

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, Saddlepoint Approximation.

First Version: September 2006. **This Version:** January 25, 2010.

*Davide La Vecchia is at the University of Lugano, Switzerland. Fabio Trojani is at the University of Lugano, Switzerland, and a Research Fellow of the Swiss Finance Institute. E-mail addresses: Davide.La.Vecchia@lu.unisi.ch and Fabio.Trojani@lu.unisi.ch. We thank the editor (Leonard Stefanski), the associate editor and three anonymous referees for many very valuable comments. We also thank Michael Sørensen and Francesco Corielli for helpful comments and suggestions. We gratefully acknowledge 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 strictly stationary and ergodic diffusion processes. The need for robust statistical methodologies is widely recognized, both in the statistical and the econometric literature. Important contributions have studied infinitesimally robust, i.e. bounded-influence, estimators and tests for i.i.d. data; see, among others, Stefanski et al. (1986), Hampel (1974), Huber (1981), Koenker (1982), Krishnakumar and Ronchetti (1997). Recent research has addressed the infinitesimal robustness problem in the time-series context. Martin and Yohai (1986) study influence functionals and different types of contaminations by outliers for linear ARMA processes. Künsch (1984) introduces a formal definition of the conditional time-series Influence Function (IF^c) for stationary time-series and constructs optimal robust estimators for linear autoregressive processes. Rieder (1994) considers conditionally unbiased robust estimators for models in which a LA(M)N expansion of the conditional likelihood function holds. Ronchetti and Trojani (2001), Mancini et al. (2005) and Ortelli and Trojani (2005) propose robust estimators and tests for strictly stationary time-series using M-type estimators. In these papers, a well-defined time-series IF^c is assumed and no general condition for its existence is stated. Künsch (1984) proves existence and uniqueness of the IF^c of linear autoregressive processes. We prove existence and uniqueness of the IF^c for square integrable estimating functions of strictly stationary and ergodic diffusions. In a second step, we introduce optimal M-estimators for general diffusions, having a bounded time-series IF^c .

Diffusion processes are used in many different areas, including Finance, Engineering, Physics, Population Genetics or Biology; see, for instance, Gallant and Tauchen (1998) (Finance), Donnelly and Stephens (1993) (Population Genetics) or Kloeden and Platen (1999) (e.g. Biology, Physics or Engineering). To our knowledge, optimal robust estimation for diffusions has been studied so far only in Yoshida (1988), who develops a Huber-type estimator for problems in which observations are collected continuously. This assumption restricts the usage of robust methods for

many applications like, e.g., those typically encountered in many financial models. We investigate the general case in which observations are collected discretely.

We study robust estimation for conditionally unbiased estimators, since these estimators exploit more conveniently the information in the conditional transition density of diffusion processes. Yoshida's (1988) estimator is defined through the invariant measure of the process. This feature simplifies the analysis of its infinitesimal robustness, which is characterized by the standard IF; see, among others, Hampel (1978) and Hampel et al. (1986). The characterization of robustness for conditionally unbiased estimators requires a formal treatment of the time-series IF^c .

An important issue in the development of robust estimators is that the discrete-time transition density of diffusions is rarely available in closed-form. Therefore, we exploit several analytical approximation methods to the discrete-time maximum likelihood score function; see, among others, Aït-Sahalia (2002), Aït-Sahalia and Yu (2006), Bibby, Jacobsen and Sørensen (2004) and Kessler and Sørensen (1999). In this way, we can avoid estimation procedures based on computationally intensive Monte Carlo methods, which are largely unfeasible in the context of robust conditionally unbiased M-estimation; see also Ortelli and Trojani (2005). We define optimal robust M-estimators for diffusions as the solution to the self-standardized version of Hampel's optimality problem, analyzed – among others – by Stefanski et al. (1986) and Mancini et al. (2005). These estimators make use of a set of Huber's weights that down-weight the impact of influential observations. To preserve conditional unbiasedness, a location correction is needed, which depends on a set of expectations under the process distribution. Using the properties of diffusions, we produce analytical approximations for these corrections that largely reduce the computational burden of the robust estimator. We show by Monte Carlo simulation that our robust estimator implies a favorable trade-off between efficiency and robustness. We then estimate the parameters of a Jacobi diffusion modeling the exchange rate dynamics in a target zone. We find that abnormal market events, typically realized during periods of turbulent financial markets, have a strong influ-

ence on the performance of classical estimators. In contrast, our robust estimator produces reliable estimation results and is able to detect such influential data points quite successfully.

The paper is structured as follows. Section 2 introduces the setup, defines the infinitesimal robustness problem and proves existence and uniqueness of the IF. It also summarizes several approaches to obtain closed-form estimating functions. In Section 3, we first define our robust conditionally unbiased estimator and derive its optimality properties. In a second step, we describe the algorithm to compute it, which exploits different approximation procedures that 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

The statistical model for discrete-time observations $\{X_0, \dots, X_n\}$ is a diffusion process $\mathcal{X} := \{X(t) : t \geq 0\}$ with state space $\mathcal{S} := (l, r)$, where $-\infty \leq l < r \leq \infty$, defined on complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$. \mathcal{X} follows the stochastic differential equation:

$$dX(t) = \alpha(X(t), \theta)dt + \sigma(X(t), \theta)dW(t), \quad X(0) = x_0, \quad (2.1)$$

where $\theta \in \Theta \subset \mathbb{R}^p$, with Θ an open set. $W := \{W(t) : t \geq 0\}$ is a standard Brownian motion and the functions $\alpha(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ satisfy the regularity conditions detailed below. Inference about parameter θ is drawn using observations $X_i := X(\Delta i)$, $i = 0, 1, \dots, n$, given a time step $\Delta = t_i - t_{i-1}$. For brevity, we mainly focus on a fix time step Δ . The case with stochastic time step is discussed at the end of Proposition 4.

2.1 Setting

The scale measure $S(\theta)$ and the speed measure $M(\theta)$ of the diffusion (2.1) are absolutely continuous measures, having densities $s(x, \theta) = \exp\left(-\int_0^x 2\alpha(y, \theta)\sigma^2(y, \theta)dy\right)$ and $m(x, \theta) = \sigma^{-2}(x, \theta)s^{-1}(x, \theta)$, for $x \in \mathcal{S}$. Both $s(x, \theta)$ and $m(x, \theta)$ are stan-

dard tools used to study the stationarity and the ergodicity of diffusions; see, e.g., Genon-Catalot, Jeantheau and Laredo (2000).

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 $|\alpha(x, \theta)| \leq C(\theta)(1 + \|x\|)$ and $\sigma^2(x, \theta) \leq C(\theta)(1 + \|x\|^2)$ for some function $C(\theta) > 0$ and all $x \in \mathcal{S}$. (II.) For every $\theta \in \Theta$ and $x^\# \in \mathcal{S}$:

$$\int_{x^\#}^r s(x, \theta) dx = \int_l^{x^\#} s(x, \theta) dx = \infty \quad A(\theta) = \int_l^r m(x, \theta) dx < \infty$$

(III.) For every $\theta \in \Theta$, the probability distribution $\mu(\cdot, \theta)$ of $X(0)$ has density defined by $\mu(dx, \theta) := A(\theta)^{-1} m(x, \theta) dx$. (IV.) $\sigma(x, \theta) m(x, \theta) \rightarrow 0$ 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 $\theta \in \Theta$.

