Robust Efficient Method of Moments*

By Claudio Ortelli and Fabio Trojani

Institute of Finance, University of Southern Switzerland, Lugano, Switzerland

Abstract

This paper focuses on the robust Efficient Method of Moments (EMM) estimation of a general parametric stationary process and proposes a broad framework for constructing robust EMM statistics in this context. This extends the application field of robust statistics to very general time series settings, including situations where the structural and the auxiliary models in the EMM estimating equations are different, models with latent non linear dynamics, and models where no closed form expressions for the robust pseudo score of the given EMM auxiliary model are available. We characterize the local robustness properties of EMM estimators for time series by computing the corresponding influence functions and propose two versions of a robust EMM (REMM) estimator with bounded IF. Two algorithms by which the two versions of a REMM estimator can be implemented are presented. We then show by Monte Carlo simulation that our REMM estimators are very successful in controlling for the asymptotic bias under model misspecification while maintaining a high efficiency under the ideal structural model.

Keywords: Efficient Method of Moments, Indirect Inference, Influence Function, Robust Estimation, Robust Statistics

JEL Classification: C1, C13, C14, C15, C22.

*We thank Hermann Garbers, Loriano Mancini, Hansruedi Künsch, Elvezio Ronchetti and seminar participants at the econometrics/statistics seminar of the University of Zurich for many valuable comments and suggestions on an earlier version of this draft. All remaining errors are ours. Correspondence address: Fabio Trojani, Institute of Finance, University of Southern Switzerland, Via Buffi 13, CH-6900 Lugano, e-mail: Fabio.Trojani@lu.unisi.ch. Both authors gratefully acknowledge the financial support of the Swiss National Science Foundation (grant 12-65196.01 and CNNR FINRISK).

1 Introduction

This paper analyzes the local robustness properties of estimators derived from the Efficient Method of Moments (EMM, Gallant and Tauchen, (1996)) and develops a new class of robust statistics for the statistical analysis of parametric models that are estimated within a general EMM framework. Precisely, we apply tools and concepts from the theory of robust statistics (see Huber (1981) and Hampel et al. (1986) for an overview) to develop a new class of EMM statistics that are robust to local misspecifications of the structural specification of some strictly stationary time series process. Since several well-known examples of an EMM estimator imply - as we show below - an unbounded influence function (IF, see Hampel (1968), (1974)), we propose a truncating algorithm by which a non robust EMM estimator can be regularized to ensure a bounded IF and local smoothness of the arising EMM functional in nonparametric neighborhoods of the relevant structural model.

The need for robust statistical procedures in estimation, testing and prediction has been stressed by many authors and is now widely recognized; some references in this respect are for instance Hampel (1974), Koenker and Basset (1978), Huber (1981), Koenker (1982), Peracchi (1990, 1991), Markatou and Ronchetti (1997), Krishnakumar and Ronchetti (1997), Ronchetti and Trojani (2001). This paper focuses on the robust EMM estimation of a general parametric stationary process where the implied stationary density may be not computed analytically, and proposes a broad framework for constructing robust EMM statistics in this context. This extends the application field of robust statistics to very general time series settings, including situations where the structural and the auxiliary models in the EMM estimating equations are different, models with latent non linear dynamics, and models where no closed form expressions for the robust pseudo score of the given EMM auxiliary model are available[1] .

Some authors already addressed some important issues related to a robust inference on time series models, starting from different perspectives. For instance, some first definitions of an influ-

[1] As a special case of our general robust EMM setting, robust M-estimators for time series models where the robust score cannot be computed analytically are also easily obtained.

ence function for times series (extending the basic definitions in Hampel's (1974) seminal work) have been developed in Künsch (1984) and Martin and Yohai (1986). Künsch (1984) also derived the optimality of a robust M-estimator for the parameters of a linear autoregressive process with normally distributed error terms. In a similar vain, de Luna and Genton (2001) proposed more recently a robust estimator for the parameters of a linear ARMA process. In the nonlinear time series context Ronchetti and Trojani (2001) introduced a robust version of a general Generalized Method of Moments (GMM) statistic for situations where a reference model for the underlying data distribution can be assumed. For such a case, an algorithm for computing robust GMM (RGMM) estimators and tests can be used to estimate fairly general models with rich time dependence structures and non linear estimating equations[2] . This was a first attempt to extend the application field of robust statistics to a general time series setting based on models with nonlinear parametric structures. Finally, in a paper more directly related to the present one Genton and Ronchetti (2003) have adressed the issue of robust estimation for models where the stationary density of the structural model cannot be computed explicitly using an indirect inference approach à la Gourieroux and Monfort (1993). In this context and for some simple model settings they show with some illustrative Monte Carlo examples that a robust estimation algorithm can be quite successful in safeguarding an indirect inference estimation procedure against local model misspecifications.

By contrast with previous research in the area we address the problem of a robust inference on general parametric models for time series from a broad perspective that allows us to extend the application field of robust procedures for time series essentially to all models that can be estimated by a classical (non robust) EMM estimator. Indeed, several robust estimation procedures, like for instance standard robust M-estimators (Huber (1981)), RGMM estimators (Ronchetti and Trojani (2001)) or robust M-estimators where the robust score cannot be compute analytically

[2] By embedding RGMM into our robust EMM setting we are able to strongly simplify the RGMM estimation algorithm proposed in Ronchetti and Trojani (2001).

are obtained as special cases of our REMM estimators. The more specific contributions to the literature are the following.

First, we compute the time series influence function (Künsch (1984)) of a general EMM estimator and illustrate the asymptotic bias approximations arising in this context; by contrast with more simple model settings, this allows to analyze the robustness impact of model quantities like for instance the dimension of the auxiliary parameter set or the lag length in the dynamics of the structural model behind the EMM.

Second, we propose two robust EMM estimators with bounded IF that ensure a bounded asymptotic bias in neighborhoods of the given structural model. Both our robust EMM estimators are obtained by truncating the auxiliary pseudo score function of the EMM in an appropriate metric. They correct simultaneously for (i) the structural bias due to the estimation of an auxiliary model in the EMM setting and (ii) the bias implied by a truncation of the initial (non robust) pseudo score function. The first of the two REMM estimators is numerically easier to compute. The second one requires some numerically more demanding computations that yield a higher robustness, especially for EMM settings where the dimensions of the auxiliary and the structural parameters sets are significantly different. This is obtained by truncating the auxiliary score function only in directions that are influential for the implied structural EMM score (see below).

Third, we propose two algorithms to compute our REMM estimators which explicitly take into account the time series properties of the given structural model and a possible dynamic misspecification of the relevant auxiliary model. This last issue is important in the REMM setting, since the metrics by which the robust EMM scores are obtained are directly related to the asymptotic covariance matrix of the auxiliary pseudo score function.

Finally, we present some Monte Carlo experiments attempting to quantify the trade-off between robustness and efficiency in the REMM estimation of an highly nonlinear model. To this end we estimate a structural ARMA(1,1)-ARCH(2) model by means of an auxiliary AR(3)-ARCH(2) model. This is a quite complex model setting from the viewpoint of robust parameter estimation, implying highly nonlinear and non Markovian structural dynamics[3] . Nevertheless, both REMM estimators proposed in the paper yield very satisfactory results in these experiments. Indeed, we find in our simulations that the efficiency loss implied by a REMM estimation under a perfectly specified structural model is virtually negligible, when compared with the results obtained by the classical EMM procedure. Further, we observe that even a quite moderate model contamination can induce a very important bias and a strong loss in efficiency of a classical EMM estimator while both our REMM estimators are very successful in bounding the induced asymptotic bias and efficiency loss. Finally, we find the outlier identification procedure implied by the robust weights of our REMM procedures to be very efficient in all our experiments.

The paper is organized as follows. Section 2 introduces the standard (non robust) EMM setting using a functional notation that highlights the functional structure of the different estimators arising in the EMM. Section 3 starts by computing the time series IF of an EMM estimator which is shown to be linearly related to the one of the estimator for the parameters of the auxiliary model. The asymptotic bias approximations implied for a general time series context are then discussed and illustrated. Section 4 introduces two REMM estimators that bound the implied IF in an appropriate metric. Both REMM estimators are obtained by a suitable truncation of the score function of the estimator for the parameters in the auxiliary model. The algorithms by which REMM estimators can be computed are then explained and some points related to the trade-off between robustness and efficiency in REMM estimation are discussed. Section 5 presents some Monte Carlo experiments where the performance of our REMM estimators is evaluated in a non trivial EMM setting. Section 6 summarizes and concludes.

2 Basic EMM Setting

Let $\mathbb{X} = \{X_t : \Omega \longrightarrow \mathbb{R}^v, v \in \mathbb{N}, t = 1, 2, \cdots\}$ be a strictly stationary and ergodic stochastic process on a complete probability space $(\Omega, \Im, Q_0)$. The goal in EMM estimation is to produce statistical

[3] Previous work in this context focused on models that are simple special cases of the one considered in our simulations (cf. for instance Martin and Yohai (1986) and de Luna and Genton (2001)).

inference on the probability $P_0 := Q_0\mathbb{X}^{-1}$ based on a *structural model* $\mathcal{P} = \{P_\rho : \rho \in \mathcal{R} \subseteq \mathbb{R}^l, l \in \mathbb{N}\}$ that defines for any *structural parameter* $\rho \in \mathcal{R}$ a probability measure P_ρ on the measurable space $(\mathbb{R}^{v\infty} := \times_{t=1}^{\infty}\mathbb{R}^v, \mathcal{B}(\mathbb{R}^{v\infty}))$. In classical EMM estimation $\mathcal{P}$ is assumed to be correctly specified for P_0.

Assumption 1 *There exists $\rho_0 \in \mathcal{R}$ such that $P_0 = P_{\rho_0}$.*

We denote by $x^n := (x_1', x_2', \ldots, x_n')'$ the observed finite history of $X^n := (X_1', X_2', \ldots, X_n')'$ and by X_{n-m+1}^n the subvector consisting of the last m components of X^n, $0 < m \leq n$, and assume all finite dimensional distributions induced by P_ρ to be absolutely continuous with respect to Lebesgue measure.

Assumption 2 *For every $\rho \in \mathcal{R}$ and $n \in \mathbb{N}$, the restriction P_ρ^n of P_ρ on the measurable space $(\mathbb{R}^{vn} := \times_{t=1}^{n}\mathbb{R}^v, \mathcal{B}(\mathbb{R}^{vn}))$ that is induced by the first n coordinates of the process $\mathbb{X}$ has a density function $p^n(x_1, \ldots, x_n;\ \rho)$ with respect to λ^{vn}, the Lebesgue measure on $\mathbb{R}^{vn}$.*

We further denote by an upper index n the unconditional density function of X^n, and by a lower index n the *conditional* density function of X_n, given the observed history X^{n-1}, i.e.

$$p_n(x^n;\ \rho) := p^n(x^n;\ \rho)/p^{n-1}(x^{n-1};\ \rho) \quad .$$

In EMM estimation the focus is on model settings where the conditional density $p_n(x^n;\ \rho)$ cannot be expressed analytically, so that direct estimation of P_{ρ_0} is not feasible. Instead, an auxiliary parametric model and an auxiliary parameter $\theta \in \Theta \subset \mathbb{R}^q$, $q \geq l$ are introduced, which can be estimated for instance by pseudo maximum likelihood. This defines an auxiliary estimation procedure that estimates in a first step of the EMM the underlying probability P_{ρ_0} in an approximate way.

Often the auxiliary model is defined using a parametric family of pseudo likelihood functions that induce a corresponding set of estimating equations, and we could do the same to develop our robust EMM (REMM) methodology. However, it is convenient for our robust analysis and straightforward from a statistical perspective to work with a larger class of auxiliary models, which can be estimated by some M-estimator (Huber (1964)) defined by a general score function

$$\psi : \mathbb{R}^{vL} \times \Theta \longrightarrow \mathbb{R}^q, q \geq l, L \geq 1 \quad . \tag{1}$$

This enables us to unify later under the REMM methodology several robust estimation methods for time series that typically require a Monte Carlo simulation of a given structural model.

A first straightforward example of an auxiliary score function associated with the corresponding structural model is the one that is implied by a pseudo maximum likelihood estimation of the auxiliary model in the standard EMM setting.

Example 3 *In the classical EMM the auxiliary model is chosen so that the proposed parametric stochastic specification* $\mathbb{S}_{f_L} := \left\{ f_L(\cdot;\theta) : \mathbb{R}^{vL} \longrightarrow \mathbb{R}^+,\ \theta \in \Theta \subset \mathbb{R}^q,\ q \in \mathbb{N},\ q \geq l \right\}$ *carefully approximate the true* $(L)-$*dimensional conditional density* $p_L(\cdot;\ \rho_0)$. *In that case we have*

$$\psi(x;\theta) = \frac{\partial \ln f_L(x;\theta)}{\partial \theta} \quad .$$

Similarly, we can also embed standard GMM estimation into an EMM setting, which is based on a score function (1) computed as the GMM score implied by a given GMM orthogonality function.

