

# Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models\*

Loriano Mancini<sup>a</sup>, Elvezio Ronchetti<sup>a,b</sup> and Fabio Trojani<sup>a,c</sup>

<sup>a</sup>Institute of Finance, University of Lugano, Switzerland

<sup>b</sup>Dept. of Econometrics, University of Geneva, Switzerland

<sup>c</sup>Dept. of Economics, University of St. Gallen, Switzerland

First Version: January 2003

Revised: June 2004

**Keywords:** Time series models,  $M$ -estimators, influence function, robust estimation and testing.

---

\***Correspondence Information:** Loriano Mancini, Institute of Finance, University of Lugano, Via Buffi 13, CH-6900 Lugano, tel: +41 (0)91 912 47 86. [Loriano.Mancini@lu.unisi.ch](mailto:Loriano.Mancini@lu.unisi.ch). E-mail addresses for Elvezio Ronchetti and Fabio Trojani: [Elvezio.Ronchetti@metri.unige.ch](mailto:Elvezio.Ronchetti@metri.unige.ch), [Fabio.Trojani@lu.unisi.ch](mailto:Fabio.Trojani@lu.unisi.ch). Loriano Mancini and Fabio Trojani gratefully acknowledge the financial support of the Swiss National Science Foundation (grant 12-65196.01 and NCCR FINRISK). The authors thank Giovanni Barone-Adesi, Franco Peracchi, Claudio Ortelli, Claudia Ravanelli, the referees, the Associate Editor, and the Editor for their valuable suggestions.

## Abstract

This paper studies the local robustness of estimators and tests for the conditional location and scale parameters in a strictly stationary time series model. We first derive optimal bounded-influence estimators for such settings under a conditionally Gaussian reference model. Based on these results, optimal bounded-influence versions of the classical likelihood-based tests for parametric hypotheses are obtained. We propose a feasible and efficient algorithm for the computation of our robust estimators, which makes use of analytical Laplace approximations to estimate the auxiliary recentering vectors ensuring Fisher consistency in robust estimation. This strongly reduces the necessary computation time by avoiding the simulation of multidimensional integrals, a task that has typically to be addressed in the robust estimation of nonlinear models for time series. In some Monte Carlo simulations of an AR(1)-ARCH(1) process we show that our robust procedures maintain a very high efficiency under ideal model conditions and at the same time perform very satisfactorily under several forms of departure from conditional normality. On the contrary, classical Pseudo Maximum Likelihood inference procedures are found to be highly inefficient under such local model misspecifications. These patterns are confirmed by an application to robust testing for ARCH.

# 1 Introduction

This paper studies the local robustness properties of estimation and testing procedures for the conditional location and scale parameters of a strictly stationary time series model.

The class of conditional location and scale time series models is quite broad and includes several well-known dynamic models largely applied in economics and empirical finance, such as pure conditional scale models (like ARCH models; Engle 1982) or models that jointly parameterize the conditional location and the scale of the given time series (like for instance ARCH in mean models; Engle, Lilien and Robins 1987). Typically, classical (non robust) estimation of the parameters of such models is obtained by means of a Pseudo Maximum Likelihood (PML) approach based on the *nominal* assumption of a conditionally Gaussian log-likelihood; see also Bollerslev and Wooldridge (1992). Such PML estimators are based on an unbounded conditional score function, implying—as shown below—an unbounded time series influence function (IF, Künsch 1984; Hampel 1974). As a consequence, PML estimators for conditional location and scale models are not robust under local departures from conditional normality. In this paper we propose a new class of inference procedures which are robust to local nonparametric misspecifications of a parametric, conditionally Gaussian, location and scale model. More specifically, we consider the class of robust, conditionally unbiased,  $M$ -estimators for the parameters of conditional location and scale models and derive the optimal (i.e. the most efficient) robust estimator within this class. Based on such estimators, several Maximum Likelihood (ML)-type bounded-influence tests for parametric hypotheses on the parameters of the conditional location and scale equations are then obtained following the general approach in Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001).

The need for robust procedures in estimation and testing has been stressed by many authors and is now widely recognized both in the statistical and the econometric literature; cf. for instance Hampel (1974), Koenker and Bassett (1978), Huber (1981), Koenker (1982), Hampel et al. (1986), Peracchi (1990), and more recently Markatou and Ronchetti (1997), Krishnakumar and Ronchetti (1997), Ronchetti and Trojani (2001), Ortelli and Trojani (2002), Gagliardini, Trojani and Urga (2004). However, the problem of the robust estimation for the parameters of conditional location and scale models has been considered

so far by a few authors and only from the specific perspective of high breakdown estimation. Even less attention has been devoted to robust inference within conditional location and scale models. High breakdown estimators resistant to large amount of contamination have been proposed by Sakata and White (1998) and Muler and Yohai (1999). These estimators are very useful at the exploratory and estimation stage. Here we focus at the inference stage, where we typically have an approximate model and we can expect small deviations from the model. Alternatives to high breakdown estimators are also needed because these estimators are computationally intensive and cannot be applied to estimate the parameters of a class of broadly applied models—such as for instance threshold ARCH or ARCH in mean models.

This paper derives optimal bounded-influence estimation and testing procedures for a general conditional location and scale model, which are computationally only slightly more demanding than those required by a classical PML estimation of such models. The more specific contributions to the current literature are the following.

First, we characterize the robustness of conditionally unbiased  $M$ -estimators for nonlinear conditional location and scale models by computing the time series IF for the implied asymptotic functional estimator. This has been defined by Künsch (1984) who applied this concept to  $AR(p)$  processes. We extend this result to our general model (1); see below. This is a first necessary step which allows us to construct robust statistical procedures which can control for (i) the local asymptotic bias on the parameter estimates and (ii) the local asymptotic distortion on the level and the power of ML-type tests.

Second, we derive the optimal bounded-influence estimator for the parameters of conditional location and scale models under a conditionally Gaussian reference model. This extends the optimality result in Künsch (1984) (obtained for  $AR(p)$  models) and the application of optimal conditionally unbiased  $M$ -estimators in Künsch, Stefanski and Carroll (1989) (obtained for generalized linear models) to general nonlinear second order dynamic models. Based on these results, optimal bounded-influence versions of the classical Wald, score and likelihood ratio tests are derived along the general lines proposed in Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001).

Third, we propose a feasible algorithm for the computation of our optimal robust estimators, which

can be easily implemented in standard packages, such as Matlab. This procedure is based on a truncating procedure which uses a set of Huber’s weights to downweight the impact of influential observations. Fisher consistency at the model is preserved by means of some auxiliary recentering vectors, which in a time series setting have generally to be computed by simulations—as for instance in a Robust Generalized Method of Moments (RGMM, Ronchetti and Trojani 2001) setting. Using the conditional unbiasedness of our estimator we provide analytical Laplace approximations for such vectors which strongly reduce the necessary computation time by avoiding the simulation of multidimensional integrals.

Fourth, we study by Monte Carlo simulation the efficiency and the robustness properties of our estimator. We estimate a simple AR(1)-ARCH(1) process under several models of local contamination of a conditionally Gaussian process. Under the Gaussian reference model the classical ML estimator and our robust estimator have essentially the same efficiency. On the contrary, under local deviations from conditional normality classical PML estimators, tests and confidence intervals are found to be strongly biased and highly inefficient, while robust procedures perform very satisfactorily.

Finally, we present an application to robust testing for ARCH where robust procedures help to identify ARCH structures which could not be detected using the classical inference approach.

The structure of the paper is as follows. Section 2 introduces conditional location-scale models and the corresponding classical  $M$ -estimation procedure. Section 3 computes the time series IF for conditionally unbiased  $M$ -estimators. The asymptotic bias on the parameter estimates induced by local deviations from the conditional Gaussian reference model is then approximated. In a second step, the optimal robust estimator is derived and the optimality of robust inference procedures based on such estimators is discussed. The section is concluded by deriving analytic approximations for the auxiliary recentering vectors in our robust estimation and by presenting an algorithm to compute our robust estimator in applications. Section 4 discusses robust inference procedures based on our robust estimators. Section 5 presents the Monte Carlo experiments where the performance of our robust estimation and inference approach is evaluated in the setting of an AR(1)-ARCH(1) model. Some advices for applications are also provided. The empirical application to testing for ARCH is presented in Section 6. Section 7 summarizes and concludes.

## 2 Conditionally Unbiased $M$ -estimators

Let  $\mathcal{Y} := (y_t)_{t \in \mathbb{Z}}$  be a real valued strictly stationary random sequence on the probability space  $(\mathbb{R}^\infty, \mathcal{F}, \mathbb{P}_*)$  and  $\mathcal{P} := \{\mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$  be some parametric model for  $\mathbb{P}_*$ . Under model  $\mathbb{P}_{\theta_0}$ , the random variable  $y_t$  has a conditionally Gaussian distribution,  $y_t | \mathcal{F}_{t-1} \sim \mathcal{N}(\mu_t(\theta_0), \sigma_t^2(\theta_0))$ . Specifically,

$$\begin{aligned} y_t &= \mu_t(\theta_0) + \varepsilon_t(\theta_0), \\ \varepsilon_t^2(\theta_0) &= \sigma_t^2(\theta_0) + \nu_t(\theta_0), \end{aligned} \tag{1}$$

where  $\mu_t(\theta_0)$  and  $\sigma_t^2(\theta_0)$  parameterize the conditional mean and the conditional variance of  $y_t$  given the information  $\mathcal{F}_{t-1}$  up to time  $t-1$ . We denote by  $y_1^m := (y_1, \dots, y_m)$  the finite random sequence of  $\mathcal{Y}$  and by  $\mathbb{P}_{\theta_0}^m$  ( $\mathbb{P}_*^m$ ) the  $m$ -dimensional marginal distribution of  $y_1^m$  induced by  $\mathbb{P}_{\theta_0}$  ( $\mathbb{P}_*$ ). Model (1) covers a broad class of well-known parametric models for time series. A general example is the following.

**Example 2.1** Double threshold AR(1)-ARCH(1) models with volatility asymmetries (see, for instance, Li and Li 1996; Glosten, Jagannathan and Runkle 1993) assume the specification

$$\begin{aligned} \mu_t(\theta_0) &= \rho_0 + (\rho_1 + \rho_2 d_{1,t-1}) y_{t-1}, \\ \sigma_t^2(\theta_0) &= \alpha_0 + (\alpha_1 + \alpha_2 d_{2,t-1})(y_{t-1} - \rho_0 - (\rho_1 + \rho_2 d_{1,t-2}) y_{t-2})^2 + \alpha_3 d_{1,t-1} \end{aligned} \tag{2}$$

with the dummy variable  $d_{1,t-1} = 1$  if  $\rho_0 + \rho_1 y_{t-1} > 0$  and zero otherwise,  $d_{2,t-1} = 1$  if  $\varepsilon_{t-1}(\theta_0) < 0$  and zero otherwise.

Model (1) includes linear autoregressive models as straightforward special cases. Künsch (1984) defined a time series influence function (IF) in this context and derived an optimal bounded-influence estimator for the parameters of an AR( $p$ ) model. Martin and Yohai (1986) provided bounded-influence estimators for AR and MA models and studied the asymptotic bias implied by additive outliers. ARCH models (cf. Engle 1982) are also special cases of model (1). Bounded-influence estimators for such models are available in the class of robust GMM (RGMM) or robust EMM (REMM) estimators; cf. Ronchetti and Trojani (2001) and Ortelli and Trojani (2002). Muler and Yohai (1999) considered explicitly the pure (no location) ARCH setting from the perspective of high breakdown estimation. Sakata and White (1998) developed high breakdown estimators for conditional location and scale models that include ARCH models

as special cases. However, their high breakdown estimators cannot be applied directly to all models of the form (1) because they assume a partitioned parameter space  $\Theta = \Theta_1 \times \Theta_2$  in order to imply  $\mu_t(\theta_0) = \mu_t(\theta_1)$ , for  $\theta_1 \in \Theta_1$ . Moreover, all the above robust estimators are much more computationally intensive than the one proposed in this paper already for simple processes, like for instance ARCH models.