Assumption 1 is standard in the diffusions literature; see among the others Ait-Sahalia and Mykland (2004). It implies easily verifiable assumptions on $\alpha(x, \theta)$ and $\sigma(x, \theta)$ in (2.1). Examples are provided in Section 4. Conditions (I.) and (II.) ensure existence and uniqueness of the solution of (2.1). Condition (III.) is needed for the strict stationarity, implying that $\mu(\theta)$ is the invariant distribution. Conditions (IV.) and (V.) ensure the ergodicity of both \mathcal{X} and the discrete-time Markov chain $\{X(i\Delta) : i \in \mathbb{N}\}$.

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

$$E_\theta[\psi(X_i, X_{i-1}; \theta) | X_{i-1}] = 0 \tag{2.2}$$

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:

$$\sum_{i=1}^T \psi(X_i, X_{i-1}; \hat{\theta}_T) = 0 ; T \in \mathbb{N}. \tag{2.3}$$

The statistical functional implied by a conditionally unbiased M-estimator takes the form $\hat{\theta}(P_\theta) = \theta \Leftrightarrow E_\theta[\psi(X_i, X_{i-1}; \theta)] = 0$. In our diffusion setting, MEF ψ is typically a highly nonlinear function, so that estimator $\hat{\theta}$ is in most cases not available in closed-form. We denote by \mathcal{M} the family of two-dimensional marginal distributions of a stationary process and consider conditionally unbiased M-functionals $\hat{\theta}(\cdot) : \text{dom}(\hat{\theta}) \subset \mathcal{M} \rightarrow \mathbb{R}^p$ that satisfy condition (2.2). The statistical model implied by diffusion process (2.1) is $\mathcal{P} := \{P_\theta : \theta \in \Theta\}$.

2.2 Infinitesimal Robustness

Let $P_{\epsilon, \nu} := \epsilon P_{\theta_0} + (1 - \epsilon)\nu$, with $\nu \in \mathcal{M}$, be a generic ϵ -contamination of the parametric diffusion probability P_{θ_0} , where $0 \leq \epsilon \leq \eta$ for fixed $0 < \eta < 1$. By $\mathcal{U}_\eta(P_{\theta_0})$ we denote the local neighborhood of such contaminations. First order infinitesimal robustness studies estimators with uniformly bounded linearized asymptotic bias, defined by $B(\epsilon, \nu) = \epsilon \partial_\epsilon \hat{\theta}(P_{\epsilon, \nu})|_{\epsilon=0}$ for all $\nu \in \mathcal{M}$ such that this derivative exists.

Definition 3 (I.) An Influence Function (IF) for conditionally unbiased estimator $\hat{\theta}(\cdot)$ of diffusion process (2.1) is any function $IF : \mathcal{S}^2 \times \Theta \rightarrow \mathbb{R}^p$ such that:

$$\left. \frac{\partial \hat{\theta}(P_{\epsilon, \nu})}{\partial \epsilon} \right|_{\epsilon=0} = \int_{\mathcal{S}^2} IF(x_1, x_0; \theta_0) \nu(dx_0, dx_1) \quad (2.4)$$

for all $\nu \in \mathcal{M}$. (II.) The Conditional Influence Function (IF^c) is any influence function $IF^c : \mathcal{S}^2 \times \Theta \rightarrow \mathbb{R}^p$ such that $E_{\theta_0}[IF^c(X_i, X_{i-1}; \theta_0)|X_{i-1}] = 0$.

If an IF exists, it can be used to describe the linearized asymptotic bias of estimator functional $\hat{\theta}(\cdot)$ over neighborhood $\mathcal{U}_\eta(P_{\theta_0})$. Several versions of the IF exist, which build a class of equivalent kernels satisfying condition (2.4). Given a kernel $IF(x_1, x_2; \theta_0)$, any other version is of the form: $IF(x_2, x_1; \theta_0) + g(x_1; \theta_0) - g(x_2; \theta_0)$, where $g : \mathcal{S} \times \Theta \rightarrow \mathbb{R}^p$ is an arbitrary function such that for all $\nu \in \mathcal{M}$:

$$\int_{\mathcal{S}^2} (g(x_1; \theta_0) - g(x_2; \theta_0)) \nu(dx_2, dx_1) = 0. \quad (2.5)$$

To construct robust estimators for diffusion processes, we first prove appropriate existence and uniqueness results for the conditional IF of these processes.

Proposition 4 *Let Assumption 1 be satisfied. (I.) If function $f : \mathcal{S}^2 \rightarrow \mathbb{R}^p$ is P_{θ_0} -square integrable and such that $E_{\theta_0}[f(X_i, X_{i-1})] = 0$, then there exists a $\mu(\theta_0)$ -square integrable function $g : \mathcal{S} \rightarrow \mathbb{R}^p$ such that $E_{\theta_0}[f(X_i, X_{i-1}) + g(X_{i-1}) - g(X_i)|X_{i-1}] = 0$. (II.) The function g is unique up to an additive constant.*

The main implication of Proposition 4 is that IF^c , when it exists, is unique. Proposition 4 holds for all diffusions satisfying Assumption 1 and extends Künsch's results for linear autoregressive processes. The fact that IF^c is a function of X_i and X_{i-1} only is a direct consequence of the Markov property of the diffusions. Moreover, from the proof of Proposition 4 it follows that our results apply generally to any strictly stationary and ergodic Markov chain. Proposition 4 can be also naturally extended to situations where the time step $\Delta_i = t_i - t_{i-1}$ is stochastic and satisfies Assumption 2 in Aït-Sahalia and Mykland (2004). More precisely, if bivariate process $Y_i := (\Delta_i, X_i)$ is strictly stationary and ergodic, and square integrable function $f : \mathcal{S}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^p$ is such that $E_{\theta_0}[f(Y_i, Y_{i-1})] = 0$, then there exists a square integrable function $g : \mathcal{S} \times \mathbb{R} \rightarrow \mathbb{R}^p$, unique up to an additive constant, such that $E_{\theta_0}[f(Y_i, Y_{i-1}) + g(Y_{i-1}) - g(Y_i)|Y_{i-1}] = 0$.