Example 4 *In a GMM setting the function* ψ *defining the auxiliary model is given by*

$$\psi(x;\theta) := E_{P_0^L}\left(\frac{\partial}{\partial \theta} h'(X^L;\theta)\right) W_0\, h(x;\theta)\,,$$

where $h(\cdot;\theta) : \mathbb{R}^{vL} \longrightarrow \mathbb{R}^H$, $H \geq q$, *is a function defining the orthogonality conditions in a GMM model and* W_0 *is a corresponding deterministic weighting matrix. Remark, that in this case*

$$\dim(\psi) = \dim(\theta) = q \leq \dim(h) = H \quad . \tag{2}$$

We will discuss the implications of (2) for the case where inequality is strict later on, when we discuss the robustness properties of EMM based inference procedures.

Notice that within the classical estimation framework embedding GMM into REMM in Example 4 can appear to be unnatural, because classical GMM does not assume any particular parametric structural model for the underlying distributions. However, in robust GMM estimation and testing the opposite happens. Indeed, RGMM is based on the existence of a structural model which defines an approximate reference distribution for RGMM inference, leading to GMM estimation and testing procedures that are robust to local misspecifications of the given structural distribution;

see Ronchetti and Trojani (2001). In fact, we will see later on that RGMM can be interpreted as a special case of REMM.

To characterize the robustness of EMM statistics we need to write EMM estimators and tests as functionals of a suitable space of distributions. For our purposes, focusing on the finite dimensional distributions induced by the strictly stationary distributions on $(\mathbb{R}^{v\infty}, \mathcal{B}(\mathbb{R}^{v\infty}))$, as in Künsch (1984), will be enough. Thus, define for $L \geq 1$ the following set of finite dimensional distributions

$$\mathcal{M}^L_{stat} = \{(L) - \text{ finite dimensional marginals of strictly stationary processes}\}.$$

Moreover, let for $L < n$ the empirical $(L)-$dimensional marginal distribution of x^n be defined by

$$P^L_n = n^{-1} \sum_{i=1}^{n} \delta_{(x'_i, \cdots, x'_{i+L-1})'} \quad ; x_i = x_{i-n} \text{ for } i > n\,, \tag{3}$$

where δ_{x^L} is the Dirac mass at $x^L \in \mathbb{R}^{vL}$. By construction $P^L_n \in \mathcal{M}^L_{stat}$.

To analyze the local robustness of statistical functionals derived from an EMM estimator, the first step is to define a statistical functional $\widehat{\theta}\,(\cdot)$ for the estimator of the auxiliary parameter θ. By definition, $\widehat{\theta}\,(\cdot)$ is the functional solution of the asymptotic estimating equations implied by the score function (1), that is

$$\widehat{\theta} : dom(\widehat{\theta}) \subset \mathcal{M}^L_{stat} \to \Theta,\ P^L \longmapsto \widehat{\theta}\left(P^L\right) := \widehat{\theta}_{P^L} \quad ,$$

is the functional solution of the implicit equation

$$E_{P^L}\left(\psi(X^L;\theta)\right) = 0 \quad , \tag{4}$$

in θ. Notice, that when restricting the auxiliary functional $\widehat{\theta}\,(\cdot)$ on the set $\mathcal{P}^L := \left\{P^L_\rho : \rho \in \mathcal{R} \subseteq \mathbb{R}^l\right\}$ it follows

$$E_{P^L_\rho}\left(\psi\left(X^L;\widehat{\theta}_{P^L_\rho}\right)\right) = 0 \quad , \tag{5}$$

so that to any $\rho \in \mathcal{R}$ a unique $\theta_{P^L_\rho}$ can be associated, provided that (4) can be uniquely solved for any $P^L_\rho \in dom(\widehat{\theta})$. Thus, in this case a well defined *binding function* (see Gourieroux, Monfort and Renault (1993)) mapping ρ to $\widehat{\theta}_{P^L_\rho}$ is obtained.

Assumption 5 *There exists a smooth injective binding function*

$$b_\psi : \mathcal{R} \longrightarrow \Theta \quad , \quad \rho \longmapsto b_\psi(\rho) := \widehat{\theta}(\rho) := \widehat{\theta}_{P_\rho^L} \quad .$$

In the new notation we thus have $\theta_0 = \widehat{\theta}(\rho_0) = b_\psi(\rho_0)$. Furthermore, notice that different auxiliary score functions ψ, ψ', of the form (1) inducing possibly different functionals $\widehat{\theta}$, $\widehat{\theta}'$ will produce different binding functions b_ψ, $b_{\psi'}$, if and only if $\widehat{\theta}\Big|_{P_\rho^L} \neq \widehat{\theta}'\Big|_{P_\rho^L}$. Later when we develop the REMM, two different alternative binding functions will be proposed to define a REMM estimator. While the second of the two REMM estimators proposed produces a higher efficiency at the model, the first one is preferable from the perspective of computational simplicity.

Some further points related to the smoothness of b_ψ and $\widehat{\theta}$ and to their well definiteness when Assumption 1 does not hold are collected in the next remark.

Remark 6 *Notice that if the parameterization $\rho \longmapsto P_\rho^L$ for the structural model is a smooth one, smoothness of b_ψ is equivalent to smoothness of $\widehat{\theta}\Big|_{P_\rho^L}$. Moreover, when inverting b_ψ in order to recover the structural parameter ρ from $\widehat{\theta}(P^L)$ in a neighborhood $\mathcal{U}(P_\rho^L)$ of P_ρ^L and by means of the inverse $b_\psi^{-1} : Im(b_\psi) \to \mathcal{R}$, it is necessary to have smoothness of $\widehat{\theta}(\cdot)$ in order to ensure smoothness of $b_\psi^{-1} \circ \widehat{\theta}(\cdot)$. Only in this case, small misspecifications of the reference model P_ρ^L will induce stable asymptotic parameter estimates $b_\psi^{-1} \circ \widehat{\theta}(P^L)$ on a full neighborhood $\mathcal{U}(\mathcal{P}) := \cup_{\rho\in\mathcal{R}}\mathcal{U}(P_\rho^L)$ of the given parametric model $\mathcal{P}$. Therefore, the local robustness properties of $\widehat{\theta}$ are crucial, in order to obtain a well defined EMM estimation procedure in the case when Assumption 1 is locally falsified, that is when $P_0 \in \mathcal{U}(\mathcal{P})\backslash\mathcal{P}$.*

The pseudo-true value θ_0 is estimated by $\widehat{\theta}_n := \widehat{\theta}(P_n^L)$ which solves (4) with respect to the empirical measure P_n^L:

$$E_{P_n^L}\left(\psi(X^L; \widehat{\theta}(P_n^L))\right) = 0 \quad .$$

Under regularity conditions (see among others White (1994)) the auxiliary estimator is consistent and asymptotically normal at the model.

Property 7 *Under regularity conditions on $\left(\widehat{\theta}_n\right)_{n\in\mathbb{N}}$ it follows*

$$\begin{cases} \lim_{n\to\infty} \| \widehat{\theta}_n - \theta_0 \| = 0 \quad , \quad a.s. - P_0 \\ \sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \mathcal{V}_0) \end{cases} , \tag{6}$$

where

$$\mathcal{V}_0 = A_0^{-1} B_0 A_0^{-1} \tag{7}$$

and

$$A_0 = E_{P_0^L}\left(\frac{\partial}{\partial\theta'}\psi(X^L;\theta)|_{\theta=\theta_0}\right) \quad , \tag{8}$$

$$B_0 = B_{0,0} + \sum_{\tau=1}^{\infty} B_{0,\tau} + \left(\sum_{\tau=1}^{\infty} B_{0,\tau}\right)' \quad . \tag{9}$$

$$B_{0,\tau} := E_{P_0^L}(\psi(X_{1+\tau}^{L+\tau};\widehat{\theta}(P_0^L))\psi(X^L;\widehat{\theta}(P_0^L))') \, , \tag{10}$$

The statistical functional $\widehat{\rho}(\cdot)$ for the EMM estimator of the structural model is defined as the minimizer of a quadratic form in (5), given the value $\widehat{\theta}_{P_L}$ of the auxiliary estimator. Thus, when denoting by $m : \mathcal{R} \times \Theta \longrightarrow \mathbb{R}^q$ the function given by

$$m(\rho, \theta) = E_{P_\rho^L}\left(\psi(X^L;\theta)\right) \quad ,$$

the functional

$$\widehat{\rho} : dom(\widehat{\rho}) \subset dom(\widehat{\theta}) \to \mathcal{R} \quad , \quad P^L \longmapsto \widehat{\rho}\left(P^L\right) := \widehat{\rho}_{P^L} \quad ,$$

is the functional solution of the implicit equation

$$m(\rho, \widehat{\theta}_{P^L})' \, \mathcal{S} \, \frac{\partial}{\partial\rho'} m(\rho, \widehat{\theta}_{P^L}) = 0 \quad , \tag{11}$$

in ρ, for some positive definite deterministic matrix $\mathcal{S}$. Under the given assumptions, we have $\widehat{\rho}\left(P_0^L\right) = \rho_0$.

Remind that we focus on structural models where P_ρ^L is not expressible in an analytical form; therefore the expectation in (11) will have to be computed by Monte Carlo simulation, using some simulated series $x^k(\rho)$, for k sufficiently large.

The structural parameter is estimated by $\widehat{\rho}_n := \widehat{\rho}\left(P_n^L\right)$, which solves (11) with respect to the empirical measure P_n^L, and for some positive definite sequence $\mathcal{S}_n$ converging $a.s. - P_0$ to $\mathcal{S}$. Specifically, $\widehat{\rho}_n$ is such that

$$m(\widehat{\rho}_n, \widehat{\theta}_n)' \, \mathcal{S}_n \, \frac{\partial}{\partial\rho'} m(\widehat{\rho}_n, \widehat{\theta}_n) = 0 \quad .$$

Weak convergence $a.s. - P_0$ of P_n^L to P_0^L implies with Assumption 7 and some further regularity conditions on m (see Gallant and Tauchen (1996) and Gourieroux, Monfort and Renault (1993)), that also the sequence $(\widehat{\rho}_n)_{n\in\mathbb{N}}$ is consistent and asymptotically normal at the model P_0.

Property 8 *Under regularity conditions on $(\widehat{\rho}_n)_{n\in\mathbb{N}}$ it follows*

$$\begin{cases} \lim_{n\to\infty} \parallel \widehat{\rho}_n - \rho_0 \parallel = 0 \quad , \quad a.s. - P_0 \\ \sqrt{n}(\widehat{\rho}_n - \rho_0) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N\left(0, \Sigma_{\mathcal{S}}\right) \end{cases} ,$$

where

$$\Sigma_{\mathcal{S}} = \left(M_\rho' \mathcal{S} M_\rho\right)^{-1} M_\rho' \mathcal{S} M_\theta \mathcal{V}_0 M_\theta' \mathcal{S} M_\rho \left(M_\rho' \mathcal{S} M_\rho\right)^{-1} , \tag{12}$$

with

$$\underset{(q\times l)}{M_\rho} := \frac{\partial}{\partial \rho'} m(\rho,\theta)|_{\rho=\rho_0,\theta=\theta_0} \quad , \quad \underset{(q\times q)}{M_\theta} := \frac{\partial}{\partial \theta'} m(\rho,\theta)|_{\rho=\rho_0,\theta=\theta_0} \quad . \tag{13}$$

Standard regularity conditions allowing the interchange of integration and differentiation in (13) imply $M_\theta = A_0$ (see 8), so that finally (12) reads

$$\Sigma_{\mathcal{S}} = \left(M_\rho' \mathcal{S} M_\rho\right)^{-1} M_\rho' \mathcal{S} B_0 \mathcal{S} M_\rho \left(M_\rho' \mathcal{S} M_\rho\right)^{-1} . \tag{14}$$

Moreover, for the same reason M_ρ is the covariance between $\psi^0 := \psi(X^L;\theta_0)$ and the unconditional score function $s_{p^L}^0 := \frac{\partial}{\partial\rho} \ln p^L \left(X^L;\rho\right)|_{\rho=\rho_0}$, i.e.

$$M_\rho = Cov\left(\psi^0, s_{p^L}^0\right) \quad . \tag{15}$$

Finally, if $E\left(\psi(X^L;\theta_0)|X^{L-1}\right) = 0$ the unconditional score $s_{p^L}^0$ in (15) may be replaced by the conditional score $s_{p_L}^0 := \frac{\partial}{\partial\rho} \ln p_L \left(X^L;\rho\right)|_{\rho=\rho_0}$.