From the point of view of robust inference, model (1) is interpreted as an “approximate” description of the true data generating process  $\mathbb{P}_*$ . Hence, our aim is to derive efficient and computationally feasible locally robust procedures for inference on the parameters of model (1) when the data distribution  $\mathbb{P}_*$  is in some nonparametric neighborhood of the reference model  $\mathbb{P}_{\theta_0}$ . To this end we consider the class of conditionally unbiased  $M$ -estimators for  $\theta_0$  and a functional  $M$ -estimator  $a(\cdot)$ ,

$$a : \text{dom}(a) \subset \mathcal{M}_{stat}^m \longrightarrow \Theta,$$

where  $\mathcal{M}_{stat}^m := \{m\text{-dimensional marginals of strictly stationary processes}\}$  and  $m \in \mathbb{N} \setminus \{\infty\}$ . In particular, this excludes GARCH models (cf. Bollerslev 1986) from our robust analysis. Strictly speaking  $a(\cdot)$  depends on  $m$  but for short we will drop this dependence in the notation.

The estimator  $a(\cdot)$  is implicitly defined by an estimating function  $\psi : \mathbb{R}^m \times \Theta \longrightarrow \mathbb{R}^p$  and some conditional moment conditions

$$E_{\theta_0}[\psi(y_1^m; a(\mathbb{P}_{\theta_0}^m)) | \mathcal{F}_{m-1}] = 0, \tag{3}$$

which hold for a unique  $\theta_0 \in \Theta$ . By construction  $(\psi(y_{1+t}^{m+t}; \theta_0))_{t \in \mathbb{Z}}$  is a martingale difference sequence under  $\mathbb{P}_{\theta_0}$ . Thus,  $a(\cdot)$  is conditionally Fisher consistent and the asymptotic estimating equation for  $\theta_0$  is

$$E_{\theta_0}[\psi(y_1^m; a(\mathbb{P}_{\theta_0}^m))] = 0. \tag{4}$$

For example, the conditionally Gaussian score function  $s$

$$s(y_1^m; \theta_0) = -k_{1,m} + k_{2,m} u_m(\theta_0) + k_{1,m} u_m(\theta_0)^2, \tag{5}$$

where  $u_m(\theta_0) = \varepsilon_m(\theta_0) \sigma_m(\theta_0)^{-1}$  and

$$k_{1,m} := \frac{1}{2\sigma_m^2(\theta_0)} \left. \frac{\partial \sigma_m^2(\theta)}{\partial \theta} \right|_{\theta=\theta_0}, \quad k_{2,m} := \frac{1}{\sigma_m(\theta_0)} \left. \frac{\partial \mu_m(\theta)}{\partial \theta} \right|_{\theta=\theta_0},$$

defines a conditionally unbiased  $M$ -estimator of  $\theta_0$ . In this setting, different specifications of  $\mu_m(\theta_0)$  and  $\sigma_m^2(\theta_0)$  are easily accommodated in the  $\mathcal{F}_{m-1}$ -measurable random vectors  $k_{1,m}$  and  $k_{2,m}$ . Notice that (4) is one of the estimating equation based on (3). One can define other equations by using any variable in  $\mathcal{F}_{m-1}$  as instrument.

The estimator  $a(\cdot)$  depends on  $m$  process coordinates which are entirely determined by the parametric reference model (1). For instance, in Example 2.1 we had  $m = 3$ . The  $M$ -estimator of  $\theta_0$  solves the finite sample estimating equations

$$n^{-1} \sum_{t=1}^n \psi(\tilde{y}_{t-m+1}^t; \hat{\theta}_n) = 0, \quad (6)$$

which are the finite sample version of the asymptotic condition (4), where  $\tilde{y}_{2-m}^n$  are sample observations of the process  $\mathcal{Y}$ . Under model  $\mathbb{P}_{\theta_0}^m$ ,  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  converges in distribution to the Gaussian distribution  $\mathcal{N}(0, V(\psi; \theta_0))$ , where  $V(\psi; \theta) := J(\theta)^{-1} I(\theta) J(\theta)^{-1}$  with

$$J(\theta) := E_{\theta} \left[ -\frac{\partial \psi(y_1^m; \theta)}{\partial \theta^{\top}} \right], \quad I(\theta) := E_{\theta} [\psi(y_1^m; \theta) \psi(y_1^m; \theta)^{\top}];$$

cf. Bollerslev and Wooldridge (1992, p. 148). In particular, under the reference model  $\mathbb{P}_{\theta_0}^m$ , the simple expression for  $I(\theta)$  is implied by condition (3).

### 3 Locally Robust Estimation

In this section we allow  $\mathbb{P}_*$  to be in a nonparametric neighborhood of the reference model  $\mathbb{P}_{\theta_0}$ . In particular,  $\mathbb{P}_{\theta_0}$  can be dynamically misspecified for  $\mathbb{P}_*$ . As we focus on local robustness we consider local deviations from  $\mathbb{P}_{\theta_0}^m$ . Therefore, we assume that  $\mathbb{P}_*^m$  is in the following nonparametric neighborhood  $\mathcal{U}^n(\mathbb{P}_{\theta_0}^m)$  of  $\mathbb{P}_{\theta_0}^m$ ,

$$\mathcal{U}^n(\mathbb{P}_{\theta_0}^m) := \{\mathbb{P}_{\eta}^m = (1 - \eta)\mathbb{P}_{\theta_0}^m + \eta\mathbb{G}^m, \quad \eta \leq b, \quad b \in [0, 1], \quad \mathbb{G}^m \in \mathcal{M}_{stat}^m\}. \quad (7)$$

The neighborhood defined in (7) is a simple way to formalize local perturbations of the model  $\mathbb{P}_{\theta_0}^m$ . Notice that  $d_k(\mathbb{P}_{\eta}^m, \mathbb{P}_{\theta_0}^m) \leq \eta$  for all  $\mathbb{G}^m \in \mathcal{M}_{stat}^m$ , where  $d_k(\cdot, \cdot)$  denotes the Kolmogorov distance.

### 3.1 Time Series Influence Function

Robust procedures aim at the estimation of the parameter  $\theta_0$  when local deviations from the reference model  $\mathbb{P}_{\theta_0}^m$  are allowed. Such deviations induce an asymptotic bias on the functional estimator  $a(\cdot)$ , defined by

$$bias := a(\mathbb{P}_*^m) - a(\mathbb{P}_{\theta_0}^m) = a(\mathbb{P}_*^m) - \theta_0.$$

For a robust inference on  $\theta_0$ , the standard robustness condition on the estimator  $a(\cdot)$  is a bounded asymptotic bias. To describe the linearized asymptotic bias of  $a(\cdot)$  under some model  $\mathbb{P}_\eta^m \in \mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m)$  we consider the first order von Mises (1947) expansion of  $a(\cdot)$  at  $\mathbb{P}_{\theta_0}^m$  (cf., for instance, Fernholz 1983),

$$a(\mathbb{P}_\eta^m) - a(\mathbb{P}_{\theta_0}^m) = \eta a'(\theta_0, \mathbb{G}^m) + o(\|\mathbb{P}_\eta^m - \mathbb{P}_{\theta_0}^m\|), \quad (8)$$

where  $a'(\theta_0, \mathbb{G}^m)$  is the Gâteaux derivative of  $a(\cdot)$  in the direction  $\mathbb{G}^m - \mathbb{P}_{\theta_0}^m$ , defined by

$$a'(\theta_0, \mathbb{G}^m) := \lim_{\eta \downarrow 0} \frac{a((1-\eta)\mathbb{P}_{\theta_0}^m + \eta\mathbb{G}^m) - a(\mathbb{P}_{\theta_0}^m)}{\eta},$$

for all  $\mathbb{G}^m$  such that this limit exists. By contrast with a simple i.i.d. setting, in a time series framework the derivative  $a'(\theta_0, \mathbb{G}^m)$  is determined by a set of equivalent kernels. The next definition introduces a natural unique representant of such kernels.

**Definition 3.1** The conditional influence function of the functional estimator  $a(\cdot)$  is a kernel  $IF : \mathbb{R}^m \times \Theta \rightarrow \mathbb{R}^p$  such that

- i)  $a'(\theta, \mathbb{G}^m) = \int_{\mathbb{R}^m} IF(y; \theta) d\mathbb{G}^m(y)$ , for all  $\mathbb{G}^m \in \mathcal{M}_{stat}^m$ ,
- ii)  $E_{\theta_0}[IF(y_1^m; \theta_0) | \mathcal{F}_{m-1}] = 0$ ,

where  $E_{\theta_0}[\cdot]$  denotes expectations with respect to the reference distribution  $\mathbb{P}_{\theta_0}^m$ .

Künsch (1984) introduced the natural additional condition *ii*) which determines a unique representant of the IF (up to additive constants). This condition simply requires that, on average at the reference model  $\mathbb{P}_{\theta_0}^m$ ,  $y_m | y_1, \dots, y_{m-1}$  has no influence on the asymptotic bias of the estimator. When the dependence of the conditional IF on the corresponding score function  $\psi$  has to be emphasized we use in the sequel the notation  $IF_\psi$ .

If the conditional IF exists then one can prove also for our more general model setting that it is unique by using the same arguments as in the first part of the proof of theorem 1.3 in Künsch (1984). Existence of the conditional IF for the case of a bounded score function  $\psi$  (which is the relevant case for our robust estimator introduced in Section 3.2) can also be proved along the lines of theorem 1.3 in Künsch (1984).

The conditional IF has some desirable properties. First, under the reference model  $\mathbb{P}_{\theta_0}^m$ , the martingale difference property *ii*) implies the simple expression  $V(\psi; \theta_0) = E_{\theta_0}[IF(y_1^m; \theta_0) IF(y_1^m; \theta_0)^\top]$  for the asymptotic covariance matrix  $V(\psi; \theta_0)$  of  $a(\cdot)$ . Second, for conditionally unbiased  $M$ -estimators of the form (3) the conditional IF can be computed as in the one dimensional case by calculating the limit

$$IF(x_1^m; \theta_0) := \lim_{\eta \downarrow 0} \frac{a((1-\eta)\mathbb{P}_{\theta_0}^m + \eta\delta_{x_{(1)}, \dots, x_{(m)}}) - a(\mathbb{P}_{\theta_0}^m)}{\eta},$$

where  $\delta_{x_{(1)}, \dots, x_{(m)}}$  is the point mass at  $\{(y_1, \dots, y_m) = (x_{(1)}, \dots, x_{(m)})\}$ . This implies

$$IF(x_1^m; a(\mathbb{P}_{\theta_0}^m)) = D(\psi; \theta_0)^{-1} \psi(x_1^m; a(\mathbb{P}_{\theta_0}^m)), \quad (9)$$

where  $D(\psi; \theta_0) := -E_{\mathbb{P}_{\theta_0}^m}[(\partial/\partial\theta)\psi(y; \theta)^\top]_{\theta=\theta_0}$ . As the conditional IF is unique and defines a martingale difference process, (9) is the only admissible representation.

A bounded conditional IF ensures a bounded linearized asymptotic bias for any contaminated distribution  $\mathbb{P}_\eta^m$  in the neighborhood  $\mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m)$

$$bias := a(\mathbb{P}_\eta^m) - a(\mathbb{P}_{\theta_0}^m) = \eta \int_{\mathbb{R}^m} IF(y; \theta_0) \frac{\partial}{\partial\eta} \mathbb{P}_\eta^m(dy) \Big|_{\eta=0} + o(\|\mathbb{P}_\eta^m - \mathbb{P}_{\theta_0}^m\|),$$

because the derivative on the right hand side is uniformly bounded for any  $\mathbb{P}_\eta^m \in \mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m)$ . Moreover, since the conditional IF is linearly related to the  $\psi$ -function of the estimating equation (4), it is bounded if and only if the  $\psi$ -function is bounded. As the Gaussian score function (5) is unbounded (at least) in  $\varepsilon_m(\theta)$ , PMLE's based on such a score function are not robust.