Corollary 5 *Let $\hat{\theta}$ be a conditionally unbiased M -estimator for diffusion process (2.1), defined by a P_{θ_0} -square integrable estimating function $\psi : \mathcal{S}^2 \times \Theta \rightarrow \mathbb{R}^p$. Then the IF^c of $\hat{\theta}$ is uniquely given by:*

$$IF_{\psi}^c(x_i, x_{i-1}; \theta_0) = -D(\psi, \theta_0)^{-1} \psi(x_i, x_{i-1}; \theta_0), \quad (2.6)$$

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

The main consequence of Corollary 5 is that infinitesimally robust conditionally unbiased M -estimators for diffusions are only those with a bounded estimating function. Their linearized asymptotic bias under contamination, which is uniquely determined by the IF^c , is bounded for all distributions in $\mathcal{U}_{\eta}(P_{\theta_0})$. Hampel et al. (1986) remark that well-known models of contamination by outliers, like isolated, replacement or additive outliers, imply a distribution that can be well approximated

by the elements of $\mathcal{U}_\eta(P_{\theta_0})$. Neighborhood $\mathcal{U}_\eta(P_{\theta_0})$ contains all distributions of the form $P_{\epsilon,\nu} = (1 - \epsilon)P_{\theta_0} + \epsilon\nu$, such that $\nu \in \mathcal{M}$ and $0 \leq \epsilon \leq \eta < 1$ and having Kolmogoroff distance from the reference model P_{θ_0} bounded by η . When the IF^c is unbounded this can motivate the construction of robust estimators. In cases where the IF^c factorizes as in models where a LA(M)N expansion holds, robust regression estimators are naturally applicable. Section 3 introduces a general robsutification procedure that applies also to models in which the IF^c does not factorize.

2.3 Martingale Estimating Functions for Diffusions

Several efficient martingale estimating functions for diffusions in the literature imply M-estimators that are not robust.

2.3.1 Exact and Approximate Maximum Likelihood

Maximum Likelihood: If the transition density $p_\theta(X_i|X_{i-1})$ is known in closed-form, the MLE is readily available, using estimating function $\psi_{ML}(X_{t_i}, X_{t_{i-1}}; \theta) = \nabla_{\theta'} \ln p_\theta(X_i|X_{i-1})$. However, this estimating function is in most cases unbounded.

Approximate likelihood score: If $p_\theta(X_i|X_{i-1})$ is unknown, it can be approximated using the Edgeworth-type expansions in Aït-Sahalia (2002) or the saddle-point expansions in Aït-Sahalia and Yu (2006). Both methods can provide accurate analytical approximations. The approach in Aït-Sahalia and Yu (2006), assumes existence of the cumulant generating function $K_\theta(\varrho|X_{i-1}) := \log E_\theta[\exp(\varrho X_i)|X_{i-1}]$. When $K_\theta(\varrho|X_{i-1})$ is known explicitly, a saddle-point approximation can be applied directly. Let $p_\theta^{(0)}(X_i|X_{i-1})$ be the approximation implied by the direct Edgeworth or saddle-point expansion. The estimating function implied by the approximate likelihood score is $\psi_{App}(X_i, X_{i-1}, \theta) = \nabla_{\theta'} \ln p_\theta^{(0)}(X_i|X_{i-1})$. When $K_\theta(\varrho|X_{i-1})$ is not explicitly available, it can be further approximated. Let $f : \mathcal{S} \rightarrow \mathbb{R}$ be a smooth function of class C^2 . It then follows; see, among others, Bibby et al. (2004) or

Aït-Sahalia and Yu (2006):

$$E_{\theta}(f(X_i)|X_{i-1}) = \sum_{i=0}^{\varsigma} \frac{\Delta^i}{i!} \mathcal{L}_{\theta}^i f(X_{i-1}) + O(\Delta^{\varsigma+1}), \quad (2.7)$$

where $\mathcal{L}_{\theta} := \alpha(x, \theta)d/dx + 0.5\sigma^2(x, \theta)d^2/dx^2$ is the infinitesimal generator of the diffusion. An approximation of the unknown cumulant generating function follows from (2.7) by setting $f(x) = \exp(\rho x)$ and truncating the expansion at the ς -th term. Let $p_{\theta}^{(0, \varsigma)}(X_i|X_{i-1})$ be the approximation of the transition density implied by this ς -th order approximation. The resulting approximate likelihood score is $\psi_{App}^{(\varsigma)}(X_i, X_{i-1}, \theta) := \nabla_{\theta'} \ln p_{\theta}^{(0, \varsigma)}(X_i|X_{i-1})$. In most diffusion settings, both ψ_{App} and $\psi_{App}^{(\varsigma)}$ are unbounded.

2.3.2 Martingale Estimating Functions

Another way to construct efficient MEFs projects the Likelihood score on a subspace of martingale differences, which can be generated by the conditional moments of the diffusion or by the solutions of the Sturm–Liouville problem.

Conditional moments: Given $k = 1, \dots, K \geq p$, consider the martingale differences $\Gamma_k(\theta) := X_i^k - E_{\theta}[X_i^k|X_{i-1}]$ and let $\Gamma(\theta) := (\Gamma_1(\theta), \dots, \Gamma_K(\theta))'$. Bibby et al. (2004) show that across functions $\psi(X_i, X_{i-1}; \theta) := \nu(X_{i-1}; \theta)\Gamma(\theta)$, the \mathcal{A} -optimal function is defined by $\nu^*(X_{i-1}; \theta) = -E_{\theta}[\nabla_{\theta'} \Gamma(\theta)|X_{i-1}]E_{\theta}[\Gamma(\theta)\Gamma(\theta)'|X_{i-1}]^{-1}$. Optimal MEF $\psi_{BJS}(X_i, X_{i-1}; \theta) = \nu^*(X_{i-1}; \theta)\Gamma(\theta)$ is the nearest function in L_2 -norm to the unknown Likelihood score. When closed-form conditional moments are unavailable, it is possible to approximate them using (2.7) applied to $f(x) = x^k$, $k = 1, \dots, K$. As in the previous settings, in most models function ψ_{BJS} is unbounded.

Sturm–Liouville problem: When eigenvalues and eigenfunctions of the conditional expectation operator $T_{\Delta}f(X_{i-1}) := E_{\theta}[f(X_i)|X_{i-1}]$ are known, we can define $\Gamma_k(\theta) := \phi_k(X_i; \theta) - \mu_k(\theta)\phi_k(X_{i-1}; \theta)$, where $\phi_k(\cdot; \theta)$ is for $k = 1, \dots, K$ the k -th eigenvector of operator T_{Δ} with eigenvalue $\mu_k(\theta)$; see Kessler and Sørensen (1999) for a related discussion. In most models, the \mathcal{A} -optimal estimating function $\nu^*(X_{i-1}; \theta)\Gamma(\theta)$ implied by the S-L solutions is again unbounded.

3 Optimal Robust Conditionally Unbiased Estimators

We derive optimal robust M-estimators and robustify unbounded estimating functions by a set of Huber-type weights that down-weight influential observations.

