Optimally Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models

Loriano Mancini, Elvezio Ronchetti and Fabio Trojani

Institute of Finance, University of Southern Switzerland, Lugano, Switzerland
Dept. of Econometrics, University of Geneva, Switzerland

May 2003

Abstract

This paper studies the local robustness of estimation and testing procedures of the conditional location and scale parameters in a strictly stationary time series model. We first derive optimally bounded-influence estimators for the parameters of conditional location and scale models under a conditionally Gaussian reference model. Based on these results, optimally bounded-influence versions of the classical likelihood-based tests for parametric hypotheses are then obtained. We propose a feasible and efficient algorithm for the computation of our robust estimators, which makes use of some 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. Finally, we show by Monte Carlo simulation based on an AR(1)-ARCH(1) process that our estimators and tests maintain a very high efficiency under ideal model conditions and at the same time perform very satisfactorily under several forms of departures from a conditionally normal AR(1)-ARCH(1) process. On the contrary, classical Pseudo Maximum Likelihood estimators are found to be strongly biased and highly inefficient under such local model misspecifications.
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; see (1) below.

First, it characterizes the local robustness of inference procedures within such a model class by considering explicitly the features of the given time series setting. Second, it derives optimal robust estimation and inference procedures starting from the class of conditionally unbiased estimators for the parameters of the conditional location and scale equations. Finally, it offers easily computable numerical procedures for robust estimation in the given setting, which are only slightly more demanding than the ones required by a classical Pseudo Maximum Likelihood (PML, see Gourieroux, Monfort and Trognon (1984))) estimation. This avoids the simulation of multidimensional integrals – a typical task in the robust estimation of nonlinear models for time series – thereby largely reducing the computation time.

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 location models (like ARCH and GARCH models; Engle (1982) and Bollerslev (1986)) 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 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 we show below – an unbounded times series influence function (IF, Künsch (1984) and Hampel (1974)) for the implied asymptotic functional estimator. As a consequence, PML estimators for conditional location and scale models are not robust with respect to local departures from conditional normality. In this paper we derive inference procedures for the parameters of conditional location and scale models which are robust to local nonparametric misspecifications of a parametric, conditionally Gaussian, location and scale reference model. More specifically, we consider the class of robust, conditionally unbiased, \( M \)-estimators for the parameters of conditional location and scale models\(^1\) 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 approach proposed in Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001).

The need for robust procedures both in estimation and testing has been stressed by many authors and is now widely recognized both in the statistical and 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

\(^1\)As we show below, this class of estimators is convenient to develop robust inference procedures for the model settings considered in the paper.
Trojani (2001), Ortelli and Trojani (2002). However, the problem of robust estimation for the parameters of conditional location and scale models has been considered so far only by very few authors and only from the specific perspective of high breakdown estimation. Even less attention has been devoted so far to robust inference within conditional location and scale models. High breakdown estimators resistant to large amount of contamination in the sample data have been proposed by Sakata and White (1998) and Muler and Yohai (1999). Such estimators are highly robust and more useful at the exploratory stage than from an inference perspective where local deviations from a reference model are more meaningful. Moreover, these estimators are very computationally intensive and cannot be applied to estimate the parameters of a class of broadly applied models – such as threshold ARCH or ARCH in mean models\(^2\).

This paper derives optimally bounded-influence estimation and testing procedures for a general conditional location and scale time series model, which are computationally only slightly more demanding than the ones 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 of Künsch (1984) for the implied asymptotic functional estimator. This is a first necessary step which allows us in a second step to construct robust statistical procedures which can control for (i) the local asymptotic bias on the implied parameter estimates and (ii) the local asymptotic distortion on the level and the power of ML-type tests in the given setting.

Second, we derive the optimally bounded-influence estimator for the parameters of conditional location and scale models under a conditionally Gaussian reference model\(^3\). This extends the optimality result in Künsch (1984) and the application of optimally conditionally unbiased M-estimators in Künsch, Stefanski and Carroll (1989) to very general nonlinear second order dynamic models. Based on these results, optimally 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 the above optimal robust estimator, 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 data. Fisher consistency at the model of our robust estimator is preserved as usual 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 pure Robust Efficient Method of Moments (REMM, Ortelli and Trojani (2002)) setting. Using the conditional unbiasedness of our estimator we provide analytical Laplace approximations

\(^2\)More precisely, Muler and Yohai (1999) proposed a class of high breakdown estimators for pure ARCH processes (having zero conditional mean). Sakata and White (1998) proposed a class of high breakdown estimators for models where the parameter space can be partitioned according to the parameters arising in the conditionally location and scale equations, respectively. For these estimators, they also compute formally the corresponding (time series) breakdown point.

\(^3\)Optimality of our estimator is defined in the sense of Stefanski, Carroll and Ruppert (1986).
for such vectors which strongly reduce the necessary computation time by avoiding the simulation of multidimensional integrals.

Finally, we study by Monte Carlo simulation the efficiency and the robustness properties of our estimator. Precisely, we estimate a simple AR(1)-ARCH(1) process under several models of local contamination of a conditionally Gaussian process. Our findings show that under the Gaussian reference model the classical ML estimator and our robust estimator are practically equivalent from the efficiency point of view. On the contrary, in the presence of 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.

The structure of the paper is the following. Section 2 introduces the conditional location-scale model and the classical PML 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 by means of the IF. The optimal robust estimator is then derived and the optimality of robust inference procedures based on such an estimator is discussed. The section is concluded by deriving an analytic approximation for the auxiliary recentering vectors in our robust estimation and by presenting the algorithm that can be used to compute our robust estimator in applications. Section 4 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. Section 5 summarizes and concludes.

2 The Standard Setting

Let $\mathcal{Y} := (y_t)_{t \in \mathbb{Z}}$ be a real valued strictly stationary random sequence on the probability space $(\mathbb{R}^\infty, \mathcal{F}, \mathbb{P}_*)$.
We start by assuming that $\mathbb{P}_*$ belongs to some parametric model $\mathcal{P} := \{\mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$, i.e. $\mathbb{P}_* = \mathbb{P}_{\theta_0}$ for some $\theta_0 \in \Theta$. Precisely, we assume that $y_t$ satisfies the second order regression model

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

where $\mu_t(\theta_0)$ and $\sigma_t^2(\theta_0)$ parameterize, respectively, the conditional mean and the conditional variance of $y_t$ given the information $\mathcal{F}_{t-1}$ up to time $t - 1$. Therefore, under $\mathbb{P}_{\theta_0}$

\begin{align*}
E[\varepsilon_t(\theta_0)|\mathcal{F}_{t-1}] = E[\nu_t(\theta_0)|\mathcal{F}_{t-1}] = 0, \tag{2}
\end{align*}

for all $t \in \mathbb{Z}$. We denote by $y^m := (y_1, \ldots, y_m)$ the finite random sequence of $\mathcal{Y}$ and by $\mathbb{P}_*^m$ the restriction of $\mathbb{P}_*$ on the $\sigma$-algebra generated by $m$ process coordinates, i.e. the $m$-dimensional marginal distribution of $y^m_{1}$ induced by $\mathbb{P}_*$. Similarly, we define $y^{m+t}_{1+t} := (y_{1+t}, \ldots, y_{m+t})$ for all $t \in \mathbb{Z}$ and we denote by $y^{1}_{-\infty} := (\ldots, y_{-1}, y_0, y_1)$ the infinite random sequence of $\mathcal{Y}$ up to time 1.

\footnote{This correct specification assumption will be relaxed in Section 3.}

\footnote{In particular, this implies that the distribution of $(\varepsilon_t(\theta_0), \nu_t(\theta_0))|\mathcal{F}_{t-1}$ is also parameterized by $\theta_0$.}
Our aim is to develop robust inference procedures on model (1) for the case where the true distribution $P^*$ belongs to some nonparametric neighborhood of a parametric reference model $P_{\theta_0}$. In this case model (1) is regarded as an “approximate” description of the true data generating process $P^*$.

Model (1) covers a broad class of well-known parametric models for time series. Some examples are presented in the sequel.

**Example 2.1** AR(1) models assume

$$\begin{align*}
\mu_t(\theta_0) &= \rho y_{t-1}, \quad |\rho| < 1, \\
\sigma^2_t(\theta_0) &= \sigma^2.
\end{align*}$$

Robust inference procedures for linear auto-regressive processes have been well studied in robust statistics. Künsch (1984) defined a time series influence function (IF) in this context and derived an optimally bounded-influence estimator for the parameters of a general AR($p$) model. Martin and Yohai (1986) provided bounded-influence estimators for AR and MA models and studied the asymptotic bias implied by additive and innovative outliers. Finally, Bustos and Yohai (1986) proposed some robust estimators for the parameters of an ARMA model using robust estimators of the residuals autocovariances.

**Example 2.2** AR(1)-ARCH(1) models (cf. Engle (1982)) assume

$$\begin{align*}
\mu_t(\theta_0) &= \rho_0 + \rho_1 y_{t-1}, \\
\sigma^2_t(\theta_0) &= \alpha_0 + \alpha_1 (y_{t-1} - \rho_0 - \rho_1 y_{t-2})^2
\end{align*}$$

for $\rho_0 \in \mathbb{R}$, $|\rho_1| < 1$, $\alpha_0 > 0$ and $0 \leq \alpha_1 < 1$.\(^6\) 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), respectively. Sakata and White (1998) developed high breakdown estimators for conditional location and scale models that include (4) as a special case. Muler and Yohai (1999) considered the pure ARCH setting.

**Example 2.3** Double threshold AR(1)-ARCH(1) models with volatility asymmetries (see for instance Li and Li (1996) and Glosten, Jagannathan and Runkle (1993)) assume

$$\begin{align*}
\mu_t(\theta_0) &= \beta_0 + (\beta_1 + \beta_2 d_{1,t-1}) y_{t-1}, \\
\sigma^2_t(\theta_0) &= \alpha_0 + \alpha_1 d_{1,t-1} + (\alpha_2 + \gamma_0 d_{2,t-1})(y_{t-1} - \beta_0 - (\beta_1 + \beta_2 d_{1,t-2}) y_{t-2})^2
\end{align*}$$

with the dummy variable $d_{1,t-1} = 1$ if $\beta_0 + \beta_1 y_{t-1} > 0$ and zero otherwise, $d_{2,t-1} = 1$ if $\varepsilon_{t-1}(\theta_0) < 0$ and zero otherwise.\(^7\) To our knowledge, so far no robust estimators have been applied to estimate such models. In principle robust EMM procedures could be applied to this model class. However, they would be highly computationally intensive. By contrast, Sakata and White (1998) high breakdown estimators

---

\(^6\)This class is very popular in financial econometrics because it accounts for two important stylized facts of financial time series, namely, fat tails and volatility clustering of asset returns; cf. for instance Mandelbrot (1963) and Fama (1965).

\(^7\)These models are broadly used in financial applications because they account for the leverage effect, i.e. the stronger impact on volatility of bad news ($\varepsilon_{t-1}(\theta_0) < 0$) than good news ($\varepsilon_{t-1}(\theta_0) \geq 0$) when $\gamma_0 > 0$. 

---

5
cannot be applied directly to threshold ARCH models of the above form because they assume that the
parameter space $\Theta$ can be partitioned as $\Theta = \Theta_1 \times \Theta_2$ in order to imply $\mu_t(\theta_0) = \mu_t(\theta_1)$, for $\theta_1 \in \Theta_1$ and all $t \in \mathbb{Z}$.

**Example 2.4** GARCH(1,1) models (cf. Bollerslev (1986)) assume

\[
\begin{align*}
\mu_t(\theta_0) &= 0, \\
\sigma_t^2(\theta_0) &= \alpha_0 + \alpha_1 y_{t-1}^2 + \delta_1 \sigma_{t-1}^2(\theta_0) \\
&= \alpha_0 / (1 - \delta_1) + \alpha_1 \sum_{j=0}^{\infty} \delta_1^j y_{t-1-j}^2,
\end{align*}
\]

where $\alpha_0, \alpha_1, \delta_1 > 0$, $\alpha_1 + \delta_1 < 1$. The last equality for $\sigma_t^2(\theta_0)$ highlights that a GARCH model is indeed an ARCH model with an infinite number of lagged $y$ variables. Formally, high breakdown estimators for GARCH models are not available. Sakata and White (1998) conjecture that their estimators should have a relatively high breakdown point also for GARCH models under appropriate assumptions. On the other hand, robust EMM procedures can be applied to obtain bounded-influence estimators for these models.

