sufficient to modify the classical point estimate of σ by about $+0.03$, which is more than the sample standard error for the volatility parameter. The same sensitivity analysis applied to our robust estimator generates virtually no change in the point estimate of σ. The maximal difference between classical and robust point estimates in our sensitivity study is about 0.042, which is approximately 1.5 times the sample standard error for the volatility parameter.

[14]This observation has been identified as influential by the low estimated Huber weight (0.48) it implies.

5 Conclusions

We developed a comprehensive framework for infinitesimal robustness in the context of discretely–observed strictly stationary and ergodic diffusion processes. We showed existence and uniqueness of the conditional influence function of conditionally unbiased M–estimators with square–integrable estimating function. Optimal conditionally unbiased robust estimators for general parametric diffusions have been derived, both for the case in which the discrete-time likelihood of the process is known explicitly and when it is not. These optimal robust estimators are often computable quite efficiently using approximation methods for diffusions that are not similarly well available in other time series settings. Monte Carlo simulation showed a good performance of our optimal robust estimator, with a very moderate efficiency loss at the parametric model and substantial improvements in terms of bias and Mean Square Error under a model deviation. An application to the robust estimation of the exchange rate dynamics in a target zone illustrated the applicability of our methodology in a relevant real–data example. The implications of our results are not limited to robust estimation issues. E.g., the development of robust inference procedures for diffusions based on our results is straightforward, using the setting in Ronchetti and Trojani [41]. Moreover, the class of robust martingale estimating functions derived in this paper can prove useful also to define first–step robust estimators of auxiliary models in the context of robust indirect inference–type of procedures for diffusions; see, for Ortelli and Trojani [39].

References

[1] Y Aït-Sahalia. Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64:527–560, 1996.

[2] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approach. Revised version of WP 467, Graduate School of Business, University of Chicago, 1998.

[3] Y. Aït-Sahalia and J. Yu. Saddlepoint approximations for continous-time Markov processes. *Journal of Econometrics*, 134:507–5551, 2006.

[4] B.M. Bibby, M. Jacobsen, and M. Sørensen. Estimating functions for discretely sampled diffusion-type models, in Aït-Sahalia's Handbook, 2004.

[5] B.A. Bodo, M.E. Thompson, and T.E. Unny. A review of stochastic differential equations for applications in hydrology. *J. Stochastic Hydrology and Hydraulics*, 1:81–100, 1987.

[6] R.J. Carroll, L.A. Stefanski and D. Ruppert Optimally bounded score functions for generalized linear models with applications to logistic regression. *JASA*, 84:460–466, 1989.

[7] R.J. Carroll, H. Künsch and L.A. Stefanski Conditionally unbiased boundend-influence estimation in general regression models, with applications to generalized linear models. *Biometrika*, 73:413–425, 1986.

[8] J.C. Cox, J.E. Ingersoll, and S. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:373–384, 1985.

[9] H.E. Daniels Saddlepoint approximations in Statistics. *The Annals of Mathematical Statistics*, Vol.25, No. 4 :631–650, 1954.

[10] F. De Jong, F.C. Drost and B.J.M. Werker. A Jump-Diffusions model for exchange rates in a target zone. *Statistica Neerlandica*, 55:270–300, 2001.

[11] P. Donnelly and M. Sthephens. Inference in molecular population genetics. *JRSS B*, 62:605–655, 1993.

[12] D. Duffie and K. Singleton. Simulated moments estimation of Markov models of asset pricing. *Econometrica*, 61:929–952, 1993.

[13] D. Florens-Zmirou. Approximate discrete-time schemes for statistics in diffusion processes. *Statistics*, 20:547–557, 1989.

[14] A.R. Gallant and G. Tauchen. Which moments to match? *Econometric Theory*, 12:657–681, 1996.

[15] A.R. Gallant and G. Tauchen. Reprojecting partially observed systems with application to interest rates diffusions. *JASA*, 93:10–24, 1998.

[16] V. Genon-Catalot, T. Jantheau, and C. Laredo. Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, 6:1051–1079, 2000.

[17] M. Genton and E. Ronchetti. Robust indirect inference. *JASA*, 98:67–76, 2003.