3.1 Estimators with Bounded Self-Standardized Sensitivity

Let $\psi_\star(X_i, X_{i-1}; \theta)$ be an unbounded MEF for diffusion (2.1). We consider standardized robust M-estimators, in Künsch (1984) terminology. Our robust estimator $\bar{\theta}$ is the M-estimator defined by the bounded MEF:

$$\begin{aligned}\psi_r(X_i, X_{i-1}; \theta) &:= A(\theta)\psi_b(X_i, X_{i-1}; \theta) \\ &:= A(\theta)(\psi_\star(X_i, X_{i-1}; \theta) - \tau(X_{i-1}; \theta))\omega(X_i, X_{i-1}; \theta)\end{aligned}\quad (3.1)$$

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

$$\omega(X_i, X_{i-1}; \theta) = \min\left(1, \frac{b}{\|A(\theta)(\psi_\star(X_i, X_{i-1}; \theta) - \tau(X_{i-1}; \theta))\|}\right)\quad (3.2)$$

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:

$$E_\theta[\psi_r(X_i, X_{i-1}; \theta)\psi_r(X_i, X_{i-1}; \theta)'] = I_p, \quad (3.3)$$

$$E_\theta[\psi_r(X_i, X_{i-1}; \theta)|X_{i-1}] = 0. \quad (3.4)$$

M-estimator $\bar{\theta}$ is conditionally unbiased because condition (3.4) preserves Fisher consistency. It implies a self-standardized sensitivity bounded by b^2 , since:

$$\begin{aligned}\Upsilon_{\psi_r}(\theta) &:= \sup_{(X_{i-1}, X_i) \in \mathcal{S}^2} |IF_{\psi_r}^c(X_i, X_{i-1}; \theta)' V_{\psi_r}^{-1}(\theta) IF_{\psi_r}^c(X_i, X_{i-1}; \theta)| \\ &= \sup_{(X_{i-1}, X_i) \in \mathcal{S}^2} |\psi_r(X_i, X_{i-1}; \theta)' \psi_r(X_i, X_{i-1}; \theta)| \leq b^2,\end{aligned}\quad (3.5)$$

using identity $V_{\psi_r}(\theta) = E_\theta[IF_{\psi_r}^c IF_{\psi_r}^{c'}]$, which follows from (3.3). Under standard conditions, estimator $\bar{\theta}$ is asymptotically normally distributed with covariance matrix $V_{\psi_r}(\theta)$.

Robust MEF ψ_b solves an optimization problem that implies admissibility of $\bar{\theta}$ in an appropriate class of estimators. A strongly efficient MEF function ψ^{opt} is such that matrix $V(\psi, \theta) - V(\psi^{opt}, \theta)$ is positive semi definite for all MEF estimating functions ψ . 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 $W_\psi(\theta) := E_\theta(\psi\psi')$. The optimality of estimating function ψ_b is characterized next.

Proposition 6 *If for given $b \geq \sqrt{p}$ equations (3.3) and (3.4) have solution $A(\theta_0)$ and $\tau(X_{i-1}; \theta_0)$, then ψ_b minimizes $\text{tr}[V_\star(\psi, \theta_0)V_\star(\psi_b; \theta_0)^{-1}]$ among all MEF ψ such that*

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

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 MEF ϕ such that $E_{\theta_0}(\phi\psi'_\star) = E_{\theta_0}[\nabla_{\theta'_0}\phi]$, $\Upsilon_\phi(\theta_0) \leq b^2$ 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.) states that if ψ_b satisfies $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$ then there cannot exist another robust MEF ϕ more efficient than ψ_b and such that $E_{\theta_0}[\phi\psi'_\star] = E_{\theta_0}[\nabla_{\theta'_0}\phi]$. Since $E_{\theta_0}[\nabla_{\theta'_0}\phi] = E_{\theta_0}[\phi(\nabla_{\theta'_0} \log p_{\theta_0})']$ these estimating functions are orthogonal to the difference between ψ_\star and the Maximum Likelihood score $\nabla_{\theta'_0} \log p_{\theta_0}$. Condition $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$ depends only on the known estimating function ψ_\star and can be verified in applications.

Corollary 7 *(I.) Assume that $V_\star(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$. Then, there cannot exist a bounded MEF ϕ strictly more efficient than ψ_b and such that both $\Upsilon_\phi(\theta_0) \leq b^2$ and $E_{\theta_0}[\phi(\psi_\star - \nabla_{\theta'_0} \log p_{\theta_0})'] = 0$ hold. (II.) Let $\psi_\star(X_i, X_{i-1}; \theta_0) = \nabla_{\theta'_0} \log p_{\theta_0}(X_i, X_{i-1}; \theta_0)$ be the Maximum Likelihood score. Then, $V_\star(\psi_b, \theta_0) = V_{\psi_b}(\theta_0)$ and there cannot exist a robust MEF ϕ strictly more efficient than ψ_b and such that $\Upsilon_\phi(\theta_0) \leq b^2$.*

Statement (II.) is the version of Corollary 1.1 in Stefanski et al. (1986) for our diffusion setting. It applies when the discrete-time Maximum Likelihood score

function is explicitly known. In this case, the robust estimator $\bar{\theta}$ is admissible in the class of conditionally unbiased robust estimators with $\Upsilon_{\psi}(\theta_0) \leq b^2$. More generally, provided that $V_{\star}(\psi_b, \theta_0) \geq V_{\psi_b}(\theta_0)$, statement (I.) implies admissibility relative to the smaller class of bounded MEF orthogonal to $\psi_{\star} - \nabla_{\theta'_0} \log p_{\theta_0}$. Precisely, if ψ_b is such that $E_{\theta_0}[\nabla_{\theta'_0} \psi_b] = E_{\theta_0}[\psi_b(\nabla_{\theta'_0} \log p_{\theta_0})']$, then ψ_b is an admissible estimating function in this class. If instead ψ_b is not orthogonal to $\psi_{\star} - \nabla_{\theta'_0} \log p_{\theta_0}$, there cannot exist a bounded MEF ϕ such that $E_{\theta_0}[\nabla_{\theta'_0} \phi] = E_{\theta_0}[\phi(\nabla_{\theta'_0} \log p_{\theta_0})']$ having a variance smaller or equal than ψ_b .

Remark 8 When ψ_{\star} is efficient in the sense of Bibby et al. (2004) it is the orthogonal projection of the Maximum Likelihood score on a k -dimensional subspace S_k in L_2 . If \mathcal{S} is bounded, it is well-known that $\cup_{k \in \mathbb{N}} S_k$ is dense in the space of L_2 -MEF. Therefore, for $k \rightarrow \infty$ the MEF ψ_{\star} converges to $\nabla_{\theta'_0} \log p_{\theta_0}$ in L_2 -norm and $E_{\theta_0}[\psi_b(\psi_{\star} - \nabla_{\theta'_0} \log p_{\theta_0})'] = 0$. If \mathcal{S} is unbounded, additional conditions are needed; see Kessler and Sørensen (1999). Moreover, the orthogonal projection of the Maximum Likelihood score satisfies the admissibility condition of Corollary 7, since $V_{\star}(\psi_{\star}, \theta_0) = V_{\psi_{\star}}(\theta_0)$; see Bibby et al. (2004). If $b \rightarrow \infty$, Lebesgue's Theorem then implies that ψ_b converges to ψ_{\star} and $V_{\star}(\psi_b, \theta_0) \rightarrow V_{\psi_b}(\theta_0)$. This limit behavior as $k \rightarrow \infty$ and $b \rightarrow \infty$ suggests that selecting low values of k and b can more likely lead to estimators for which the admissibility result in Corollary 7 does not hold.

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

The \mathcal{F}_{i-1} -measurable random vector $\tau(X_{i-1}, \theta)$ solves equation (3.4), which ensures conditional unbiasedness at the parametric diffusion model. Solving (3.4) for τ , we obtain:

$$\tau(X_{i-1}; \theta) = \frac{E_{\theta}[\psi_{\star}(X_i, X_{i-1}; \theta) \omega(X_i, X_{i-1}; \theta) | X_{i-1}]}{E_{\theta}[\omega(X_i, X_{i-1}; \theta) | X_{i-1}]}. \quad (3.6)$$

(3.6) is a fixed-point equation that depends on the ratio of two conditional expectations with respect to the discrete-time transition density of the diffusion process.