Our goal is to derive efficient, locally robust estimators and testing procedures for the joint inference on the conditional mean and the conditional variance parameters of model (1)–(2), that do not require heavy computational methods.

In the sequel we first briefly review the classical Pseudo Maximum Likelihood estimation procedure of model (1)–(2). In a later section we will derive the optimal robust version (see Hampel (1974) and Stefanski, Carroll and Ruppert (1986)) of such estimators.

Under the assumption that $P_\star = P_{\theta_0} \in \mathcal{P}$, a Maximum Likelihood (ML) estimator of $\theta_0$ is available. This happens for instance when under $P_{\theta_0}$ the random variable $y_t$ has a conditionally Gaussian distribution, $y_t | \mathcal{F}_{t-1} \sim \mathcal{N}(\mu_t(\theta_0), \sigma_t^2(\theta_0))$. For many applications, the assumption of normality is not satisfied. Generalizations of ML then only assume that the dynamic equations (1) are satisfied by $y_t$ under $P_\star$, i.e. that the model is dynamically correctly specified. Let $P_0$ denote some probability distribution on $(\mathbb{R}^\infty, \mathcal{F})$ which implies the dynamic model (1)–(2) for $y_t$. When $P_\star = P_0$, ML estimators of $\theta_0$ are no longer available as the distribution of $\varepsilon_t(\theta_0)$ is unknown. Therefore, the conditional location and scale model (1)–(2) is estimated by Pseudo Maximum Likelihood (PML) (cf. Gourieroux, Monfort and Trognon (1984)) based on a conditionally Gaussian score function; cf. also Bollerslev and Wooldridge (1992) for GARCH-type models.

Specifically, such a PML estimator (PMLE) $\hat{\theta}_n^{pml}$ is the solution of the maximization problem

\[
\max_{\theta \in \Theta} \sum_{t=1}^n l(\hat{y}_t^i; \theta),
\]

where

\[
l(\hat{y}_t^i; \theta) := -\frac{1}{2} \ln \sigma_t^2(\theta) - \frac{1}{2} \frac{(\hat{y}_t - \mu_t(\theta))^2}{\sigma_t^2(\theta)}
\]
and $y^n_i := (y_1, y_2, \ldots, y_n)$ are sample observations of the process. Under $P_0$ and by means of standard regularity conditions, $I_n(\theta) := n^{-1} \sum_{i=1}^n l(\hat{y}^i_n; \theta)$ converges almost surely, uniformly on $\Theta$, to $I(\theta) := E_0[l(y; \theta)]$, where $E_0[\cdot]$ is the expectation under $P_0$ and either $y = y^n_i$ for some $m < \infty$ (as in Example 2.1, 2.2 and 2.3) or $y = y_1^\infty$ (as in Example 2.4). Such a distinction between the cases $m < \infty$ and $m = \infty$ entirely depends on the specification of $\mu_i(\theta_0)$ and $\sigma^2_1(\theta_0)$ in model (1).

As the Gaussian pseudo true density is in the class of quadratic exponential densities, $\hat{\theta}_{pml}^n$ is consistent under $P_0$. Precisely, consistency is implied by the unique maximum of $l(\theta)$ at $\theta_0$. Then, the following asymptotic first order condition holds

$$E_0[s(y_{-\infty}; \theta_0)] = 0,$$

where $s(y_{-\infty}; \theta) := \partial l(y_{-\infty}; \theta)/\partial \theta$ denotes the pseudo score function based on the Gaussian pseudo true density, i.e.

$$s(y_{-\infty}; \theta) = \frac{1}{\sigma^2_1(\theta)} \frac{\partial \mu_1(\theta)}{\partial \theta} \varepsilon_1(\theta) + \frac{1}{2\sigma^2_1(\theta)} \frac{\partial \sigma^2_1(\theta)}{\partial \theta} \left( \frac{\varepsilon^2_1(\theta)}{\sigma^2_1(\theta)} - 1 \right).$$

Notice that the asymptotic condition (7) is equivalent to equation (2) and (8). The PMLE of $\theta_0$ then solves the finite sample estimating equations

$$n^{-1} \sum_{i=1}^n s(\hat{y}^i_n; \hat{\theta}_{pml}^n) = 0,$$

which are the finite sample version of the asymptotic condition (7). Under the model $P_0$, the usual Taylor expansion of (9) implies that $\sqrt{n}(\hat{\theta}_{pml}^n - \theta_0)$ converges in distribution to the Gaussian distribution $N(0, V(s; \theta_0))$, where $V(s; \theta_0) := J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}$ and

$$J(\theta_0) := E_0 \left[ - \frac{\partial s(y_{-\infty}; \theta)}{\partial \theta} \right]_{\theta = \theta_0}, \quad I(\theta_0) := E_0 \left[ s(y_{-\infty}; \theta_0) s(y_{-\infty}; \theta)^T \right];$$


When $P_0 = P_{\theta_0}$ the conditional distribution of $\varepsilon_i(\theta_0)$ is truly Gaussian, $y_t|F_{t-1} \sim N(\mu_i(\theta_0), \sigma^2_1(\theta_0))$ and the PMLE is indeed the MLE of $\theta_0$. In this case, $I(\theta_0) = J(\theta_0)$ and $V(s; \theta_0)$ attains the Cramér-Rao lower bound $I(\theta_0)^{-1}$. Notice that the Gaussianity of $y_t$ only affects the efficiency of the PMLE and not the functional form of the estimating equation (9). Hence, the bias in the estimate of $\theta_0$ induced by contaminated distributions will have the same functional form for the PMLE and the MLE (and will be given by the influence function (18)).

It is well-known that PMLE’s are in the class of $M$-estimators since they are defined by the roots of the implicit equation (9). This class is very convenient to develop robust estimators; cf. Huber (1981) and Hampel et al. (1986). The next section proposes a class of robust $M$-estimators and an optimal robust version of the PMLE for time series models of the form (1)–(2).

---

8As usual, when $\mu_i(\theta)$ and/or $\sigma^2_1(\theta)$ depend on pre-sample values $y_0, y_{-1}, \ldots$, such values are replaced, for estimation purposes, by ‘in-sample estimates’.

9See for instance Gourieroux, Monfort and Trognon (1984) Appendix 1, Section 1.

3 Locally Robust Estimation and Inference

In this section we relax the assumption that the true underlying distribution \( P_* \) is given by the parametric distribution \( P_{\theta_0} \). Specifically, we allow \( P_* \) to be in a nonparametric neighborhood of the reference model specified by the parametric distribution \( P_{\theta_0} \) for \( Y \). For instance, the case \( P_* = P_0 \) falls in this setting when the distance between the error distribution under \( P_0 \) and \( P_{\theta_0} \) is “small” and the PMLE can be used for inference on \( \theta_0 \). However, as we will show, the PMLE is not robust and if \( P_* \) is indeed “slightly” different from \( P_0 \) the arising asymptotic bias can be very large. In general, we allow the model \( P_{\theta_0} \) to be dynamically misspecified with respect to \( P_* \).

Here, we are interested in robust \( M \)-estimators which are “resistant” to local deviations of \( P_* \) from the given reference model \( P_{\theta_0} \). We recall that under \( P_{\theta_0} \) one has \( \varepsilon_t(\theta_0)|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2(\theta_0)) \), where \( \varepsilon_t(\theta_0) \) and \( \sigma_t^2(\theta_0) \) are defined by (1). Hence, the reference distribution of \( Y \) is indirectly specified by the uncorrelated, standard Gaussian, random sequence \( (u_t(\theta_0))_{t \in \mathbb{Z}} \), where \( u_t(\theta_0) := \varepsilon_t(\theta_0) \sigma_t(\theta_0)^{-1} \).

In order to analyze the local stability properties of \( M \)-estimators, we introduce the following functional notation for the asymptotic value of an estimator \( a(\cdot) \) (cf. for instance Martin and Yohai (1986), p. 786),

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

where \( \mathcal{M}_{\text{stat}}^m := \{m\text{-dimensional marginals of strictly stationary processes}\} \) and \( m \in \mathbb{N}^* \). Then, the \( M \)-functional \( a(\cdot) \) is implicitly defined by some estimating function \( \psi : \mathbb{R}^m \times \Theta \longrightarrow \mathbb{R}^p \) such that

\[
E[\psi(y_1, \ldots, y_m; a(P_*^m))] = 0,
\]

where \( P_*^m \) is the restriction of \( P_* \) on \( m \) process coordinates. As already pointed out, the number of process coordinates \( m \) entirely depends on the parametric model (1) of interest. For instance, in Example 2.1, 2.2 and 2.3 we have \( m = 2, 3, 3 \), respectively, while in Example 2.4 \( m = \infty \). In robust statistics, only the finite dimensional case \( m < \infty \) has been well studied in the setting of linear AR models; cf. Künsch (1984). For the infinite dimensional case \( m = \infty \) Martin and Yohai (1986) proposed a specific influence function to study the asymptotic bias implied by a specific contamination, namely the replacement model, for the estimates of MA models.

In the rest of the paper, we consider the finite dimensional case \( m < \infty \). Hence, the asymptotic value estimates of \( \theta_0 \) is obtained by evaluating \( a(\cdot) \) at the \( m \)-dimensional distribution \( \mathcal{P}_{\theta_0}^m \). The reference model distribution is given by \( \mathcal{P}_{\theta_0}^m \). As we focus on local robustness we consider local deviations of \( \mathcal{P}_{\theta_0}^m \) from \( \mathcal{P}_{\theta_0}^m \). Therefore, we assume that \( \mathcal{P}_{\theta_0}^m \) is in the following nonparametric neighborhood \( \mathcal{U}(\mathcal{P}_{\theta_0}^m) \) of the reference model \( \mathcal{P}_{\theta_0}^m \),

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

The neighborhood defined in (11) is a simple way to formalize local perturbations11 of the model \( \mathcal{P}_{\theta_0}^m \).

11Notice that \( d_\varepsilon(P_{\eta}^m, P_{\theta_0}^m) \leq \eta \) for all \( G^m \in \mathcal{M}_{\text{stat}}^m \), where \( d_\varepsilon(\cdot, \cdot) \) denotes the Kolmogorov distance.
We recall that the reference model \( P_{\theta_0}^m \) needs not be the true model for \( Y \) but has rather to be interpreted as an “approximate” model for the true data generating process \( P_{\star}^m \), where the notion of “approximate” model is formalized implicitly in terms of “small” distributional distances.

Remark 3.1 The true underlying model \( P_{\star}^m \) is well approximated by some element of \( U^\eta(P_{\theta_0}^m) \) in all cases where we have \( \epsilon \) percent of contaminated sample data as (cf. Künsch (1984), p. 846)

\[
P_{\star}^m = (1 - \epsilon c(m)) P_{\theta_0}^m + \epsilon c(m) G^m + o(\epsilon),
\]

where \( c(m) := \partial p(m, \epsilon)/\partial \epsilon |_{\epsilon=0} \) and

\[
p(m, \epsilon) = P \{ \text{at least one outlier in a block of length } m \}.
\]

However, since \( p(m, 0) = 0 \) for all \( m \in \mathbb{N}^* \) and \( \lim_{m \to \infty} p(m, \epsilon) = 1 \) for all \( \epsilon > 0 \), we have \( \lim_{m \to \infty} c(m) = \infty \). Therefore, when \( m = \infty \), (12) cannot be applied\(^{12}\) to define a neighborhood for the reference distribution \( P_{\theta_0} \) of \( Y \). In general, any mixture distribution \( P_{\epsilon} = (1 - \epsilon) P_{\theta_0} + \epsilon G, \epsilon \in [0, 1] \), on an infinite dimensional state space does not correspond to any interesting distribution for time series models as it implies that any sample path of the series is generated either by \( P_{\theta_0} \) or by \( G \); cf. Martin and Yohai (1986), p. 791. One possible way to study the robustness of a statistic in this setting is to focus on a specific contamination model. For instance, additive outlier models have been studied in Martin and Yohai (1986) in the context of (linear) MA models. In this case, the “neighborhood” of the reference model \( P_{\theta_0} \) is implicitly defined by all possible distributions induced by the assumed contamination models.