The next section analyzes the local robustness properties of EMM estimators and statistics derived from EMM estimators.

3 Local Robustness Properties of EMM Estimators

In this section we derive the IF of the EMM estimator. The EMM methodology is mainly applied to estimate dependent stochastic processes. Therefore, based on previous works of Künsch (1984), Martin and Yohai (1986) and Ronchetti and Trojani (2001), we also discuss the peculiarities arising from approximating the asymptotic bias in an EMM framework when observations are not *iid*.

3.1 The Influence Function of the EMM Estimator

We analyze the local behaviour of the EMM functional $\widehat{\rho}$ in a nonparametric neighborhood

$$\mathcal{U}_\eta\left(P^L_{\rho_0}\right) = \left\{P^L_{0,\varepsilon,G^L} := (1-\varepsilon)\,P^L_{\rho_0} + \varepsilon G^L \quad | \quad 0 \le \varepsilon < \eta \le 1\ ;\ G^L \in \mathcal{M}^L_{stat}\right\} \tag{16}$$

of a parametric reference model $P^L_{\rho_0} \in \mathcal{P}$ by means of the IF; see Hampel (1974) for basic definitions and properties and Künsch (1984) for the time series context. In the sequel we denote by $dom\,(T)$ the domain[4] of a statistical functional T.

Definition 9 *The influence function* $IF\left(\cdot;T,P^L_0\right)$ *of a statistical functional* $T : dom\,(T) \subset \mathcal{M}^L_{stat} \to \mathbb{R}^m$ *at* P^L_0 *is defined by*

$$IF\left(x^L;T,P^L_0\right) = \lim_{\varepsilon\downarrow 0}\frac{T\left(P^L_{0,\varepsilon,\delta_{x^L}}\right) - T\left(P^L_0\right)}{\varepsilon} \quad ,$$

for all $x^L \in \mathbb{R}^{vL}$ *such that this limit exists.*

For an M-estimator defined by a score function ψ the IF is given by the standard expression (see also Hampel et al. (1986) pp. 230)

$$IF(x^L;\widehat{\theta},P^L_0) = -M_\theta^{-1}\,\psi(x^L;\theta_0). \tag{17}$$

The IF of the EMM functional $\widehat{\rho}$ is obtained by implicitly differentiating the first order condition

$$m\left(\widehat{\rho}(P^L),\widehat{\theta}(P^L)\right)'\,\mathcal{S}\,\frac{\partial}{\partial\rho'}m\left(\widehat{\rho}(P^L),\widehat{\theta}(P^L)\right) = 0 \quad , \tag{18}$$

in direction δ_{x^L}. We then have

$$IF(x^L;\widehat{\rho},P^L_0) = -\left(M'_\rho \mathcal{S} M_\rho\right)^{-1} M'_\rho \mathcal{S} M_\theta\, IF(x^L;\widehat{\theta},P^L_0) \quad . \tag{19}$$

Thus, the IF of the EMM estimator depends linearly on the IF of the estimator of the auxiliary parameters, implying that the IF of the EMM estimator is bounded if and only if the estimator of the auxiliary parameter has a bounded IF. Furthermore, remark that a model deviation G^L

[4] $dom\,(T)$ does not generally contain $P^L_{0,\varepsilon,\delta_{X^L}}$ for arbitrary directions of contamination x^L. However, it is possible to extend the domain of T to all measures P^L_{0,ε,G^L} such that $E_{G^L}\left(\psi(X^L;\theta_0)\right)$ is well defined. The same applies to the EMM functionals $\widehat{\theta}$ and $\widehat{\rho}$.

having a large influence on the estimated auxiliary parameters does not have necessarily the same large influence on the estimates of the structural parameters. Specifically, the final effect depends on some scalar products between the score of the auxiliary model and the columns of M_ρ. Indeed, replacing (17) in (19) it follows

$$IF(x^L; \widehat{\rho}, P_0^L) = \left(M_\rho' \mathcal{S} M_\rho\right)^{-1} M_\rho' \mathcal{S}\, \psi(x^L; \theta_0) \quad . \tag{20}$$

Hence, for a $\xi \in \mathbb{R}^{vL}$ such that $\psi(\xi; \theta_0)$ is large, ξ may have no influence on the structural parameters if the vector $\psi(\xi; \theta_0)$ belongs to the kernel[5] of $M_\rho' \mathcal{S}$.

Two examples of non robust auxiliary score functions are presented in the sequel. The first example considers seminonparametric auxiliary score functions.

Example 10 *The seminonparametric (SNP) class of conditional densities presented in Gallant and Nychka (1987) has become the standard auxiliary model used in EMM estimation. For example, Gallant and Tauchen (1998) estimate the parameters of a dynamic nonlinear system with partially observed variables by means of an auxiliary score obtained from a SNP-AR(1)-ARCH(4) model with conditional density*

$$f\left(x_t \mid x_{t-5}^{t-1}; \theta\right) = \frac{\left[P(z_t, x_{t-1}; \theta)\right]^2 \phi(z_t)}{\int \left[P(u, x_{t-1}; \theta)\right]^2 \phi(u) du} \tag{21}$$

where

- $\phi(z_t) = (2\pi)^{-1/2} \exp\left(-z_t^2/2\right)$
- e_t *denotes the non standardized innovation, i.e.* $e_t = x_t - \mu_0 - \mu_1 x_{t-1}$,
- z_t *denotes the standardized innovation, i.e.* $z_t = R_t^{-1} e_t$,
- R_t^{-1} *denotes the conditional scale function*

$$R_t = \gamma_0 + \gamma_1 |e_{t-1}| + \gamma_2 |e_{t-2}| + \gamma_3 |e_{t-3}| + \gamma_4 |e_{t-4}|$$

- $P(z_t, x_{t-1}; \theta) = \sum_{i=0}^{4} \left(a_{0i} + a_{1i} x_{t-1}\right) z_t^i$, $a_{00} = 1$.

Clearly, the score implied by (21) is unbounded since it is the one of an AR-ARCH model.

The next example considers auxiliary score functions induced by a GMM auxiliary model, as in RGMM estimation.

[5] The $l \times q$ matrix M_ρ' has full row rank l so that the dimension of its kernel is equal to $q - l$.

Example 11 *Consider the case where* $\dim(\rho) < \dim(\theta)$. *If the auxiliary parameters are estimated by GMM as in Example 4 and if* $\dim(\theta) < \dim(h)$, *then the influence function of* $\widehat{\theta}$ *is given by*

$$IF(x^L;\widehat{\theta},P_0^L) = -\left(C_0'W_0C_0\right)^{-1} C_0'W_0h(x^L;\theta_0)\ ,$$

where

$$C_0 = E_{P_0^L}\left(\frac{\partial}{\partial\theta'}h(X^L;\theta)|_{\theta=\theta_0}\right).$$

Note the similarities with formula (20) when h, W_0 *and* C_0 *replace* ψ, $\mathcal{S}$ *and* M_ρ, *respectively. Furthermore, using the notation* $K_0 = Cov\left(h(X^L;\theta_0), s_{p^L}\left(X^L;\rho_0\right)\right)$, M_ρ *can be written as*

$$M_\rho = C_0'\,W_0\,K_0\,.$$

Consequently, the influence function of $\widehat{\rho}$ *reads*

$$\begin{aligned} IF(x^L;\widehat{\rho},P_0^L) &= \left(K_0'W_0C_0\mathcal{S}C_0'W_0K_0\right)^{-1}\,K_0'\,W_0\,C_0\,\mathcal{S}\,C_0'\,W_0\,h(x^L;\theta_0) \\ &= \left(K_0'\,Q_0\,K_0\right)^{-1}\,K_0'\,Q_0\,h(x^L;\theta)\,, \end{aligned} \tag{22}$$

where

$$Q_0 := W_0C_0\mathcal{S}C_0'W_0\,.$$

Formula (22) preserves the same structure of formula (20) but expresses the sensitivity of the estimator $\widehat{\rho}$ *in terms of the orthogonality function* h *and its covariance with the score function* s_{p^L}. *Finally, note that by contrast with* $\mathcal{S}$ *the weighting matrix* Q_0 *is not invertible.*

Finally, notice that the standard result of the IF of an M-estimator can be recovered in the EMM setting when assuming a correctly specified auxiliary model identical to the structural one. In fact, for this case one obtains from (20) when the structural and the auxiliary models are identical and the auxiliary estimator $\widehat{\theta}$ is Fisher consistent[6] , the expression

$$IF(x^L;\widehat{\rho},P_0^L) = -\left[E\left(\nabla_{\rho'}\psi(X^L;\rho)|_{\rho=\rho_0}\right)\right]^{-1}\psi(x^L;\rho_0)\quad ,$$

which is Huber's IF formula for M-estimators; see also Huber (1981).

3.2 Asymptotic Bias approximation in the EMM Framework

When observations are *iid*, given a local deviation P^1_{0,ε,G^1} (see (16)) the asymptotic bias of the estimator $\widehat{\rho}$

$$\mathcal{B}(\varepsilon) := \widehat{\rho}(P^1_\varepsilon) - \widehat{\rho}(P^1_0)\,, \tag{23}$$

[6] We refer to Hampel et al. (1986), p. 83, and the references therein for the notion of Fisher consistency of a statistical functional.

can be approximated by an integral of the IF of $\widehat{\rho}$ with respect to G^1 (Hampel et. al., 1986), i.e.

$$\mathcal{B}(\varepsilon) = \varepsilon\, E_{G^1}\left[IF(X_1;\widehat{\rho},P_0^1)\right] + o(\varepsilon)\,. \tag{24}$$

We remarked at the beginning of this section that EMM estimation is mainly concerned with dependent processes. By contrast with the *iid* setting, in the time series context it is not always possible to approximate a contaminated measure as a mixture of P_0^L and a further *probability* measure $G^L \in \mathcal{M}_{stat}^L$, even for $\varepsilon \downarrow 0$. However, Künsch (1984) shows that if the curve $\varepsilon \mapsto P_{0,\varepsilon}^L \in \mathcal{M}_{stat}^L$ is sufficiently smooth for $\varepsilon^{-1}(P_{0,\varepsilon}^L - P_0^L)$ to converge weakly as $\varepsilon \downarrow 0$ to a finite signed measure $\widetilde{P}^L$, then

$$\mathcal{B}'(0) = \int IF(x^L;\widehat{\rho},P_0^L)\, d\widetilde{P}^L(x^L),$$

and the asymptotic bias of $\widehat{\rho}$ can be approximated similarly to (24) as

$$\widehat{\rho}(P_{0,\varepsilon}^L) - \widehat{\rho}(P_0^L) = \varepsilon \int IF(x^L;\widehat{\rho},P_0^L)\, d\widetilde{P}^L(x^L) + o(\varepsilon). \tag{25}$$

For specific models of contamination such that $\gamma := \lim \varepsilon^{-1} P\left[\text{at least one outlier in } X^L\right]$ exists, it is possible to approximate the contaminated measure $P_{0,\varepsilon}^L$ by a suitable linear combination of the form

$$P_{0,\varepsilon}^L = (1-\gamma\varepsilon)\, P_0^L + \gamma\varepsilon\mu^L + o\,(\varepsilon) \quad , \tag{26}$$

where μ^L is a finite signed measure that depends on the distribution of the outliers and on P_0^L. Therefore, in (25) and (26) the lag length L also has an impact on the asymptotic bias and on the robustness properties of the given EMM estimator. We further illustrate this last point by an example that helps us to discuss in more detail these robustness aspects related to the EMM methodology.

