

ROBUST INFERENCE WITH GMM ESTIMATORS

Elvezio Ronchetti	Fabio Trojani
Dept. of Econometrics	Institute of Finance
University of Geneva	University of Southern
Switzerland*	Switzerland
	Lugano, Switzerland [†]

First Version: August, 1999

Revised: April 2000

Keywords: ARCH models, GMM estimators and tests, influence function, robust model selection, robustness of validity.

JEL Classification: C12, C13, C14.

*Blv. Carl Vogt 102, CH-1211 Geneva, tel: +41 22 705 81 31, e-mail: Elvezio.Ronchetti@metri.unige.ch

[†]Via Ospedale 13, CH-6900 Lugano, tel: +41 91 912 47 23, e-mail: Fabio.Trojani@lu.unisi.ch.

Abstract

The local robustness properties of Generalized Method of Moments (GMM) estimators and of a broad class of GMM based tests are investigated in a unified framework. GMM statistics are shown to have bounded influence if and only if the function defining the orthogonality restrictions imposed on the underlying model is bounded. Since in many applications this function is unbounded, it is useful to have procedures that modify the starting orthogonality conditions in order to obtain a robust version of a GMM estimator or test. We show how this can be obtained when a reference model for the data distribution can be assumed. We develop a flexible algorithm for constructing a robust GMM (RGMM) estimator leading to stable GMM test statistics. The amount of robustness can be controlled by an appropriate tuning constant. We relate by an explicit formula the choice of this constant to the maximal admissible bias on the level or (and) the power of a GMM test and the amount of contamination that one can reasonably assume given some information on the data. Finally, we illustrate the RGMM methodology with some simulations of an application to RGMM testing for conditional heteroscedasticity in a simple linear autoregressive model. In this example we find a significant instability of the size and the power of a classical GMM testing procedure under a non-normal conditional error distribution. On the other side, the RGMM testing procedures can control the size and the power of the test under non-standard conditions while maintaining a satisfactory power under an approximatively normal conditional error distribution.

1 Introduction

THIS PAPER ANALYZES the local robustness properties of estimators based on the GMM (cf. Hansen (1982)) and of test statistics based on a GMM estimator. We characterize the local robustness of GMM estimators, of Hansen's specification test and of GMM based tests that are GMM versions of the classical Wald, score, and likelihood-ratio test (cf. Newey and West (1987a) and Gourieroux et Monfort (1989)) by a single property: the boundedness of the underlying orthogonality function. Since many available econometric models are based on an unbounded orthogonality function, we propose a simple unified setting for constructing a robust GMM (RGMM) estimator yielding at once the local robustness of all GMM based tests.

The need for robust statistical procedures for estimation, testing and prediction has been stressed by many authors 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, 1991), Markatou and Ronchetti (1997), Krishnakumar and Ronchetti (1997). This paper focuses on locally robust GMM estimation and testing and contributes to the current literature in the following directions.

First of all, our results extend the application of robust instrumental variables estimators proposed by Krasker and Welsch (1985), Krasker (1986) and Lucas, van Dijk and Kloek (1994) to a general GMM setting with nonlinear orthogonality conditions and where some stationary ergodic dependence in the underlying data generating process is admitted.

Secondly, the paper generalizes the robust testing framework developed by Heritier and Ronchetti (1994) to a general GMM setting. It unifies and simplifies the theory by proposing a RGMM estimator leading to robust Wald, score and likelihood-ratio type tests for general nonlinear parameter

restrictions.

Finally, the paper provides some robust versions of Hansen's specification test. This yields RGMM model selection procedures that were not available before.

RGMM analysis focuses implicitly on econometric models defined by some sets of orthogonality conditions that are at best ideal approximations to the reality. This task is accomplished by defining a reference distribution for the data playing the role of a reference model for the underlying data generating process. Of course, this entails a lack of generality compared to a standard nonparametric GMM situation. However, in many applications of the GMM, the reference model distribution is already implied by the problem under investigation (for instance in the case of normality of the error distribution). Furthermore, when no natural reference model is supplied we claim that it is often useful to impose one in order to obtain GMM statistics that behave sufficiently well at least over a restricted set of relevant model distributions. The implied orthogonality conditions are then approximate in the sense that they should be satisfied by any model distribution "near" - in some appropriate sense - to the given reference model. When translating this argument in terms of the empirical distribution of the data, this means that in a RGMM framework a small fraction of the observations can deviate from the rest of the sample without affecting the empirical moment conditions in a dramatic way. Therefore, the derived parameter estimates and statistics are representative for the structure of the "majority" of the data. In other words, robust GMM procedures pay a small "insurance premium" in terms of efficiency at the reference model in order to be robust in a neighbourhood of it.