[18] C. Gourieroux, A. Monfort, and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118, 1993.

[19] F.R. Hampel. The influence curve and its role in robust estimation. *JASA*, 69:383–393, 1974.

[20] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics: The Approach based on influence functions*. John Wiley, New York, 1986.

[21] L.P. Hansen. Large sample properties of GMM estimators. *Econometrica*, 50:1029–1054, 1982.

[22] L.P. Hansen, J.A. Scheinkman, and N. Touzi Spectral methods for identifyng scalar diffusions. *Journal of Econometrics*, 86 (1):1–32, 1998.

[23] Huber, P.J. *Robust Statistics*. New York,1981, Wiley.

[24] C.C. Heyde. *Quasi-likelihood and its Application*. Springer-Verlag, New York, 1997.

[25] J.L. Jensen. *Saddlepoint Approximations*. Oxford Science Pubblications, New York, 1995.

[26] S. Karlin and H.M. Taylor. *A Second Course in Stochastic Procesess*. Orlando, FL, 1981.

[27] M. Kessler and M. Sørensen. Estimating functions based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5:299–314, 1999.

[28] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Spriger-Verlag, Berlin, 1999.

[29] R.W. Koencher and G. Basset. Regression quantiles. *Econometrica*, 46:33–50, 1978.

[30] R.W. Koenker. Robust methods in Econometrics. *Econometric Review*, 1:213–255, 1982.

[31] J.E. Kolassa. *Series approximation methods in Statistics*. Springer, Lecture Notes in Statistics, 2006.

[32] W.S. Krasker and R.E. Welsh Efficient bounded-influence regression estimation. *JASA*, 77:595–604, 1982.

[33] J. Krishnakumar and E. Ronchetti Robust estimators for simulateneous equations models. *Journal of Econometrics*, 1:295–314, 1997.

[34] H. Künsch. Infinitesimal robustness for autoregressive processes. *The Annals of Statistics*, 12:843–863, 1984.

[35] K.S. Larsen and M. Sørensen. Diffusions models for exchange rates in a target zone. *Mathematical Finance*, 17(No.2):286–305, April 2007.

[36] L. Mancini, E. Ronchetti, and F. Trojani. Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models. *JASA*, pages 628–641, 2005.

[37] M. Markatou and E. Ronchetti. Robust inference: the approach based on influence functions. *Handbook of Statistcs* eds. G.S. Maddala and C.R. Rao, Amsterdan, North–Holland, pp 49–75, 1997.

[38] B.K. Øksendal Stochastic differential equations. An introduction with applications (5th ed.). Springer, Universitext, 1998.

[39] C. Ortelli and F. Trojani. Robust efficient method of moments. *Journal of Econometrics*, 128:69–97, 2005.

[40] R. Poulsen. Approximate maximum likelihood estimation of discretely observed diffusion processes. Working Paper 29, Center of Analytical Finance, Aahrus, 1999.

[41] E. Ronchetti and F. Trojani. Robust inference with GMM estimators. *Journal of Econometrics*, 101:37–69, 2001.

[42] K. Sobcyzk. *Stochastic differential equations: with applications to Physics and Engineering.* Volume 40 of *Math. Appl. (East European Ser.)*, Kluwer, 1991.

[43] M. Sørensen. On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics*, 13:111–136, 1999.

[44] N. Yoshida. Robust M-estimators in diffusion processes. *Ann. Inst. Statist. Math.*, 4:799–820, 1988.

Appendix A: Proofs

In this Appendix, we provide the proofs of the Propositions in the paper.

Appendix A.1. Proof of Proposition 4

Preliminaries. First, recall that under Assumption 1 the diffusion process defined by equation (1) and its discrete–time Markov chain $\mathcal{X}(\Delta) := \{X(i\Delta) : i \in \mathbb{N}\}$ are strictly stationary and ergodic. Theorem 2.3 in [16] then implies that 1 is a simple eigenvalue of operator T_Δ, i.e., the space $\{h \in L_2(\mu(\theta_0)) : T_\Delta h = h\}$ is a one dimensional space spanned by constants. Therefore, the resolvent operator $(I - \lambda T_\Delta)^{-1}$ at $\lambda = 1$ is well defined; see, e.g., Yoshida (1970, p. 209) and Bibby, Jacobsen and Sørensen [4, p. 9]:

$$(I - T_\Delta)^{-1} h(x) = \sum_{k=0}^{\infty} T_\Delta^k h(x) \ , \tag{1}$$

for any $h \in L_2(\mu(\theta_0))$ with $E_{\theta_0}[h(X_i)] = 0$, where convergence is in the space $L_2(\mu(\theta_0))$.