Example 12 *We assume that the observed process $\mathbb{X}$ is generated according to the following pure replacement model*

$$X_t = (1 - H_t^\varepsilon)\, Y_t + H_t^\varepsilon \xi\,. \tag{27}$$

where the clean process $\mathbb{Y}$ is given by the Gaussian ARMA(1,1) process

$$Y_t = \rho_1 + \rho_2 Y_{t-1} + \epsilon_t + \rho_3 \epsilon_{t-1}, \quad \epsilon_t \ iid \ \sim \ N\,(0,\rho_4)\,. \tag{28}$$

For simplicity we take ξ to be a constant. $\mathbb{H}^{\varepsilon}$ represents an iid 0-1 process independent of $\mathbb{Y}$ with the property that $P(H_t^{\varepsilon}=1)=\varepsilon$, i.e. at every date t the clean observation Y_t is replaced by ξ with probability ε. In order to estimate the structural ARMA(1,1) model (28) we specify an auxiliary Gaussian AR(3) process, i.e.

$$X_t = \theta_1 + \theta_2 X_{t-1} + \theta_3 X_{t-2} + \theta_4 X_{t-3} + \varepsilon_t, \quad \varepsilon_t \text{ iid } N(0, \theta_5). \tag{29}$$

In this model, the number L of coordinates used to define the functional estimator $\widehat{\theta}$ is equal to 4. Therefore, let us define $H^{4,\varepsilon} := (H_1^{\varepsilon}, H_2^{\varepsilon}, H_3^{\varepsilon}, H_4^{\varepsilon})$ and denote by $h_i \in \{0,1\}^4$, $i=1,\ldots,16$, the possible realizations of $H^{4,\varepsilon}$, where $h_1 = 0$. For every $B \in \mathcal{B}(\mathbb{R}^4)$ and $\varepsilon > 0$ we have

$$P_{0,\varepsilon}^4(B) := P\left(X^4 \in B\right) = \sum_{i=1}^{16} P\left(X^4 \in B \mid H^{4,\varepsilon} = h_i\right) P\left(H^{4,\varepsilon} = h_i\right).$$

We denote by $P'\left(H^{4,0}=h_i\right)$ the derivative of $P\left(H^{4,\varepsilon}=h_i\right)$ in 0. Because conditional probabilities $\mu_i(B) := P\left(X^4 \in B \mid H^{4,\varepsilon}=h_i\right)$ depend on the joint distribution of Y^4 and ξ but not on ε

$$\widetilde{P}^4(B) := \lim_{\varepsilon \downarrow 0} \frac{P_{0,\varepsilon}^4(B) - P_0^4(B)}{\varepsilon} \quad \textit{exists}$$

and is equal to

$$\widetilde{P}^4(B) = \sum_{i=2}^{16} \left(\mu_i(B) - P_0^4(B)\right) P'\left(H^{4,0}=h_i\right) = \gamma\left(\mu^4(B) - P_0^4(B)\right)$$

where $P_0^4(B) := P\left(Y^4 \in B\right)$, $\gamma := \sum_{i=2}^{16} P'\left(H^{4,0}=h_i\right)$, $p_i := \gamma^{-1} P'\left(H^{4,0}=h_i\right) \geq 0$ for $i = 2,\ldots 16$ and $\mu^4(B) := \sum_{i=2}^{16} \mu_i(B)\, p_i$ is a probability measure. In this example the value of the constant γ in (26) is equal to 4, i.e. the number of coordinates on which the score function ψ is defined, and $p_i = 0$ for all i such that h_i has two or more components equal 1. Thus, formula (26) reads

$$P_{0,\varepsilon}^4 = (1-4\varepsilon) P_0^4 + 4\varepsilon\mu^4 + o(\varepsilon).$$

While the asymptotic bias approximation's formulae (24), (25) in the iid case and in a time series context, respectively, are at least formally identical, for the latter the smoothness of $P_{0,\varepsilon}^L$ can be not generally granted in the case where a time series functional depends on distributions defined on an infinite dimensional space, i.e. when $L \to \infty$. This causes serious problems in obtaining robust estimation procedures for non Markovian processes[7] . However, using the EMM a non Markovian process can be also quite efficiently estimated by means of an auxiliary model where the auxiliary functional estimator $\widehat{\theta}$ (and consequently $\widehat{\rho}$) depends on distributions defined on a finite number L of coordinates of the process[8] $\mathbb{X}$. Thus, using REMM one can avoid in a natural

[7] Martin and Yohai (1986) propose a definition of IF for time series functionals defined on the entire distribution of the process.

[8] Gallant and Long (1997) show that the efficiency of the EMM estimator can approach that of maximum likelihood as both $\dim\psi$ and L increase. However, in this paper both $\dim\psi$ and L have to be considered fixed.

way the problems arising in the robust estimation of non Markovian time series processes[9] . This feature makes EMM a particularly suitable methodology for the robust estimation of time series models that include the non Markovian case. This last assertion has however to be specified further. Indeed, recall that the IF of $\widehat{\rho}$ was derived under the implicit assumption that the weighting matrix $\mathcal{S}$ in (18) was known. In applications $\mathcal{S}$ is unknown and is replaced by a suitable estimator $\widehat{\mathcal{S}}_n$, whose asymptotic limit depends on $\widehat{\theta}$ and on a probability distribution defined on given (possibly infinite) number of coordinates of the process. For instance, in Example 12 for $\varepsilon = 0$, the dynamic misspecification of the auxiliary model implies an optimal weighting matrix given by

$$\mathcal{S}_{opt}\left(P_0, \widehat{\theta}(P_0^L)\right) = (B_0)^{-1} , \tag{30}$$

so that the best EMM functional $\widehat{\rho}$ actually depends through $\mathcal{S}$ on the whole distribution of the process $\mathbb{X}$ (at least theoretically). The opposite situation where the functional $\widehat{\rho}$ depends on the same marginal distribution P_0^L as $\widehat{\theta}$ is given in Gallant and Tauchen (1996) who propose a weighting matrix

$$\mathcal{S}\left(P_0^L, \widehat{\theta}(P_0^L)\right) = E_{P_0^L}(\psi(X^L; \widehat{\theta}(P_0^L))\psi(X^L; \widehat{\theta}(P_0^L))')$$

whenever the auxiliary model is a good statistical approximation of the data generation process (DGP). Indeed, under the assumption that the derivative $D = \frac{\partial}{\partial\varepsilon}\mathcal{S}|_{\varepsilon=0}$ exists, from the chain rule it follows that D has no impact on the IF of $\widehat{\rho}$ because, at the model P_0, $m\left(\widehat{\rho}(P_0), \widehat{\theta}(P_0^L)\right) = 0$. The problem of a non differentiability of $\mathcal{S}$ can be avoided in at least two related ways in practice. Both are based on the fact that the choice of the weighting matrix $\mathcal{S}$ is important from an efficiency perspective, but is not necessary to ensure consistency of the corresponding EMM estimator:

1. A first solution is to use an approximate optimal matrix $\mathcal{S}$ that depends on a finite dimensional marginal distribution of $\mathbb{X}$ and for which the derivative D exists[10] .

[9] See also Martin and Yohai (1986) for a discussion related to this point within the setting of a linear MA model.

[10] The existence of the derivative $\frac{\partial}{\partial\varepsilon}\mathcal{S}_{opt}|_{\varepsilon=0}$ is also necessary for the existence of the change of variance function (cf. Hampel et. al., 1986, Chapter 2) of the auxiliary estimator $\widehat{\theta}$. A sufficient condition for this is the boundedness of the auxiliary score ψ.

2. A second straightforward way to solve the problem is to use the same deterministic matrix for all n, as for example the identity matrix.

Besides the choice of the lag length L for the auxiliary model, the above one is a second trade-off between efficiency and robustness in robust EMM estimation.

4 Robust EMM estimators

In this section we first define two robust EMM estimators with bounded IF. Both of them are defined by a truncation algorithm that makes use of the Huber function (see for instance Huber (1981)). However, they differ in the metric used to measure the length of the score function ψ and consequently in the way how the corresponding robust weights are constructed. It is a remarkable feature of the REMM estimators presented below that they correct simultaneously for two sources of asymptotic bias arising in the REMM setting: (i) the standard bias arising in the EMM estimation of an auxiliary model and (ii) the usual bias induced by a robust M-estimation of the auxiliary model via the truncation of an unbounded auxiliary score function. In a second step we present the estimation algorithms and discuss some final points on the trade-off between efficiency and robustness in REMM estimation.

4.1 Bounding the IF

Following Ronchetti and Trojani (2001), the construction of a REMM estimator is performed by bounding the IF of $\widehat{\rho}$ with respect to the metric induced by its variance-covariance matrix. Formally, we want the IF of $\widehat{\rho}$ to satisfy

$$\| IF(x^L; \widehat{\rho}, P_0^L) \|_{\Sigma_{\mathcal{S}}^{-1}} := \| \Sigma_{\mathcal{S}}^{-1/2} IF(x^L; \widehat{\rho}, P_0^L) \| \leq c, \tag{31}$$

where c is an a priori positive bound on the self-standardized sensitivity of $\widehat{\rho}$ (see also Hampel et al. (1986), Chapter 4). By (31) this gives

$$\| IF(x^L; \widehat{\rho}, P_0^L) \|^2_{\Sigma_{\mathcal{S}}^{-1}} = \psi(x^L; \theta_0)' \mathcal{S} M_\rho \left(M_\rho' \mathcal{S} M_\rho \right)^{-1} \Sigma_{\mathcal{S}}^{-1} \left(M_\rho' \mathcal{S} M_\rho \right)^{-1} M_\rho' \mathcal{S} \, \psi(x^L; \theta_0),$$

and, using (14),

$$\| IF(x^L;\widehat{\rho},P_0^L) \|^2_{\Sigma_{\mathcal{S}}^{-1}} = \psi(x^L;\theta_0)'\,\mathcal{S}M_\rho\left(M_\rho'\mathcal{S}B_0\mathcal{S}M_\rho\right)^{-1}M_\rho'\mathcal{S}\,\psi(x^L;\theta_0). \tag{32}$$

In this case, the matrix M_ρ must be calculated numerically even when interchange of integration and differentiation in (13) is permitted because both the structural density and the structural score functions cannot be generally expressed in closed form.

As a simple alternative, the properties of orthogonal projections for the matrix

$$K := B_0^{1/2}\mathcal{S}M_\rho\left(M_\rho'\mathcal{S}B_0\mathcal{S}M_\rho\right)^{-1}M_\rho'\mathcal{S}B_0^{1/2}$$

imply

$$\begin{aligned}
\| IF(x^L;\widehat{\rho},P_0^L) \|^2_{\Sigma_{\mathcal{S}}^{-1}} &= \psi(x^L;\theta_0)'\,B_0^{-1/2}\,K\,B_0^{-1/2}\,\psi(x^L;\theta_0) \\
&\leq \psi(x^L;\theta_0)'\,B_0^{-1}\,\psi(x^L;\theta_0) \\
&= \| \psi(x^L;\theta_0) \|^2_{B_0^{-1}} \\
&= \psi(x;\theta_0)'\,M_\theta^{-1}\left(M_\theta^{-1}B_0M_\theta^{-1}\right)^{-1}M_\theta^{-1}\,\psi(x;\theta_0) \\
&= \| IF(x;\widehat{\theta},P_0^L) \|^2_{\left(M_\theta^{-1}B_0M_\theta^{-1}\right)^{-1}} \\
&= \| IF(x;\widehat{\theta},P_0^L) \|^2_{\mathcal{V}_0^{-1}}\,.
\end{aligned}$$

Hence:

$$\| IF(x^L;\widehat{\rho},P_0^L) \|_{\Sigma_{\mathcal{S}}^{-1}} \leq \| IF(x^L;\widehat{\theta},P_0^L) \|_{\mathcal{V}_0^{-1}} = \| \psi(x^L;\theta_0) \|_{B_0^{-1}}\,. \tag{33}$$

The implications of this last result are very useful. Firstly, bounding the self-standardized norm of the score of the auxiliary model by a positive constant c is sufficient in order to bound the self-standardized IF of the EMM estimator of the structural parameters. Secondly, *this bound does not depend on the weighting matrix* $\mathcal{S}$ used in the second step of the EMM. Finally, by contrast with $\| IF(x^L;\widehat{\rho},P_0^L) \|_{\Sigma_{\mathcal{S}}^{-1}}$ in order to bound ψ in the metric defined by B_0^{-1} it is not necessary to compute numerically the matrix M_ρ.

4.2 Definition of Robust EMM estimators

Construction of robust EMM estimators having a self-standardized IF bounded by c can be performed by truncating the function $\psi(x^L;\theta_0)$ with an appropriate algorithm; see also Hampel et al. (1986), pp. 238 ff. We now present two procedures by which this can be achieved. In both procedures we replace the (unbounded) function ψ of a classical (nonrobust) EMM estimator by a new bounded one, denoted by $\psi_c^{\widetilde{A}}$ and ψ_c^A, respectively. Computation of M_ρ is required for $\psi_c^{\widetilde{A}}$ but not for ψ_c^A as the latter makes use of inequality (33). Because of its simplicity, we start with the construction of ψ_c^A.