At least in linear models with normal serially independent errors, the

effects of different kinds of distributional deviations from the assumptions are well studied and known to have an important impact on the asymptotic properties of a GMM estimator; cf. Krasker and Welsch (1985), Krasker (1986) and Lucas, van Dijk and Kloek (1994). For the time series context important work has been done by Künsch (1984) and Martin and Yohai (1986) within the framework of (linear) autoregressive models. Since there is a priori no reason to generally believe that in a nonlinear model with stochastic time dependence these effects should be less serious, a general RGMM framework can offer a powerful complement to the classical GMM in many applications.

In this paper we focus on locally robust GMM estimation and particularly on GMM testing, that is on smooth GMM functionals that can be locally approximated by means of their influence function (IF); see Hampel (1968, 1974) and Hampel et al. (1986) for basic definitions and Künsch (1984) and Martin and Yohai (1986) for the time series context. Boundedness of the IF implies that in a neighbourhood of the model the bias of an estimator cannot become arbitrarily large. In the testing framework this implies that in a neighborhood of the model the level of the test does not become arbitrarily close to 1 (robustness of validity) and the power does not become arbitrarily close to 0 (robustness of efficiency). Hence, a bounded influence function is a desirable local stability property of a statistic.

Since in applications the IF of a GMM statistical functional is often unbounded (some examples are listed in Section 2), we propose a robustified version of a GMM estimator that is shown to induce at the same time (locally) robust GMM testing procedures. The RGMM estimator is constructed by applying a basic truncation argument of the theory of robust statistics modified to the particular GMM situation. An important feature of the pro-

posed estimator is that the amount of robustness imposed can be controlled by a tuning constant which is related by an explicit formula to both the maximal local bias in the level and the power of a GMM test and to the magnitude of the given model deviation; see Section 4.

As an illustration of these general principles consider for instance a simple AR(1) model with ARCH(1) (cf. Engle (1982)) errors for a random sequence $(y_t)_{t \in \mathbb{N}}$:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \sqrt{h_t} u_t \quad , \quad h_t = \alpha_0 + \alpha_1 u_{t-1}^2 \quad , \quad (1)$$

where $(u_t)_{t \in \mathbb{N}}$ is a standardized i.i.d sequence with unknown distribution.

A natural set of orthogonality conditions for a GMM estimation of the parameters $(\alpha_0, \alpha_1, \beta_0, \beta_1)$ is given by:

$$E[\epsilon_t] = 0 \quad , \quad E[\epsilon_t y_{t-1}] = 0 \quad , \quad E[\eta_t - h_t] = 0 \quad , \quad E[\eta_t \eta_{t-1}] = 0 \quad , \quad (2)$$

where

$$\epsilon_t = y_t - \beta_0 - \beta_1 y_{t-1} \quad , \quad \eta_t = \epsilon_t^2 \quad . \quad (3)$$

We will see that the unboundedness of the orthogonality function defining these orthogonality conditions implies a lack of robustness of GMM estimators and tests. For this case we propose a small simulated application to testing for a conditional heteroscedasticity specification in the linear autoregressive model. In this simple experiment we observe that for relevant sample sizes the classical GMM procedure is unstable even under relatively small distributional deviations from the normality of the error distribution. Specifically, the GMM specification test often produces sizes that are higher than theoretically expected and power curves that are already much flatter

than under normality for conditional error distributions very near to the normal (for instance a t_9 distribution). On the other side, when introducing a normal reference model for the error distribution it is possible to control for the empirical bias in the level and the power of a RGMM test under non-standard situations. Of course, imposing more robustness on a RGMM test has an impact on the power of the test at the reference model. However, in the proposed application it seems that the loss in power of the RGMM test at the reference model is quickly compensated by a strong gain in power under non-normality of the error distribution.

One could argue that local robustness is not sufficient and global reliability should be the target in order to guarantee resistance in the presence of a large amount of contamination. This would require the development of high-breakdown estimators, but it seems difficult to extend the available results for regression models (see for instance Rousseeuw and Leroy (1987)) to a general GMM setting. Indeed, the latter does not show the high degree of structure which is exploited in the definition of high-breakdown estimators for regression models. Moreover, although we believe that high-breakdown estimators play a useful role in the exploratory and estimation part of any data analysis, we feel that small deviations from the model are more meaningful for inference.