(II.) Uniqueness of g. Assume that there exist two square integrable functions g_1 and g_2 satisfying equation (17). We show that the difference $g := g_1 - g_2$ is constant under the given assumptions. To this end, it is sufficient to prove that if $f = 0$ in equation (17) then g is constant. For $f = 0$, equation (17) reads:

$$\int_{\mathcal{S}} \left(g(x_{i-1}) - g(x_i) \right) p_\theta(x_i | x_{i-1}) dx_i = 0 \ , \tag{2}$$

which implies that $\{g(X_i) : i \in \mathbb{N}\}$ is a martingale.[15] From the convergence theorem of martingales, there exists a square integrable random variable Z

[15] The martingale property follows also under the weaker assumption:

$$\int_{\mathcal{S}} f(x_i, x_{i-1}) p_\theta(x_i | x_{i-1}) dx_i = 0 \ .$$

such that $Z := \lim_{i\to\infty} g(X_i)$, P-almost surely. The ergodicity of process $\{X_i : i \in \mathbb{N}\}$ then implies that Z is constant P–almost surely. Furthermore:

$$g(X_i) = E\left(Z|X_i\right) = E\left(Z|X_0\right) = g(X_0) = Z\ .$$

This concludes the proof of the uniqueness part.

(III.) Existence of g. We show that for any $f \in L_2(P_{\theta_0})$ satisfying equation (16) there exists $g \in L_2(\mu(\theta_0))$ satisfying equation (17). Consider the following subspace of $L_2(\mu(\theta_0))$:

$$J = \left\{ h \in L_2(\mu(\theta_0)) : \int_{\mathcal{S}} h(x)m(x,\theta_0)dx = 0 \right\}.$$

For simplicity of notation, we rewrite integral equation (17) as:

$$g - T_\Delta g = (I - T_\Delta)g = -T_\Delta f. \tag{3}$$

Applying iterated expectations, we observe that $\tilde{f} := -T_\Delta f \in J$. Consider now the operator $T_\Delta : J \to J$ restricted to the class J, and mapping J in J. The resolvent operator $(I - T_\Delta)^{-1}$ is well–defined under Assumption 1. Therefore, $g := (I - T_\Delta)^{-1} T_\Delta f$ satisfies equation (3). This concludes the existence part of the proof. □

Appendix A.2. Proof of Proposition 8

(I.) Let $D_\star(\psi,\theta) = E_\theta[\psi\psi_\star']$ and $V_\star(\psi,\theta) = D_\star(\psi,\theta)^{-1} W_\psi(\theta) D_\star(\psi,\theta)'^{-1}$. We show that ψ_b is, up to multiplication by a constant matrix, the solution of the optimization problem

$$\inf_{\psi} tr[V_\star(\psi,\theta) V_\star(\psi_b,\theta)^{-1}] \tag{4}$$

in the class of martingale estimating functions ψ such that

$$\sup_{(X_i,X_{i-1})\in\mathcal{S}^2} \psi(X_i,X_{i-1};\theta)' V_\star(\psi_b,\theta)^{-1} \psi(X_i,X_{i-1};\theta) \le b^2. \tag{5}$$

Without loss of generality, let ψ be such that

$$D_\star(\psi, \theta) = I_p \ . \tag{6}$$

It then follows, for any $\mathcal{F}_{t-1}$–measurable vector τ:

$$\begin{aligned} V_\star(\psi,\theta) &= E_\theta\left[(D_\star(\psi_b,\theta)^{-1}(\psi_\star-\tau)-\psi)(D_\star(\psi_b,\theta)^{-1}(\psi_\star-\tau)-\psi)'\right] \\ &\quad -D_\star(\psi_b,\theta)^{-1}E_\theta\left[(\psi_\star-\tau)(\psi_\star-\tau)'\right]D_\star(\psi_b,\theta)'^{-1}+D_\star(\psi_b,\theta)'^{-1} \\ &\quad +D_\star(\psi_b,\theta)^{-1} \ . \end{aligned}$$