When this density is unavailable in closed-form, Monte Carlo methods can be applied to compute these expectations. However, the state dependent feature of the solution can render this approach too computationally demanding already for simple models. We propose different ways to circumvent this problem.

3.2.1 Computation by Eigenexpansions

When the spectrum $\{\lambda_n(\theta) : n \in \mathbb{N}\}$ of the diffusion process is discrete, it is possible to produce a good approximation using two eigenexpansions for the numerator and the denominator of (3.6). Since $\psi_\star \omega$ and ω are bounded and continuous, their conditional expectation $u(\Delta, X_{i-1}, \theta) := E_\theta[f(X_i)|X_{i-1}]$, with $f(X_i) := \omega(X_i, X_{i-1}; \theta)$ and $f(X_i) := \psi_{\star k}(X_i, X_{i-1}; \theta)\omega(X_i, X_{i-1}; \theta)$, with subscript k indicating the k -th component of ψ_\star , solves a Kolmogorov backward differential equation with initial condition $u(0, X_{i-1}, \theta_0) = f(X_{i-1})$ and boundary conditions $u(\Delta, l, \theta) = u(\Delta, r, \theta) = 0$. Given eigenfunctions $\{\phi_n(\cdot; \theta) : n \in \mathbb{N}\}$, it follows:

$$E_\theta[f(X_i)|X_{i-1}] = \sum_{n=0}^{\infty} c_{f,n}(\theta) \exp(-\lambda_n(\theta)\Delta) \phi_n(X_{i-1}, \theta) \quad (3.7)$$

with Fourier coefficients given by:

$$c_{f,n}(\theta) = \frac{\int_l^r f(x_i) \phi_n(x_i, \theta) m(x_i, \theta) dx_i}{\int_l^r \phi_n^2(x_i, \theta) m(x_i, \theta) dx_i}.$$

These coefficients have usually to be computed numerically, but their computation is typically fast. To define an approximation for (3.6), we truncate after $q > 0$ terms the series (3.7) and replace these approximations in the numerator and the denominator of equation (3.6). Since coefficients $c_{n,f}(\theta)$ have a weight that decreases exponentially with n , the convergence of the approximation can be often achieved with a moderate number of terms. In the Monte Carlo simulations of Section 4, a choice $q = 5$ already produces accurate approximations.

3.2.2 Computation by Edgeworth or Saddlepoint Methods

An accurate approximation for $\tau(X_{i-1}; \theta)$ can be obtained by computing the two integrals in (3.6) with respect to an Edgeworth or a saddlepoint expansion $p_\theta^{(0)}(X_i|X_{i-1})$

for unknown density $p_\theta(X_i|X_{i-1})$; see again Aït-Sahalia (2002) and Aït-Sahalia and Yu (2006):

$$\tau^{(0)}(X_{i-1}; \theta) = \frac{\int_l^r \omega(x_i, X_{i-1}; \theta) \psi_\star(x_i, X_{i-1}; \theta) p_\theta^{(0)}(x_i|X_{i-1}) dx_i}{\int_l^r \omega(x_i, X_{i-1}; \theta) p_\theta^{(0)}(x_i|X_{i-1}) dx_i} \quad (3.8)$$

This expression has to be evaluated numerically, but its computation is often fast with standard methods when the dimension of Θ is not too large. Higher order approximations are also possible, using higher order approximations of $p_\theta(X_i|X_{i-1})$. If the cumulant generating function is not explicitly available, we can use an additional expansion in the time step as in (2.7) and compute (3.6) based on the approximate saddlepoint expansion $p_\theta^{(0,\varsigma)}(x_i|X_{i-1})$. In applications where the time step is small, a moderate order of this expansion can often produce good results. Depending on the dimension of Θ and the form of ψ_\star , further simplifications might arise. For instance, in the case of an Ornstein–Uhlenbeck process, it is known that $\tau(X_{i-1}; \theta) = 0$ when $\psi_\star(X_i, X_{i-1}; \theta)$ is the discrete–time Maximum Likelihood score function; see Künsch (1984). Alternatively, the tails integrals in (3.6) might be approximated by other analytical methods using, e.g., Lugannani and Rice–type formulas; see, e.g., Jensen (1995).

3.3 Selection of the Bounding Constant

An important issue for applications is the selection of the clipping constant b in our robust approach. This can be based, e.g., on efficiency or testing accuracy considerations. A first standard possibility is to select b in a way that controls the loss in efficiency of the robust estimator at the reference parametric model. A second approach can consider the maximal bias under contamination for the level of a test based on the robust estimator. Given a maximal contamination size ϵ Ronchetti and Trojani (2001) show that the constant b can be uniquely determined in order to ensure a maximal allowed asymptotic size distortion under a contamination. For instance, given a nominal size $\alpha = 5\%$ of such a test constant b is determined as $b = 3(maxbias)^{1/2}p^{0.3}/\epsilon$, where $maxbias$ is the maximal allowed size distortion

under a model contamination of size ϵ . An alternative method based on a radius minimax procedure, which does not require the specification of ϵ , has been recently proposed by Rieder et al. (2008).

3.4 Algorithm

To compute robust estimator $\bar{\theta}$, an iterative procedure is applied. Given a constant $b \geq \sqrt{p}$ (see, e.g., Hampel et al (1986), p. 228), robust estimator $\bar{\theta}$ is computed by the following algorithm.

1. Set initial values $\theta^{(0)}, \tau_i^{(0)} := \tau^{(0)}(X_{i-1}, \theta^{(0)}) := 0$ and $A^{(0)}$, by solving equation (3.3) for the given (unbounded) estimating function ψ_\star :

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

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

$$\omega_i^{(0)} := \min \left(1; \frac{b}{\left\| A^{(0)} \left(\psi_{\star i}^{(0)} - \tau_i^{(0)} \right) \right\|} \right). \quad (3.9)$$

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

$$\tau_i^{(1)} = \frac{E_{\theta^{(0)}} \left[\omega_i^{(0)} \psi_{\star i}^{(0)} | X_{i-1} \right]}{E_{\theta^{(0)}} \left[\omega_i^{(0)} | X_{i-1} \right]} \quad (3.10)$$

and $A^{(1)}$, using equation (3.3):

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

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

$$0 = \sum_{i=1}^n A^{(1)} \left(\psi(X_i, X_{i-1}; \theta^{(1)}) - \tau_i^{(1)} \right) \min \left(1; \frac{b}{\left\| A^{(1)} \left(\psi_{\star i}^{(0)} - \tau_i^{(1)} \right) \right\|} \right) \quad (3.11)$$

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

$$\omega_i^{(1)} := \min \left(1; \frac{b}{\|A^{(1)}(\psi_{\star i}^{(1)} - \tau_i^{(1)})\|} \right). \quad (3.12)$$

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 (3.10). As mentioned above, a Monte Carlo computation of these expectations in the diffusion setting can be very time-consuming already for simple models. Therefore, it is convenient to circumvent this problem by exploiting the approximation procedures in Section 3.2 for computing $\tau(X_{i-1}; \theta)$ in equation (3.6). Note that in Step 3 of the algorithm the solution of equation (3.11) 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 Applications

We study by Monte Carlo simulation the performance of our robust estimator. In a second step, we consider a real-data application.