In the next section we introduce the influence function for time series to characterize the asymptotic bias and the asymptotic variance of an \( M \)-functional estimator \( a(\cdot) \) defined by (10). This will motivate, in a second step, our optimal robust estimator for models of the form (1)-(2).

### 3.1 Time Series Influence Functions

Robust procedures aim at the estimation of the parameter \( \theta_0 \) in the reference model \( P_{\theta_0}^m \) when local deviations from such a reference model are allowed. Such deviations induce an asymptotic bias on the functional estimator \( a(\cdot) \) defined by\(^{13}\)

\[
\text{bias} := a(P_{\star}^m) - a(P_{\theta_0}^m) = a(P_{\star}^m) - \theta_0.
\]

For robust inference on \( \theta_0 \), the standard robustness condition on the corresponding estimator is a bounded asymptotic bias. In order to describe the linearized asymptotic bias of \( a(\cdot) \) induced by some model \( P_{\eta}^m \in U^\eta(P_{\theta_0}^m) \), one can consider the first order von Mises (1947) expansion of \( a(\cdot) \) at \( P_{\eta_0}^m \) (cf. for instance Fernholz (1983)),

\[
a(P_{\eta}^m) - a(P_{\eta_0}^m) = \eta a'(\theta_0, G^m) + o(\|P_{\eta}^m - P_{\eta_0}^m\|),
\]

\(^{12}\)This problem is related to the second ‘open question’ in Künsch (1984), p. 859.

\(^{13}\)We assume that \( a(\cdot) \) is Fisher consistent.
where $a'(\theta_0, G^m)$ is the Gâteaux derivative of $a(\cdot)$ in the direction $G^m - P^m_{\theta_0}$, i.e.

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

provided the limit exists. By contrast with a simple i.i.d. setting, in a time series framework $a'(\theta_0, G^m)$ is determined by a set of equivalent kernels. To characterize such kernels, we introduce the following concept from the theory of robust statistics; cf. Künsch (1984), p. 847.

**Definition 3.1** The influence function of the functional estimator $a(\cdot)$ is an equivalent class of kernels $IF : R^m \times \Theta \rightarrow R^p$ such that

$$a'(\theta, G^m) = \int_{R^m} IF(y; \theta) dG^m(y), \text{ for all } G^m \in M^m_{\text{stat}}.$$

We use the short notation $IF(y; \theta)$ for $IF(y; a, P_{\theta})$ as in Künsch (1984). In the one dimensional case, $m = 1$, the set of kernels is a singleton. Hence, in this case the influence function (IF) is unique and can be directly computed by taking functional Gâteaux derivatives of $a(\cdot)$ in the direction $\delta_x - P_{\theta_0}$, where $\delta_x$ is the Dirac mass at $x \in R$; cf. Hampel (1974) and Hampel et al. (1986).

In the $m$-dimensional case, $m \geq 2$, the IF can no longer be computed just by taking derivatives of $a(\cdot)$ in some suitable singular directions, namely because two different extremal measures in $M^m_{\text{stat}}$ are not mutually singular. Therefore, any differentiation in such directions gives a corresponding kernel conforming to Definition 3.1; cf. Künsch (1984), p. 847. As a consequence, in this case the kernel is not unique. Moreover, any function

$$IF(y_1, \ldots, y_m; \theta) + g(y_1, \ldots, y_{m-1}; \theta) - g(y_2, \ldots, y_m; \theta), \quad (14)$$

where $g : R^{m-1} \times \Theta \rightarrow R^p$ is an integrable function, is again a kernel satisfying Definition 3.1, because

$$E_{P^m_{\theta_0}}[g(y_1, \ldots, y_{m-1}; \theta) - g(y_2, \ldots, y_m; \theta)] = 0,$$

by the strictly stationarity of $G^m$.

Künsch (1984) introduced a natural additional condition on a kernel satisfying Definition 3.1, which determines a unique representant of the IF. This condition simply requires that, at the reference model $P^m_{\theta_0}$, $y_m|y_1, \ldots, y_{m-1}$ has no influence on the asymptotic bias of the estimator.

**Definition 3.2** The conditional influence function, $IF^{\text{cond}}(y^m_1; \theta_0)$, of the functional estimator $a(\cdot)$ is a kernel satisfying Definition 3.1, such that

$$E_{\theta_0}[IF^{\text{cond}}(y_1, \ldots, y_m; \theta_0)\mid \mathcal{F}_{m-1}] = 0,$$

where $E_{\theta_0}[\cdot]$ denotes the expectation with respect to the reference distribution $P^m_{\theta_0}$.

---

14For instance, for the functional estimator $a(\cdot)$ of the AR(1) model in Example 2.1, two different kernels are obtained when differentiating in the directions $(\delta_x(1), y(2) + \delta_{\nu(1), x(2)})/2 \text{ and } (\delta_x(1), y(2) + \delta_{\nu(1), x(2)} + \delta_{\nu(1), x(2)})/3$, respectively; cf. Künsch (1984), p. 847.
\(IF^{cond}\) is unique\(^{15}\) (cf. Künsch (1984), Th. 1.3) and has some desirable properties. First, under the reference model \(P_{\theta_0}^m\), it implies the following simple expression for the asymptotic covariance matrix \(V(\psi; \theta_0)\) of \(a(\cdot)\)

\[
V(\psi; \theta_0) := E_{\theta_0} [IF^{cond}(y_1^m; \theta_0) IIF^{cond}(y_1^m; \theta_0)^\top],
\]

which in application does not need to be estimated by a Newey-West (1987) covariance matrix estimator, because of the martingale difference property (15). Moreover, as we will show in the sequel, conditional IF for MLE’s has a natural interpretation in terms of the information matrix and a straightforward derivation; cf. Künsch (1984), (1.25).

Since under the model \(P_{\theta_0}\) the process \((IF^{cond}(y_{1+t}; \theta_0))_{t \in \mathbb{Z}}\) is a martingale difference sequence and the conditional mean-variance of \(y_t\) is correctly specified, it is natural to introduce the class of conditionally unbiased \(M\)-estimators for \(\theta_0\). These functional estimators \(a(\cdot)\) are implicitly defined by some function \(\psi: \mathbb{R}^m \times \Theta \rightarrow \mathbb{R}^p\), such that the conditional moment conditions

\[
E_{\theta_0} [\psi(y_1, \ldots, y_m; a(P_{\theta_0}^m))] | \mathcal{F}_{m-1} = 0
\]  

(16)

hold for a unique \(\theta_0 \in \Theta\). Such estimators have several desirable properties that will be presented below. Notice that by construction \((\psi(y_{1+t}, \ldots, y_{m+t}; \theta_0))_{t \in \mathbb{Z}}\) is also a martingale difference sequence under the model \(P_{\theta_0}\). Thus, by definition \(a(\cdot)\) is conditionally Fisher consistent and the asymptotic estimating equation for \(\theta_0\) is

\[
E_{\theta_0} [\psi(y_1, \ldots, y_m; a(P_{\theta_0}^m))] = 0.
\]  

(17)

The Gaussian score function (8) is an example of a function \(\psi\) defining a conditionally unbiased estimator of \(\theta_0\).

**Example 3.1** The PMLE for the AR(1)-ARCH(1) process in Example 2.2 is defined by the \(\psi\) function

\[
\psi(y_1^3; \theta_0) = -k_{1,3} + k_{2,3}\frac{y_3 - \rho_0 - \rho_1 y_2}{\alpha_0 + \alpha_1(y_2 - \rho_0 - \rho_1 y_1)^2} + k_{1,3}\frac{(y_3 - \rho_0 - \rho_1 y_2)^2}{\alpha_0 + \alpha_1(y_2 - \rho_0 - \rho_1 y_1)^2}
\]

where

\[
k_{1,3} := \frac{1}{\sigma_3^2} \begin{pmatrix} -2\alpha_1(y_2 - \rho_0 - \rho_1 y_1) \\ -2\alpha_1(y_2 - \rho_0 - \rho_1 y_1)y_1 \\ (y_2 - \rho_0 - \rho_1 y_1)^2 \end{pmatrix}, \quad k_{2,3} := \frac{1}{\sigma_3^3} \begin{pmatrix} 1 \\ y_2 \\ 0 \end{pmatrix}.
\]

More generally, the function \(\psi\) for the PMLE of model (1)–(2) has the functional form

\[
\psi(y_1^m; \theta_0) = -k_{1,m} + k_{2,m} u_m(\theta_0) + k_{1,m} u_m(\theta_0)^2,
\]

\(^{15}\)Under condition (15) the \(g\)-function in (14) is unique up to an additive constant.
where
\[ k_{1,m} := \frac{1}{2\sigma_m^4(\theta_0)} \left. \frac{\partial \sigma_m^2(\theta)}{\partial \theta} \right|_{\theta=\theta_0}, \quad k_{2,m} := \frac{1}{\sigma_m(\theta_0)} \left. \frac{\partial \mu_m(\theta)}{\partial \theta} \right|_{\theta=\theta_0}. \]

Therefore, 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 \( k_{2,m} \). Since \( E_{\theta_0}[u_m(\theta_0) | \mathcal{F}_{m-1}] = 0 \) and \( E_{\theta_0}[u_m(\theta_0)^2 | \mathcal{F}_{m-1}] = 1 \), \( \psi \) defines a conditionally unbiased estimator.

The main property of a conditionally unbiased \( M \)-estimator defined by (16) is that the corresponding conditional IF is computed as in the one dimensional case by calculating the limit
\[
IF^{\text{cond}}(x_1^m; \theta_0) := \lim_{\eta \to 0} \frac{a(1-\eta)P_{\theta_0}^m + \eta \delta_{x_1}^{(1)}\ldots,x_{(m)}) - a(P_{\theta_0}^m)}{\eta},
\]
where \( \delta_{x_1}^{(1)}\ldots,x_{(m)} \) is the Dirac mass at \( \{(y_1, \ldots, y_m) = (x_1^{(1)}, \ldots, x_{(m)})\} \) and provided the limit exists. Precisely, by defining \( P_{\eta}^m := (1-\eta)P_{\theta_0}^m + \eta \delta_{x_1}^{(1)}\ldots,x_{(m)} \) and implicitly differentiating equation (17) in the direction \( \delta_{x_1}^{(1)}\ldots,x_{(m)} - P_{\theta_0}^m \) one gets
\[
0 = \left. \frac{\partial}{\partial \eta} E_{P_{\eta}^m}[\psi(y; a(P_{\eta}^m))] \right|_{\eta=0} = \left. E_{P_{\theta_0}^m} \left[ \frac{\partial \psi(y; a(P_{\theta_0}^m))}{\partial a} \right] \right|_{\theta=\theta_0} \left. \frac{\partial}{\partial \eta} a(P_{\eta}^m) \right|_{\eta=0} + \int_{\mathbb{R}^m} \psi(y; a(P_{\theta_0}^m)) \left. \frac{\partial}{\partial \eta} P_{\eta}^m \right|_{\eta=0}
\]
and, as \( \left. \frac{\partial}{\partial \eta} P_{\eta}^m \right|_{\eta=0} = \delta_{x_1}^{(1)}\ldots,x_{(m)} - P_{\theta_0}^m \), using (17)
\[
IF^{\text{cond}}(x_1^m; a(P_{\theta_0}^m)) = D(\psi; a(P_{\theta_0}^m))^{-1} \psi(x_1^m; a(P_{\theta_0}^m)), \quad (18)
\]
where
\[
D(\psi; \theta_0) := -E_{P_{\theta_0}^m} \left[ \frac{\partial \psi(y; \theta)}{\partial \theta} \right] \left. \right|_{\theta=\theta_0}.
\]

As the conditional IF is unique and defines a martingale difference process, equation (18) is the only admissible representation. When the dependence of the conditional IF on the corresponding score function \( \psi \) has to be emphasized we use in the sequel the notation \( IF^{\text{cond}}_{\psi} \).

Notice that if condition (16) does not hold, the conditional IF is not given by (18), as the following simple example shows.