Therefore, problem (4) is equivalent to the problem:

$$\inf_\psi E_\theta\left[(D_\star(\psi_b,\theta)^{-1}(\psi_\star-\tau)-\psi)V_\star(\psi_b,\theta)^{-1}(D_\star(\psi_b,\theta)^{-1}(\psi_\star-\tau)-\psi)'\right] \tag{7}$$

where ψ is a martingale estimating function such that constraint (5) holds. Let $\phi = V_\star(\psi_b, \theta)^{-1/2}\psi$ and note that

$$\|\phi\|^2 = \psi' V_\star(\psi_b, \theta)^{-1}\psi \ . \tag{8}$$

Therefore, subject to (5), problem (4) is minimized in terms of ϕ by:

$$\phi = A(\theta)(\psi_\star - \tau)\min\left(1, \frac{b}{|(\psi_\star-\tau)'A'(\theta)A(\theta)(\psi_\star-\tau)|^{1/2}}\right) \tag{9}$$

where

$$A(\theta) = V_\star(\psi_b,\theta)^{-1/2}D_\star(\psi_b,\theta)^{-1} \tag{10}$$

Condition (6) ensures that ϕ is unique almost surely. Moreover:

$$A(\theta)'A(\theta) = W_{\psi_b}(\theta)^{-1} \ .$$

It follows that the solution in terms of ψ is $\psi = D_\star(\psi_b,\theta)^{-1}\psi_b$, which implies:

$$\psi' V_\star(\psi_b,\theta)^{-1}\psi = \psi_b W_{\psi_b}(\theta)^{-1}\psi_b \ . \tag{11}$$

To ensure that ψ is conditionally unbiased, we define $\mathcal{F}_{t-1}$–measurable random vector τ implicitly as the solution of the equation:

$$E_{\theta_0}[\psi_b(X_i, X_{i-1}; \theta_0)] = 0 \ . \qquad (12)$$

(II.) Assume that $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$. If there exists a strongly efficient robust martingale estimating function ψ such that $\Upsilon_\psi(\theta_0) \leq b^2$ and $D_\star(\psi, \theta_0) = D(\psi, \theta_0)$, then it follows $V_\psi(\theta_0) \leq V_{\psi_b}(\theta_0) \leq V_\star(\psi_b, \theta_0)$ and

$$\psi' V_\star(\psi_b, \theta_0)^{-1} \psi \leq \psi' V_{\psi_b}(\theta_0)^{-1} \psi \leq \psi' V_\psi(\theta_0)^{-1} \psi \leq b^2 \ .$$

Therefore, ψ satisfies the constraint (5) and we obtain:

$$tr[V_\psi(\theta_0) V_\star(\psi_b, \theta_0)^{-1}] = tr[V_\star(\psi, \theta_0) V_\star(\psi_b, \theta_0)^{-1}] \geq tr[V_\star(\psi_b, \theta_0) V_\star(\psi_b, \theta_0)^{-1}] \ .$$

This implies:

$$tr[(V_\psi(\theta_0) - V_{\psi_b}(\theta_0)) V_\star(\psi_b, \theta_0)^{-1}] \geq tr[(V_\psi(\theta_0) - V_\star(\psi_b, \theta_0)) V_\star(\psi_b, \theta_0)^{-1}] \geq 0 \ .$$

However, since ψ is strongly efficient this last equality can hold only if ψ and ψ_b are equivalent. □

True $\theta_0 = 2$	KS (Cl. Pr.)	KS (Cn. Pr.)	Rob. (Cl. Pr.)	Rob (Cn. Pr.)
q_{25}	1.92	2.06	1.89	1.95
Median	2.00	2.17	1.98	2.05
q_{75}	2.09	2.27	2.08	2.16
Mean	2.01	2.17	1.99	2.06
SD	0.132	0.154	0.140	0.147
Mean bias (%)	0.5%	8.5%	0.5%	3.0%
MSE	0.0177	0.0533	0.0200	0.0254