4.1 Monte Carlo Setting

The trigonometric diffusion satisfies $dX(t) = -\theta \tan X(t)dt + dW(t)$, $X(0) = 0$. If $\theta \geq 0.5$, this process satisfies Assumption 1 and has a bounded state space $\mathcal{S} = (-\pi/2, \pi/2)$. The transition density is not known explicitly, but the solution of its S-L problem is known. The \mathcal{A} -optimal MEF in Kessler and Sørensen (1999) is:

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

This estimating function can be also Taylor–approximated by a simpler expression:

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

Even if for small Δ these estimating functions tend to have similar efficiency properties, they lead to different robustness implications since ψ_\star is bounded but $\tilde{\psi}_\star$ is not. Note, however, that ψ_\star grows quite fast at the boundaries as $X_i \rightarrow \pi/2$ or $X_{i-1} \rightarrow -\pi/2$, which can imply an excessive sensitivity of the corresponding M-estimator to potential influential points. Thus, a robustification also of estimating function ψ_\star can prove useful in applications.

4.2 Monte Carlo Results

We simulate discrete–time trajectories of the trigonometric diffusion 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 model:

$$Y(t) = H_t^\eta X(t) + (1 - H_t^\eta) \xi(t) \tag{4.1}$$

where $X(t)$ is the clean (diffusion) process as in (2.1) and H_t^η is a binary ergodic, stationary 0/1 process, with $\eta := P(H_t^\eta = 1)$. Specification (4.1) generates several different types of outliers like additive, isolated and patchy outliers, as a function of the assumed dependence structure of H_t , $\xi(t)$ and the clean process. An example of other types of outliers not included in (4.1) are endogenous outliers. Moreover, Hampel et al. (1986) remark that for a large class of processes obtained as in (4.1), the distribution of the contaminated process $Y(t)$ can be approximated by a convex linear combination of the form $P_{\epsilon, \nu} := \epsilon P_{\theta_0} + (1 - \epsilon) \nu$, $\nu \in \mathcal{M}$, as in the neighborhood $\mathcal{U}_\eta(P_{\theta_0})$. We set $\eta := P(H_t^\eta = 1) = 0.005$ and $\xi(t) = 1$. Table 1 summarizes the results. The comparison is between the M-estimator based on estimating function (4.1) and the robust version implied by a bounding constant $b = 3$. This value of the bounding constant is selected to ensure a 90% efficiency under the parametric reference model. To reduce the computation time, we use the

approximation methods of Section 3.2 based on a truncated eigenexpansion as in (3.7). In our application, we set $q = 5$ which we find to provide a good approximation of $\tau(X_{i-1}; \theta)$ on the whole relevant state space, as illustrated in Figure 1. We find that the performance of classical and robust estimators under the clean process is quite similar (see column 2 and 4, in Table 1), with the classical estimator having a slightly lower Mean Square Error (MSE): In absence of contamination, the MSE for the robust estimator is 0.020, a value close to the MSE of 0.018 for the \mathcal{A} -optimal estimator. This moderate cost in terms of efficiency ensures a stronger stability in presence of contaminations, where the MSE of the robust estimator is almost half the MSE for the classical estimator.

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; See Werker et al. (2001) and Larsen and Sørensen (2007) for related applications. The Jacobi Diffusion satisfies the SDE:

$$dX(t) = -\beta(X(t) - m)dt + \sigma\sqrt{Z^2 - (X(t) - m)^2}dW(t). \quad (4.2)$$

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. The process satisfies Assumption 1 if $\beta \geq \sigma^2$; see De Jong et al. (2001) and Larsen and Sørensen (2007). 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. Even if the implied estimating functions are bounded, they can grow quite fast close to the boundaries. Thus, a robustification of the M-estimator implied by quadratic MEF can prove useful in this setting. To reduce the computation time of the robust estimator, we compute random vector τ using a saddlepoint approximation, as described in (3.8). Unreported Monte Carlo evidence shows that the resulting robust estimator provides a satisfactory tradeoff between

efficiency and robustness in the Jacobi diffusion setting as well. In the EMS, each currency had official fluctuation bands around a central parity, fixed by bilateral agreements of the Central Banks. In equation (4.2), this is modeled by parameter Z fixing 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 Krone (DK) versus the German Mark (DM). 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. Table 2 presents estimation results for the classical and the robust M-estimator. The main differences are a lower volatility estimated with the robust method for all currencies and a lower mean reversion for the French Franc, similar to the results in the unreported Monte Carlo simulations. To interpret these different results, it is instructive to study the Huber weights estimated by the robust M-estimator for each observation in the sample, plotted in Figure 2. We present the picture for the Danish Krone, since the other currencies have a similar behavior. The Danish Krone highlights two periods, from June '92 to December '92 and toward the end of our sample, in which low estimated Huber weights cluster. 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 Maastricht 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 the structural instability detected by the Huber

weights estimated with our robust procedure. Figure 3 highlights the consequences of these particular events for the stability properties of the classical estimator. We perturb observation X_{89} in the interval $(0, 1.5)$, which includes the actual value 1.052 of this observation. This observation has been identified as influential by the low estimated Huber weight of 0.48 it implies. Such a perturbation of only one observation is sufficient to modify the classical point estimate of σ by about +0.03, which is more than the sample standard error of the volatility parameter estimate. 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.

5 Conclusions

We developed a comprehensive framework for infinitesimal robustness in the context of discretely-observed strictly stationary and ergodic diffusions. We showed existence and uniqueness of the conditional influence function of M-estimators with square-integrable estimating function. Optimal conditionally unbiased robust M-estimators for diffusions have been derived. These estimators are often computable quite efficiently using available approximation methods for diffusions. Monte Carlo simulation showed a good performance of our robust estimator, with a moderate efficiency loss at the parametric model and substantial improvements in terms of bias and MSE 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. Future research can address several potential extensions of our robust approach. A natural area is the development of tractable robust M-estimators for multivariate diffusions, which could be addressed starting from the class of MEF studied in Bibby et al. (2004) or using the multivariate Likelihood approximation methods in Aït-Sahalia (2008). Another interesting

direction suggested by an anonymous referee could investigate robust estimators for general semi-martingales and conditions for existence of a time series influence function in such settings.