**Example 3.2** For an AR(1) process in Example 2.1, an unconditionally unbiased \( M \)-estimator for \( \theta_0 = \begin{pmatrix} \rho \\ \sigma^2 \end{pmatrix} \) can be defined by the following \( \psi \)-function
\[
\psi(y_1, y_2; \theta_0) = \begin{pmatrix} y_2y_1 - \sigma^2 \rho/(1 - \rho^2) \\ y_2^2 - \sigma^2/(1 - \rho^2) \end{pmatrix}.
\]
However, the conditional IF of such an estimator is not given by (18) because \( \psi \) is not a martingale difference.
A bounded conditional IF ensures a bounded linearized asymptotic bias induced by any contaminated distribution $P^m_\eta$ in the neighborhood $U^\eta(P^m_{\theta_0})$

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

and the derivative on the right hand side is uniformly bounded for any $P^m_\eta \in U^\eta(P^m_{\theta_0})$ when $m < \infty$. Moreover, since the conditional IF is linearly related to the $\psi$-function of the asymptotic estimating equation (17), it is bounded if and only if the $\psi$-function is bounded. As the Gaussian score function (8) is unbounded (at least) in $\varepsilon_1(\theta)$, PMLE’s based on such a score function are not robust. For MLE’s the conditional IF is given by

$$IF^{cond}(y_1^m; \theta_0) = I(\theta_0)^{-1}s(y_1^m; \theta_0),$$

where $I(\theta_0)$ is the information matrix. For instance, the conditional IF of the MLE in Example 2.1 is given explicitly by

$$IF^{cond}(x_1, x_2; \theta_0) = \begin{bmatrix} (1 - \rho^2)^{-1} & 0 \\ 0 & 0.5\sigma^{-4} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{-2}x_2 - \rho x_1 \\ -0.5\sigma^{-2} + 0.5\sigma^{-4}(x_2 - \rho x_1)^2 \end{bmatrix}.$$ 

For a contaminated distribution $P^2_\eta = (1 - \eta)P^2_{\theta_0} + \eta G^2$ we can then compute the implied linearized asymptotic bias as

$$a(P^2_\eta) - a(P_{\theta_0}) = \eta \begin{bmatrix} (1 - \rho^2)^{-1} & 0 \\ 0 & 0.5\sigma^{-4} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{-2}(\zeta^{(1,2)} - \rho\zeta^{(2)}) \\ -0.5\sigma^{-2} + 0.5\sigma^{-4}((1 + \rho^2)\zeta^{(2)} - 2\rho\zeta^{(1,2)}) \end{bmatrix},$$

where

$$\zeta^{(1,2)} := E_{G^2}[y_1y_2] \quad \text{and} \quad \zeta^{(2)} := E_{G^2}[y_1^2] = E_{G^2}[y_2^2] \quad \text{because} \quad G^2 \in \mathcal{M}_{stat}^2.$$ 

Hence, the asymptotic bias can be arbitrarily large, depending on the values of $\zeta^{(1,2)}$ and $\zeta^{(2)}$ on the neighborhood $U^\eta(P^2_{\theta_0})$. In Figure 1 we plot $\|V(s; \theta_0)^{-1/2} IF^{cond}(x_1, x_2; \theta_0)\|$, that is the self-standardized norm of the IF of the MLE for $\theta_0$ under $P^2_{\theta_0}$ (cf. also Section 3.2 below), as a function of $x_1$ and $x_2$ for the parameter choice $\rho = 0.8$ and $\sigma^2 = 2$. In the left and right region of the grid (where, respectively, $x_1$ is ‘low’, $x_2$ is ‘high’ and vice versa), the self-standardized sensitivity is clearly unbounded denoting high sensitivity of the MLE to possible “jumps” in the data.\textsuperscript{16}

As already mentioned, a bounded conditional IF ensures a bounded asymptotic bias for any distribution in the neighborhood $U(P^m_{\theta_0})$ of the reference model $P^m_{\theta_0}$. In the next section we define such a robust estimator for the parameters of model (1).

\textsuperscript{16}In this example the most robust estimator has a self standardized sensitivity no larger than 1; cf. Hampel et al. (1986), p. 228.
3.2 Optimal robust estimator

In the following we derive the optimally bounded-influence estimator (see Proposition 3.1 and Corollary 3.1) for models of the form (1)–(2), under the conditionally Gaussian reference model \( P_{\theta_0}^m \), in the class of conditionally unbiased \( M \)-estimators for \( \theta_0 \).

The most common approach to derive bounded-influence estimators is to impose a bound on the ‘self-standardized sensitivity’ \( \gamma \) of the estimator (cf. for instance Krasker and Welsch (1982)) defined by

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

where \( V(\psi; \theta_0) = E_{\theta_0}[IF_{\psi}^{\text{cond}}(z; \theta_0) IF_{\psi}^{\text{cond}}(z; \theta_0)^\top] \). This sensitivity measure has some desirable properties. Firstly, the bound on \( \gamma \) does not depend on the scaling of observations, a feature that improves the algorithm convergence. Secondly, the maximal bias for level and power of several ML-type tests can be controlled by imposing a bound on \( \gamma \); cf. Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001).

By definition, nonrobust 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.

3.2.1 Optimality results

Under a conditionally Gaussian reference model \( P_{\theta_0}^m \), the MLE for \( \theta_0 \) is most efficient but not robust. Then, robustness can be ensured only by “paying” a small loss in efficiency at the reference model \( P_{\theta_0}^m \).

Here we propose an estimator of \( \theta_0 \) that achieves the same optimality result within the class of conditionally unbiased \( M \)-estimators as in Künsch, Stefanski and Carroll (1989).

Consider the functional estimator \( \hat{\pi}(\cdot) \) of \( \theta_0 \) implicitly defined by

\[
E_{\theta_0}[\psi_c(y_m^m; \pi(P_{\theta_0}^m))] = 0, \tag{19}
\]

where \( \psi_c(y_1^m; \theta) := A(\theta) \psi_{bf}(y_1^m; \theta), \)

\[
\psi_{bf}(y_1^m; \theta) := (s(y_1^m; \theta) - \tau(y_1^{m-1}; \theta)) w(y_1^m; \theta), \tag{20}
\]

and \( w(y_1^m; \theta) := \min(1, \ c \| A(\theta) (s(y_1^m; \theta) - \tau(y_1^{m-1}; \theta)) \|^{-1}) \). The non singular 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, respectively, by solving the implicit equations

\[
E_{\theta_0}[\psi_c(y_1^m; \theta_0) \psi_c(y_1^m; \theta_0)^\top] = I, \tag{21}
\]

and

\[
E_{\theta_0}[\psi_c(y_1^m; \theta_0)|\mathcal{F}_{m-1}] = 0. \tag{22}
\]

We will discuss below the interpretation and the computation of the \( A \) matrix and the \( \tau \)-vectors. Notice that the estimating function \( \psi_c \) (or the unscaled version \( \psi_{bf} \)) is conditionally unbiased at the reference
model and is a truncated version of the ML score as, by construction, \( \|\psi_c(y_1^m; \theta)\| \leq c \). Moreover, as \((\psi_c(y_{1+t}^m; \theta_0))_{t \in \mathbb{Z}}\) is a martingale difference sequence under \(P_{\theta_0}\), the conditional IF of the functional estimator \(\bar{\pi}(\cdot)\) is given by (18),

\[
IF_{\psi_c}^{\text{cond}}(y_1^m; \bar{\pi}(P_{\theta_0}^m)) = D(\psi_c; \theta_0)^{-1} \psi_c(y_1^m; \bar{\pi}(P_{\theta_0}^m)).
\]

The estimating function \(\psi_{\text{bif}}\) satisfies the following optimality criterion.

**Proposition 3.1** If for a given constant of \(c \geq \sqrt{p}\), equations (21) and (22) have solutions \(A(\theta_0)\) and \(\tau(y_1^{m-1}; \theta_0)\), respectively, then \(\psi_{\text{bif}}\) minimizes \(\text{tr}(V(\psi; \theta_0) V(\psi_{\text{bif}}; \theta_0)^{-1})\) among all \(\psi\) satisfying (16) and

\[
\sup_{z \in \mathbb{R}^m} \left( IF_{\psi}^{\text{cond}}(z; \theta_0) \right) \left( V(\psi_{\text{bif}}; \theta_0)^{-1} IF_{\psi}^{\text{cond}}(z; \theta_0) \right)^{1/2} \leq c. \tag{23}
\]

Up to multiplication by a constant matrix, \(\psi_{\text{bif}}\) is unique almost surely.

Any score function \(\psi_{\text{opt}}\) such that \(V(\psi; \theta_0) = V(\psi_{\text{opt}}; \theta_0)\) is positive semi-definite for all \(\psi\) satisfying (16) is called strongly efficient and the following corollary holds; cf. also Stefanski et. al. (1986), p. 416.

**Corollary 3.1** If there exists an unbiased, strongly efficient score function \(\psi_{\text{opt}}\) satisfying \(\gamma(\psi_{\text{opt}}) \leq c < \infty\), then \(\psi_{\text{opt}}\) is equivalent almost surely to \(\psi_{\text{bif}}\) whenever the latter is defined.

Proofs are given in Appendix B. Under standard conditions\(^\text{17}\), the optimal robust estimator \(\bar{\pi}(\cdot)\) is consistent and asymptotically normally distributed at the reference model \(P_{\theta_0}^m\) with an asymptotic covariance matrix given by

\[
V(\psi_c; \theta_0) = E_{\theta_0} \left[ \frac{\partial}{\partial \theta} \psi_c(y_1^m; \theta_0) \right]^\top E_{\theta_0} \left[ \frac{\partial}{\partial \theta} \psi_c(y_1^m; \theta_0) \right]^{-1} = D(\psi_c; \theta_0)^{-1} D(\psi_c; \theta_0)^{-\top}.
\]

Notice that, under the reference model \(P_{\theta_0}^m\), the simple expression for \(V(\psi_c; \theta_0)\) is implied by condition (22) and the Newey-West (1987) covariance matrix estimator is not necessary. We point out that, although \((s(y_{1+t}^m; \theta_0))_{t \in \mathbb{Z}}\) is a martingale difference sequence under the reference model \(P_{\theta_0}^m\), not any truncated version of the Gaussian score function \(s\) is a martingale difference sequence. Such a property has to be explicitly imposed as in condition (22).

### 3.2.2 Computation of \(A\) and \(\tau\)

The \(A\) matrix is a scaling matrix ensuring that the upper bound \(c\) on the function \(\psi_c\) is also the upper bound on the self-standardized influence function. Indeed, under the scaling condition (21), \(V(\psi_c; \theta_0)^{-1} = D(\psi_c; \theta_0)^\top D(\psi_c; \theta_0)\) and

\[
\|V(\psi_c; \theta_0)^{-1/2} IF_{\psi_c}^{\text{cond}}(y; \theta_0)\|^2 = IF_{\psi_c}^{\text{cond}}(y; \theta_0)^\top V(\psi_c; \theta_0)^{-1} IF_{\psi_c}^{\text{cond}}(y; \theta_0)
\]

\[
= \psi_c(y; \theta_0)^\top D(\psi_c; \theta_0)^{-\top} V(\psi_c; \theta_0)^{-1} D(\psi_c; \theta_0)^{-1} \psi_c(y; \theta_0)
\]

\[
= \|\psi_c(y; \theta_0)\|^2.
\]

\(^{\text{17}}\)See for instance Heritier and Ronchetti (1994), Appendix A.2.
Hence, the self-standardized sensitivity of $\pi(\cdot)$ is equal to the Euclidian norm of the robust score function $\psi_c$, which is bounded by $c$. The $A$ matrix is computed by a simple iterative procedure given in Section 3.3, where we propose an algorithm to compute the optimal estimator (19)–(22).

Further, to satisfy the conditional Fisher consistency condition (22), each truncated score function has to be shifted by some corresponding $\psi$.

Therefore, in virtually all cases relevant for this paper, such moments are unknown. Therefore, in this case the computation of (24) requires computing some unconditional moments under $m$-dimensional integrals by Monte Carlo simulation. Indeed, in the unconditionally unbiased case, the centering $\tau$-vector defining an unconditionally unbiased robust $M$-estimators is implicitly defined by the condition

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

(24)

In general, the expectations involved in (24) cannot be expressed analytically and numerical procedures must be applied to compute (24). However, this makes robust estimation procedures of models of the form (1) very time consuming. For the conditionally Gaussian reference model $P_{\theta_0}^m$ we therefore provide a simple and easy implementable accurate approximation for (24). Under the reference distribution $P_{\theta_0}^m$ and for models of the form (1)–(2), we provide a simple and easy implementable accurate approximation for $\tau(y_1^{m-1}; \theta)$.