### 3.2 Optimal Conditionally Unbiased Robust Estimators

We derive the most efficient estimator with bounded self-standardized sensitivity (see Proposition 3.1 and Corollary 3.1 below) in the class of conditionally unbiased  $M$ -estimators for  $\theta_0$ . The self-standardized

sensitivity  $\gamma$  of the estimator is

$$\gamma(\psi) := \sup_{z \in \mathbb{R}^m} \|V(\psi; \theta_0)^{-1/2} IF_\psi(z; \theta_0)\|;$$

cf., for instance, Krasker and Welsch (1982). By definition, non robust estimators have  $\gamma = \infty$ , while bounded influence estimators have  $\gamma \leq c < \infty$ , for some positive constant  $c \geq \sqrt{p}$ ; cf. Hampel et al. (1986, p. 228). Some more concrete advices for the way of choosing the tuning constant  $c$  in applications are given in Section 5.4.

### 3.2.1 Optimality Results

Under the reference model  $\mathbb{P}_{\theta_0}^m$ , the classical estimator of  $\theta_0$  defined by the score function (5) is the most efficient but not robust. Therefore, we propose a robust estimator of  $\theta_0$  that achieves for models of the form (1) an optimality result, which is the direct extension of the one in Künsch, Stefanski and Carroll (1989).

Let  $\psi_c(y_1^m; \theta) := A(\theta) \psi^{bif}(y_1^m; \theta)$ , where  $\psi^{bif}(y_1^m; \theta) := (s(y_1^m; \theta) - \tau(y_1^{m-1}; \theta)) w(y_1^m; \theta)$ , where  $w(y_1^m; \theta) := \min(1, c \|A(\theta) (s(y_1^m; \theta) - \tau(y_1^{m-1}; \theta))\|^{-1})$ . We define a robust functional  $M$ -estimator  $\bar{a}(\cdot)$  of  $\theta_0$  implicitly by

$$n^{-1} \sum_{t=1}^n \psi_c(\tilde{y}_{t-m+1}^t; \bar{a}(\mathbb{P}_{\theta_0}^m)) = 0. \quad (10)$$

The nonsingular matrix  $A(\theta) \in \mathbb{R}^p \times \mathbb{R}^p$  and the  $\mathcal{F}_{m-1}$ -measurable random vectors  $\tau(y_1^{m-1}; \theta) \in \mathbb{R}^p$  are determined by the implicit equations

$$E_{\theta_0}[\psi_c(y_1^m; \theta_0) \psi_c(y_1^m; \theta_0)^\top] = I, \quad (11)$$

$$E_{\theta_0}[\psi_c(y_1^m; \theta_0) | \mathcal{F}_{m-1}] = 0. \quad (12)$$

The estimating function  $\psi_c$  (or the unscaled version  $\psi^{bif}$ ) is conditionally unbiased at the reference model and is a truncated version of the ML score (5) as, by construction,  $\|\psi_c(y_1^m; \theta)\| \leq c$ . Moreover, as  $(\psi_c(y_{1+t}^{m+t}; \theta_0))_{t \in \mathbb{Z}}$  is a martingale difference sequence under  $\mathbb{P}_{\theta_0}$ , the conditional IF of the functional estimator  $\bar{a}(\cdot)$  is given by (9),

$$IF_{\psi_c}(y_1^m; \bar{a}(\mathbb{P}_{\theta_0}^m)) = D(\psi_c; \theta_0)^{-1} \psi_c(y_1^m; \bar{a}(\mathbb{P}_{\theta_0}^m)).$$

The estimating function  $\psi^{bif}$  satisfies the following optimality criterion.

**Proposition 3.1** If for a given constant  $c \geq \sqrt{p}$  equations (11) and (12) have solutions  $A(\theta_0)$  and  $\tau(y_1^{m-1}; \theta_0)$ , then  $\psi^{bif}$  minimizes  $tr(V(\psi; \theta_0) V(\psi^{bif}; \theta_0)^{-1})$  among all  $\psi$  satisfying (3) and such that

$$\sup_{z \in \mathbb{R}^m} (IF_\psi(z; \theta_0)^\top V(\psi^{bif}; \theta_0)^{-1} IF_\psi(z; \theta_0))^{1/2} \leq c. \quad (13)$$

Up to multiplication by a constant matrix,  $\psi^{bif}$  is unique almost surely.

Any score function  $\psi^{opt}$  such that  $V(\psi; \theta_0) - V(\psi^{opt}; \theta_0)$  is positive semi-definite for all  $\psi$  satisfying (3) is called strongly efficient. The following corollary holds.

**Corollary 3.1** If  $\psi^{bif}$  exists and there exists a conditionally unbiased, strongly efficient score function  $\psi^{opt}$  satisfying  $\gamma(\psi^{opt}) \leq c < \infty$ , then the two estimators coincide.

The proofs follow from Stefanski et al. (1986, pp. 422–423), using the property

$$E_{\theta_0}[\tau(y_1^{m-1}; \theta_0) \psi(y_1^m; \theta_0)] = E_{\theta_0}[\tau(y_1^{m-1}; \theta_0) E_{\theta_0}[\psi(y_1^m; \theta_0) | \mathcal{F}_{m-1}]] = 0,$$

for any conditionally unbiased score function  $\psi$ .

Under general conditions (see Clarke 1983, 1986; Heritier and Ronchetti 1994), the optimal robust estimator  $\bar{a}(\cdot)$  is consistent and asymptotically normally distributed at the reference model  $\mathbb{P}_{\theta_0}^m$ , with an asymptotic covariance matrix given by  $V(\psi_c; \theta_0) = D(\psi_c; \theta_0)^{-1} D(\psi_c; \theta_0)^{-\top}$ .

**Remark 3.1** The conditions given in the references above guarantee the Fréchet differentiability of the functional corresponding to the estimator  $\bar{a}(\cdot)$  which in turn imply the asymptotic normality in a  $\epsilon n^{-1/2}$ -neighborhood of the reference model; see also Bednarski (1993). These conditions are satisfied for  $M$ -estimators with a score function  $\psi$  which is bounded, continuous, and a.e. differentiable. Notice in particular that the PMLE is defined by an unbounded score function and is not Fréchet differentiable.

### 3.2.2 Interpretation of $A$ and $\tau$

The  $A$  matrix ensures that the normed self-standardized IF of  $\bar{a}(\cdot)$  is equal to the norm of the robust score function  $\psi_c$ , which is bounded by  $c$ . Indeed, under the scaling condition (11),

$$\|V(\psi_c; \theta_0)^{-1/2} IF_{\psi_c}(y; \theta_0)\| = \|\psi_c(y; \theta_0)\|.$$

The  $A$  matrix can be computed by a simple iterative procedure given explicitly in Section 3.2.4.

Further, to satisfy the conditional Fisher consistency condition (12), each truncated score function has to be shifted by some corresponding  $\tau$ -vector. This implicitly defines the random sequence of  $\tau$ -vectors  $(\tau(y_{1+t}^{m-1+t}; \theta_0))_{t \in \mathbb{Z}}$  associated to  $(\psi_c(y_{1+t}^{m+t}; \theta_0))_{t \in \mathbb{Z}}$ . The existence of such a sequence is guaranteed by the continuity of the mapping  $\tau(y_1^{m-1}; \theta) \mapsto (s(y_1^m; \theta) - \tau(y_1^{m-1}; \theta)) w(y_1^m; \theta)$  and by the mean value theorem; cf. also Lemma 2.1 in Künsch et al. (1989). As  $\tau(y_1^{m-1}; \theta_0)$  is  $\mathcal{F}_{m-1}$ -measurable,

$$\tau(y_1^{m-1}; \theta_0) = \frac{E_{\theta_0}[s(y_1^m; \theta_0) w(y_1^m; \theta_0) | \mathcal{F}_{m-1}]}{E_{\theta_0}[w(y_1^m; \theta_0) | \mathcal{F}_{m-1}]}.$$
 (14)

In the next section we provide an accurate analytical approximation of  $\tau(y_1^{m-1}; \theta_0)$ . This approximation makes use crucially of the conditionally unbiasedness of the robust score function  $\psi_c$ . For an unconditionally unbiased robust  $M$ -estimator the centering  $\tau$ -vector is implicitly defined by

$$\tau(\theta_0) = \frac{E_{\theta_0}[s(y_1^m; \theta_0) w(y_1^m; \theta_0)]}{E_{\theta_0}[w(y_1^m; \theta_0)]}.$$
 (15)

In general, the expectations in (15) cannot be expressed analytically, except in some very particular cases like AR( $p$ ) model settings where  $\tau(\theta_0) = 0$ , because the computation of  $\tau$  requires computing some unconditional moments under  $\mathbb{P}_{\theta_0}^m$ . Unfortunately, in virtually all cases relevant for this paper such moments are unknown and the functional dependence of  $\tau$  on  $\theta_0$  and  $A$  in (15) must be computed by solving some  $m$ -dimensional integrals by Monte Carlo simulation.

### 3.2.3 Analytical Approximation for $\tau(y_1^{m-1}; \theta_0)$

We briefly explain the analytical approximation of the  $\tau$ -vectors in (14). Detailed calculations are given in Appendix A. We proceed in two steps.

**Step 1.** Given  $\tau^{(0)}$  as initial value for  $\tau(y_1^{m-1}; \theta_0)$ , we compute the real roots of the following quartic equation, with respect to the real variable  $u_m(\theta_0)$ ,

$$\begin{aligned} 0 &= \|A(\theta_0) (s(y_1^m; \theta_0) - \tau^{(0)})\|^2 - c^2 \\ &:= \|A(\theta_0) (-k_{1,m} + k_{2,m} u_m(\theta_0) + k_{1,m} u_m^2(\theta_0) - \tau^{(0)})\|^2 - c^2. \end{aligned}$$
 (16)

In almost all simulations and all empirical estimations in Sections 5 and 6, equation (16) had only two real roots. Therefore, we only consider this case for brevity. The case of four real roots is discussed in

Appendix A.

**Step 2.** We ‘split’ the integrals in (14) according to the roots determined by (16). Denoting such roots by  $\underline{u}_m$  and  $\bar{u}_m$ , with  $\underline{u}_m \leq \bar{u}_m$ , the denominator in (14) is

$$E_{\theta_0}[w(y_1^m; \theta_0) | \mathcal{F}_{m-1}] = \int_{-\infty}^{\underline{u}_m} \frac{c}{\|A(\theta_0)(s(v; \theta_0) - \tau^{(0)})\|} d\Phi(u) + [\Phi(\bar{u}_m) - \Phi(\underline{u}_m)] + \int_{\bar{u}_m}^{+\infty} \frac{c}{\|A(\theta_0)(s(v; \theta_0) - \tau^{(0)})\|} d\Phi(u), \quad (17)$$

where  $v := (y_1^{m-1}, \mu_m(\theta_0) + \sigma_m(\theta_0)u)$  and  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the standard Gaussian density and cumulative distribution function. In our applications, typical values of  $\bar{u}_m$  range from 2.7 to 3.5 (the opposite for  $\underline{u}_m$ ). Hence, the ‘main contribution’ to the expectation on the left hand side of (17) comes from the term in the square brackets. Since  $\underline{u}_m$  and  $\bar{u}_m$  are ‘quite far’ in the tails of a standard Gaussian distribution, the remaining integrals can be well approximated using Laplace’s method; cf. for instance Jensen (1995), theorem 3.1.1. The integral in the numerator of (14) is split in the same way as in (17) and the relevant integrals are again approximated using Laplace’s method. The resulting formula for the computation of  $\tau$  is given in the next proposition.

**Proposition 3.2** Given model (1) and the conditionally Gaussian reference model  $\mathbb{P}_{\theta_0}^m$ , if the quartic equation (16) has only two real roots  $\underline{u}_m \leq \bar{u}_m$ , then

$$\tau(y_1^{m-1}; \theta_0) = \frac{-L_n(\underline{u}_m) - k_{1,m} [\Phi(\bar{u}_m) - \Phi(\underline{u}_m)] + k_{2,m} M_{1,m} + k_{1,m} M_{2,m} + L_n(\bar{u}_m)}{-L_d(\underline{u}_m) + [\Phi(\bar{u}_m) - \Phi(\underline{u}_m)] + L_d(\bar{u}_m)} + O(\underline{u}_m^{-3}) + O(\bar{u}_m^{-3}),$$

where  $M_{1,m} := \phi(\underline{u}_m) - \phi(\bar{u}_m)$ ,  $M_{2,m} := \underline{u}_m \phi(\underline{u}_m) - \bar{u}_m \phi(\bar{u}_m) + \Phi(\bar{u}_m) - \Phi(\underline{u}_m)$ .  $L_n(\cdot)$  and  $L_d(\cdot)$  are defined in Appendix A and correspond to some Laplace approximations for the integrals in the numerator and in the denominator of (14).