APPENDIX A: Proofs

Appendix A.1. Proof of Proposition 4

Preliminaries. First, recall that under Assumption 1 the diffusion process defined by equation (2.1) 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 Genon-catalot et al. (2001) 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. Bibby, Jacobsen and Sørensen (2004):

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

for any $h \in L_2(\mu(\theta_0))$ with $E_{\theta_0}[h(X_i)] = 0$, where convergence is in the space $L_2(\mu(\theta_0))$. (I.) *Uniqueness of g .* Assume that there exist two square integrable functions g_1 and g_2 satisfying

$$E_{\theta_0}[f(X_i, X_{i-1}) + g(X_{i-1}) - g(X_i)|X_{i-1}] = 0. \quad (2)$$

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 the previous equation, then g is constant. For $f = 0$, the equation reads:

$$\int_{\mathcal{S}} (g(x_{i-1}) - g(x_i)) p_{\theta_0}(x_i|x_{i-1}) dx_i = 0 , \quad (3)$$

which implies that $\{g(X_i) : i \in \mathbb{N}\}$ is a martingale. The martingale property follows also under the weaker assumption:

$$\int_{\mathcal{S}} f(x_i, x_{i-1}) p_{\theta_0}(x_i|x_{i-1}) dx_i = 0 .$$

From the convergence theorem of martingales, there exists a square integrable random variable Z such that $Z := \lim_{i \rightarrow \infty} g(X_i)$, P_{θ_0} -almost surely. The ergodicity of process $\{X_i : i \in \mathbb{N}\}$ then implies that Z is constant P_{θ_0} -almost surely. Furthermore:

$$g(X_i) = E_{\theta_0}(Z|X_i) = E_{\theta_0}(Z|X_0) = g(X_0) = Z.$$

This concludes the proof of the uniqueness part.

(II.) *Existence of g .* We show that for any $f \in L_2(P_{\theta_0})$ such that $E_{\theta_0}f(x_i, x_{i-1}; \theta) = 0$ there exists $g \in L_2(\mu(\theta_0))$ satisfying equation (2). 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 the integral equation as:

$$g - T_{\Delta}g = (I - T_{\Delta})g = -T_{\Delta}f. \quad (4)$$

Applying iterated expectations, we observe that $\tilde{f} := -T_{\Delta}f \in J$. Consider now the operator $T_{\Delta} : J \rightarrow J$ 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 (4). This concludes the existence part of the proof.

□

Appendix A.2. Proof of Proposition 6

(I.) Let $D_{\star}(\psi, \theta_0) = E_{\theta}[\psi\psi'_{\star}]$ and $V_{\star}(\psi, \theta_0) = D_{\star}(\psi, \theta_0)^{-1}W_{\psi}(\theta_0)D_{\star}(\psi, \theta_0)'^{-1}$. We show that ψ_b is, up to multiplication by a constant matrix, the solution of the optimization problem

$$\inf_{\psi} \text{tr}[V_{\star}(\psi, \theta_0)V_{\star}(\psi_b, \theta_0)^{-1}] \quad (5)$$

in the class of martingale estimating functions ψ such that

$$\sup_{(X_i, X_{i-1}) \in \mathcal{S}^2} \psi(X_i, X_{i-1}; \theta_0)' V_{\star}(\psi_b, \theta_0)^{-1} \psi(X_i, X_{i-1}; \theta_0) \leq b^2. \quad (6)$$

Without loss of generality, let ψ be such that $D_\star(\psi, \theta_0) = I_p$. It then follows, for any \mathcal{F}_{t-1} -measurable vector τ :

$$\begin{aligned} V_\star(\psi, \theta_0) &= E_{\theta_0} [(D_\star(\psi_b, \theta_0)^{-1}(\psi_\star - \tau) - \psi)(D_\star(\psi_b, \theta_0)^{-1}(\psi_\star - \tau) - \psi)'] \\ &\quad - D_\star(\psi_b, \theta_0)^{-1} E_{\theta_0} [(\psi_\star - \tau)(\psi_\star - \tau)'] D_\star(\psi_b, \theta_0)'^{-1} + D_\star(\psi_b, \theta_0)'^{-1} \\ &\quad + D_\star(\psi_b, \theta_0)^{-1} . \end{aligned}$$

Therefore, problem (5) is equivalent to the problem:

$$\inf_{\psi} E_{\theta_0} [(D_\star(\psi_b, \theta_0)^{-1}(\psi_\star - \tau) - \psi)V_\star(\psi_b, \theta_0)^{-1}(D_\star(\psi_b, \theta_0)^{-1}(\psi_\star - \tau) - \psi)'] \quad (7)$$

where ψ is a martingale estimating function such that constraint (6) holds. Let $\phi = V_\star(\psi_b, \theta_0)^{-1/2}\psi$ and note that $\|\phi\|^2 = \psi'V_\star(\psi_b, \theta_0)^{-1}\psi$. Therefore, subject to (6), problem (5) is minimized in terms of ϕ by:

$$\phi = A(\theta_0)(\psi_\star - \tau) \min \left(1, \frac{b}{|(\psi_\star - \tau)'A'(\theta_0)A(\theta_0)(\psi_\star - \tau)|^{1/2}} \right) \quad (8)$$

where $A(\theta_0) = V_\star(\psi_b, \theta_0)^{-1/2}D_\star(\psi_b, \theta_0)^{-1}$. Since $D_\star(\psi, \theta_0) = I_p$, ϕ is unique almost surely. Moreover $A(\theta_0)'A(\theta_0) = W_{\psi_b}(\theta_0)^{-1}$, it then follows that the solution in terms of ψ is $\psi = D_\star(\psi_b, \theta)^{-1}\psi_b$, which implies:

$$\psi'V_\star(\psi_b, \theta_0)^{-1}\psi = \psi_b W_{\psi_b}(\theta_0)^{-1}\psi_b . \quad (9)$$

To ensure that ψ is conditionally unbiased, we define \mathcal{F}_{t-1} -measurable random vector τ implicitly as the solution of the equation $E_{\theta_0}[\psi_b(X_i, X_{i-1}; \theta_0)] = 0$. (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 (6) 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. \square

APPENDIX B: Figures and Tables

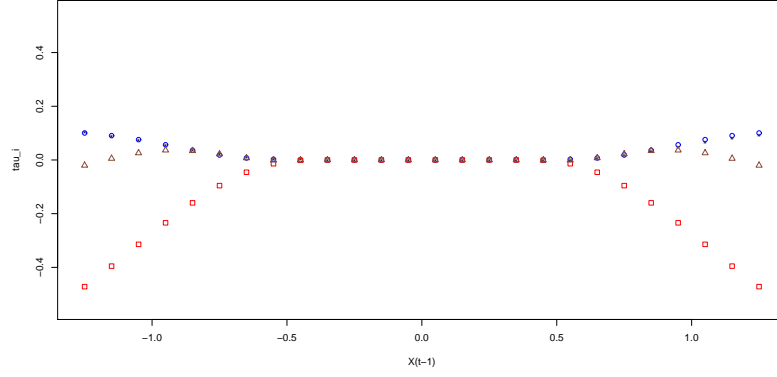


Figure 1: Values of $\tau(X_{i-1}; \theta_0)$ obtained using different methods, for $\theta_0 = 2$ and $X_{i-1} \in (-\pi/2; \pi/2)$. The * represents the true values obtained using a long MC simulation having size of 50.000. The circles provide the values obtained by means of the eigenexpansion with $q = 5$; the triangles are for $q = 2$ and the squares are for $q = 1$.