In the following we briefly explain the procedure; detailed calculations are given in Appendix A. We proceed in two steps.

In the first step, 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 $u_m(\theta_0)$,

$$0 = \|A(\theta_0) \left( s(y_1^m; \theta_0) - \tau^{(0)} \right) \|^2 - c^2$$

$$:= \|A(\theta_0) \left( -k_{1,m} + k_{2,m} u_m(\theta_0) + k_{1,m} u_m^2(\theta_0) - \tau^{(0)} \right) \|^2 - c^2,$$

where

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

By the continuity of the mapping $\tau(y_1^{m-1}; \theta) \rightarrow (s(y_1^m; \theta) - \tau(y_1^{m-1}; \theta) w(y_1^m; \theta)$ and by the mean value theorem.

19In the i.i.d. setting the class of conditionally unbiased $M$-estimators is equivalent to the unconditional ones and $\tau(y_1^{m-1}; \theta_0) \equiv \tau(\theta_0)$ as $E_{\theta_0}[|F_{m-1}|] = E_{\theta_0}[\cdot]$.  

20Since the $y$ variables are not i.i.d., sufficiently long simulations are necessary in order to adequately capture the time dependence in the $Y$ process, such as in the RGMM estimator case in Ronchetti and Trojani (2001).
Notice that \( k_{1,m} \) and \( k_{2,m} \) are \( F_{m-1} \)-measurable. This allows to solve equation (26) with respect to \( u_m(\theta_0) \). In almost all simulations and all empirical estimations equation (26) had only two real roots. Therefore, we only consider that case for brevity. The case of four real roots is discussed in Appendix A.

In the second step, we ‘split’ the integrals in (24) according to the roots determined by (26). Precisely, denoting the roots by \( \underline{u}_m \) and \( \overline{u}_m \), with \( \underline{u}_m \leq \overline{u}_m \), the denominator in (24) is given by

\[
E_{\theta_0} [w(y^n_1; \theta_0)|F_{m-1}] = \int_{-\infty}^{\underline{u}_m} \frac{c}{\|A(\theta_0)(s(v; \theta_0) - \tau(0))\|} d\Phi(u) + [\Phi(\overline{u}_m) - \Phi(\underline{u}_m)] + \int_{\overline{u}_m}^{+\infty} \frac{c}{\|A(\theta_0)(s(v; \theta_0) - \tau(0))\|} d\Phi(u),
\]

(27)

where \( v := (y_1, \ldots, y_{m-1}, \mu_m(\theta_0) + \sigma_m(\theta_0) u) \) and \( \phi(\cdot) \) and \( \Phi(\cdot) \) denote the standard Gaussian density and cumulative function, respectively. Typical values of \( \overline{u}_m \) range from 2.7 to 3.5 (the opposite for \( \underline{u}_m \)), so that both \( \underline{u}_m \) and \( \overline{u}_m \) are quite far in the tails of a standard normal distribution. Therefore, the ‘main contribution’ to the expectation on the left hand side of (27) comes from the term in the square brackets\(^{21}\). Since the integrals on the right hand side of (27) are ‘symmetric’, we can consider only the integral on the right tail. As mentioned, full analytical solutions are not available. However, since \( \overline{u}_m \) is ‘quite far’ in the right tail of a standard Gaussian distribution, the integral can be well approximated using the Laplace’s method; cf. for instance Jensen (1995), Th. 3.1.1. This gives

\[
\int_{\overline{u}_m}^{+\infty} \frac{c}{\|A(\theta_0)(s(v; \theta_0) - \tau(0))\|} d\Phi(u) \bigg|_{\theta_0 = \theta_0} = \int_{\overline{u}_m}^{+\infty} q_d(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) du = \frac{1}{\sqrt{2\pi}} \exp(-0.5\overline{u}_m^2) \frac{1}{\overline{u}_m} \left( \overline{q}_d(0) + \overline{q}_d'(0) \right) + O \left( \frac{1}{\overline{u}_m^3} \right)
\]

(28)

with the obvious notation for \( q_d(u) \) and where \( \overline{q}_d(z) := q_d(\overline{u}_m + z) \exp(-0.5z^2) \).\(^{22}\)

The integral in the numerator of (24) is split in the same way as in (27). Then the integrals on the tails are again approximated using the Laplace method. Detailed calculations are presented in Appendix A. The resulting formula for the computation of \( \tau \) is given in the next proposition.

**Proposition 3.2** Given the second order regression model (1)–(2) and the conditionally Gaussian reference model \( P^{\mu_m}_{\theta_0} \), if the quartic equation (26) has only two real roots \( \underline{u}_m \leq \overline{u}_m \), then

\[
\tau(y^{n-1}_1; \theta_0) = \frac{-L_n(\underline{u}_m) - k_{1,m} [\Phi(\overline{u}_m) - \Phi(\underline{u}_m)] + k_{2,m} M_{1,m} + k_{1,m} M_{2,m} + L_n(\overline{u}_m) + O \left( \frac{1}{\underline{u}_m^3} \right) + O \left( \frac{1}{\overline{u}_m^3} \right)}{\overline{q}_d(0) + \overline{q}_d'(0) + \left| L_d(\overline{u}_m) \right|},
\]

where \( M_{1,m} := \phi(\overline{u}_m) - \phi(\overline{u}_m) \), \( M_{2,m} := \underline{u}_m \phi(\underline{u}_m) - \overline{u}_m \phi(\overline{u}_m) + \Phi(\overline{u}_m) - \Phi(\underline{u}_m) \). \( L_n(\cdot) \) and \( L_d(\cdot) \) are defined in Appendix A and denote, respectively, the Laplace’s approximations for the integrals in the numerator and in the denominator of (24).

---

\(^{21}\)The Gaussian density \( \phi(u) \) is roughly zero for \( u \not\in (-4, 4) \) and \( c \|A(\theta_0)(s(v; \theta_0) - \tau(0))\|^{-1} \rightarrow 0 \) for \( u \rightarrow \pm\infty \).

\(^{22}\)For the Laplace’s approximation of the integral on the left tail replace \( \overline{u}_m \) by \( \underline{u}_m \) and \( + \) by \( - \) in (28).
We recall that Proposition 3.2 is available because \( \tau(y_1^{m-1}; \theta) \), \( k_{1,m} \) and \( k_{2,m} \) are \( \mathcal{F}_{m-1} \)-measurable. This allows to ‘split’ the integrals involved in (24) and to approximate the Gaussian integrals.

**Remark 3.2** Intuitively, the real roots \( u_m \) and \( \pi_m \) in equation (26) determine the range where the standardized innovation \( u_m(\theta_0) \) is ‘not influential’ (in terms of self-standardized sensitivity of \( \pi(\cdot) \)) for the arising asymptotic bias because

\[
\|A(\theta_0) (s(y_1^m; \theta_0) - \tau^{(0)})\| \leq c, \iff u_m(\theta_0) \in [u_m, \pi_m],
\]

\[
> c, \iff u_m(\theta_0) \in (-\infty, u_m) \cup (\pi_m, +\infty),
\]

and the self-standardized sensitivity of the functional estimator \( \pi(\cdot) \) is equal to the Euclidian norm of the \( \psi_c \)-function.

**Example 3.3** For the AR(1)-ARCH(1) process of Example 2.2 and the parameters \( \rho_0 = 0.01, \rho_1 = 0.8, \alpha_0 = 0.02 \) and \( \alpha_1 = 0.8 \) Figure 2 shows \( \|A(\theta_0) s(y_1^2; \theta_0)\| \), i.e. the self-standardized sensitivity of the PMLE, as a function of \( u_3(\theta_0) \) and for some values of \( y_1 \) and \( y_2 \). Clearly, \( \|A(\theta_0) s(y_1^2; \theta_0)\| \) is unbounded and increases very rapidly for some values of \( y_1 \) and \( y_2 \). For the PMLE (\( \tau^{(0)} = 0 \) and \( w(y_1^2; \theta_0) = 1 \)), the real roots in equation (26) are given by the intersections of \( \|A(\theta_0) s(y_1^2; \theta_0)\| \) with a constant line at some given value of \( c \). For instance, the most robust estimator has \( c = 2 \) and even in such a case equation (26) has two real roots for the two cases depicted in Figure 2. In our Monte Carlo Simulations in Section 4 we consider a robust estimator of the AR(1)-ARCH(1) model with \( c = 9 \).

### 3.2.3 Robust testing procedures

The robust estimator defined in (19)–(22) allows us to derive the optimal robust version of several ML-type tests to control the maximal bias on the level and the power of the tests induced by local distributional misspecifications of a null or an alternative hypothesis. Precisely, the robust versions of the classical Wald, score and likelihood ratio tests based on the robust estimator in (19)–(22) 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. In the following, we present the robust testing procedure and we discuss in detail only the robust version of the Wald test. The score and likelihood ratio test can be similarly derived.

Consider a general null hypothesis of the form

\[
g(\pi(P_{\theta_0}^m)) = 0, \tag{29}
\]

for a smooth function \( g : \Theta \rightarrow \mathbb{R}^r \) such that \( (\partial/\partial \theta) g(\pi(P_{\theta}^m)) \) is of full column rank for all \( \theta \in \Theta \). The aim is to analyze the local stability properties of a ML-type test based on the robust estimator defined in Section 3.2.1. We consider test statistics \( nQ \) that are quadratic form of a functional \( U \),

\[
nQ(P_{\theta_n}^m) := n U(P_{\theta_n}^m)^\top U(P_{\theta_n}^m), \quad n \in \mathbb{N}^*, \tag{30}
\]
where $P^m_{\theta_0}$ is the empirical $m$-dimensional distribution of the observations $\tilde{y}_1, \ldots, \tilde{y}_n$. The functional $U$ associated with the Wald test is

$$U^W(P^m_{\theta_0}) := \left[ \frac{\partial g(\theta)}{\partial \theta} V(\psi_c; \theta) \frac{\partial g(\theta)}{\partial \theta} \right]^{-1/2} g(p(P^m_{\theta_0})).$$

To apply the methodology in Heritier and Ronchetti (1994), we assume

$$\sqrt{n} \left( p(P^m_{\theta_0}) - p(P^m_{\eta(n)}) \right) \rightarrow \mathcal{N}(0, V(\psi; \theta_0)), \quad n \rightarrow \infty$$

in distribution, uniformly over the sequence $(U^m(\epsilon, n)(P^m_{\theta_0}))_{n \in \mathbb{N}}$ of $\eta(\epsilon, n)$-neighborhoods of $P^m_{\theta_0}$ defined in (11) by setting $\eta := \eta(\epsilon, n) = \epsilon / \sqrt{n}$ and $G \in \text{dom}(\pi)$. Then, the following Proposition holds.

**Proposition 3.3** Let $\tilde{a}$ be the robust estimator defined by (19)–(22) and denote by $\alpha$ the level functional of the test based on the functional $Q(\cdot)$ in (30). Let further $(P^m_{\eta(n)})_{n \in \mathbb{N}}$ be a sequence of $\eta(\epsilon, n)$-contaminations of the underlying null distribution $P^m_{\theta_0}$, each of them belonging to the corresponding neighborhood $U^m(\epsilon, n)(P^m_{\theta_0})$ as defined in (11). Then,

$$\lim_{n \rightarrow \infty} \alpha(P^m_{\eta(n)}) = \alpha_0 + \epsilon^2 \mu \left\| \int_{\mathbb{R}^m} IF(z; U, P^m_{\theta_0}) dG^m(z) \right\|^2 + o(\epsilon^2),$$

for all $G^m \in \text{dom}(\alpha)$, where $U(\cdot)$ is the $U$-functional associated with the corresponding test,

$$\mu := - \frac{\partial}{\partial \beta} H_r(q_1 - \alpha_0; \beta) \bigg|_{\beta = 0}.$$ 

$H_r(\cdot; \beta)$ 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 $\alpha_0 = \alpha(P^m_{\theta_0})$ is the nominal level of the test. Moreover, the bias of $\alpha(P^m_{\eta(n)})$ is uniformly bounded by the inequality

$$\lim_{n \rightarrow \infty} \left| \alpha(P^m_{\eta(n)}) - \alpha_0 \right| \leq \epsilon^2 \mu \sup_{z \in \mathbb{R}^m} \left\| V(\psi_c; \theta_0)\right\|^{-1/2} \left\| IF^c_{\psi_c}(z; \tilde{a}) \right\| + o(\epsilon^2).$$

**Proof.** The proof follows from Heritier and Ronchetti (1994), p. 903.

As a consequence of Proposition 3.3, the maximal asymptotic bias for the level of the test based on $\tilde{a}$ is bounded by

$$\lim_{n \rightarrow \infty} \left| \alpha(P^m_{\eta(n)}) - \alpha_0 \right| \leq \mu (\epsilon c)^2 + o(\epsilon^2).$$

The “power” counterpart of Proposition 3.3 can also be obtained. Hence, also the maximal asymptotic bias of the power induced by local contaminations of the alternative distribution can be controlled by imposing a bound on the self-standardized sensitivity of $\tilde{a}$; cf. Ronchetti and Trojani (2001), Th. 2.