Intuitively, the real roots  $\underline{u}_m$  and  $\bar{u}_m$  in equation (16) determine the range where the standardized innovation  $u_m(\theta_0)$  is ‘not influential’ (in terms of the self-standardized sensitivity of  $\bar{a}(\cdot)$ ) for the arising asymptotic bias. Indeed,

$$\begin{aligned} \|A(\theta_0)(s(y_1^m; \theta_0) - \tau^{(0)})\| \leq c, & \iff u_m(\theta_0) \in [\underline{u}_m, \bar{u}_m], \\ > c, & \iff u_m(\theta_0) \in (-\infty, \underline{u}_m) \cup (\bar{u}_m, +\infty), \end{aligned}$$

and the normed self-standardized IF of  $\bar{a}(\cdot)$  is equal to the norm of the  $\psi_c$ -function.

### 3.2.4 Algorithm

To compute the robust estimator defined by (10)–(12) an iterative algorithm has to be adopted because the weights  $w(y_1^m; \bar{a}(\mathbb{P}_{\theta_0}^m))$ , the matrix  $A(\bar{a}(\mathbb{P}_{\theta_0}^m))$  and the random vectors  $\tau(y_1^{m-1}; \bar{a}(\mathbb{P}_{\theta_0}^m))$  depend on the value of the estimator itself in a nonlinear way. Given a constant  $c \geq \sqrt{p}$  (cf. Hampel et al. 1986, p. 228), the robust estimator is computed by the following four steps algorithm.

1. Fix a starting value  $\theta^{(0)}$  for  $\theta_0$ , and initial values  $\tau_t^{(0)} := \tau(\tilde{y}_{t-m+1}^{t-1}; \theta^{(0)}) = 0$ , for all  $t = 1, \dots, n$  and  $A^{(0)}$  such that

$$A^{(0)\top} A^{(0)} = \left[ n^{-1} \sum_{t=1}^n s(\tilde{y}_{t-m+1}^t; \theta^{(0)}) s(\tilde{y}_{t-m+1}^t; \theta^{(0)})^\top \right]^{-1}.$$

2. Compute, for all  $t = 1, \dots, n$ , the real roots of equations (16), and the associated new values  $\tau_t^{(1)} := \tau(\tilde{y}_{t-m+1}^{t-1}; \theta^{(0)})$  for  $\tau_t$  and the new matrix  $A^{(1)}$  for  $A$ , defined by

$$\tau_t^{(1)} := \frac{-L_n(\underline{u}_t) - k_{1,t} [\Phi(\bar{u}_t) - \Phi(\underline{u}_t)] + k_{2,t} M_{1,t} + k_{1,t} M_{2,t} + L_n(\bar{u}_t)}{-L_d(\underline{u}_t) + [\Phi(\bar{u}_t) - \Phi(\underline{u}_t)] + L_d(\bar{u}_t)},$$

$$\begin{aligned} (A^{(1)\top} A^{(1)})^{-1} &:= n^{-1} \sum_{t=1}^n \left( s(\tilde{y}_{t-m+1}^t; \theta^{(0)}) - \tau_t^{(0)} \right) \left( s(\tilde{y}_{t-m+1}^t; \theta^{(0)}) - \tau_t^{(0)} \right)^\top \times \\ &\quad \min^2(1, c \|A^{(0)} (s(\tilde{y}_{t-m+1}^t; \theta^{(0)}) - \tau_t^{(0)})\|^{-1}). \end{aligned}$$

3. Compute the robust estimator  $\theta^{(1)}$  implied by (10) for given  $A^{(1)}$  and  $\tau_t^{(1)}$  as the solution of the implicit equation

$$\sum_{t=1}^n \left( s(\tilde{y}_{t-m+1}^t; \theta^{(1)}) - \tau_t^{(1)} \right) \min(1, c \|A^{(1)} (s(\tilde{y}_{t-m+1}^t; \theta^{(1)}) - \tau_t^{(1)})\|^{-1}) = 0.$$

4. Replace  $A^{(0)}$  by  $A^{(1)}$  and  $\tau_t^{(0)}$  by  $\tau_t^{(1)}$  for all  $t = 1, \dots, n$  and iterate steps 2 and 3 above until convergence of the sequence  $(\theta^{(i)})_{i \in \mathbb{N}}$  of estimators associated with (10) and with the sequence  $(A^{(i)}, \bar{\tau}^{(i)})_{i \in \mathbb{N}}$ , where  $\bar{\tau}^{(i)} := (\tau_1^{(i)}, \dots, \tau_n^{(i)})$ .

Starting values for  $\theta^{(0)}$  could be the PML estimate of  $\theta_0$  or the result of a grid search algorithm. We wrote a Matlab code to implement the algorithm and we used the Matlab function ‘roots’ to compute the real roots of equation (16). Analytical expressions for the  $\tau$ -vectors avoid “internal” simulations to compute

the robust estimator. This largely reduces the computation time. For comparison, we implemented a second algorithm in which the  $\tau$  integrals were computed numerically using the Matlab function ‘quad1’. This algorithm is unfeasible as the computation time of  $\bar{a}$  is almost two hours already for a simple AR(1)-ARCH(1) model. For further comparison, we also implemented a robust GMM estimator as in Ronchetti and Trojani (2001) with moment conditions  $A(\theta)(s(y_1^m; \theta) - \tau(\theta)) w(y_1^m; \theta)$ , where  $\tau(\theta)$  is given by (15). In our simulations of Section 5, the computation time of the estimator (10)–(12) was about 20% the one of such a robust GMM estimator. The Matlab code is available from the authors upon request.

## 4 Robust Testing Procedures

Robust versions of the classical Wald, score and likelihood ratio tests based on the robust estimator in (10)–(12) can be derived following the general approach proposed by Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001). Such robust tests satisfy the optimality criterion of maximizing the asymptotic power subject to a bound on the asymptotic bias of the level and the power test. Such biases can be controlled by imposing a bound on  $\gamma$ ; cf. Ronchetti and Trojani (2001, p. 54). For brevity, we focus on the robust version of the Wald test. Robust score and likelihood ratio tests can be handled in a similar way.

Consider a parametric null hypothesis of the form

$$g(\bar{a}(\mathbb{P}_{\theta_0}^m)) = 0, \tag{18}$$

for a smooth function  $g : \Theta \rightarrow \mathbb{R}^r$  such that  $(\partial/\partial\bar{a})g(\bar{a}(\mathbb{P}_{\theta}^m))^\top$  is of full column rank  $r$  for all  $\theta \in \Theta$ .

We consider for any  $n \in \mathbb{N}$  test statistics  $nQ$  that are quadratic forms of a Wald functional  $U$ ,

$$nQ(\mathbb{P}_n^m) := nU(\mathbb{P}_n^m)^\top U(\mathbb{P}_n^m); \quad U(\mathbb{P}_n^m) := \left[ \frac{\partial g(\theta)}{\partial \theta^\top} V(\psi_c; \theta) \frac{\partial g(\theta)^\top}{\partial \theta} \right]_{\theta=\bar{a}(\mathbb{P}_n^m)}^{-1/2} g(\bar{a}(\mathbb{P}_n^m)), \tag{19}$$

where  $\mathbb{P}_n^m$  is the empirical  $m$ -dimensional distribution of the observations  $\tilde{y}_1^n$ . Under the reference model  $\mathbb{P}_{\theta_0}^m$  and the null hypothesis (18),  $nQ(\mathbb{P}_n^m)$  converges in distribution to a  $\chi^2$  distribution with  $r$  degrees of freedom. To apply the methodology in Heritier and Ronchetti (1994), we make the following assumption.

**Assumption 4.1** Let a bounded-influence estimator  $\bar{a}$  of  $a(\mathbb{P}_{\theta_0}^m)$  be given. Then,

$$\sqrt{n} \left( \bar{a}(\mathbb{P}_n^m) - \bar{a}(\mathbb{P}_{\eta(\epsilon, n)}^m) \right) \rightarrow \mathcal{N}(0, V(\psi; \theta_0)), \quad n \rightarrow \infty, \quad (20)$$

in distribution, uniformly over the sequence  $(\mathcal{U}^{\eta(\epsilon, n)}(\mathbb{P}_{\theta_0}^m))_{n \in \mathbb{N}}$  of  $\eta(\epsilon, n)$ -neighborhoods defined by (7) for  $\eta := \eta(\epsilon, n) = \epsilon n^{-1/2}$  and  $\mathbb{G} \in \text{dom}(\bar{a})$ .

Assumption 4.1 is implied by the Fréchet differentiability of the functional  $\bar{a}(\cdot)$ ; cf. Remark 3.1 after Corollary 3.1. Under Assumption 4.1 the following proposition holds.

**Proposition 4.1** Let  $\bar{a}$  be the robust estimator defined by (10)–(12) and denote by  $\alpha$  the asymptotic level functional of the Wald statistic (19). Further let  $(\mathbb{P}_{\eta(\epsilon, n)}^m)_{n \in \mathbb{N}}$  be a sequence of  $\eta(\epsilon, n)$ -contaminations of the null distribution  $\mathbb{P}_{\theta_0}^m$ , i.e.  $\mathbb{P}_{\eta(\epsilon, n)}^m \in \mathcal{U}^{\eta(\epsilon, n)}(\mathbb{P}_{\theta_0}^m)$ . Then, the bias of the asymptotic level  $\lim_{n \rightarrow \infty} \alpha(\mathbb{P}_{\eta(\epsilon, n)}^m)$  is uniformly bounded by

$$\lim_{n \rightarrow \infty} |\alpha(\mathbb{P}_{\eta(\epsilon, n)}^m) - \alpha_0| \leq \mu \epsilon^2 \sup_{z \in \mathbb{R}^m} \|\psi_c(z; \bar{a})\|^2 + o(\epsilon^2),$$

where  $\mu := -(\partial/\partial\beta)H_r(q_{1-\alpha_0}; \beta)|_{\beta=0}$ ,  $H_r(\cdot; \beta)$  is the cumulative distribution function of a noncentral  $\chi^2(r; \beta)$  distribution with  $r$  degrees of freedom and noncentrality parameter  $\beta \geq 0$ ,  $q_{1-\alpha_0}$  is the  $1 - \alpha_0$  quantile of a  $\chi^2(r; 0)$  distribution and  $\alpha_0 = \alpha(\mathbb{P}_{\theta_0}^m)$  is the nominal asymptotic level of the test at the reference model.

The proof follows from Heritier and Ronchetti (1994, p. 903) and Ronchetti and Trojani (2001, p. 64).

As a consequence of Proposition 4.1, the maximal asymptotic bias in the level of the test based on the robust  $M$ -estimator  $\bar{a}$  in (10)–(12) is bounded by

$$\lim_{n \rightarrow \infty} |\alpha(\mathbb{P}_{\eta(\epsilon, n)}^m) - \alpha_0| \leq \mu (\epsilon c)^2 + o(\epsilon^2). \quad (21)$$

The “power” counterpart of Proposition 4.1 can also be obtained; cf. Ronchetti and Trojani (2001), theorem 2.

## 5 Monte Carlo Simulations

We compare by Monte Carlo simulations the performance of the classical PMLE (cf. Gouriéroux, Monfort and Trognon 1984) defined by the score function (5) with the one of our robust estimator, both at the reference model and in the presence of local model contaminations. We estimate an AR(1)-ARCH(1) model and simulate the following contaminated models “near” the reference model  $\mathbb{P}_{\theta_0}$ .