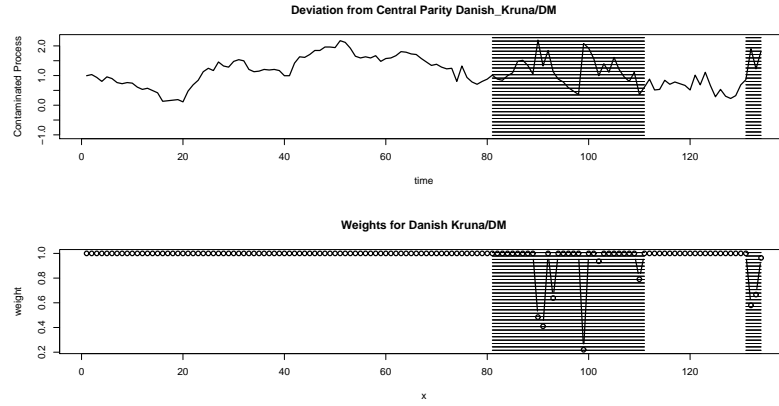


Figure 2: Time-series and Huber weights for DK. The gray zone highlights periods of instability for the currencies in the EMS.

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(x100)	1.77	5.33	2.00	2.54

Table 1: Classical and robust estimators of parameter θ . The first and second columns summarize the results of KS estimator. The third and fourth columns refer to our robust M-estimator. The bounding constant for the robust estimator is $b = 3$.

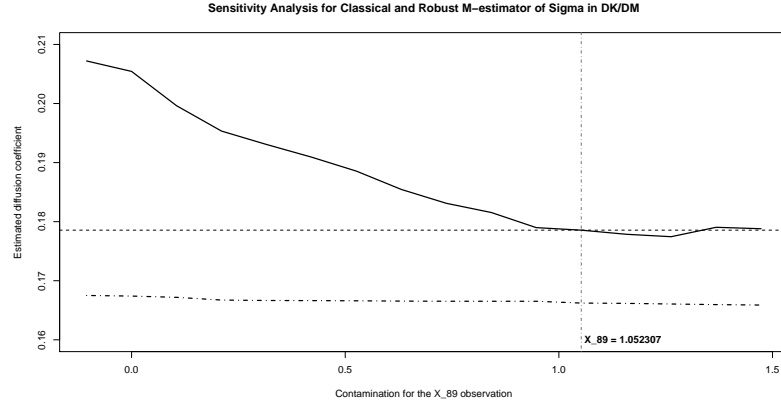


Figure 3: Sensitivity of M-estimators for σ in DK. Continuous line: values of classical estimator obtained by moving the influential observation X_{89} . Dot-dashed line: estimates implied by robust estimator.

Currency	Classical		Robust	
	$\hat{\beta}_{clas}$	$\hat{\sigma}_{clas}$	$\hat{\beta}_{rob}$	$\hat{\sigma}_{rob}$
Dfl/DM	0.059 (0.03602)	0.073 (0.00169)	0.064 (0.04960)	0.057 (0.00222)
Ffr/DM	0.135 (0.03125)	0.200 (0.00993)	0.086 (0.02079)	0.142 (0.01289)
DK/DM	0.056 (0.01927)	0.178 (0.02681)	0.072 (0.01144)	0.166 (0.02747)

Table 2: Classical and robust estimates of β and σ in (4.2) for the real-data example. The first two columns give the classical estimates for β and σ , obtained by means of a quadratic MEF. The third and fourth columns refer to our robust M-estimator. The bounding constant for Dfl/DM is $b = 2.75$ and the one for the other currencies is $b = 2.25$. Standard errors are given in parentheses.

References

- Aït-Sahalia, Y. (1996). “Nonparametric pricing of interest rate derivative securities,” *Econometrica*, 64:527–560.
- (2002). “Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach,” *Econometrica*, 70:223–262.
- (2008). “Closed-form likelihood expansions for multivariate diffusions,” *The Annals of Statistics*, 36(2):906–937.
- Aït-Sahalia, Y. and Mykland, P. (2004). “Estimating diffusions with discretely and possibly randomly spaced data: a general theory,” *The Annals of Statistics*, 32:2186–2222.
- Aït-Sahalia, Y. and Yu, J. (2006). “Saddlepoint approximations for continuous-time Markov processes,” *Journal of Econometrics*, 134:507–551.
- Bibby, B.M., J. M. and Sørensen, M. (2004). “*Estimating functions for discretely sampled diffusion-type models*”, Handbook of Financial Econometrics.
- De Jong, F., D. F. and Werker, B. (2001). “A jump-diffusions model for exchange rates in a target zone,” *Statistica Neerlandica*, 55:270–300.
- Donnelly, P. and Stephens, M. (1993). “Inference in molecular population genetics,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62:605–655.
- Gallant, A. and Tauchen, G. (1998). “Reprojecting partially observed systems with application to interest rates diffusions”, *Journal of the American Statistical Association*, 93:10–24.
- Hampel, F. (1974). “The influence curve and its role in robust estimation”, *Journal of the American Statistical Association*, 69:383–393.
- Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986). *Robust statistics: the approach based on influence functions*, Wiley, New York.
- Huber, P. (1981). *Robust Statistics*. New York, Wiley.

- Jensen, J. (1995). *Saddlepoint Approximations*. Oxford, U.K.
- Kessler, M. and Sørensen, M. (1999). “Estimating equations based on eigenfunctions for a discretely observed diffusion process”, *Bernoulli*, 5:299–314.
- Kloeden, P. and Platen, E. (1999). *Numerical solution of stochastic differential equations*, Springer-Verlag, Berlin.
- Koenker, R. (1982). “Robust methods in econometrics”, *Econometric Reviews*, 1:213–255.
- Krishnakumar, J. and Ronchetti, E. (1997). “Robust estimators for simultaneous equations models”, *Journal of Econometrics*, 78:259–314.
- Künsch, H. (1984). “Infinitesimal robustness for autoregressive processes”, *Annals of Statistics*, 12:843–863.
- Larsen, K. and Sørensen, M. (2007). “Diffusions models for exchange rates in a target zone”, *Mathematical Finance*, 17:286–305.
- Mancini, L., Ronchetti, E., and Trojani, F. (2005). “Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models”, *Journal of the American Statistical Association*, 100(470):628–641.
- Martin, D. and Yohai, V. (1986). “Influence functionals for time series”, *Annals of Statistics*, 14:781–818.
- Ortelli, C. and Trojani, F. (2005). “Robust efficient method of moments”, *Journal of Econometrics*, 128:69–97.
- Rieder, H. (1994). *Robust asymptotic statistics*. Springer, New York.
- Rieder, H., K. M. and Ruckdeschel, P. (2008). “The cost of not knowing the radius”, *Statistical Methods and Applications*, 17:13–40.
- Ronchetti, E. and Trojani, F. (2001). “Robust inference with GMM estimators”, *Journal of Econometrics*, 101:37–69.

- Stefanski, L.A., C. R. and Ruppert, D. (1986). “Optimally bounded score functions for generalized linear models with applications to logistic regression”, *Biometrika*, 73:413–424.
- Yoshida, N. (1988). “Robust M-estimators in diffusion processes”, *Annals of the Institute of Statistical Mathematics*, 40:799–820.