---

23 For the functional $U$ associated with the score and the likelihood ratio test see Heritier and Ronchetti (1994), p. 898.
3.3 Algorithm

In order to compute the robust estimator defined in (19)–(22) an iterative algorithm has to be adopted because the weights \( w(y_t^n; \pi(P_{\theta_0}^n)) \), the matrix \( A(\pi(P_{\theta_0}^n)) \) and the random vectors \( \tau(y_t^{n-1}; \pi(P_{\theta_0}^n)) \) 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 \( \tau^{(0)} := \tau(y_{t-m+1}^{(0)}; \theta(0)) = 0 \), for all \( t = 1, \ldots, n \) and \( A^{(0)} \) such that

\[
A^{(0)} = \left[ n^{-1} \sum_{t=1}^{n} s(y_{t-m+1}; \theta^{(0)}) s(y_{t-m+1}; \theta^{(0)})^T \right]^{-1}
\]

2. Compute, for all \( t = 1, \ldots, n \), the real roots of equations (26), and the associated new values \( \tau^{(1)}_t := \tau(y_{t-m+1}^{(1)}; \theta^{(1)}) \) for \( \tau_t \) and the new matrix \( A^{(1)} \) for \( A \) defined by

\[
\frac{\tau^{(1)}_t := \frac{-L_n(u_t) - k_{1,t}[\Phi(u_t) - \Phi(u_{t-1})] + k_{2,t}M_{1,t} + k_{1,t}M_{2,t} + L_n(\tau_{t-1})}{-L_n(u_t) + [\Phi(u_t) - \Phi(u_{t-1})] + L_n(\tau_{t-1})},}{(A^{(1)})^T A^{(1)}: = n^{-1} \sum_{t=1}^{n} \left( s(y_{t-m+1}; \theta^{(0)}) - \tau^{(0)}_t \right) \left( s(y_{t-m+1}; \theta^{(0)}) - \tau^{(0)}_t \right)^T \times \min^2(1, c \| A^{(0)} (s(y_{t-m+1}; \theta^{(0)}) - \tau^{(0)}_t) \|^{-1}).}
\]

3. Compute the robust estimator \( \theta^{(1)} \) implied by (19) for given \( A^{(1)} \) and \( \tau^{(1)}_t \) as the solution of the implicit equation

\[
\sum_{t=1}^{n} \left( s(y_{t-m+1}; \theta^{(1)}) - \tau^{(1)}_t \right) \min(1, c \| A^{(1)} (s(y_{t-m+1}; \theta^{(0)}) - \tau^{(1)}_t) \|^{-1}) = 0.
\]

4. Replace \( A^{(0)} \) by \( A^{(1)} \) and \( \tau^{(0)}_t \) by \( \tau^{(1)}_t \) for all \( t = 1, \ldots, n \) and iterate step 2 and 3 above until convergence of the sequence \( (\theta(i))_{i \in \mathbb{N}} \) of estimators associated to (19) and to the sequence \( (A^{(i)}, \tau^{(i)})_{i \in \mathbb{N}} \), where \( \tau^{(i)} := (\tau^{(1)}_1, \ldots, \tau^{(1)}_n) \).

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 (26). Closed form solution for \( \tau \)-vectors avoid “internal” simulations to compute the robust estimator, reducing the computation time from almost two hours to about one minute.\(^{24}\)

Alternatively, in Step 3 the GMM quadratic criterion

\[
\sum_{t=1}^{n} (s(y_{t-m+1}; \theta^{(1)}) - \tau^{(1)}_t)^T w(y_{t-m+1}; \theta^{(0)}) A^{(1)} (s(y_{t-m+1}; \theta^{(1)}) - \tau^{(1)}_t) w(y_{t-m+1}; \theta^{(0)})
\]

\(^{24}\)More precisely, we also implemented a second algorithm for the robust estimator defined in (19)–(22), based on the previous four steps, but in which the \( \tau \) integrals in (24) were computed numerically using the Matlab function ‘quadl’. Then we simulated samples of 1’000 observations from an AR(1)-ARCH(1) process with Gaussian innovations and computed the time elapsed for estimations with the Matlab function ‘etime’. The first algorithm described in the text took about 70 seconds, while the second algorithm took about one hour and 50 minutes on a PC Pentium 4 1GHz running Windows XP. The estimation results were the same.
can be minimized to compute $\theta^{(1)}$ to improve the algorithm convergence.

4 Monte Carlo Simulations

In this section we compare by Monte Carlo simulations the performance of the classical PMLE and the robust estimator at the reference model and in the presence of model contaminations. We estimate the AR(1)-ARCH(1) model presented in Example 2.2. We simulate the following contaminated models “near” the reference model $P_{\theta_0}$.

1. **Standard Gaussian innovations.** In this experiment, the innovation $u_t(\theta_0)$ has a standard Gaussian distribution, the PMLE is the MLE and we compare the efficiency of the robust and the MLE under the reference model $P_{\theta_0}$.

2. **Replacement model** (cf. for instance Martin and Yohai (1986)). Under such a model the observed process $X := (x_t)_{t \in \mathbb{Z}}$ is generated according to the following data generating process,

$$x_t = (1 - \vartheta^2_t) y_t + \vartheta_t \xi,$$

(33)

where the clean process $Y := (y_t)_{t \in \mathbb{Z}}$ is generated by the reference model $P_{\theta_0}$ and $(\vartheta^2_t)_{t \in \mathbb{Z}}$ is an i.i.d. 0-1 random sequence independent of $Y$ with the property $P(\vartheta_t^2 = 1) = \eta$. Hence, at a time $t \in \mathbb{Z}$, the clean observation $y_t$ is replaced by $\xi$ with probability $\eta$. In our simulations we set $\eta = 0.5\%$ and $\xi = 1.5$. 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 the model (1)–(2) 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 $u_t(\theta_0) = \tilde{u}_t(\theta_0) \operatorname{Var}[^t(\theta_0)]^{-1/2}$, where $\tilde{u}_t(\theta_0)$ is distributed as the following mixture distribution

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

(34)

We set $\epsilon = 1\%$ and $\varrho = 3$. Contamination (34) describes situations where a given shock (or outlier) affects also future realizations of the process $Y$. Furthermore, as $u_t(\theta_0) \sim i.i.d.(0, 1)$, the dynamic equations in model (1)–(2) are satisfied and the model is dynamically correctly specified. Hence, this is a typical situation in which the PMLE is applied (and it is not the MLE) and there are no theoretical efficiency reasons to prefer one estimator to the other.

The simulation design covers a good range of local deviations from the reference model $P_{\theta_0}$. The tail indices (cf. Gasko and Rosenberger (1983), p. 322) of the innovation $u_t$ under the given distributions are $1$.
for the standard Gaussian distribution, approximately 1.08 for the replacement model (33) and 1.03 for the innovative outlier model (34). For comparison, a standard Student-\(t_5\) distribution has a tail index of 1.34. Therefore, all simulated distributions are very close and samples from different processes are virtually indistinguishable.

We simulate the AR(1)-ARCH(1) model (4) for the following parameter choice: \(\rho_0 = 0.01, \rho_1 = 0.8, \alpha_0 = 0.02, \alpha_1 = 0.8\), under the different distributions for \(y_t\) presented above and for the sample size \(n = 1'000\). The tuning constant for the robust estimator was set at \(c = 9\). Such a rather large value implies that very few observations were downweighted\(^25\).

Each model is simulated 5'000 times. For each simulation we compute the PML and the robust estimates for \(\theta_0\) and the corresponding covariance matrices. Then, for each parameter we compute the corresponding confidence interval at the 95% confidence level.

Estimation results are presented in Table 1–3. For each estimated parameter, the first row contains summary statistics for the PML estimates and the second row for the robust estimates. In Figure 3–5 we plot the estimated densities of the classical and robust estimator. Table 1 shows that the efficiency loss of the robust estimator at the reference model \(P_{\theta_0}\) is almost negligible. Specifically, the mean squared errors of all parameter estimates are very close. This is confirmed by Figure 3. We recall that in this experiment the PMLE is the MLE. Table 2 and Figure 4 show instead large biases and mean squared errors of PML estimates. By contrast, robust estimates maintain low mean squared errors. It is somehow surprising that such bias in PML estimates is induced by contaminating (on average) only 0.5% of the sample observations\(^26\). Finally, Table 3 and Figure 5 shows that, in terms of mean squared error, both estimators correctly estimate the conditional mean parameters \(\rho_0\) and \(\rho_1\), but the robust estimator always outperforms the PMLE, especially in the estimation of the conditional variance parameters \(\alpha_0\) and \(\alpha_1\). We recall that, under the innovative outlier model (34), the PMLE should “perform well” as model (1)–(2) is correctly specified.

We now turn to the confidence intervals. Figure 6–8 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 95% in both cases. (An exception is the confidence interval of the parameter \(\alpha_0\) which is 78% for PML and 92% for the robust version under the replacement model (33).) Moreover, Figure 6 shows that, under the reference model \(P_{\theta_0}\), the confidence intervals lengths for both techniques are almost identical. However, Figure 7 shows that, under the replacement model (33), the PML confidence intervals are much larger than the robust ones, denoting large inaccuracy on the inference results. Moreover, PML confidence intervals are not centered around \(\theta_0\) as the parameter estimates are biased; cf. Figure 4. Robust

\(^25\)For instance, under the reference model \(P_{\theta_0}\) only 3 or 4 (out of 1’000) observations were slightly downweighted with weights of 0.8–0.9.

\(^26\)In the estimation of linear ARMA models Bustos and Yohai (1986) observed that replacement outliers can cause a large bias in the classical ML estimates and “the only way to control them is through robust estimation”; cf. Bustos and Yohai (1986), p. 156.
confidence intervals are much more concentrated around $\theta_0$. Finally, Figure 8 shows that confidence intervals are tighter for robust estimates than for PML estimates, especially for the conditional variance parameters $\alpha_0$ and $\alpha_1$.

To analyze the performance, outside the contamination model, of the classical PML estimator and our robust estimator from the perspective of hypothesis testing we also simulated 1000 sample paths of an AR(1)-ARCH(1) model for the parameter choice $\rho_0 = 0$ and $\rho_1 = 0, 0.05, 0.10$ under scaled Student $t_3$ and scaled Student $t_5$ innovations, respectively. We do not necessarily believe that the innovations follow these distributions but we take $t_3$ and $t_5$ as examples of distributions which are very close to the normal model. Notice that under scaled Student $t$ innovations the model (1)–(2) is dynamically correctly specified and hence the PMLE should perform well. In our experiment we tested the joint null hypothesis $\rho_0 = 0$ and $\rho_1 = 0$ by means of a corresponding classical and robust Wald statistic, respectively. The empirical rejection frequencies of a Wald test based on the classical PML estimator and a Wald test based on our robust estimator are calculated for a fixed nominal level of the test of 5%. Results are presented in Table 4. The estimated standard error of the empirical rejection frequency $\hat{p}$ (using the binomial distribution) is 0.7%, 1.4% and 1.5% for $\hat{p} = 5\%, 30\%, 60\%$, respectively. Table 4 shows that the robust Wald test performs very well across all models, while the classical test is oversized and shows a lower power than the robust ones.

5 Conclusions

We derived optimally bounded-influence estimators for the parameters of conditional location and scale models under a conditionally Gaussian reference model. Based on these results, we obtained optimally 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 very 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 strongly biased and highly inefficient even under small departures from conditional Gaussianity.
A Computation of $\tau(y_1^{m-1}; \theta_0)$

In this appendix we describe the computation of the correction factor $\tau(y_1^{m-1}; \theta_0)$ implicitly defined by equation (22).