Table 1: Comparison of classical and robust estimators for θ in the trigonometric diffusion (47). The first and second columns summarize the results of Kessler-Sørensen estimator. The third and fourth columns give the results for our robust M-estimator. Every discrete–time trajectory (both clean and contaminated) has been simulated by the Milstein scheme (simulation step $\delta = 0.001$) with $\beta = 2$, sample size $T = 2000$ and a discrete–time sampling frequency $\Delta = 0.2$. The contaminated samples are obtained by means of the replacement model (51), using the parameter choice $\eta = 0.005$ and $\xi = 1$. The bounding constant for the robust estimator is $b = 3$.

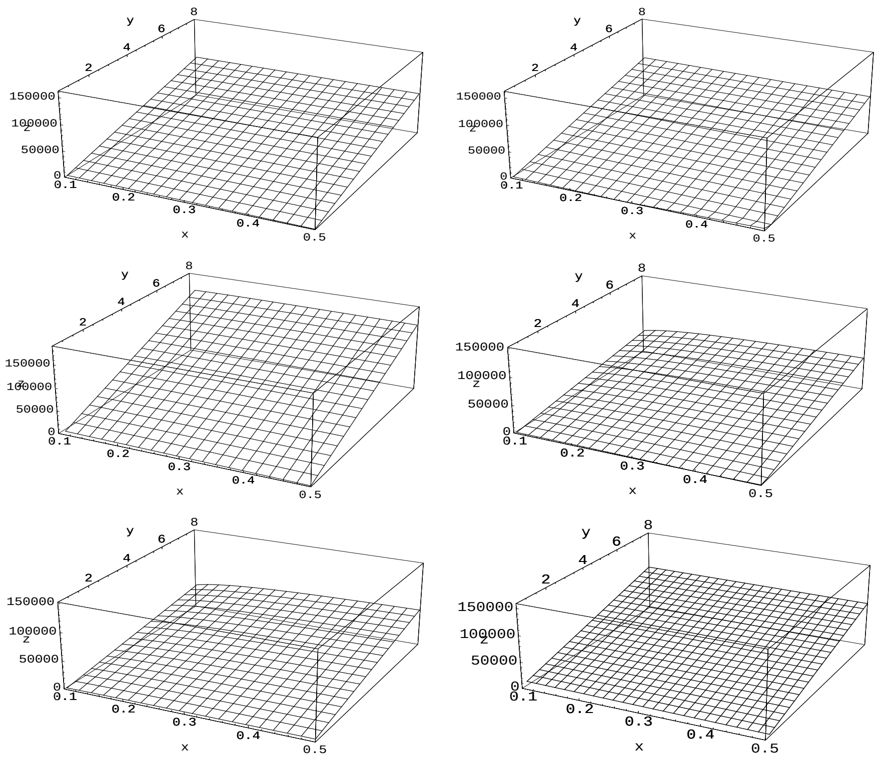

Figure 1: **Different estimating functions for β in the CIR model:** Maximum Likelihood score (left top Panel), Maximum Likelihood score with saddlepoint approximation (right top Panel), Maximum Likelihood score with saddlepoint and cumulant approximation (left middle Panel), Martingale Estimating function implied by first conditional moments (right middle Panel), optimal martingale estimating function (left bottom Panel), optimal martingale estimating function with Taylor approximation in Δ (right bottom Panel). For all Panels, we have $x = X_{i-1}$, $y = X_i$, $\Delta = 0.2$, $\beta = 1$ and $\alpha = 0.0305$.

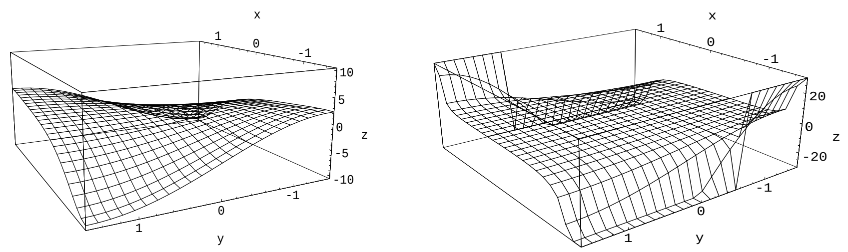

Figure 2: **Different estimating functions for θ in the trigonometric diffusion:** Optimal estimating function $\psi_\star$ implied by the first pair of eigenfunctions of the process (left Panel) and approximate optimal estimating function $\tilde{\psi}_\star$ implied by a Taylor expansion of $\psi_\star$ in Δ (right Panel). For both graphs, we have $x = X_{i-1}$, $y = X_i$, $\Delta = 0.2$ and $\theta = 2$.