For a non-singular matrix $A \in \mathbb{R}^{l_\theta} \times \mathbb{R}^{l_\theta}$ we define a new function $\psi^A : \mathbb{R}^{vL} \times \Theta \to \mathbb{R}^{l_\theta}$ as the scale transformation of ψ, i.e.

$$\psi^A(x^L;\theta) := A\psi(x^L;\theta).$$

Further, using the Huber function

$$\mathcal{H}_c : \mathbb{R}^{l_\theta} \to \mathbb{R}^{l_\theta}\,,\ x \longmapsto x\, w_c(x)$$

where

$$w_c(x) = \begin{cases} \min\left(1, c/ \parallel x \parallel\right) & x \neq 0 \\ 1 & x = 0 \end{cases},$$

and $\parallel \cdot \parallel$ denotes the Euclidean norm, we introduce the truncated auxiliary score function $\psi_c^A(x^L;\theta) : \mathbb{R}^{vL} \times \Theta \to \mathbb{R}^{l_\theta}$ defined as

$$\psi_c^A(x^L;\theta) := \mathcal{H}_c\left(\psi^A(x^L;\theta)\right) = A\psi(x^L;\theta)\, w_c(A\psi(x^L;\theta))\,, \tag{34}$$

so that the new binding function

$$b_{\psi_c^A} : \mathcal{R} \longrightarrow \Theta,\ \rho \mapsto \theta = b_{\psi_c^A}(\rho) \tag{35}$$

and the functional estimator $\widehat{\theta}$ are now implicitly defined as the solution of the nonlinear system of equations

$$E_{P_\rho^L}\left[\psi_c^A(X^L;\theta)\right] = 0 \tag{36}$$

$$E_{P^L}\left[\psi_c^A(X^L;\theta)\right] = 0\ ,$$

respectively. The binding function $b_{\psi_c^A}$ is generally different from that implied by the score function ψ. However, Fisher consistency of the structural parameters is naturally maintained in the EMM framework because of the second step[11] . The non-singular matrix $A \in \mathbb{R}^{l_\theta} \times \mathbb{R}^{l_\theta}$ is determined by solving

$$\begin{aligned} I &= B_0 = E_{P_0^L}\left(\psi_c^A(X^L;\theta_0)\psi_c^A(X^L;\theta_0)'\right) \\ &\quad + \sum_{\tau=1}^{\infty} E_{P_0}(\psi_c^A(X^L;\theta_0)\psi_c^A(X_{\tau+1}^{L+\tau};\theta_0)') \\ &\quad + \sum_{\tau=1}^{\infty} E_{P_0}(\psi_c^A(X_{\tau+1}^{L+\tau};\theta_0)\psi_c^A(X^L;\theta_0)'). \end{aligned} \tag{37}$$

The system of equations (37) ensures $B_0 = I$ for the robust EMM estimator implied by the score ψ_c^A so that, because of (33), the self standardized IF of $\widehat{\rho}$ is automatically bounded by c.

Remark 13 *In applications, B_0 is replaced by a consistent estimator $\widehat{B}_n$ (c.f. Andrews, 1991; Gallant, 1987; Newey and West, 1987).*

Construction of the second version of a robust EMM estimator is similar to the first one. Let us define the non-singular matrix $\widetilde{A} \in \mathbb{R}^{l_\rho} \times \mathbb{R}^{l_\rho}$ and the truncated auxiliary score function $\psi_c^{\widetilde{A}}(x^L;\theta)$: $\mathbb{R}^{vL} \times \Theta \to \mathbb{R}^{l_\theta}$ as

$$\psi_c^{\widetilde{A}}(x^L;\theta) := \psi(x^L;\theta)\, w_c(\widetilde{A}M_\rho'\mathcal{S}\,\psi(x^L;\theta))\,. \tag{38}$$

The interpretation of (38) is straightforward: observations x^L which cause the score function to be very large and hence have a large influence on the estimates of θ_0 and ρ_0 must be downweighted. However, the length of ψ is measured with respect to the semi-metric induced by the matrix $\mathcal{S}M_\rho\widetilde{A}'\widetilde{A}M_\rho'\mathcal{S}$ which takes into account the fact that we are interested in the robustness of the structural estimator and not necessarily in the one of the auxiliary estimator. Indeed, the IF of $\widehat{\rho}$ depends on the matrix product $M_\rho'\mathcal{S}$ (see formula (20)). Because $rank\,(M_\rho) = l_\rho \leq l_\theta = rank\,(\mathcal{S})$, observations with large influence on the auxiliary parameters may have small or even zero influence

[11] We will return to this point later when we discuss REMM in the context of robust ML estimation.

on the structural parameters and, therefore, do not have to be downweighted. Furthermore, (31) is true whenever

$$\widetilde{A}'\widetilde{A} = \left(M_\rho' \mathcal{S} B_0 \mathcal{S} M_\rho\right)^{-1} \tag{39}$$

is satisfied. In fact, substituting (39) in (32) we obtain

$$\| \, IF(x^L; \widehat{\rho}, P_0^L) \, \|^2_{\Sigma_{\mathcal{S}}^{-1}} = \psi_c^{\widetilde{A}}(x^L;\theta_0)' \, \mathcal{S} M_\rho \widetilde{A}' \widetilde{A} M_\rho' \mathcal{S} \, \psi_c^{\widetilde{A}}(x^L;\theta_0) \leq c^2$$

Remark 14 *Note that because both matrices M_ρ and B_0 depend on $\psi_c^{\widetilde{A}}$, matrix $\widetilde{A}$ enters also in the right hand side of equation (39). However, we prefer not to show explicitly this dependence for convenience of notation. In applications, the matrices B_0, $\mathcal{S}$ and M_ρ are replaced by consistent estimators $\widehat{B}_n$, $\widehat{\mathcal{S}}_n$ and $\widehat{M}_{\rho,n}$, respectively.*

4.3 The Estimation Algorithms

Before explaining the algorithms in detail, we summarize for easy reference the sets of implicit equations that have to be solved to compute a REMM estimator. The empirical $(L)-$dimensional marginal distribution of a sample $x^k(\rho) := (x_1'(\rho), x_2'(\rho), \ldots, x_k'(\rho))'$ simulated according to P_ρ is denoted by $P^L_{\rho,k}$.

a. Equation for the $(l_\theta)-$ dimensional vector of **auxiliary parameters**

$$E_{P_n^L}\left(\psi_c^A(X^L;\theta)\right) = 0 \; , \text{ or} \tag{40}$$

$$E_{P_n^L}\left(\psi_c^{\widetilde{A}}(X^L;\theta)\right) = 0 \; , \tag{41}$$

respectively.

b. Equation for the $(l_\rho)-$ dimensional vector of **structural parameters**

$$\frac{\partial}{\partial\rho} E_{P^L_{\rho,k}}\left(\psi_c^A\left(X^L;\theta\right)\right)' E_{P^L_{\rho,k}}\left(\psi_c^A\left(X^L;\theta\right)\right) = 0 \; , \text{ or} \tag{42}$$

$$\frac{\partial}{\partial\rho} E_{P^L_{\rho,k}}\left(\psi_c^{\widetilde{A}}\left(X^L;\theta\right)\right)' E_{P^L_{\rho,k}}\left(\psi_c^{\widetilde{A}}\left(X^L;\theta\right)\right) = 0 \; , \tag{43}$$

respectively.

c. Equation for the $(l_\theta \times l_\theta)-$ dimensional matrix A defined in (37). B_0 is replaced by a suitable consistent estimator $\widehat{B}_n$, i.e. the Newey and West estimator (see Newey and West, 1987)

$$\begin{aligned} I &= E_{P_n^L}\left(\psi_c^A(X^L;\theta)\psi_c^A(X^L;\theta)'\right) \\ &+\sum_{\tau=1}^{q}\left[1-\frac{\tau}{q+1}\right] E_{P_n^{L+\tau}}(\psi_c^A(X^L;\theta)\psi_c^A(X_{\tau+1}^{L+\tau};\theta)') \\ &+\sum_{\tau=1}^{q}\left[1-\frac{\tau}{q+1}\right] E_{P_n^{L+\tau}}(\psi_c^A(X_{\tau+1}^{L+\tau};\theta)\psi_c^A(X^L;\theta)') \,. \end{aligned} \tag{44}$$

or equivalently

$$\begin{aligned} (A'A)^{-1} &= E_{P_n^L}\left(\psi(X^L;\theta)\,\psi(X^L;\theta)' w_c^2\left(\psi^A(X^L;\theta)\right)\right) \qquad (45) \\ &+\sum_{\tau=1}^{q}\left(1-\frac{\tau}{q+1}\right) E_{P_n^{L+\tau}}\left(\psi(X^L;\theta)\,\psi(X_{\tau+1}^{L+\tau};\theta)' w_c\left(\psi^A(X^L;\theta)\right) w_c\left(\psi^A(X_{\tau+1}^{L+\tau};\theta)\right)\right) + \\ &+\sum_{\tau=1}^{q}\left(1-\frac{\tau}{q+1}\right) E_{P_n^{L+\tau}}\left(\psi(X_{\tau+1}^{L+\tau};\theta)\,\psi(X^L;\theta)' w_c\left(\psi^A(X_{\tau+1}^{L+\tau};\theta)\right) w_c\left(\psi^A(X^L;\theta)\right)\right) . \end{aligned}$$

The $(l_\rho \times l_\rho)-$ dimensional matrix $\widetilde{A}$ defined in (39) is computed by solving

$$\left(\widetilde{A}'\widetilde{A}\right)^{-1} = \widehat{M}_{\rho,n}'\,\widehat{\mathcal{S}}_n\,\widehat{B}_n\,\widehat{\mathcal{S}}_n\,\widehat{M}_{\rho,n} \,. \tag{46}$$

It is not possible to simultaneously solve all sets of equations. Therefore, we solve the systems of equations **a-c** recursively, by updating the parameters at the end of each iteration. The econometrician fixes the degree of robustness by means of the bound $c > \sqrt{l_\theta}$ for the first version of a REMM estimator or $c > \sqrt{l_\rho}$ for the second one. Once a value has been assigned to c, it is held fixed until the end of the procedure. Index j denotes the j-th iteration. We first describe the updating steps of the first version of a REMM estimator.

Let us denote the starting values of the auxiliary and structural parameters by $\widehat{\theta}_{n,0}$, $\widehat{\rho}_{n,0}$. A starting value A_0 for the matrix A is determined by solving

$$\begin{aligned} (A'A)^{-1} &= E_{P_n^L}\left(\psi(X^L;\widehat{\theta}_n)\,\psi(X^L;\widehat{\theta}_n)'\right) \\ &+\sum_{\tau=1}^{q}\left[1-\frac{\tau}{q+1}\right] E_{P_n^{L+\tau}}\left(\psi(X^L;\widehat{\theta}_n)\,\psi(X^L;\widehat{\theta}_n)'\right) \\ &+\sum_{\tau=1}^{q}\left[1-\frac{\tau}{q+1}\right] E_{P_n^{L+\tau}}\left(\psi(X^{L+\tau};\widehat{\theta}_n)\,\psi(X^L;\widehat{\theta}_n)'\right) . \end{aligned}$$

Since A_0 solves the set of equations (45) for $c = \infty$ we can expect A_0 to be a reasonable starting point (see for instance Hampel et al., 1986, pp. 251). The REMM estimation is performed by iteration of the following four steps.

1. *Update step.* Formula (45) is particularly appropriate for a recursive update of A. In fact, once the right hand side of (45), denoted by D_{j-1}, has been computed by means of A_{j-1} and $\widehat{\theta}_{n,j-1}$, the new value A_j is obtained by solving $\left(A_j' A_j\right)^{-1} = D_{j-1}$. Specifically, since D_{j-1} is positive-definite we can apply the Cholesky decomposition of D_{j-1}, e.g. $LL' = D_{j-1}$, to get $A_j = L^{-1}$.

2. *Estimation of* θ_0. With the new matrix A_j perform the first step of the EMM as in (40) using the truncated score function $\psi_c^{A_j}$. The starting value of this minimization problem is $\widehat{\theta}_{n,j-1}$ and its solution is denoted by $\widehat{\theta}_{n,j}$.

3. *Estimation of* ρ_0. Solve the minimization problem (c.f. equation (42)):

$$\widehat{\rho}_{n,j} := \arg\min_{\rho \in \mathcal{R}} E_{P^L_{\rho,k}} \left(\psi_c^{A_j}\left(X^L; \widehat{\theta}_{n,j}\right)\right)' E_{P^L_{\rho,k}} \left(\psi_c^{A_j}\left(X^L; \widehat{\theta}_{n,j}\right)\right),$$

   using $\widehat{\rho}_{n,j-1}$ as starting value.

4. If $\widehat{\theta}_{n,j} - \widehat{\theta}_{n,j-1}$, $\widehat{\rho}_{n,j} - \widehat{\rho}_{n,j-1}$ and $A_j - A_{j-1}$ are small enough stop, otherwise go back to step 1.

The second version of the robust EMM estimator is similar to the first one with two additional updating steps regarding the matrices $\mathcal{S}$ and M_ρ. Given starting values $\widehat{\theta}_{n,0}$, $\widehat{\rho}_{n,0}$, we compute some starting values $\widehat{\mathcal{S}}_{n,0}$, $\widehat{M}_{\rho,n,0} = \frac{\partial}{\partial \rho'} m(\rho, \theta)|_{\rho=\widehat{\rho}_{n,0}, \theta=\widehat{\theta}_{n,0}}$ and $\widetilde{A}_0$ setting $c = \infty$. Specifically, $\widehat{M}_{\rho,n,0}$ generally has to be calculated numerically while $\widetilde{A}_0$ solves

$$\left(\widetilde{A}'\widetilde{A}\right)^{-1} = \widehat{M}'_{\rho,n,0} \, \widehat{\mathcal{S}}_{n,0} \, \widehat{B}_{n,0} \, \widehat{\mathcal{S}}_{n,0} \, \widehat{M}_{\rho,n,0} \, . \tag{47}$$

The REMM estimation is performed by iteration of the following four steps.