We recall that, according to (26), 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 $\mathcal{P}_0^m$. We write $A$ for $A(\theta_0)$.

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

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

where

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

and

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

Clearly, the difficult part in the computation of these integrals is the weighting function $w(y_1^m; \theta_0)$, defined by (20). However, as the weighting function just implies that $\|\psi_v(y_1^m; \theta_0)\|^2 \leq c^2$, 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 (26), i.e.

$$0 = \|A(s(y_1^m; \theta_0) - \tau^{(0)})\|^2 - c^2$$

$$:= \|A(-k_{1,m} + k_{2,m} u_m(\theta_0) + k_{1,m} u_m^2(\theta_0) - \tau^{(0)})\|^2 - c^2$$

$$:= a_4 u_m^4(\theta_0) + a_3 u_m^3(\theta_0) + a_2 u_m^2(\theta_0) + a_1 u_m(\theta_0) + a_0 - c^2,$$

where

$$a_4 := k_{1,m} \frac{\gamma_t}{\gamma} A_k_{1,m}, \quad a_3 := 2 k_{1,m} \frac{\gamma_t}{\gamma} A_k_{2,m},$$

$$a_2 := k_{2,m} \frac{\gamma_t}{\gamma} A_k_{2,m} - \frac{\gamma_t}{\gamma} A_k_{1,m} - 2 k_{1,m} \frac{\gamma_t}{\gamma} A_k_{1,m} - 2 k_{1,m} \frac{\gamma_t}{\gamma} A_k^{(0)},$$

$$a_1 := -a_3 - 2 k_{2,m} \frac{\gamma_t}{\gamma} A_k^{(0)},$$

$$a_0 := a_4 + 2 k_{1,m} \frac{\gamma_t}{\gamma} A_k^{(0)} + \tau^{(0)} \frac{\gamma_t}{\gamma} A_k^{(0)}.$$
We recall that 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 we have

$$\|A(s(y_1^m; \theta_0) - \tau^{(0)})\| \leq c, \quad u_m(\theta_0) \in [\underline{u}_m, \overline{u}_m],$$

$$> c, \quad u_m(\theta_0) \in (-\infty, \underline{u}_m) \cup (\overline{u}_m, +\infty)$$

denoting by $\underline{u}_m \leq \overline{u}_m$ the real roots. In the second case, with real roots $\underline{u}_m \leq \overline{u}_m \leq \overline{u}_m$, we have

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

Notice that in almost all simulations and all empirical estimations we had two real roots.

**Step 2**

In the second step we ‘split’ the integral in equation (35) and (36) according to the roots determined in Step 1. Let us assume first that we have two real roots, then

$$\tau_{num}(y_1^{m-1}; \theta_0)$$

$$= \int_{\underline{u}_m}^{\overline{u}_m} \frac{\exp(-0.5u^2) \exp(-k_{1,m} u + k_{1,m}^2 u^2)}{\sqrt{2\pi}} \frac{1}{2\pi} \exp(-0.5u^2) du - k_{1,m} [\Phi(\overline{u}_m) - \Phi(\underline{u}_m)] + k_{2,m} M_{1,m} + k_{1,m} M_{2,m}$$

$$+= \int_{\overline{u}_m}^{+\infty} \frac{\exp(-0.5u^2) \exp(-k_{1,m} u + k_{1,m}^2 u^2)}{\sqrt{2\pi}} \frac{1}{2\pi} \exp(-0.5u^2) du.$$

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

Notice that $M_{1,m}$ and $M_{2,m}$ are the truncated first and second moment of a standard Gaussian random variable and integration by parts yields

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

The remaining univariate integrals are approximated ‘componentwise’ using the Laplace’s method. Under standard regularity conditions$^{27}$ on the real function $q(\cdot)$, for $\alpha \to \infty$ one has

$$\int_0^\infty \alpha \exp(-\alpha u) q(u) du = q(0) + \frac{q'(0)}{\alpha} + \frac{q''(0)}{\alpha^2} + O\left(\frac{1}{\alpha^3}\right)$$

$$=: \mathcal{L}(q, \alpha) + O\left(\frac{1}{\alpha^3}\right)$$

by iterating integration by parts. \( L(q, \alpha) \) is the Laplace’s approximation of the integral up to the third order\(^{28}\). Therefore,
\[
\begin{align*}
\int_{\pi_m}^{+\infty} q_n(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) \, du &= \frac{1}{\sqrt{2\pi}} \exp(-0.5\pi_m^2) \frac{1}{\pi_m} \int_{0}^{+\infty} \pi_m \exp(-\pi_m z) q_n(\pi_m + z) \exp(-0.5z^2) \, dz \\
&= \frac{1}{\sqrt{2\pi}} \exp(-0.5\pi_m^2) \frac{1}{\pi_m} \left( L(q_n, \pi_m) + O \left( \frac{1}{\pi_m^3} \right) \right) \\
&= L_n(\pi_m) + O \left( \frac{1}{\pi_m^3} \right),
\end{align*}
\]
where the first equality follows from the substitution \( z = u - \pi_m \) and \( q_n(z) := q_n(\pi_m + z) \exp(-0.5z^2) \).

Similarly,
\[
\begin{align*}
\int_{-\infty}^{\pi_m} q_n(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) \, du &= -\frac{1}{\sqrt{2\pi}} \exp(-0.5\pi_m^2) \frac{1}{\pi_m} \left( L(q_n, \pi_m) + O \left( \frac{1}{\pi_m^3} \right) \right) \\
&= -L_n(\pi_m) + O \left( \frac{1}{\pi_m^3} \right),
\end{align*}
\]
where \( q_n(z) := q_n(\pi_m + z) \exp(-0.5z^2) \) and using the substitution \( z = u - \pi_m \).

The procedure for computing the denominator of \( \tau \) in (36) is completely analogous. Specifically,
\[
\tau_{\text{den}}(y_1^{m-1}; \theta_0) \]
\[
\begin{align*}
&= \int_{-\infty}^{\pi_m} \frac{c}{\|A(s(u; \theta_0) - \tau(0))\|} \, d\Phi(u) + \int_{\pi_m}^{\tau_m} \frac{c}{\|A(s(u; \theta_0) - \tau(0))\|} \, d\Phi(u) \\
&= \int_{-\infty}^{\pi_m} q_d(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) \, du + \left[ \Phi(\pi_m) - \Phi(\pi_m) \right] \\
&\quad + \int_{\pi_m}^{+\infty} q_d(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) \, du.
\end{align*}
\]
The Laplace’s approximations of the remaining integrals are
\[
\begin{align*}
\int_{\pi_m}^{+\infty} q_d(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) \, du &= \frac{1}{\sqrt{2\pi}} \exp(-0.5\pi_m^2) \frac{1}{\pi_m} \left( L(q_d, \pi_m) + O \left( \frac{1}{\pi_m^3} \right) \right) \\
&= L_d(\pi_m) + O \left( \frac{1}{\pi_m^3} \right),
\end{align*}
\]
where \( q_d(z) := q_d(\pi_m + z) \exp(-0.5z^2) \) and
\[
\begin{align*}
\int_{-\infty}^{\pi_m} q_d(u) \frac{1}{\sqrt{2\pi}} \exp(-0.5u^2) \, du &= -\frac{1}{\sqrt{2\pi}} \exp(-0.5\pi_m^2) \frac{1}{\pi_m} \left( L(q_d, \pi_m) + O \left( \frac{1}{\pi_m^3} \right) \right) \\
&= -L_d(\pi_m) + O \left( \frac{1}{\pi_m^3} \right),
\end{align*}
\]
where \( q_d(z) := q_d(\pi_m + z) \exp(-0.5z^2) \). Collecting the terms one gets the expression for \( \tau \) stated in Proposition 3.2.

\(^{28}\)We use third order Laplace’s approximations as the contribution of higher order terms is negligible.
In the general case where the quartic equation (26) has four real roots \( u_{\frac{m}{n}} \leq u_1 \leq \tau_m \leq \tau_m \), for instance the integral in (36) becomes

\[
\tau_{\text{den}}(y_1^{m-1}; \theta_0) = \int_{-\infty}^{u_m} \frac{c}{\|A(s; \theta_0) - \tau(0)\|} \, d\Phi(u) + \int_{-\infty}^{u_m} \, d\Phi(u) + \int_{u_m}^{\tau_m} \frac{c}{\|A(s; \theta_0) - \tau(0)\|} \, d\Phi(u) + \int_{\tau_m}^{+\infty} \, d\Phi(u)
\]

and Laplace’s approximation holds only for the first and the last integral. However, numerical results (not reported here) show that the error when neglecting the weighting function in the central integral is very small. When \( u_m \) and \( \tau_m \) are close enough and/or \( \|A(s; \theta_0) - \tau(0)\| \) is not very large, the error is practically zero. Finally, we note that the implementation of such a case (which happens very rarely) in the algorithm would largely increase the computational time, without improving the inference results.

B Proof of Proposition 3.1 and Corollary 3.1

The proofs of Proposition 3.1 and Corollary 3.1 follow from Theorem 1 and Corollary 1 in Stefanski et al. (1986).

In order to prove Proposition 3.1, let \( \psi \) be any competitor with \( \psi^{\text{bif}} \). Without loss of generality assume that \( \psi = IF_{\psi}^{\text{cond}} \), i.e. \( \psi \) is in canonical form in the sense of Hampel (1974). Hence, \( D(\psi; \theta_0) = I \) and

\[
E_{\theta_0}[\psi(y_1^n; \theta_0) s(y_1^n; \theta_0)^\top] = I
\]

(37)
as differentiating (17) with respect to \( \theta \) gives \( -D(\psi; \theta_0) + E_{\theta_0}[\psi(y_1^n; \theta_0) s(y_1^n; \theta_0)^\top] = 0 \) assuming that integration and differentiation can be interchanged. Moreover, \( V(\psi; \theta_0) = E_{\theta_0}[\psi(y_1^n; \theta_0) \psi(y_1^n; \theta_0)^\top] \). Now write \( s \) for \( s(y_1^n; \theta_0) \), \( \tau \) for \( (y_1^{m-1}; \theta_0) \), \( \psi \) for \( (y_1^n; \theta_0) \), \( V_\psi \) for \( V(\psi; \theta_0) \), \( D_{\text{bif}} \) for \( D(\psi^{\text{bif}}; \theta_0) \) and \( V_{\psi^{\text{bif}}} \) for \( V(\psi^{\text{bif}}; \theta_0) \).

Then

\[
\min_\psi \text{tr}(V_\psi^{-1} V_{\psi^{\text{bif}}}^{-1}) = \min_\psi \text{tr}(E_{\theta_0}[\psi \psi^\top] V_{\psi^{\text{bif}}}^{-1})
\]

\[
= \min_\psi \text{tr}(E_{\theta_0}[(D_{\psi\psi}(s - \tau) - \psi) (D_{\psi\psi}(s - \tau) - \psi)^\top] V_{\psi^{\text{bif}}}^{-1})
\]

\[
= \min_\psi E_{\theta_0}[(D_{\psi\psi}(s - \tau) - \psi)^\top V_{\psi^{\text{bif}}}^{-1} (D_{\psi\psi}(s - \tau) - \psi)]
\]

\[
= \min_\psi E_{\theta_0}[(V_{\psi^{\text{bif}}}^{-1/2} D_{\psi\psi}(s - \tau) - V_{\psi^{\text{bif}}}^{-1/2} \psi)^\top (V_{\psi^{\text{bif}}}^{-1/2} D_{\psi\psi}(s - \tau) - V_{\psi^{\text{bif}}}^{-1/2} \psi)],
\]

(38)
where, using (16) and (37), the second equality follows from

\[
E_{\theta_0}[(D_{\psi\psi}(s - \tau) - \psi) (D_{\psi\psi}(s - \tau) - \psi)^\top] = D_{\psi\psi} E_{\theta_0}[(s - \tau) (s - \tau)^\top] D_{\psi\psi} - D_{\psi\psi}^{-1} - D_{\psi\psi}^{-1} + V_\psi,
\]
and the first three terms in the right hand side are independent of $\psi$. Under condition (23), the objective function (38) is minimized by

$$\psi^* = D_{\text{bif}}^{-1}(s - \tau) \min(1, c (s - \tau) \mathbb{I} D_{\text{bif}}^{-1} V_{\text{bif}}^{-1} (s - \tau)^{-1/2})$$

and, as $D_{\text{bif}}^{-1} V_{\text{bif}}^{-1} = E_{\theta_0}[\psi^{\text{bif}} \psi^{\text{bif} \top}]^{-1} = A(\theta_0)^\top A(\theta_0)$, $\psi^* = D_{\text{bif}}^{-1} \psi^{\text{bif}}$. Condition (37) ensures that $\psi^*$ is unique almost surely. □

To prove Corollary 3.1, again assume that all score functions are in canonical form and satisfy (16). Define

$$S := \{\psi: \sup_{z \in \mathbb{R}^m} \psi^\top V^{-1} \psi \leq c^2\}, \quad S_{\text{bif}} := \{\psi: \sup_{z \in \mathbb{R}^m} \psi^\top V_{\text{bif}}^{-1} \psi \leq c^2\}.$$  

We must show that if there exists $\psi^{\text{opt}}$ in $S$ such that $V_{\psi^{\text{opt}}} \leq V_{\psi}$ (in the strong sense of positive-definiteness) for all $\psi$ in $S$, then $\psi^{\text{opt}}$ is equivalent to $D_{\text{bif}}^{-1} \psi^{\text{bif}}$.