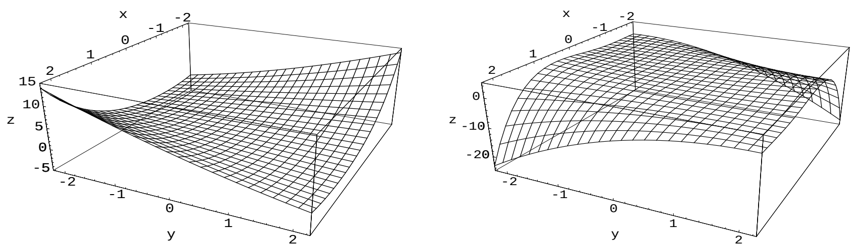

Figure 3: **Estimating functions for β and σ in the Jacobi diffusion:** Quadratic estimating function for β implied by the first two conditional moments of the process (left Panel) and quadratic estimating function for σ implied by the first two conditional moments of the process (right Panel). For both graphs, we have $x = X_{i-1}$, $y = X_i$, $\Delta = 0.2$, $\beta = 0.05$, $\sigma = 0.1$, $Z = 2.25$ and $m = 0$.

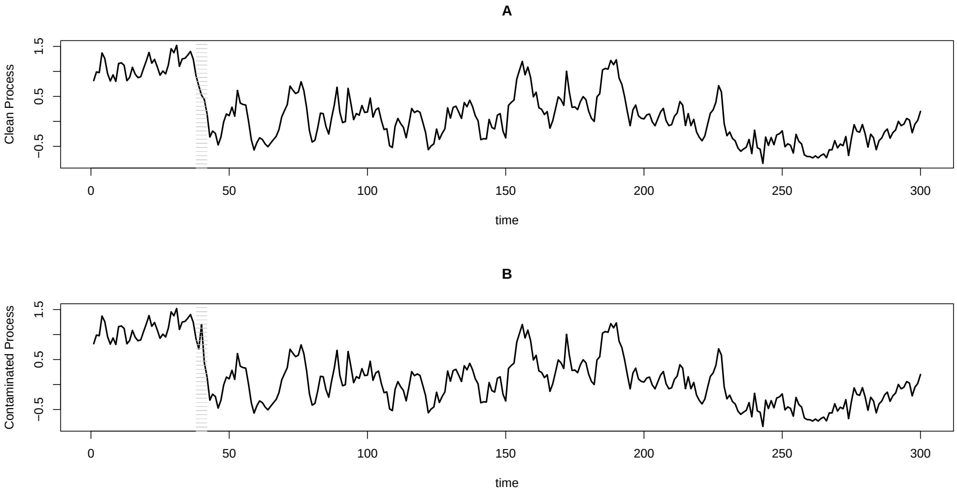

Figure 4: Clean and contaminated trajectory for the Jacobi diffusion (Panel A and B, respectively). The trajectories are simulated by means of Euler scheme with a simulation step $\delta = 0.1$ and the parameter choice $\beta = 0.05$, $\sigma = 0.1$. The contaminated trajectory is obtained by means of the replacement model (51), with a probability of contamination $\eta = 0.005$ and contaminating values of $\xi = 1.2$. The grey zone highlights the difference between the clean and the contaminated trajectory. Discrete–time observations are obtained for $\Delta = 1$ and sample size $T = 300$.