1. *Update step.* By means of $\widehat{\theta}_{n,j-1}$, $\widehat{\rho}_{n,j-1}$, $\widehat{M}_{\rho,n,j-1}$, $\widetilde{A}_{j-1}$, $\widehat{\mathcal{S}}_{n,j-1}$ or their just updated values, update $\mathcal{S}$, B, M_ρ and $\widetilde{A}$.

2. *Estimation of* θ_0. Perform the first step of the EMM as in (41) using the truncated score function $\psi_c^{\widetilde{A}_j}$.

3. *Estimation of* ρ_0. Solve the minimization problem

$$\widehat{\rho}_{n,j} := \arg\min_{\rho\in\mathcal{R}} E_{P^L_{\rho,k}}\left(\psi_c^{\widetilde{A}_j}\left(X^L;\widehat{\theta}_{n,j}\right)\right)' E_{P^L_{\rho,k}}\left(\psi_c^{\widetilde{A}_j}\left(X^L;\widehat{\theta}_{n,j}\right)\right).$$

4. If $\widehat{\theta}_{n,j}-\widehat{\theta}_{n,j-1}$, $\widehat{\rho}_{n,j}-\widehat{\rho}_{n,j-1}$ and $\widetilde{A}_j-\widetilde{A}_{j-1}$ are small enough stop, otherwise go back to step 1.

It is important to remark that both algorithms proposed above correct simultaneously for the standard bias arising in the EMM estimation of an auxiliary model and the usual bias induced by a robust M-estimation via the truncation of an unbounded auxiliary score function. This avoids an extra simulation based bias correction of the auxiliary model estimator $\widehat{\theta}$ (as necessary for instance in some RGMM estimation procedures for time series) and renders the above algorithms for computing REMM estimators computationally only slightly more demanding than the ones used in standard (non-robust) EMM estimation.

4.4 Robustness vs Efficiency

The main advantage of the EMM estimation method consists in its relative efficiency compared to other estimation methods for time series, see Gallant and Tauchen (1999), Andersen et al. (1999). Therefore, we discuss the different aspects underlying the trade-off between local robustness and asymptotic efficiency in the EMM framework. An important element for the whole discussion is the robustness tuning constant c. So far we have argued with a fixed but not specified value for c. In applications the econometrician has to select c on the basis of some prior information (or degree of confidence) about the kind and/or the extent of the deviation from the reference model[12]

. Although there is no a priori upper bound for the choice of c, Hampel et al. (1986), Section 4.3 showed that equations (37) and (39) have no solution if $c < \sqrt{l_\theta}$ or $c < \sqrt{l_\rho}$, respectively. A similar feature applies to our REMM setting. Indeed, within the REMM framework we are interested in the computation of

$$E_{P_0} \parallel \psi(X^L;\theta_0) \parallel^2_{(B_0)^{-1}} \text{ and } E_{P_0} \parallel IF(x^L;\widehat{\rho},P_0^L) \parallel^2_{\Sigma_S^{-1}} . \tag{48}$$

If $\left\{\psi(X^{L+\tau}_{\tau+1};\theta_0)\right\}_{\tau\geq 0}$ is a sequence of uncorrelated random variables then

$$B_0 = B_{0,0} \ ,$$

see (9) and (10), and we are essentially as in the iid. case so that we can use the same results of Hampel et al. (1986). However, because of a possible misspecification of the auxiliary model or because of the truncation in our REMM algorithm we cannot generally expect equality between the variance $B_{0,0}$ of $\psi(X^L;\theta_0)$ and B_0. Therefore, let us assume for illustration purposes that the following inequality holds

$$B_0 \leq B_{0,0} \ ,$$

so that

$$E_{P_0}\left[\parallel (B_0)^{-1/2} \ \psi(x^L;\theta_0) \parallel^2\right] \geq E_{P_0}\left[\parallel (B_{0,0})^{-1/2} \ \psi(x^L;\theta_0) \parallel^2\right] = l_\theta. \tag{49}$$

From (49) we obtain a lower bound for c. In fact, assuming

$$\parallel (B_0)^{-1/2} \ \psi(x^L;\theta_0) \parallel^2 < l_\theta \quad a.s. - P_0$$

then

$$E_{P_0}\left[\parallel (B_0)^{-1/2} \ \psi(x^L;\theta_0) \parallel^2\right] < l_\theta.$$

But this contradicts (49) so that we obtain the necessary condition $c \geq \sqrt{l_\theta}$. Notice that this lower bound depends on the number of auxiliary parameters and not on that of the structural ones.

[12] For a more objective way of selecting c in the context of RGMM testing see Ronchetti and Trojani (2001), p. 53-54.

Applying the same arguments to the second expectation in (49) we obtain a similar inequality, i.e.

$$E_{P_0} \parallel IF(x^L;\widehat{\rho},P_0^L) \parallel^2_{\Sigma_{\mathcal{S}}^{-1}} \geq l_\rho \ ,$$

so that $c \geq \sqrt{l_\rho}$.

In a related iid context Maronna (1976) proved existence and uniqueness of a solution to systems of equations similar to those defined by (36) and (37) when $B_0 = B_{0,0}$. However, the case where $B_0 \neq B_{0,0}$ has still to be solved. If one is not willing to accept existence of a solution in that case, she can modify the norm used in (31). In fact, recall that[13] (see formula (12))

$$\Sigma_{\mathcal{S}} = \left(M_\rho'\mathcal{S}M_\rho\right)^{-1} M_\rho'\mathcal{S}M_\theta\mathcal{V}_0M_\theta'\mathcal{S}M_\rho \left(M_\rho'\mathcal{S}M_\rho\right)^{-1} .$$

Now, if we replace the correct asymptotic covariance matrix $\mathcal{V}_0 = M_\theta^{-1}B_0M_\theta^{-1}$ by

$$\mathcal{V}_{0,0} := M_\theta^{-1}B_{0,0}M_\theta^{-1}, \tag{50}$$

by the same arguments used to derive (33) we obtain similarly to (33) the expression

$$\parallel IF(x^L;\widetilde{T},P_0^L) \parallel_{\widetilde{\Sigma}_{\mathcal{S}}^{-1}} \leq \parallel IF(x^L;T,P_0^L) \parallel_{(\mathcal{V}_{0,0})^{-1}} = \parallel \psi(x^L;\theta_0) \parallel_{(B_{0,0})^{-1}} \ , \tag{51}$$

where

$$\widetilde{\Sigma}_{\mathcal{S}} := \left(M_\rho'\mathcal{S}M_\rho\right)^{-1} M_\rho'\mathcal{S}M_\theta\mathcal{V}_{0,0}M_\theta'\mathcal{S}M_\rho \left(M_\rho'\mathcal{S}M_\rho\right)^{-1} . \tag{52}$$

The advantage of such replacement is twofold. Firstly, the same lower bounds $\sqrt{l_\theta}$ and $\sqrt{l_\rho}$ apply to c independently of B_0. Secondly, the estimating equations (37) and (39) become identical to the case when $\left\{\psi(X_{\tau+1}^{L+\tau};\theta_0)\right\}_{\tau=0}^{\infty}$ is a martingale difference sequence, i.e. $B_0 = B_{0,0}$. If the auxiliary score ψ is a good statistical approximation of the DGP then for moderate truncations implied by a constant c one would expect $B_0 \approx B_{0,0}$. Consequently $\Sigma_{\mathcal{S}}$ and $\widetilde{\Sigma}_{\mathcal{S}}$ should be also very similar in this case.

[13] We prefer formula (12) to its simplified version (cf. (14))

$$\Sigma_{\mathcal{S}} = \left(M_\rho'\mathcal{S}M_\rho\right)^{-1} M_\rho'\mathcal{S}B^0\mathcal{S}M_\rho \left(M_\rho'\mathcal{S}M_\rho\right)^{-1}$$

because (12) contains the asymptotic variance-covariance matrix $\mathcal{V}^0$ of $\widehat{\theta}$.

Similarly to robust ML estimation, a low level of c reduces the effect that a local misspecification may have on the bias but it also decreases the estimator's efficiency at the model because some information is "discarded". Furthermore, the resulting efficiency loss depends on the metric used to truncate the score function ψ and hence on which version of a REMM estimator is used for estimation. Although the versions are identical with respect to the nature of the constant c, there is an important point which clearly favourises the second one in some model settings. In fact, suppose that a low level of c is required and that the number of auxiliary parameters l_θ is significantly larger than l_ρ. Because for the first version of a REMM estimator $\sqrt{l_\theta}$ is a lower bound for c, we have to relinquish in this case the possibility of dealing with robust estimates for large parameterized auxiliary models. For instance, this is typically the case with a SNP auxiliary model where with a growing number of observations the degree of the SNP-polynomial and hence the number of auxiliary parameters may also be increased to improve the accuracy of the density approximation. Similar considerations apply to the auxiliary models of a RGMM estimation if the number of orthogonality conditions is significantly larger than l_θ. In that case the lower bound for c is given by $\sqrt{\dim(h)}$; see again Example 11 above. By contrast with the previous considerations, the dimension l_ρ of the structural model is fix, so that the second version of a REMM estimator does not suffer from this dimensionality problem[14] .

We mentioned above that the REMM methodology is well suited to estimate also non Markovian structural models. In order to achieve high efficiency it is then important to consider a sufficient number L of coordinates in the definition of the auxiliary model, see Gallant and Long (1997). However, an increase in L generally implies that more auxiliary parameters have to be estimated or that more ortogonality conditions are used in estimation. As we have just described this can imply an important robustness problem for the first version of our REMM estimator. A second point related to the choice of the lag length L concerns both versions of our REMM

[14] A similar problem arises in the context of high break down variance-covariance matrix estimators where the upper bound for the break down point is $1/\sqrt{H}$, with H the number of rows of the covariance matrix.

estimators. Indeed, we have seen that for the proposed model of contamination in Example 12 the number of coordinates involved in the definition of the functional estimator $\widehat{\rho}$ also enters in the derivative of the contaminated measure. This suggests that for the same amount ε of contamination, EMM estimators with estimating equations defined on a large number of coordinates of the process $\mathbb{X}$ may be more unstable than those defined on a smaller set.

A last consideration regards the choice of the weighting matrix $\mathcal{S}$. Because of a functional or dynamic misspecification of the auxiliary model, an optimal choice usually involves a functional form of $\mathcal{S}$ which depends on an infinite number of coordinates. However, as noted such a functional matrix $\mathcal{S}$ may not be differentiable. The first version of a REMM estimator is independent of $\mathcal{S}$. Hence, only the second version of robust estimator is effectively affected by the choice of $\mathcal{S}$. This is at least formally a new trade-of between efficiency and robustness not appearing in standard ML estimation.

5 Monte Carlo Simulations

The purposes of this section are the following. First, we show that by means of REMM even highly non linear and non Markovian models for time series can be robustly estimated in a very efficient way. Second, we demonstrate with a concrete example that when dealing with complex model dynamics it is very difficult (if not virtually impossible) to identify outliers by means of simple outliers detection procedures or by a visual inspection. On the other hand, we show that also for complex time series models REMM yields a very efficient and model consistent outliers identification using the estimated robust weights. Third, we are interested in quantifying the trade-off between efficiency and robustness in REMM estimation. Therefore, we compare by Monte Carlo simulation the performance of EMM and REMM in the presence and in the absence of model contamination. Specifically, we compare the classical EMM estimator and the first and second version of a robust EMM estimator defined above, denoted by REMM1 and REMM2, respectively.

We extend the simple MA(1) model analyzed in Martin and Yohai (1986) to an ARMA(1,1)-ARCH(2) process given by

$$\begin{aligned} Y_t &= 0.9Y_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}, \quad \epsilon_t = \sqrt{h_t}\nu_t, \quad \nu_t \text{ iid } N(0,1) \qquad (53) \\ h_t &= 0.04 + 0.6\epsilon_{t-1}^2 + 0.3\epsilon_{t-2}^2 \,. \end{aligned}$$

Notice that standard (non robust) ML estimation of this process is straightforward. However, a robust ML estimation is not as simple. Indeed, to achieve this goal one would first have to correct the bias induced by a truncation of the ML score in a robust ML estimation. This is already a difficult task in time series models where the process unconditional distribution is not given in closed form and would require numerical procedures, as for instance in many RGMM models for time series. Furthermore, the ML score of model (53) is a function of all coordinates of the process, implying an unbounded estimator's IF even after truncation with a kind of Huber's weights (see Martin and Yohai (1986) for a discussion of this point). Therefore, we are forced to estimate the structural model (53) by means of an auxiliary model whose estimating functions depend only on a finite number of process coordinates. In this case, a natural choice is an AR(3)-ARCH(2) auxiliary model. Since we include a constant term in the estimation of the structural and the auxiliary model the total number of auxiliary parameters l_θ is equal to 7 while l_ρ is equal to 6. For both robust estimators REMM1 and REMM2 the robustness tuning constant c is set equal to 10.