Clearly, $D_{\text{bif}}^{-1} \psi^{\text{bif}}$ is in $S$; thus by assumption $V_{\psi^{\text{opt}}} \leq V_{\text{bif}}$ (and $V_{\psi^{\text{opt}}}^{-1} \geq V_{\text{bif}}^{-1}$). From this follows that

$$\psi^{\text{opt} \top} V_{\text{bif}}^{-1} \psi^{\text{opt}} \leq \psi^{\text{opt} \top} V_{\psi^{\text{opt}}}^{-1} \psi^{\text{opt}} \leq c^2,$$

and hence $\psi^{\text{opt}}$ is in $S_{\text{bif}}$. Let $\mathcal{I} = S \cap S_{\text{bif}}$. The set $\mathcal{I}$ is nonempty; it contains $D_{\text{bif}}^{-1} \psi^{\text{bif}}$ and $\psi^{\text{opt}}$. For any $\psi$ in $\mathcal{I}$ we know $V_{\psi^{\text{opt}}} \leq V_{\psi}$ and hence

$$\text{tr}(V_{\psi^{\text{opt}}} V_{\text{bif}}^{-1}) \leq \text{tr}(V_{\psi} V_{\text{bif}}^{-1})$$

for all $\psi$ in $\mathcal{I}$. But Proposition 3.1 proves that $D_{\text{bif}}^{-1} \psi^{\text{bif}}$, when defined, is the almost everywhere unique minimizer of $\text{tr}(V_{\psi} V_{\text{bif}}^{-1})$ among all $\psi$ in $\mathcal{I}$. Hence, the equivalence of $\psi^{\text{opt}}$ and $\psi^{\text{bif}}$ follows. □
<table>
<thead>
<tr>
<th>true</th>
<th>mean</th>
<th>median</th>
<th>$q_{25}$</th>
<th>$q_{75}$</th>
<th>Stdv</th>
<th>$q_{75} - q_{25}$</th>
<th>MSE%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_0$</td>
<td>0.0100</td>
<td>0.0099</td>
<td>0.0066</td>
<td>0.0133</td>
<td>0.0051</td>
<td>0.0067</td>
<td>0.0026</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0066</td>
<td>0.0134</td>
<td>0.0051</td>
<td>0.0067</td>
<td>0.0026</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.7983</td>
<td>0.7989</td>
<td>0.7896</td>
<td>0.8074</td>
<td>0.0140</td>
<td>0.0178</td>
<td>0.0199</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7977</td>
<td>0.7985</td>
<td>0.7890</td>
<td>0.8070</td>
<td>0.0142</td>
<td>0.0181</td>
<td>0.0208</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.0200</td>
<td>0.0199</td>
<td>0.0189</td>
<td>0.0210</td>
<td>0.0015</td>
<td>0.0021</td>
<td>0.0002</td>
</tr>
<tr>
<td>0.02</td>
<td>0.0200</td>
<td>0.0200</td>
<td>0.0189</td>
<td>0.0211</td>
<td>0.0015</td>
<td>0.0021</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.7976</td>
<td>0.7986</td>
<td>0.7461</td>
<td>0.8496</td>
<td>0.0758</td>
<td>0.1034</td>
<td>0.5756</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8007</td>
<td>0.8016</td>
<td>0.7490</td>
<td>0.8525</td>
<td>0.0765</td>
<td>0.1035</td>
<td>0.5850</td>
</tr>
</tbody>
</table>

Table 1: Summary statistics for each estimated parameter $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$ for the MLE (first row) and the robust estimator (second row) based on 5'000 simulations of 1'000 observations of the AR(1)-ARCH(1) model in Example 2.2 under the reference model $P_{\theta_0}$, i.e. a standard Gaussian distribution for innovations; cf. Figure 3.

<table>
<thead>
<tr>
<th>true</th>
<th>mean</th>
<th>median</th>
<th>$q_{25}$</th>
<th>$q_{75}$</th>
<th>Stdv</th>
<th>$q_{75} - q_{25}$</th>
<th>MSE%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_0$</td>
<td>0.0166</td>
<td>0.0161</td>
<td>0.0102</td>
<td>0.0227</td>
<td>0.0094</td>
<td>0.0125</td>
<td>0.0132</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0112</td>
<td>0.0111</td>
<td>0.0074</td>
<td>0.0150</td>
<td>0.0057</td>
<td>0.0076</td>
<td>0.0034</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.7930</td>
<td>0.7952</td>
<td>0.7787</td>
<td>0.8089</td>
<td>0.0280</td>
<td>0.0302</td>
<td>0.0832</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7959</td>
<td>0.7965</td>
<td>0.7864</td>
<td>0.8063</td>
<td>0.0154</td>
<td>0.0200</td>
<td>0.0254</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.0298</td>
<td>0.0290</td>
<td>0.0246</td>
<td>0.0341</td>
<td>0.0070</td>
<td>0.0095</td>
<td>0.0145</td>
</tr>
<tr>
<td>0.02</td>
<td>0.0222</td>
<td>0.0218</td>
<td>0.0204</td>
<td>0.0236</td>
<td>0.0027</td>
<td>0.0032</td>
<td>0.0012</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.8037</td>
<td>0.8052</td>
<td>0.7069</td>
<td>0.9125</td>
<td>0.1337</td>
<td>0.2055</td>
<td>1.7880</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8081</td>
<td>0.8064</td>
<td>0.7486</td>
<td>0.8704</td>
<td>0.0896</td>
<td>0.1218</td>
<td>0.8097</td>
</tr>
</tbody>
</table>

Table 2: Summary statistics for each estimated parameter $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$ for the PMLE (first row) and the robust estimator (second row) based on 5'000 simulations of 1'000 observations of the AR(1)-ARCH(1) model in Example 2.2 under the replacement model (33); cf. Figure 4.
<table>
<thead>
<tr>
<th>true</th>
<th>mean</th>
<th>median</th>
<th>$q_{25}$</th>
<th>$q_{75}$</th>
<th>Stdv</th>
<th>$q_{75} - q_{25}$</th>
<th>MSE%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_0$</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0065</td>
<td>0.0135</td>
<td>0.0055</td>
<td>0.0070</td>
<td>0.0030</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0067</td>
<td>0.0132</td>
<td>0.0051</td>
<td>0.0064</td>
<td>0.0026</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.7982</td>
<td>0.7988</td>
<td>0.7882</td>
<td>0.8086</td>
<td>0.0159</td>
<td>0.0204</td>
<td>0.0256</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7978</td>
<td>0.7983</td>
<td>0.7889</td>
<td>0.8074</td>
<td>0.0143</td>
<td>0.0185</td>
<td>0.0209</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.0199</td>
<td>0.0199</td>
<td>0.0189</td>
<td>0.0210</td>
<td>0.0016</td>
<td>0.0022</td>
<td>0.0003</td>
</tr>
<tr>
<td>0.02</td>
<td>0.0194</td>
<td>0.0194</td>
<td>0.0184</td>
<td>0.0203</td>
<td>0.0014</td>
<td>0.0019</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.7989</td>
<td>0.7992</td>
<td>0.7405</td>
<td>0.8594</td>
<td>0.0895</td>
<td>0.1189</td>
<td>0.8006</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7727</td>
<td>0.7748</td>
<td>0.7225</td>
<td>0.8221</td>
<td>0.0747</td>
<td>0.0995</td>
<td>0.6328</td>
</tr>
</tbody>
</table>

Table 3: Summary statistics for each estimated parameter $\hat{\theta}_0 := (\hat{\rho}_0, \hat{\rho}_1, \hat{\alpha}_0, \hat{\alpha}_1)^T$ for the PMLE (first row) and the robust estimator (second row) based on 5'000 simulations of 1'000 observations of the AR(1)-ARCH(1) model in Example 2.2 under the innovative outlier model (34); cf. Figure 5.

<table>
<thead>
<tr>
<th>$\rho_1$</th>
<th>$t_3$ PML</th>
<th>$t_3$ ROB</th>
<th>$t_5$ PML</th>
<th>$t_5$ ROB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.08</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>0.05</td>
<td>0.17</td>
<td>0.24</td>
<td>0.22</td>
<td>0.26</td>
</tr>
<tr>
<td>0.10</td>
<td>0.46</td>
<td>0.65</td>
<td>0.62</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Table 4: 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 and based on 1'000 simulations of 1'000 observations of the AR(1)-ARCH(1) model in Example 2.2 under scaled $t_3$ and scaled $t_5$ innovations, respectively.
Figure 1: Self standardized norm of the IF $\|V(s; \theta_0)^{-1/2} IF_{\text{cond}}(x_1, x_2; \theta_0)\|$ of the MLE, under the reference model $P_{\theta_0}^s$, for the AR(1) model: $y_2 = 0.8y_1 + \varepsilon_2$, where $\varepsilon_2 \sim \mathcal{N}(0, 2)$, as a function of $x_1$ and $x_2$. 
Figure 2: Self-standardized norm of the IF $\|A(\theta_0) s(y_1^3; \theta_0)\|$ of the MLE, under the reference model $P^{3}_{\theta_0}$, for the AR(1)-ARCH(1) model: $y_3 = 0.01 + 0.8y_2 + \varepsilon_3(\theta_0)$, $\varepsilon_3(\theta_0) = \sigma_3(\theta_0) u_3(\theta_0)$, $\sigma_3^2(\theta_0) = 0.02 + 0.8\varepsilon_2^2(\theta_0)$, as a function of $u_3(\theta_0)$. 
Figure 3: Estimated densities of $\hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$ based on 5’000 simulations of 1’000 observations of the AR(1)-ARCH(1) process in Example 2.2 under the reference model $P_{\theta_0}$, i.e. Gaussian distribution for innovations; cf. Table 1.
Figure 4: Estimated densities of $\hat{\theta} := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^\top$ based on 5’000 simulations of 1’000 observations of the AR(1)-ARCH(1) process in Example 2.2 under the replacement model (33); cf. Table 2.
Figure 5: Estimated densities of \( \hat{\theta}_0 := (\hat{\rho}_0, \hat{\rho}_1, \hat{\alpha}_0, \hat{\alpha}_1)^T \) based on 5'000 simulations of 1'000 observations of the AR(1)-ARCH(1) process in Example 2.2 under the innovative outlier model (34); cf. Table 3.
Figure 6: 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 3) under the reference model $P_{\theta_0}$. 

\[
\rho_0 = 0.01 \quad \rho_1 = 0.8 \quad \alpha_0 = 0.02 \quad \alpha_1 = 0.8
\]
Figure 7: Boxplot of the lengths of PML (column 1) and robust (column 2) confidence intervals for $\hat{\theta}_0 := (\hat{\rho}_0, \hat{\rho}_1, \hat{\alpha}_0, \hat{\alpha}_1)^T$ (cf. Figure 4) under the replacement model (33).
Figure 8: Boxplot of the lengths of PML (column 1) and robust (column 2) confidence intervals for \( \hat{\theta}_0 := (\hat{\rho}_0 \ \hat{\rho}_1 \ \hat{\alpha}_0 \ \hat{\alpha}_1)^T \) (cf. Figure 5) under the innovative outlier model (34).
References