1. *Standard Gaussian innovations.* In this experiment, the innovation  $u_t(\theta_0)$  has a standard Gaussian distribution. Hence, the PMLE is the MLE and we compare the efficiency of our robust estimator and the MLE at the reference model  $\mathbb{P}_{\theta_0}$ .
2. *Replacement model* (cf., for instance, Martin and Yohai 1986). Under such a model the observed process  $\mathcal{X} := (x_t)_{t \in \mathbb{Z}}$  is generated according to the data generating process,

$$x_t = (1 - \vartheta_t^\eta) y_t + \vartheta_t^\eta \xi_t, \quad (22)$$

where the clean process  $\mathcal{Y} := (y_t)_{t \in \mathbb{Z}}$  is generated by the reference model  $\mathbb{P}_{\theta_0}$  and  $(\vartheta_t^\eta)_{t \in \mathbb{Z}}$  is an i.i.d. 0-1 random sequence, independent of  $\mathcal{Y}$ , such that  $\mathbb{P}(\vartheta_t^\eta = 1) = \eta$ . Hence, at a time  $t \in \mathbb{Z}$ , the clean observation  $y_t$  is replaced by  $\xi_t$  with probability  $\eta$ . In our simulations we set  $\eta = 0.5\%$  and  $\xi_t = 1.5$  for all  $t$ . Such a low probability of contaminations is motivated by some difficulties of the standard PMLE to converge when higher probabilities of contaminations occur (for e.g.  $\eta = 1\%$ ). In this experiment model (1) is dynamically “slightly” misspecified as the dynamic equations (1) are not satisfied. This experiment allows to compare the performances of the PMLE and the robust estimator when very few observations deviate from the assumed model.

3. *Innovative outlier model* (cf., for instance, Bustos and Yohai 1986). Under such a contamination the innovations are given by  $u_t(\theta_0) = \check{u}_t(\theta_0) [(1 - \epsilon) + \epsilon \varrho^2]^{-1/2}$ , where  $\check{u}_t(\theta_0)$  is distributed as the following mixture distribution

$$\check{u}_t(\theta_0) \sim (1 - \epsilon) \mathcal{N}(0, 1) + \epsilon \mathcal{N}(0, \varrho^2). \quad (23)$$

We set  $\epsilon = 1\%$  and  $\varrho = 3$ . (23) describes situations where a given shock (or outlier) affects also future realizations of the process  $\mathcal{Y}$ . Furthermore, as  $u_t(\theta_0) \sim i.i.d.(0, 1)$ , the dynamic equations (1)

are satisfied and the model is dynamically correctly specified. Hence, this is a typical situation in which the PMLE is applied and there are no theoretical efficiency reasons to prefer one estimator to the other.

We simulate an AR(1)-ARCH(1) model implied by (2) for the parameter choice:  $\rho_0 = 0.01$ ,  $\rho_1 = 0.8$ ,  $\alpha_0 = 0.02$ ,  $\alpha_1 = 0.8$  and  $\rho_2 = \alpha_2 = \alpha_3 = 0$  under the above distributions for  $y_t$  and for a sample size  $n = 1,000$ . The tuning constant for the robust estimator  $\bar{a}$  was set at  $c = 9$ . A formal criterion for choosing  $c$  based on efficiency is given in Section 5.4. In particular, the choice of  $c = 9$  in our Monte Carlo simulations implies a relative efficiency of about 95% for our robust estimator under the reference model  $\mathbb{P}_{\theta_0}$ . Each model is simulated 5,000 times. For each simulation we compute the PML and the robust estimates of  $\theta_0$  with the corresponding estimated asymptotic covariance matrices. Then, for each parameter we compute the corresponding confidence interval at the 95% confidence level.

## 5.1 Point Estimation

Estimation results are presented in Table 1. For each estimated parameter, the first (second) column contains summary statistics for the PML estimates (the robust estimates). In Figures 1–3 we plot the estimated densities of the classical and the robust estimators. The first panel in Table 1 shows that the efficiency loss of the robust estimator at the reference model  $\mathbb{P}_{\theta_0}$  is almost negligible. Specifically, the mean squared errors of all parameter estimates for the two estimation procedures are very close. This is confirmed by Figure 1. The second panel in Table 1 and Figure 2 show instead large mean squared errors of PML estimates under the replacement model (22). By contrast, robust estimates maintain low mean squared errors. It is somehow surprising that such large inefficiencies in the PML estimates are induced by contaminating (on average) only 0.5% of the sample. Finally, the third panel in Table 1 and Figure 3 show that, in terms of mean squared error, both estimators estimate correctly the conditional mean parameters  $\rho_0$  and  $\rho_1$ . However, the robust estimator always outperforms the PMLE in the estimation of the conditional variance parameters  $\alpha_0$  and  $\alpha_1$ .

## 5.2 Interval Estimation

Figures 4–6 show the boxplots of the estimated confidence interval lengths for the PML and the robust estimates. Actual confidence interval coverages are close to the nominal level 95% (between 93% and 96%) in both cases and for all parameters. Two exceptions are the confidence intervals of the parameter  $\alpha_0$  which are 78.4% for the PMLE and 92.4% for our robust estimator and the confidence intervals of the parameter  $\rho_0$  which are 90.6% for the PMLE and 93.6% for our robust estimator under the replacement model (22). Moreover, Figure 4 shows that, *under the reference model*  $\mathbb{P}_{\theta_0}$ , the confidence interval lengths for both estimation techniques are almost identical. However, Figure 5 shows that, under the replacement model (22), the PML confidence intervals are much larger than the robust ones, denoting a large inaccuracy of the inference results. Robust confidence intervals are much more concentrated around  $\theta_0$ . Finally, Figure 6 shows that confidence intervals under the innovative outlier model (23) are tighter for the robust than for the PML estimates, especially for the conditional variance parameters  $\alpha_0$  and  $\alpha_1$ .

## 5.3 Hypothesis Testing

To compare the performance of the classical PMLE and our robust estimator from the perspective of hypothesis testing we also simulated 1,000 sample paths of an AR(1)-ARCH(1) model for the parameter choices  $\rho_0 = 0$ ,  $\rho_1 = 0, 0.05, 0.10$  and  $\alpha_0 = 0.02$ ,  $\alpha_1 = 0.8$  under scaled Student  $t_3$  and scaled Student  $t_5$  innovations. We do not necessarily believe that in applications the innovations will generally follow these distributions. We rather take the Student  $t_3$  and  $t_5$  distributions as further examples of distributions which are close to the normal one. Under scaled Student  $t$  innovations model (1) is dynamically correctly specified and hence the PMLE should perform well. In our experiments we tested the joint null hypothesis  $\rho_0 = \rho_1 = 0$  by means of a classical and a robust Wald statistic. The empirical rejection frequencies of a Wald test based on the classical PMLE and a Wald test based on our robust estimator are calculated for a fixed nominal level 5% of the test. The results are presented in Table 2. The estimated standard error of the empirical rejection frequency  $\hat{p}$  is 0.7%, 1.4%, 1.5% for  $\hat{p} = 5\%$ , 30%, 60%, respectively. Table 2 shows that the robust Wald test performs very well across all models, while the classical test is oversized

in finite samples and shows a lower power than the robust one.

The low power of classical tests under even slight departures from conditional normality suggests that robust tests could be useful in application to unmask some possible ‘dynamics’ in the data hidden by the presence of influential observations.

## 5.4 Choice of the Robustness Level $c$ and Comparison with RGMM

An important issue in the application of our robust inference procedures is the selection of an appropriate tuning constant  $c$  in (10). It appears from the optimality result in Section 3.2.1 and from (21) that such a constant controls the degree of robustness of our robust procedure. Usually one chooses the tuning constant  $c$  in order to achieve a given asymptotic efficiency of the robust estimates under the reference model  $\mathbb{P}_{\theta_0}$ . Figure 7 shows the relative efficiency of the robust estimator  $\bar{a}$  and the RGMME defined in Section 3.2.4 with respect to the MLE. The relative efficiency is measured as  $tr(V(s; \hat{\theta}_n))/tr(V(\psi; \hat{\theta}_n))$ , where the asymptotic covariance matrix  $V$  is estimated using  $n = 10,000$  observations simulated under the reference model  $\mathbb{P}_{\theta_0}$  and  $\psi$  is the corresponding estimating function for the robust estimators. For  $c = 9$ , our robust estimator  $\bar{a}$  has about 95% asymptotic relative efficiency; cf. Figure 7. Alternatively, one can choose the tuning constant by taking an inference point of view and by using (21) to set  $c$  according to the expected degree of contamination  $\epsilon$  and the maximal bias allowed on the level of the test. A table providing guidance on how to select  $c$  according to this criterion is given in Ronchetti and Trojani (2001, p. 54).

In a simulation study we compared the performances of the PMLE, our robust estimator  $\bar{a}$  and the RGMME introduced in Section 3.2.4 for the same parameter choices and sample size as in the previous Monte Carlo simulation, under the reference model  $\mathbb{P}_{\theta_0}$ , the replacement model (22) for  $\eta = 5\%$ ,  $\xi_t \sim \mathcal{N}(0, 1)$  and the innovative outlier model for  $\epsilon = 1\%$ ,  $\varrho = 5$ . Notice the larger amount of contamination in these experiments. Each model is simulated 1,000 times, given the large computation time of the RGMME which is about five times that of our robust estimator  $\bar{a}$ . For both estimators we set  $c = 4$ . Table 3 summarizes the simulation results. As expected from Figure 7, under the reference model  $\mathbb{P}_{\theta_0}$ , our robust estimator  $\bar{a}$  is more efficient than the RGMME and has a lower MSE in three out of four

estimated parameters. Under the contaminated models the robust estimator  $\bar{a}$  tends to outperform the RGMME in terms of biases and MSE's. Again, under such model misspecifications the PML estimates are found to be highly inefficient.

## 6 Empirical Application

We apply classical and robust Wald tests for ARCH to weekly exchange rate returns of the Swedish krona versus the US dollar over the period November 29<sup>th</sup>, 1993 until November 17<sup>th</sup>, 2003. The data were downloaded from Datastream and consist of 522 observations. The first ten sample autocorrelations of squared and absolute returns are not significantly different from zero. Moreover, the Jarque-Bera test has a  $p$ -value of 0.47 not rejecting normality. Classical PML estimates for the parameters  $\rho_0$ ,  $\rho_1$ ,  $\alpha_0$  and  $\alpha_1$  of an AR(1)-ARCH(1) model (Wald test  $p$ -values for the hypothesis that the corresponding parameter is zero) are 0.02 (0.73),  $-0.030$  (0.53), 1.86 (0), 0.06 (0.22). The robust estimates under a tuning constant  $c = 4$  are 0.01 (0.88), 0.014 (0.75), 1.64 (0), 0.47 (0). Therefore, as in typical financial return series, the conditional mean parameters are not significantly different from zero. Moreover, the PML estimate of the ARCH parameter  $\alpha_1$  is also not significant. Hence, the classical Wald test does not reject the homoscedasticity hypothesis. By contrast, the robust estimate of this ARCH parameter is highly significant, showing that ARCH effects in the data are possibly obscured by some outlying observations detected by the robust weights presented in the bottom panel of Figure 8. These results are consistent with the low power of PML tests under non normal conditional returns in Section 5.3. Finally, it is interesting to notice that one would expect outliers to *enhance* the ARCH structure. Instead, because the estimation of the volatility by classical techniques is inflated, the potential ARCH structure is hidden by the presence of a few outlying observations.

## 7 Conclusions

We derived optimal bounded-influence estimators for the parameters of conditional location and scale models under a conditionally Gaussian reference model. Based on these results, we obtained optimal

bounded-influence versions of the classical likelihood-based tests for parametric hypotheses. We proposed an efficient algorithm for the computation of our robust estimators, which strongly reduces the necessary computation time by avoiding the simulation of multidimensional integrals. Monte Carlo simulations show that our robust estimators maintain a high efficiency under ideal model conditions and have good robustness properties under local departures from conditional normality, both in estimation and inference. On the contrary, classical PML estimators are highly inefficient even under small departures from conditional Gaussianity. An application to exchange rate data confirms these patterns.