For illustration purposes a typical (uncontaminated) path from model (53) is presented in Figure 1. Note there the infrequent large positive and negative movements which typically occur in periods of higher volatility. At first sight they may easily be identified as outliers in a naive model misspecification analysis. Indeed, Figure 2 shows the same path contaminated with the replacement outlier model defined in Example 12, where we fixed $\xi = 3$ and $\epsilon = 1\%$. For this particular path only 4 observations have been replaced. The two figures show that even when knowing the form of the outliers generating process, their exact identification in Figure 2 could be a very difficult task.

Table 1 summarizes the results obtained from 1000 uncontaminated simulations of model (53) by presenting corresponding summary statistics and mean squared errors. Figure 4 presents the densities of the resulting structural parameter estimates $\widehat{\rho}_i$. For each $\widehat{\rho}_i$ in Table 1 the first row contains the summary statistics for the parameter estimates under a classical EMM estimation. The summary statistics for the REMM1 and REMM2 estimators are given, respectively, in the second and the third row for each $\widehat{\rho}_i$ in Table 1. The sample size behind all simulations is 800.

From Table 1 it is evident that the efficiency loss at the model when using REMM is virtually negligible. Indeed, the mean squared errors of all parameters estimates are very similar across the different EMM and REMM estimators. As one could intuitively expect REMM2 does not provide a clearly higher efficiency than REMM1 in this example, since the parameter dimension of the auxiliary and the structural models are almost equal. These impressions are visually confirmed by the estimated densities given in Figure 4.

Model estimations have been then repeated based on the paths contaminated by outliers according to the contaminating model in Example 12. The effects of the contamination on the resulting parameter estimates are summarized in Table 2 with corresponding summary statistics and in Figure 5 with the implied estimated densities. Table 2 is organized exactly as the previous Table 1. The results for the classical EMM estimator in Table 2 highlight some large biases and mean squared errors of a standard EMM model estimation. Indeed, we observe that the mean squared error of all EMM parameter estimates in Table 2 is highly inflated by the presence of contamination (compare for instance the mean squared errors for the estimates of ρ_1 and ρ_4 given in Table 1 and 2) and that some large biases are obtained especially in the parameter estimates of the conditional variance equation. The REMM procedures, on the other side, are very successful in controlling both for bias and efficiency in the presence of contamination. Indeed, when compared with EMM, all mean squared errors in Table 2 are much smaller and the estimates in the conditional variance equation present a quite reasonable bias. Again, both REMM1 and REMM2 perform in this second experiment similarly. These findings are even more apparent from the estimated densities in Figure 5.

Finally, it is important to stress the ability of REMM in identifying possible model deviations present in the data. In all our simulations we have observed REMM to identify correctly outliers when they were generated by the contaminating model used. To illustrate this point Figure 3 presents again the contaminated path of Figure 2 (in the top panel) with the estimated REMM weights (in the bottom panel). Weights clearly below one indicate an influential observation and can be used to identify outlying observations or more general model misspecifications. In this series, outliers were generated at the observations 180, 548, 660 and 797. As shown in the bottom panel of Figure 3 all these influential observations were clearly indicated with the lowest REMM weights. On the other hand, even some model generated very erratic movements, as for instance the ones at the observation 294, 480 and 614, were correctly identified as observations generated by the underlying structural model.

6 Conclusions

We characterized the local robustness properties of EMM estimators for time series and proposed two versions of a REMM estimator with bounded IF. We then presented some algorithms by which REMM estimators can be implemented, essentially with only a minor further computational effort when compared with the standard EMM. We finally verified in a Monte Carlo simulation study that REMM estimators are successful in controlling for the asymptotic bias under model misspecification while maintaining a high efficiency under the ideal structural model. REMM extends the application field of robust statistics to very general times series models and permits in a natural way a robust estimation of non Markovian processes. Further research on REMM requires applications to models with complex nonlinear dynamics, possible latent factors and switching regimes.

References

Andersen T.G., Chung H.-J. and B.E. Sørensen, 1999. *Efficient Method of Moments estimation of a stochastic volatility model: A Monte Carlo study*, Journal of Econometrics, 91, 61-87.

Andrews, D.W.K, 1991. *Heteroskedasticity and autocorrelation consistent covariance matrix estimation*, Econometrica, 59, 817-858.

de Luna, X and Genton, M. G. 2001: *Robust simulation-based estimation of ARMA models*, Journal of Computational and Graphical Statistics, Vol.10, No. 2, 370-387.

Gallant, A.R., 1987. *Nonlinear statistical models.* (Wiley, New York).

Gallant, A.R. and Nychka, D.W., 1987. *Semi-Nonparametric Maximum Likelihood estimation*, Econometrica, 55, 363-390.

Gallant, A.R. and Long, J.R., 1997. *Estimating stochastic differential equations efficiently by minimum chi-squared*, Biometrica, 84, 125-141.

Gallant, A.R. and G. Tauchen, 1996. *Which moments to match?*, Econometric Theory, 12, 657-681.

Gallant, A.R. and G. Tauchen, 1998. *Reprojecting partially observed systems with application to interest rate diffusions*, Journal of the American Statistical Association, 93, 10-24.

Gallant, A.R. and G. Tauchen, 1999. *The relative efficiency of method of moments estimators*, Journal of Econometrics, 92, 149-172.

Genton M. G. and E. Ronchetti 2003. *Robust Indirect Inference* , to appear in the Journal of the American Statistical Association.

Gouriéroux, C. and A. Monfort, 1997. *Simulation-Based Econometric Methods.* (Oxford Univerity Press, New York).

Gouriéroux, C., A. Monfort and E. Renault, 1993. *Indirect Inference*, Journal of Applied Econometrics, 8, 85-118.

Hampel, F.R., 1968. *Contribution to the theory of robust estimation.* Ph.D. Thesis, University of California, Berkeley.

Hampel, F.R., 1974. *The influence curve and its role in robust estimation*, Journal of the American Statistical Association, 68, 383-393.

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, 1986. *Robust Statistics: The Approach Based on Influence functions.* (Wiley, New York).

Huber, P.J., 1964. *Robust estimation of a location parameter*, Ann. Math. Statist., 35, 73-101.

Huber, P.J., 1981. *Robust Statistics.* (Wiley, New York).

Koenker, R.W. 1982. *Robust methods in econometrics.* Econometric Review 1, 213-255.

Koenker, R.W., Bassett, G., 1978. *Regression Quantiles.* Econometrica 46, 33-50.

Krishnakumar, J., Ronchetti E., 1997. *Robust-estimators for simultaneous equations models.* Journal of Econometrics 78, 295-314.

Künsch, H., 1984. *Infinitesimal robustness for autoregressive processes*, The Annals of Statistics, 12, 843-863.

Markatou, M., Ronchetti, E., 1997. *Robust inference: the approach based on influence functions.* In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, Vol. 15. North-Holland. Amsterdam, pp. 49-75.

Maronna, R. A., 1976. *Robust M-estimators of multivariate location and scatter*, The Annals of Statistics, 4, 51-67.

Martin, R. D. and V. J. Yohai, 1986. *Influence functionals for time series*, The Annals of Statistics, 14, 836-837.

Newey, W.K. and K.D. West, 1987. *A simple, positive simi-definite, heteroskedasticity and autocorrelation consistent covariance matrix*, Econometrica, 55, 703-708.

Peracchi, F. 1990. *Robust M-estimators.* Econometric Review 9, 1-30.

Peracchi, F. 1991. *Robust M-tests.* Econometric Theory 7, 69-84.

Ronchetti, E. and F. Trojani, 2001. *Robust inference with GMM estimators*, Journal of Econometrics, 101, 37-69.

White, H., 1994. *Estimation, Inference and Specification Analysis.* (Cambridge University Press, New York).

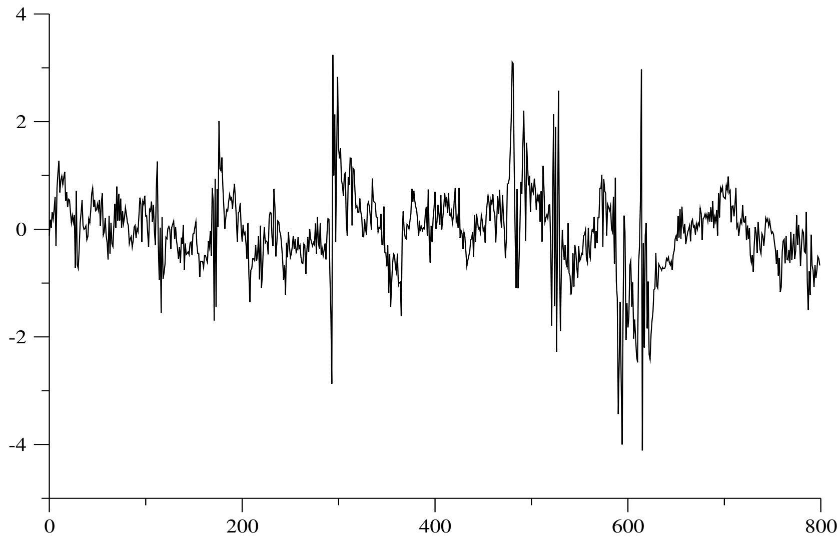

Figure 1: Uncontaminated simulation of model (53).

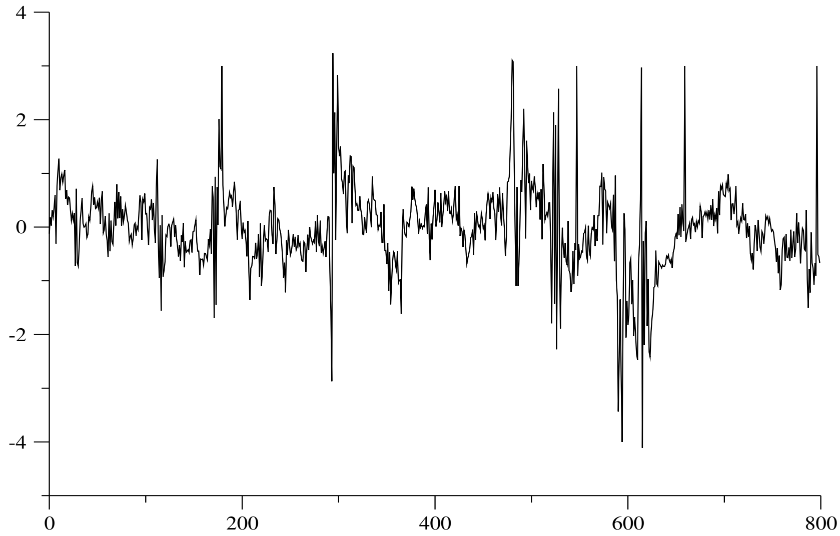

Figure 2: Contaminated simulation of model (53) using the replacement model in Example 12.

True	Mean	Median	q_{25}	q_{75}	Stdv	$q_{75} - q_{25}$	MSE
0	0.0005	0.0005	-0.0025	0.0036	0.0048	0.0061	2.33E-5
	0.0008	0.0010	-0.0022	0.0039	0.0049	0.0061	2.43E-5
	0.0008	0.0009	-0.0023	0.0037	0.0048	0.0060	2.38E-5
0.9	0.8963	0.8978	0.8842	0.9098	0.0199	0.0255	4.10E-4
	0.8954	0.8974	0.8837	0.9083	0.0196	0.0246	4.06E-4
	0.8964	0.8985	0.8847	0.9088	0.0197	0.0241	4.02E-4
-0.5	-0.5036	-0.5033	-0.5408	-0.4653	0.0563	0.0755	3.19E-3
	-0.5014	-0.5000	-0.5360	-0.4644	0.0556	0.0716	3.09E-3
	-0.5054	-0.5034	-0.5401	-0.4697	0.0556	0.0704	3.11E-3
0.04	0.0397	0.0393	0.0363	0.0429	0.0048	0.0066	2.33E-5
	0.0400	0.0396	0.0365	0.0431	0.0048	0.0066	2.30E-5
	0.0403	0.0399	0.0369	0.0434	0.0048	0.0065	2.33E-5
0.6	0.5846	0.5844	0.5339	0.6389	0.0745	0.1050	5.78E-3
	0.5900	0.5923	0.5379	0.6436	0.0761	0.1057	5.90E-3
	0.5744	0.5746	0.5256	0.6257	0.0730	0.1001	5.98E-3
0.3	0.2922	0.2923	0.2513	0.3332	0.0607	0.0818	3.75E-3
	0.2959	0.2950	0.2539	0.3375	0.0617	0.0836	3.82E-3
	0.2986	0.2992	0.2575	0.3414	0.0606	0.0839	3.67E-3

Table 1: Summary statistics for EMM, REMM1 and REMM2 under the uncontaminated model (53). For each parameter ρ_i that has to be estimated the first row in the table gives summary statistics for the EMM, the second one statistics for the REMM1 and the third one statistics for the REMM2.