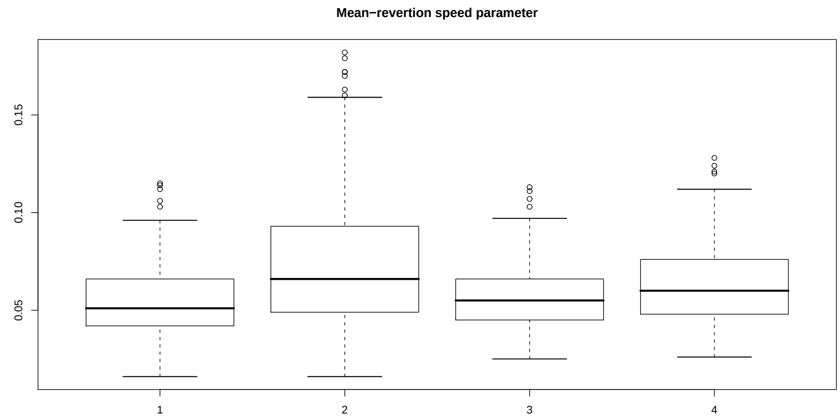

Figure 5: Box Plots for classical and robust estimates of β, both under the true model distribution (50) and the replacement model (51): (1) Classical est. for clean process, (2) Classical est. for cont. process, (3) Robust est. for clean process, (4) Robust est. for cont. process.

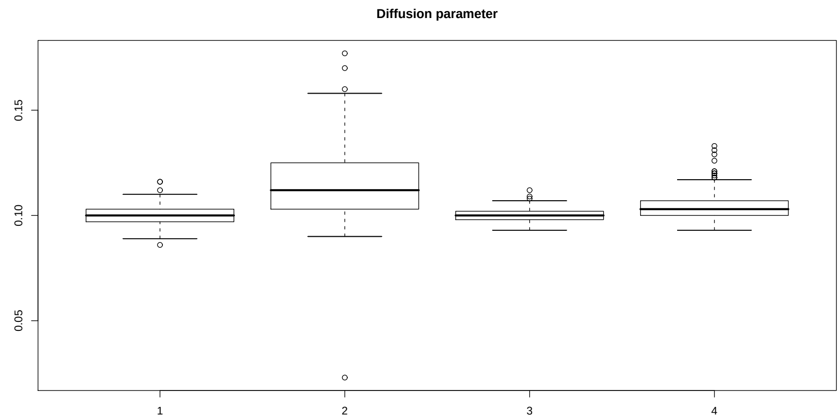

Figure 6: Box Plots for classical and robust estimates of σ both under the true model distribution (50) and the replacement model (51): (1) Classical est. for clean process, (2) Classical est. for cont. process, (3) Robust est. for clean process, (4) Robust est. for cont. process.