## A Computation of $\tau(y_1^{m-1}; \theta_0)$

This appendix describes the computation of the correction factor  $\tau(y_1^{m-1}; \theta_0)$  in equation (12). According to (5), the Gaussian score function can be written as

$$s(y_1^m; \theta_0) = -k_{1,m} + k_{2,m}u_m(\theta_0) + k_{1,m}u_m^2(\theta_0),$$

where  $u_m(\theta_0) \sim \mathcal{N}(0, 1)$  under the reference model  $\mathbb{P}_{\theta_0}^m$ . For brevity we write  $A$  instead of  $A(\theta_0)$ . Formally, the problem is to compute  $\tau(y_1^{m-1}; \theta_0)$  such that

$$0 = A \int_{-\infty}^{+\infty} (-k_{1,m} + k_{2,m}u + k_{1,m}u^2 - \tau(y_1^{m-1}; \theta_0)) w(y_1^{m-1}, \mu_m(\theta_0) + \sigma_m(\theta_0)u; \theta_0) d\Phi(u).$$

As  $\tau(y_1^{m-1}; \theta_0)$ ,  $k_{1,m}$  and  $k_{2,m}$  are  $\mathcal{F}_{m-1}$ -measurable, we have

$$\tau(y_1^{m-1}; \theta_0) = \frac{\tau_{num}(y_1^{m-1}; \theta_0)}{\tau_{den}(y_1^{m-1}; \theta_0)},$$

where

$$\tau_{num}(y_1^{m-1}; \theta_0) := \int_{-\infty}^{+\infty} (-k_{1,m} + k_{2,m}u + k_{1,m}u^2) w(y_1^{m-1}, \mu_m(\theta_0) + \sigma_m(\theta_0)u; \theta_0) d\Phi(u), \quad (24)$$

$$\tau_{den}(y_1^{m-1}; \theta_0) := \int_{-\infty}^{+\infty} w(y_1^{m-1}, \mu_m(\theta_0) + \sigma_m(\theta_0)u; \theta_0) d\Phi(u). \quad (25)$$

Clearly, the difficulties in the computation of these integrals derive from the presence of the weighting function  $w(y_1^m; \theta_0)$  defined in (10). However, as the weighting function implies that  $\|\psi_c(y_1^m; \theta_0)\|^2 \leq c^2$ , we can equivalently express such an inequality in terms of ‘admissible’ values of the standardized innovation  $u_m(\theta_0)$ . Specifically, we compute  $\tau(y_1^{m-1}; \theta_0)$  by means of the following two steps procedure.

**Step 1.** In the first step we compute the real roots in the real variable  $u_m(\theta_0)$  of the quartic equation (16), i.e.

$$\begin{aligned} 0 &= \|A(s(y_1^m; \theta_0) - \tau^{(0)})\|^2 - c^2 \\ &= \|A(-k_{1,m} + k_{2,m}u_m(\theta_0) + k_{1,m}u_m^2(\theta_0) - \tau^{(0)})\|^2 - c^2 \\ &= a_4u_m^4(\theta_0) + a_3u_m^3(\theta_0) + a_2u_m^2(\theta_0) + a_1u_m(\theta_0) + a_0 - c^2, \end{aligned}$$

where

$$a_4 := k_{1,m}^\top A^\top A k_{1,m}, \quad a_3 := 2k_{1,m}^\top A^\top A k_{2,m},$$

$$a_2 := k_{2,m}^\top A^\top A k_{2,m} - 2k_{1,m}^\top A^\top A k_{1,m} - 2k_{1,m}^\top A^\top A \tau^{(0)},$$

$$a_1 := -a_3 - 2k_{2,m}^\top A^\top A \tau^{(0)},$$

$$a_0 := a_4 + 2k_{1,m}^\top A^\top A \tau^{(0)} + \tau^{(0)\top} A^\top A \tau^{(0)}.$$

Existence of a solution is guaranteed by Lemma 2.1 in Künsch et al. (1989) when choosing  $c \geq \sqrt{p}$ . In general, we have either two or four real roots. As  $a_4 > 0$ , in the first case

$$\begin{aligned} \|A(s(y_1^m; \theta_0) - \tau^{(0)})\| &\leq c, \quad u_m(\theta_0) \in [\underline{u}_m, \bar{u}_m], \\ &> c, \quad u_m(\theta_0) \in (-\infty, \underline{u}_m) \cup (\bar{u}_m, +\infty), \end{aligned}$$

denoting by  $\underline{u}_m \leq \bar{u}_m$  the real roots. In the second case, with four real roots  $\underline{\underline{u}}_m \leq \underline{u}_m \leq \bar{u}_m \leq \bar{\bar{u}}_m$ ,

$$\begin{aligned} \|A(s(y_1^m; \theta_0) - \tau^{(0)})\| &\leq c, \quad u_m(\theta_0) \in [\underline{\underline{u}}_m, \underline{u}_m] \cup [\bar{u}_m, \bar{\bar{u}}_m] \\ &> c, \quad u_m(\theta_0) \in (-\infty, \underline{\underline{u}}_m) \cup (\underline{u}_m, \bar{u}_m) \cup (\bar{\bar{u}}_m, +\infty). \end{aligned}$$

In almost all simulations and all empirical estimations in the paper we found only two real roots.

**Step 2.** In the second step we ‘split’ the integral in equation (24) and (25) according to the roots determined in Step 1. Assume first that there are two real roots, then

$$\begin{aligned} &\tau_{num}(y_1^{m-1}; \theta_0) \\ &= \int_{-\infty}^{\underline{u}_m} \overbrace{(-k_{1,m} + k_{2,m}u + k_{1,m}u^2)}^{q_n(u) := c} \frac{c}{\|A(s(v; \theta_0) - \tau^{(0)})\|} d\Phi(u) \\ &\quad + \int_{\underline{u}_m}^{\bar{u}_m} (-k_{1,m} + k_{2,m}u + k_{1,m}u^2) d\Phi(u) \\ &\quad + \int_{\bar{u}_m}^{+\infty} (-k_{1,m} + k_{2,m}u + k_{1,m}u^2) \frac{c}{\|A(s(v; \theta_0) - \tau^{(0)})\|} d\Phi(u) \\ &:= \int_{-\infty}^{\underline{u}_m} q_n(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) du - k_{1,m} [\Phi(\bar{u}_m) - \Phi(\underline{u}_m)] + k_{2,m}M_{1,m} + k_{1,m}M_{2,m} \\ &\quad + \int_{\bar{u}_m}^{+\infty} q_n(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) du. \end{aligned}$$

Notice that  $q_n : \mathbb{R} \rightarrow \mathbb{R}^p$  with the same functional form in each component.  $M_{1,m}$ ,  $M_{2,m}$  are defined in Proposition 3.2 and  $v := (y_1^{m-1}, \mu_m(\theta_0) + \sigma_m(\theta_0)u)$ .

Specifically,  $M_{1,\cdot}$  and  $M_{2,\cdot}$  are the truncated first and second moments of a standard Gaussian random variable and integration by parts yields

$$M_{1,\cdot} = \int_a^b u d\Phi(u) = \phi(a) - \phi(b), \quad M_{2,\cdot} = \int_a^b u^2 d\Phi(u) = a\phi(a) - b\phi(b) + \Phi(b) - \Phi(a).$$

The remaining univariate integrals are approximated ‘componentwise’ using Laplace’s method. Under standard regularity conditions (cf. for instance Jensen (1995), p. 58) on the real function  $q(\cdot)$ , for  $\alpha \rightarrow \infty$

$$\begin{aligned} \int_0^\infty \alpha \exp(-\alpha u) q(u) du &= q(0) + \frac{q'(0)}{\alpha} + \frac{q''(0)}{\alpha^2} + O\left(\frac{1}{\alpha^3}\right) \\ &=: \mathcal{L}(q, \alpha) + O\left(\frac{1}{\alpha^3}\right). \end{aligned}$$

$\mathcal{L}(q, \alpha)$  is the Laplace approximation of the integral up to the third order. We use third order Laplace approximations as the contribution of higher order terms is negligible. Therefore,

$$\begin{aligned} &\int_{\bar{u}_m}^{+\infty} q_n(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) du \\ &= \frac{1}{\sqrt{2\pi}} \exp(-0.5\bar{u}_m^2) \frac{1}{\bar{u}_m} \int_0^{+\infty} \bar{u}_m \exp(-\bar{u}_m z) q_n(\bar{u}_m + z) \exp(-0.5z^2) dz \\ &= \frac{1}{\sqrt{2\pi}} \exp(-0.5\bar{u}_m^2) \frac{1}{\bar{u}_m} \left( \mathcal{L}(\bar{q}_n, \bar{u}_m) + O\left(\frac{1}{\bar{u}_m^3}\right) \right) \\ &=: L_n(\bar{u}_m) + O\left(\frac{1}{\bar{u}_m^3}\right), \end{aligned}$$

where  $\bar{q}_n(z) := q_n(\bar{u}_m + z) \exp(-0.5z^2)$ . Similarly,

$$\begin{aligned} \int_{-\infty}^{\underline{u}_m} q_n(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) du &= -\frac{1}{\sqrt{2\pi}} \exp(-0.5\underline{u}_m^2) \frac{1}{\underline{u}_m} \left( \mathcal{L}(\underline{q}_n, \underline{u}_m) + O\left(\frac{1}{\underline{u}_m^3}\right) \right) \\ &=: -L_n(\underline{u}_m) + O\left(\frac{1}{\underline{u}_m^3}\right), \end{aligned}$$

where  $\underline{q}_n(z) := q_n(\underline{u}_m + z) \exp(-0.5z^2)$ . The procedure for computing the denominator of  $\tau$  in (25) is analogous.

In the general case where the quartic equation (16) has four real roots  $\underline{\underline{u}}_m \leq \underline{u}_m \leq \bar{u}_m \leq \bar{\bar{u}}_m$ , for instance the integral in (25) becomes

$$\begin{aligned} &\tau_{den}(y_1^{m-1}; \theta_0) \\ &= \int_{-\infty}^{\underline{\underline{u}}_m} \frac{c}{\|A(s(v; \theta_0) - \tau^{(0)})\|} d\Phi(u) + \int_{\underline{\underline{u}}_m}^{\underline{u}_m} d\Phi(u) + \int_{\underline{u}_m}^{\bar{u}_m} \frac{c}{\|A(s(v; \theta_0) - \tau^{(0)})\|} d\Phi(u) \\ &\quad + \int_{\bar{u}_m}^{\bar{\bar{u}}_m} d\Phi(u) + \int_{\bar{\bar{u}}_m}^{+\infty} \frac{c}{\|A(s(v; \theta_0) - \tau^{(0)})\|} d\Phi(u) \end{aligned}$$

and Laplace approximation could be applied to the first and the last integral. Numerical results (not reported here) show that the error when neglecting the weighting function in the central integral is very small. In fact  $\underline{u}_m$  and  $\bar{u}_m$  are typically very close and  $\|A(s(v; \theta_0) - \tau^{(0)})\|$  is quite small, so that the error is essentially zero.

	$\rho_0 = 0.01$		$\rho_1 = 0.8$		$\alpha_0 = 0.02$		$\alpha_1 = 0.8$	
	PML	ROB	PML	ROB	PML	ROB	PML	ROB
mean	0.0100	0.0100	0.7983	0.7977	0.0200	0.0200	0.7976	0.8007
median	0.0099	0.0100	0.7989	0.7985	0.0199	0.0200	0.7986	0.8016
MSE%	0.0026	0.0026	0.0199	0.0208	0.0002	0.0002	0.5756	0.5850
mean	0.0166	0.0112	0.7930	0.7959	0.0298	0.0222	0.8037	0.8081
median	0.0161	0.0111	0.7952	0.7965	0.0290	0.0218	0.8052	0.8064
MSE%	0.0132	0.0034	0.0832	0.0254	0.0145	0.0012	1.7880	0.8097
mean	0.0100	0.0100	0.7982	0.7978	0.0199	0.0194	0.7989	0.7727
median	0.0100	0.0100	0.7988	0.7983	0.0199	0.0194	0.7992	0.7748
MSE%	0.0030	0.0026	0.0256	0.0209	0.0003	0.0002	0.8006	0.6328

Table 1: Summary statistics based on 5,000 simulations for the PMLE and the robust estimator  $\bar{a}$  ( $c = 9$ ) of the AR(1)-ARCH(1) model under the reference model  $\mathbb{P}_{\theta_0}$  (first panel; cf. also Figure 1), the replacement model (22) for  $\eta = 0.5\%$ ,  $\xi_t = 1.5$  (second panel; cf. also Figure 2), the innovative outlier model (23) for  $\epsilon = 1\%$ ,  $\varrho = 3$  (third panel; cf. also Figure 3).