True	Mean	Median	q_{25}	q_{75}	Stdv	$q_{75} - q_{25}$	MSE
0	0.0100	0.0091	0.0018	0.0172	0.0127	0.0154	2.59E-4
	0.0009	0.0010	-0.0025	0.0045	0.0056	0.0071	3.17E-5
	0.0010	0.0011	-0.0024	0.0046	0.0056	0.0070	3.28E-5
0.9	0.8793	0.8852	0.8585	0.9096	0.0464	0.0511	2.58E-3
	0.8861	0.8889	0.8730	0.9023	0.0225	0.0293	7.02E-4
	0.8865	0.8895	0.8732	0.9031	0.0235	0.0299	7.33E-4
-0.5	-0.4913	-0.4877	-0.5484	-0.4266	0.0979	0.1218	9.65E-3
	-0.4854	-0.4857	-0.5219	-0.4488	0.0541	0.0731	3.14E-3
	-0.4855	-0.4856	-0.5216	-0.4495	0.0570	0.0721	3.46E-3
0.04	0.1095	0.1023	0.0703	0.1422	0.0474	0.0719	7.08E-3
	0.0482	0.0472	0.0424	0.0531	0.0088	0.0107	1.45E-4
	0.0488	0.0480	0.0430	0.0537	0.0082	0.0106	1.44E-4
0.6	0.5499	0.5473	0.4216	0.6672	0.1722	0.2456	3.21E-2
	0.5996	0.5995	0.5375	0.6671	0.0925	0.1296	8.55E-3
	0.5795	0.5825	0.5180	0.6490	0.0973	0.1310	9.88E-3
0.3	0.2415	0.2187	0.1277	0.3254	0.1499	0.1977	2.59E-2
	0.2554	0.2567	0.2003	0.3053	0.0748	0.1050	7.59E-3
	0.2651	0.2648	0.2078	0.3174	0.0797	0.1096	7.57E-3

Table 2: Summary statistics for EMM, REMM1 and REMM2 under model (53) contaminated by the replacement model in Example 12. For each parameter ρ_i that has to be estimated the first row in the table gives summary statistics for the EMM, the second one statistics for the REMM1 and the third one statistics for the REMM2.

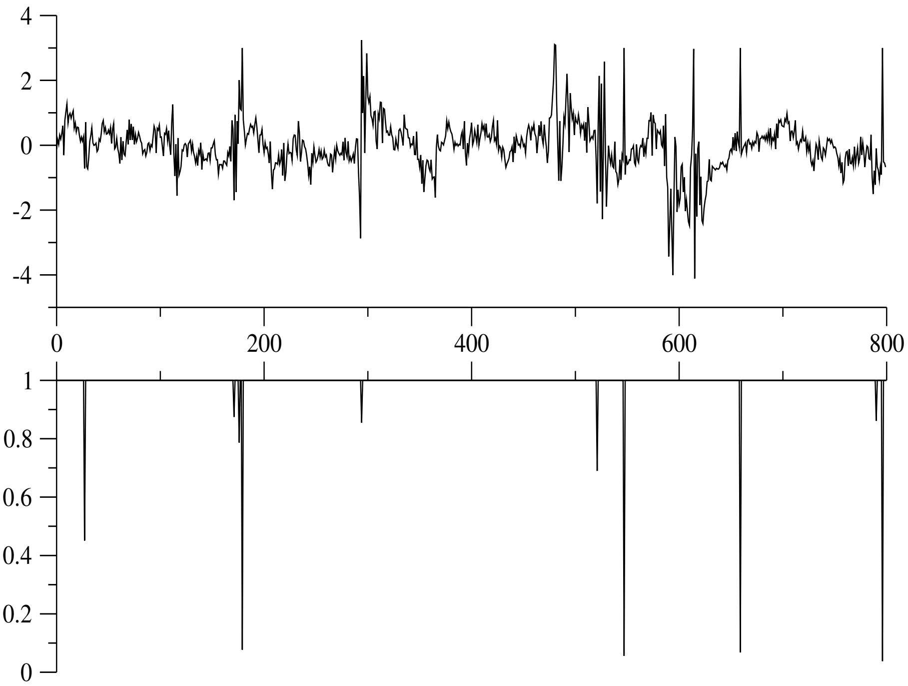

Figure 3: Contaminated series (top panel) and weights implied by the REMM (bottom panel).

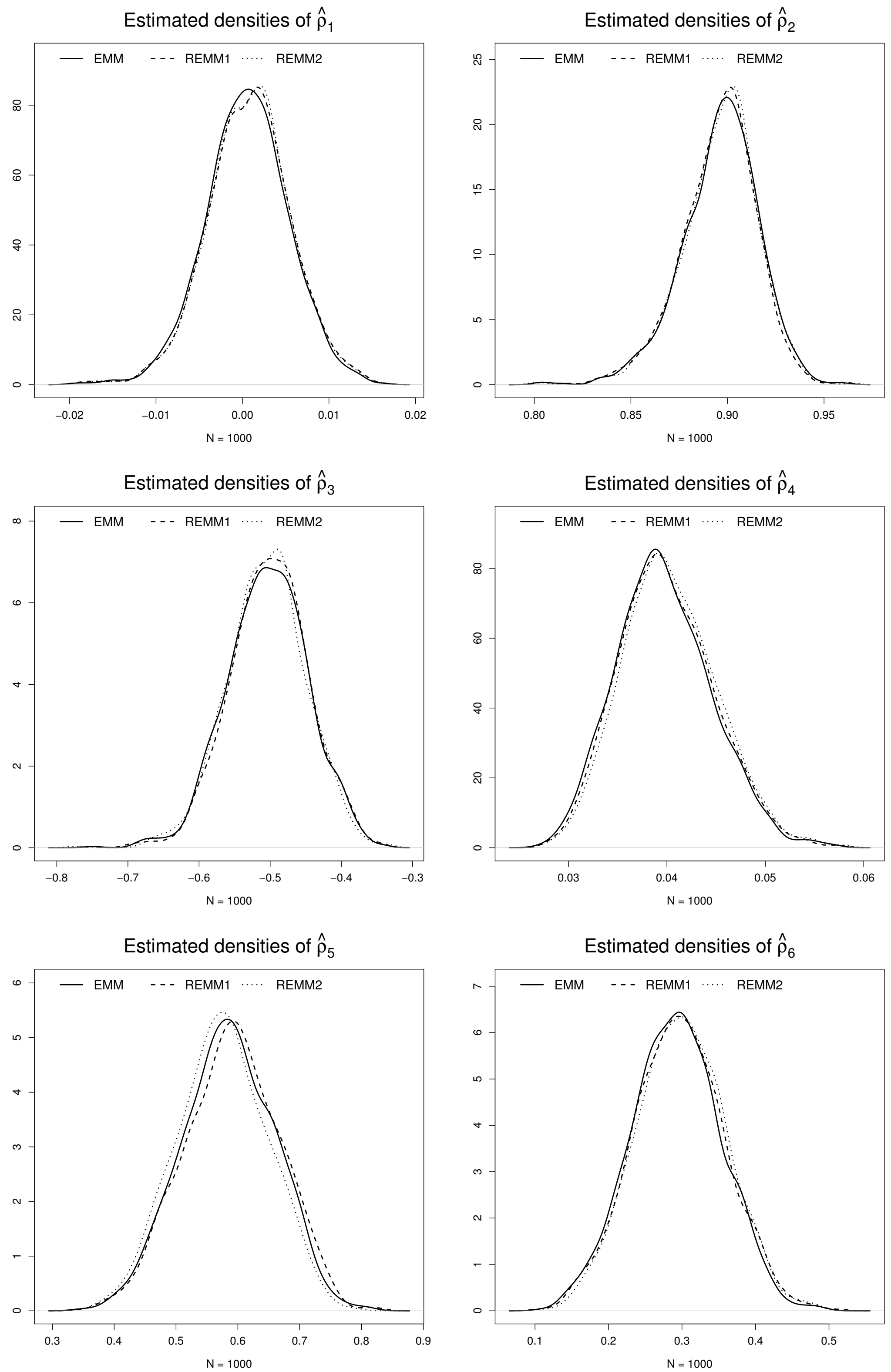

Figure 4: Estimated densities of $\widehat{\rho}_i$ under non contaminated simulations of (53).

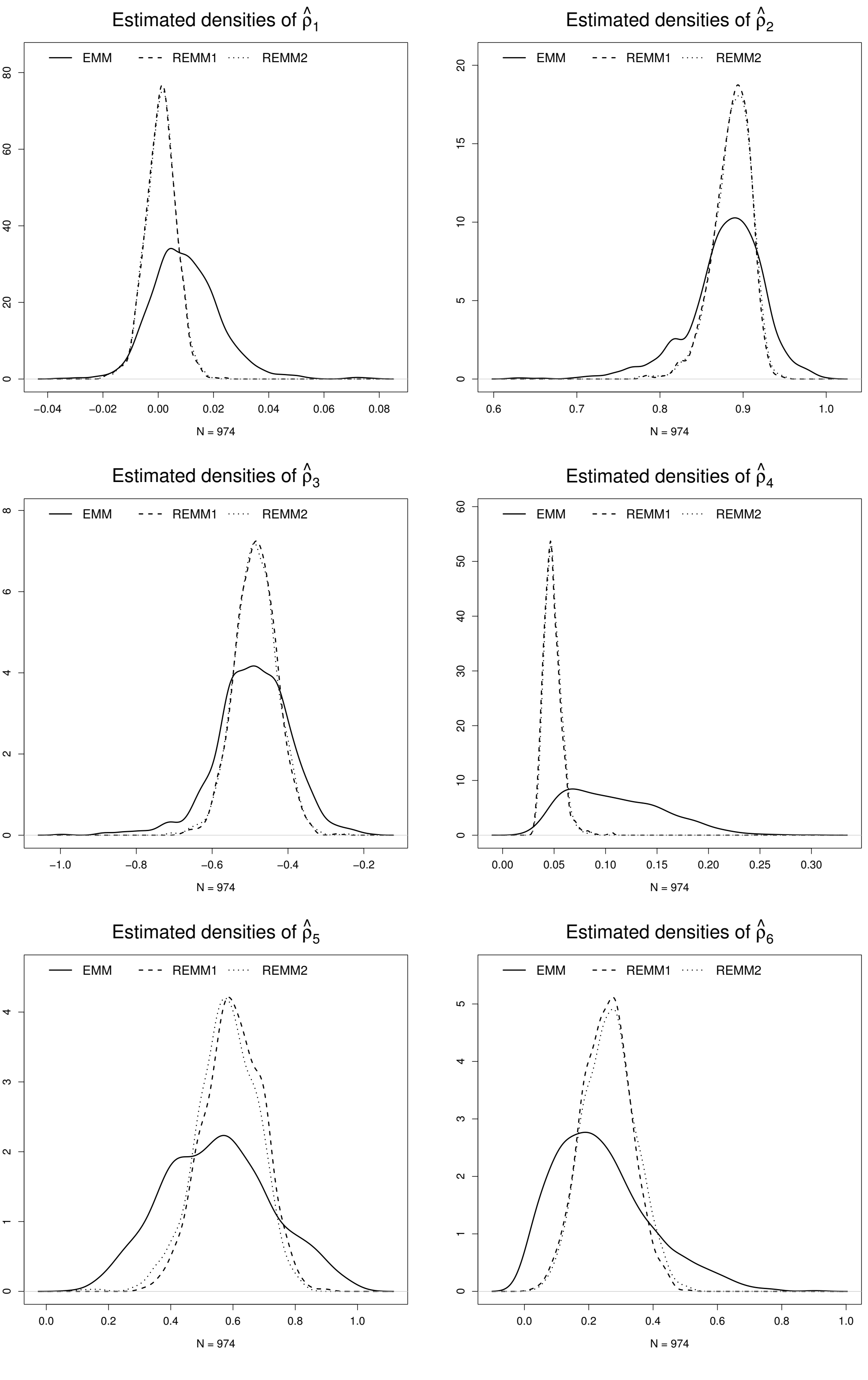

Figure 5: Estimated densities of $\widehat{\rho}_i$ under contaminated simulations of (53).