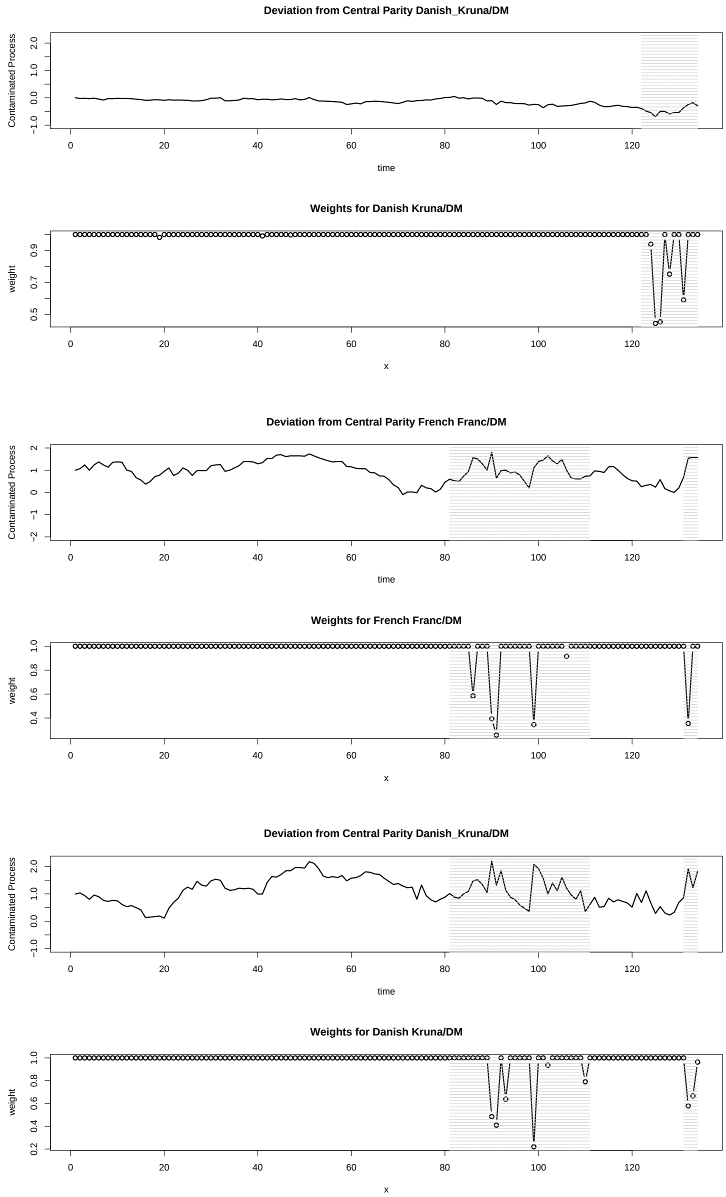

Figure 7: Time series and Huber's type weights for Dutch Guilder (first and second panel from the top), French Franc (third and fourth panel from the top) and Danish Kruna (last two bottom panels). The gray zone highlights a period of strong instability for the currencies in the EMS, because of several speculative attacks and political turmoil (as in the months from June '92 to (late) December'92)

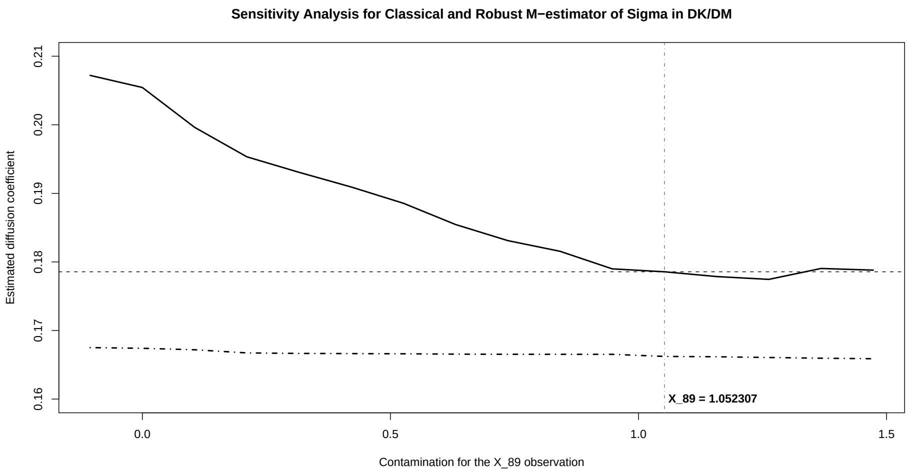

Figure 8: Sensitivity analysis of classical and robust M–estimator for the volatility parameter (σ). The continuous line represents the values of the classical M–estimator obtained by moving the influential observation X_{89} from its original value. The dashed horizonal line gives the original fitted value of σ obtained by using the classical M–estimator when X_{89} is at its original value (dashed vertical line). The dot–dashed bottom line gives the estimated values of σ implied by our robust M–estimator.

Currency	Classical		Robust	
	$\hat{\beta}_{clas}$	$\hat{\sigma}_{clas}$	$\hat{\beta}_{rob}$	$\hat{\sigma}_{rob}$
Dfl/DM	0.059 (0.03602)	0.0735 (0.00169)	0.064 (0.04960)	0.0571 (0.00222)
Ffr/DM	0.135 (0.03125)	0.2002 (0.00993)	0.086 (0.02079)	0.1421 (0.01289)
DK/DM	0.056 (0.01927)	0.1785 (0.02681)	0.0719 (0.01144)	0.1662 (0.02747)

Table 2: Classical and robust point estimates of β and σ in SDE (50) for our real-data example. The first two columns give the classical point estimates for β and σ, obtained by means of a quadratic MEF. The third and fourth columns give the robust point estimates of β and σ. The bounding constant for Dfl is $b = 2.75$ and the one for the other currencies is $b = 2.25$. The estimated standard errors, both for classical and robust M–estimators, are given in parentheses.