$\rho_1$	$t_3$		$t_5$	
	PML	ROB	PML	ROB
0.00	0.08	0.05	0.07	0.05
0.05	0.17	0.24	0.22	0.26
0.10	0.46	0.65	0.62	0.74

Table 2: Each entry in the Table corresponds to the empirical rejection frequency of the joint hypothesis  $\rho_0 = 0$  and  $\rho_1 = 0$  obtained using 5% critical values for the  $\chi^2$  test of the AR(1)-ARCH(1) model under scaled  $t_3$  and scaled  $t_5$  innovations, respectively.

	$\rho_0 = 0.01$			$\rho_1 = 0.8$			$\alpha_0 = 0.02$			$\alpha_1 = 0.8$		
	PML	ROB	RGMM	PML	ROB	RGMM	PML	ROB	RGMM	PML	ROB	RGMM
mean	0.0100	0.0103	0.0099	0.7986	0.7946	0.7898	0.0201	0.0205	0.0204	0.7947	0.8151	0.7995
median	0.0099	0.0101	0.0098	0.7996	0.7953	0.7905	0.0200	0.0205	0.0203	0.7960	0.8177	0.8018
MSE%	0.0028	0.0031	0.0034	0.0197	0.0271	0.0367	0.0002	0.0004	0.0003	0.5963	0.6687	0.6928
mean	0.0113	0.0114	0.0113	0.7281	0.7688	0.7580	0.0742	0.0320	0.0320	0.6857	0.7602	0.7393
median	0.0105	0.0113	0.0113	0.7336	0.7701	0.7589	0.0724	0.0312	0.0310	0.6685	0.7582	0.7329
MSE%	0.0192	0.0051	0.0056	0.8731	0.1483	0.2396	0.3264	0.0172	0.0172	4.7878	1.1301	1.5891
mean	0.0103	0.0103	0.0103	0.7980	0.7968	0.7895	0.0202	0.0176	0.0173	0.7758	0.6912	0.6843
median	0.0103	0.0101	0.0101	0.7991	0.7970	0.7902	0.0197	0.0175	0.0172	0.7646	0.6921	0.6823
MSE%	0.0060	0.0026	0.0027	0.0509	0.0255	0.0381	0.0012	0.0008	0.0009	2.5792	1.8753	2.1094

Table 3: Summary statistics based on 1,000 simulations for the PML estimator, the robust estimator  $\bar{a}$  ( $c = 4$ ) and the robust GMM estimator ( $c = 4$ ) of the AR(1)-ARCH(1) model under the reference model  $\mathbb{P}_{\theta_0}$  (first panel), the replacement model (22) for  $\eta = 5\%$ ,  $\xi_t \sim \mathcal{N}(0, 1)$  (second panel), the innovative outlier model (23) for  $\epsilon = 1\%$ ,  $\varrho = 5$  (third panel).

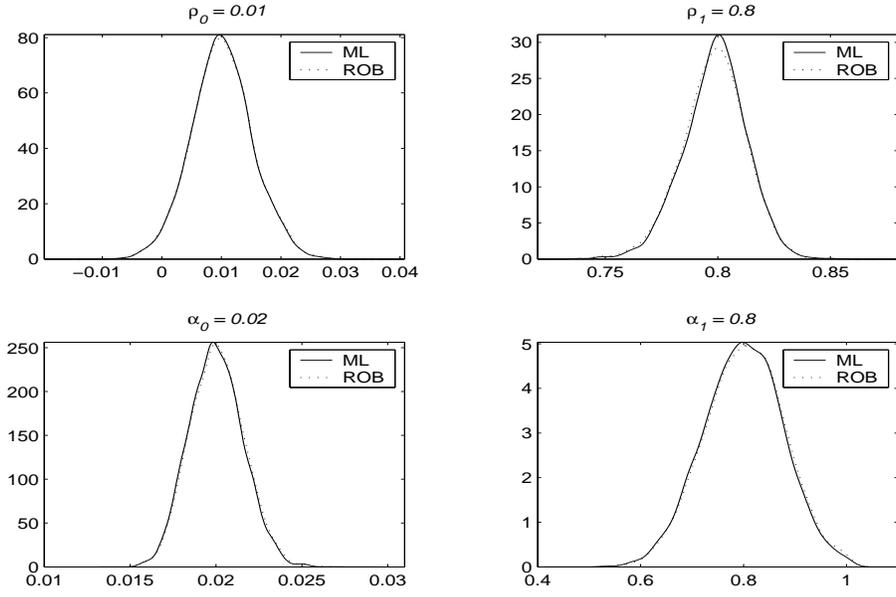


Figure 1: Estimated densities of  $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$  for the AR(1)-ARCH(1) process under the reference model  $\mathbb{P}_{\theta_0}$ , i.e. under Gaussian distribution for the innovations; cf. also the first panel of Table 1.

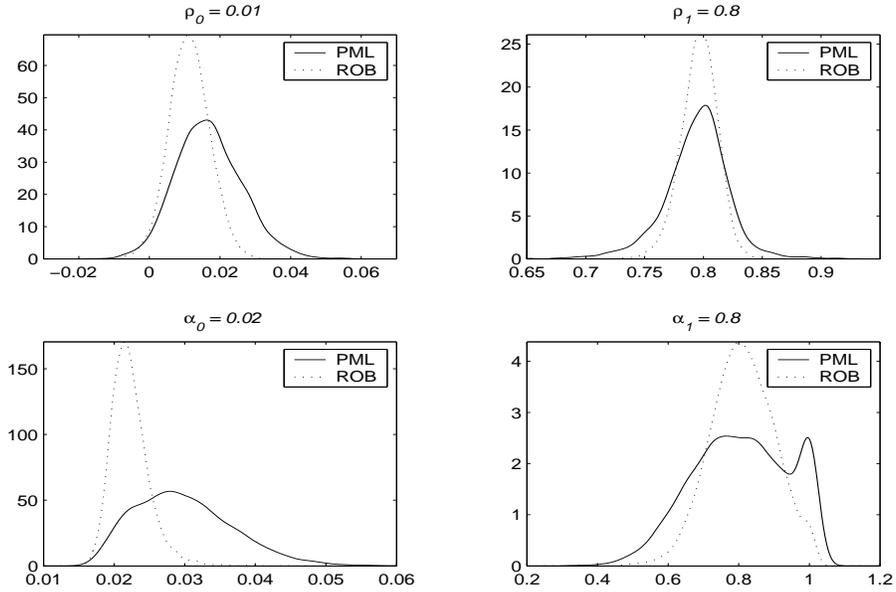


Figure 2: Estimated densities of  $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$  of the AR(1)-ARCH(1) process under the replacement model (22); cf. also the second panel of Table 1.

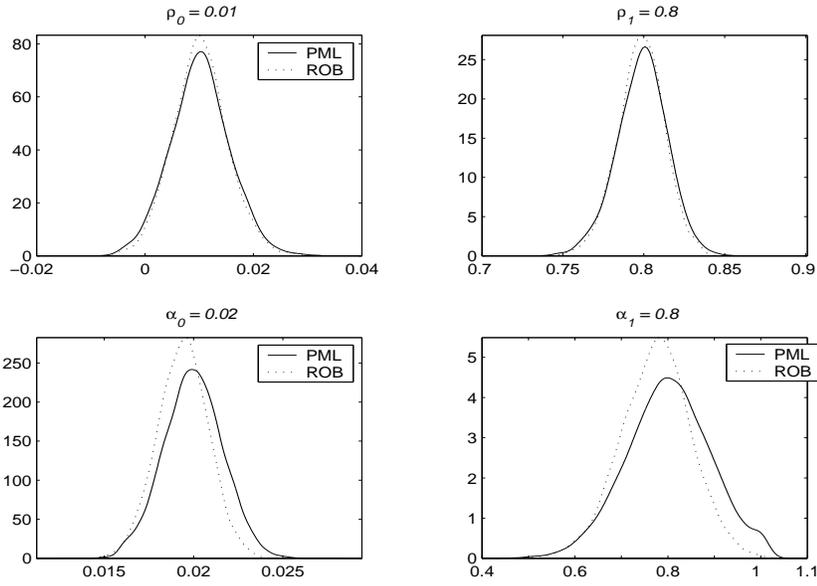


Figure 3: Estimated densities of  $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$  of the AR(1)-ARCH(1) process under the innovative outlier model (23); cf. also the third panel of Table 1.

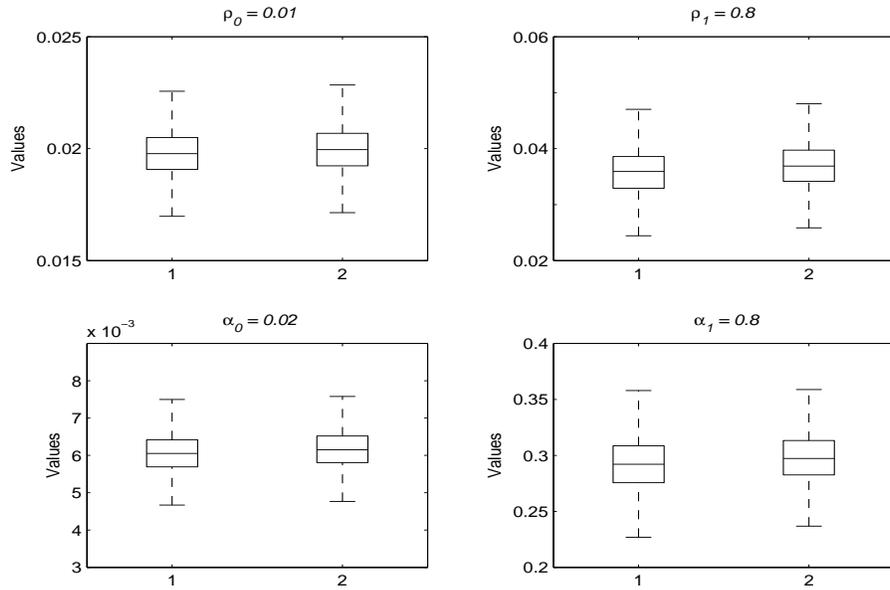


Figure 4: Boxplot of the lengths of ML (column 1) and robust (column 2) confidence intervals for  $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$  (cf. Figure 1) under the reference model  $\mathbb{P}_{\theta_0}$ .

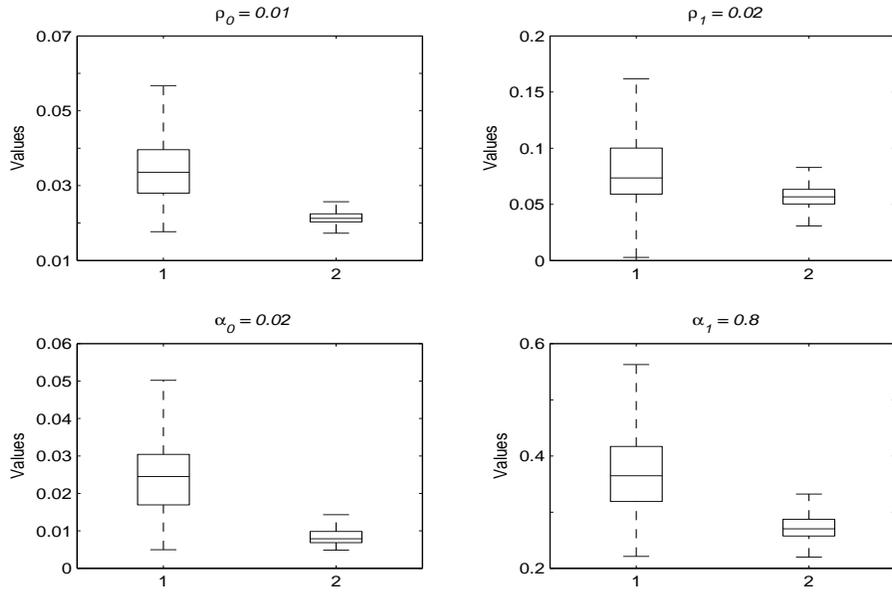


Figure 5: Boxplot of the lengths of PML (column 1) and robust (column 2) confidence intervals for  $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$  (cf. Figure 2) under the replacement model (22).

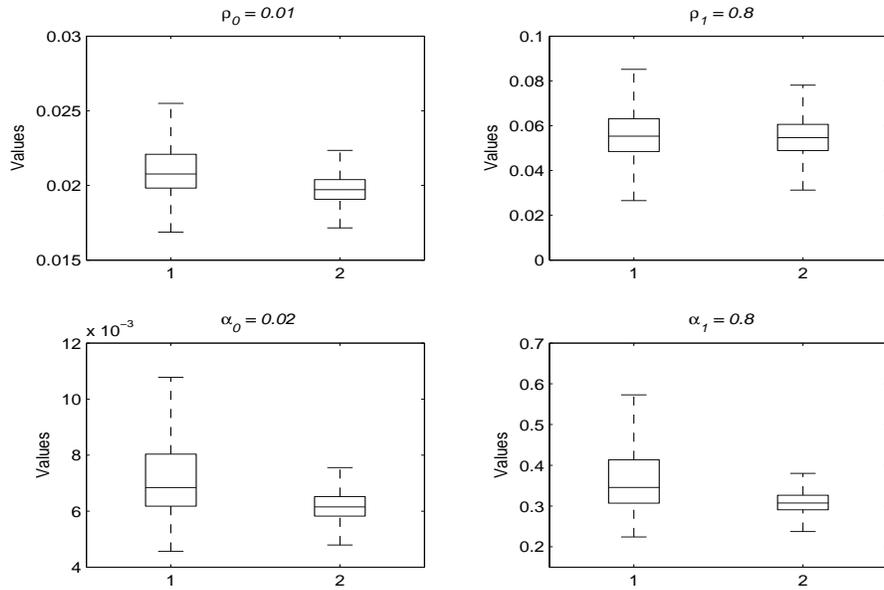


Figure 6: Boxplot of the lengths of PML (column 1) and robust (column 2) confidence intervals for  $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$  (cf. Figure 3) under the innovative outlier model (23).

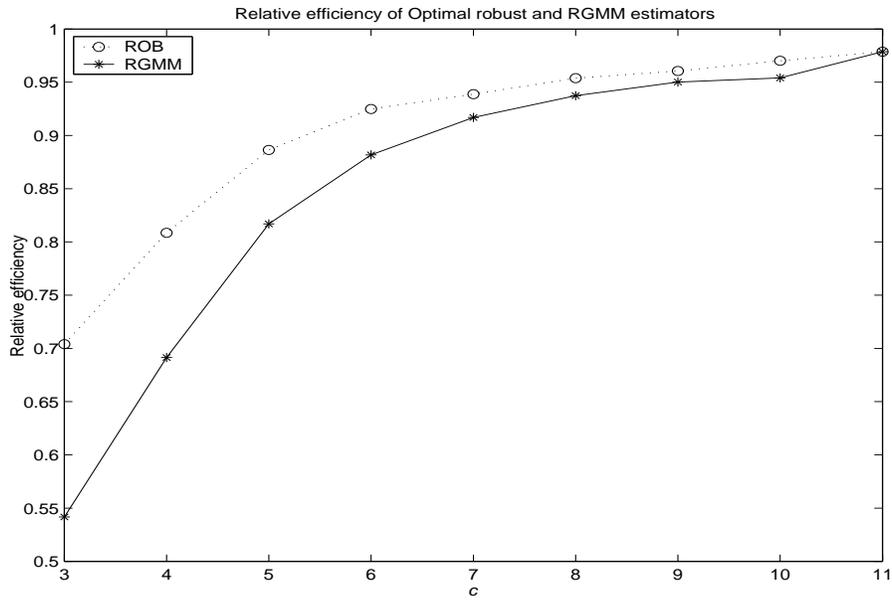


Figure 7: Relative efficiency under the reference model  $\mathbb{P}_{\theta_0}$  (relatively to the MLE) of the robust estimator  $\bar{a}$  and the robust GMME for different values of the robustness level  $c$ .

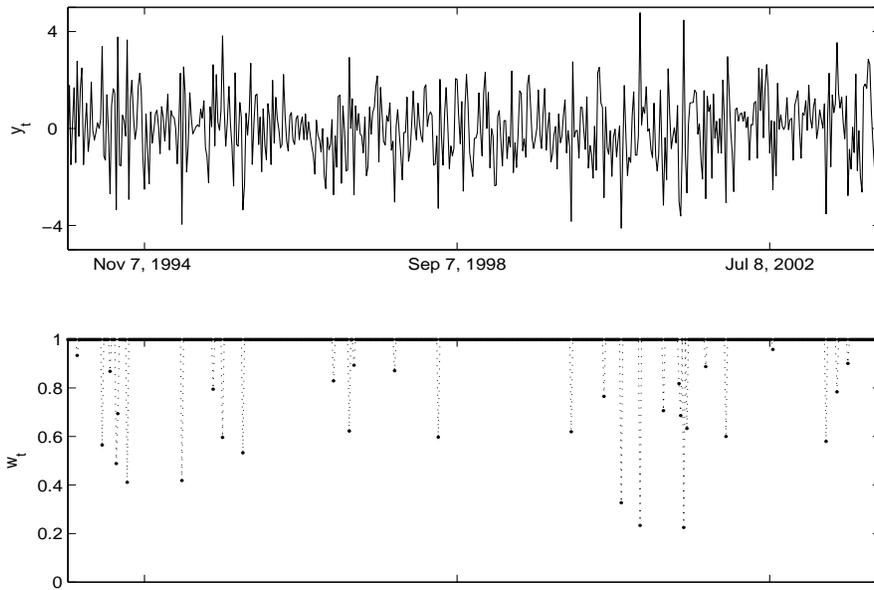


Figure 8: Weekly exchange rate returns of the Swedish krona versus the US dollar for the period 11/29/1993 until 11/17/2003 (top panel) and the weights implied by the robust estimate of the AR(1)-ARCH(1) model with  $c = 4$  (bottom panel).

## References

- [1] Bednarski, T. (1993), “Fréchet Differentiability of Statistical Functionals and Implications to Robust Statistics,” in *New Directions in Statistical Data Analysis and Robustness*, eds. S. Morgenthaler, E. Ronchetti, and W. A. Stahel, Basel: Birkhäuser, pp. 26–34.
- [2] Bollerslev, T. P. (1986), “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.
- [3] Bollerslev, T. P., and Wooldridge, J. M. (1992), “Quasi-Maximum Likelihood Estimates and Inference in Dynamic Models with Time Varying Covariances,” *Econometric Reviews*, 11, 143–172.
- [4] Bustos, O., and Yohai, V. J. (1986), “Robust Estimates for ARMA Models,” *Journal of the American Statistical Association*, 81, 155–168.
- [5] Clarke, B. R. (1983), “Uniqueness and Fréchet Differentiability of Functional Solutions to Maximum Likelihood-Type Equations,” *Annals of Statistics*, 11, 1196–1205.
- [6] ——— (1986), “Nonsmooth Analysis and Fréchet Differentiability of  $M$ -Functionals,” *Probability Theory and Related Fields*, 73, 197–209.
- [7] Engle, R. F. (1982), “Autoregressive Conditional Heteroskedasticity with Estimate of the Variance of United Kingdom Inflation,” *Econometrica*, 50, 987–1007.
- [8] Engle, R. F., Lilien, D. M., and Robins, R. P. (1987), “Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model,” *Econometrica*, 55, 391–407.
- [9] Fernholz, L. T. (1983), *Von Mises Calculus for Statistical Functionals*, New York: Springer.
- [10] Gagliardini, P., Trojani, F., and Urga, G. (2004), “Robust GMM Tests for Structural Breaks,” *Journal of Econometrics*, To appear.
- [11] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks,” *Journal of Finance*, 48, 1779–1801.

- [12] Gourieroux, C., and Monfort, A. (1989), “A General Framework for Testing a Null Hypothesis in a Mixed Form,” *Econometric Theory*, 5, 63–82.
- [13] Gourieroux, C., Monfort, A., and Trognon, A. (1984), “Pseudo Maximum Likelihood Methods: Theory,” *Econometrica*, 52, 681–700.
- [14] Hampel, F. R. (1968), “Contributions to the Theory of Robust Estimation,” Ph.D. thesis, University of California, Berkeley.
- [15] ——— (1974), “The Influence Curve and its Role in Robust Estimation,” *Journal of the American Statistical Association*, 69, 383–393.
- [16] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley.
- [17] Heritier, S., and Ronchetti, E. (1994), “Robust Bounded-Influence Tests in General Parametric Models,” *Journal of the American Statistical Association*, 89, 897–904.
- [18] Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.
- [19] Jensen, J. L. (1995), *Saddlepoint Approximations*, Oxford: Clarendon Press.
- [20] Koenker, R. W. (1982), “Robust Methods in Econometrics,” *Econometric Review*, 1, 213–255.
- [21] Koenker, R. W., and Bassett, G. (1978), “Regression Quantiles,” *Econometrica*, 46, 33–50.
- [22] Krasker, W. S., and Welsh, R. E. (1982), “Efficient Bounded-Influence Regression Estimation,” *Journal of the American Statistical Association*, 77, 595–604.
- [23] Krishnakumar, J., and Ronchetti, E. (1997), “Robust Estimators for Simultaneous Equations Models,” *Journal of Econometrics*, 78, 295–314.
- [24] Künsch, H. (1984), “Infinitesimal Robustness for Autoregressive Processes,” *Annals of Statistics*, 12, 843–863.

- [25] Künsch, H., Stefanski, L. A., and Carroll, R. J. (1989), “Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models,” *Journal of the American Statistical Association*, 84, 460–466.
- [26] Li, C. W., and Li, W. K. (1996), “On a Double Threshold Autoregressive Heteroscedastic Time Series Model,” *Journal of Applied Econometrics*, 11, 253–274.
- [27] Markatou, M., and Ronchetti, E. (1997), “Robust Inference: The Approach Based on Influence Functions,” in *Handbook of Statistics*, eds. G. S. Maddala and C. R. Rao, Amsterdam: North Holland, pp. 49–75.
- [28] Martin, R. D., and Yohai, V. J. (1986), “Influence Functionals for Time Series,” *Annals of Statistics*, 14, 781–818.
- [29] Muler, N., and Yohai, V. J. (1999), “Robust Estimates for ARCH Processes,” *Journal of Time Series Analysis*, 23, 341–375.
- [30] Ortelli, C., and Trojani, F. (2002), “Robust Efficient Method of Moments,” *Journal of Econometrics*, To appear.
- [31] Peracchi, F. (1990), “Robust  $M$ -Estimators,” *Econometric Review*, 9, 1–30.
- [32] Ronchetti, E., and Trojani, F. (2001), “Robust Inference with GMM Estimators,” *Journal of Econometrics*, 101, 37–69.
- [33] Sakata, S., and White, H. (1998), “High Breakdown Point Conditional Dispersion Estimation with Application to S&P 500 Daily Returns Volatility,” *Econometrica*, 66, 529–567.
- [34] Stefanski, L. A., Carroll, R. J., and Ruppert, D. (1986), “Optimally Bounded Score Functions for Generalized Linear Models with Applications to Logistic Regression,” *Biometrika*, 73, 413–425.
- [35] von Mises, R. (1947), “On the Asymptotic Distribution of Differentiable Statistical Functions,” *Annals of Mathematical Statistics*, 18, 309–348.