The paper is structured as follows. In Section 2 we derive the influence function of a GMM estimator and show that GMM estimators have bounded IF if and only if the function inducing the natural orthogonality conditions of the model is bounded. We then give some examples of GMM estimators with unbounded IF. Section 3 is devoted to the definition and construction of a RGMM estimator suited to induce stable GMM testing procedures. Section 4 analyses the local robustness of tests constructed from a GMM estimator and

derives some basic expansions for the power and level functionals of a GMM test. These expansions provide a useful asymptotic bound for the asymptotic bias of level and power of a GMM test under small deviations from the model distribution. The bound is proportional to a particular supremum norm of the underlying orthogonality function. Therefore, it can be used to obtain RGMM estimators that explicitly control the maximal bias of level and power of a GMM test under deviations from the assumptions. It is this bound that allows us to derive the explicit link between the "degree" of robustness of the RGMM estimator of Section 3, the amount of contamination that can be reasonably assumed given some information on the data, and the maximal bias in level and power of a RGMM test. Section 5 presents the results of our simulations of a RGMM test for conditional heteroscedasticity in the errors of an autoregressive model and Section 6 concludes the paper with some summarizing remarks and suggestions for further research.

2 Robustness Properties of GMM Estimators

Let $\mathcal{X} := (X_n)_{n \in \mathbb{N}}$ be a stationary ergodic sequence defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in \mathbb{R}^N . Without loss of generality, we index the family $\mathcal{P} := \{P_\theta, \theta \in \Theta\}$ of distributions on \mathbb{R}^N by a parameter vector $\theta \in \Theta$. θ_0 is the parameter vector corresponding to the model distribution (the reference model) for X_1 . This notation is used in order to write several GMM statistics used in the paper as functionals on a subset of \mathcal{P} .

The GMM consists in estimating indirectly some functional $a : \mathcal{P} \rightarrow \mathcal{A} := a(\mathcal{P}) \subset \mathbb{R}^k$ of parameters of interest by introducing a function $h : \mathbb{R}^N \times \mathcal{A} \rightarrow$

\mathbb{R}^H enforcing a set of orthogonality conditions

$$E_{\theta_0} h(X_1; a(P_{\theta_0})) = 0 \quad , \quad (4)$$

on the structure of the underlying model. Let $\mathcal{W} := (W_n)_{n \in \mathbb{N}}$ be a sequence of weighting symmetric positive definite matrices converging a.s to a positive definite matrix W_0 .

A *generalized method of moments estimator* (GMME) associated with \mathcal{W} is a sequence $(\tilde{a}(P_{\theta_n}))_{n \in \mathbb{N}}$ of (functional) solutions to the optimization problem:

$$\tilde{a}(P_{\theta_n}) = \arg \min_{a \in \mathcal{A}} E_{\theta_n} h^\top(X_1; a) W_n E_{\theta_n} h(X_1; a) \quad ; \quad n \in \mathbb{N} \quad , \quad (5)$$

where $P_{\theta_n} := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ is the empirical distribution of X_1, \dots, X_n , and δ_x denotes the point mass distribution at $x \in \mathbb{R}^N$. This functional notation of the GMM minimization problem is useful for investigating the functional structure of a general GMM statistic later on.

Under appropriate regularity conditions (see Hansen (1982)) the GMME exists, is strongly consistent and asymptotically normally distributed at the model with an asymptotic covariance matrix given by

$$\Sigma_{\theta_0}(W_0) = S_{\theta_0}(W_0) E_{\theta_0} \frac{\partial h^\top(X_1; a(P_{\theta_0}))}{\partial a} W_0 V_0 W_0 E_{\theta_0} \frac{\partial h(X_1; a(P_{\theta_0}))}{\partial a^\top} S_{\theta_0}(W_0),$$

where

$$V_0 := E_{\theta_0} [h(X_1; a(P_{\theta_0})) h^\top(X_1; a(P_{\theta_0}))] \quad , \quad (6)$$

(the covariance matrix of $h(X_1, a(P_{\theta_0}))$ at the model P_{θ_0}) and

$$S_{\theta_0}(W_0) := \left[E_{\theta_0} \frac{\partial h^\top(X_1; a(P_{\theta_0}))}{\partial a} W_0 E_{\theta_0} \frac{\partial h(X_1; a(P_{\theta_0}))}{\partial a^\top} \right]^{-1} . \quad (7)$$

The GMME associated to a sequence \mathcal{W} such that

$$W_0 = V_0^{-1}, \quad (8)$$

is asymptotically best in the sense of Hansen (1982) and yields a "smallest" asymptotic covariance matrix given by:

$$\Sigma_{\theta_0}(W_0) = \left[E_{\theta_0} \frac{\partial h^\top(X_1; a(P_{\theta_0}))}{\partial a} V_0^{-1} E_{\theta_0} \frac{\partial h(X_1; a(P_{\theta_0}))}{\partial a^\top} \right]^{-1}. \quad (9)$$

We will adopt in the sequel the following shortened notation:

$$S_0 := S_{\theta_0}(W_0), \quad \Sigma_0 := \Sigma_{\theta_0}(W_0), \quad a(\theta_0) := a(P_{\theta_0}).$$

To analyze the asymptotic local stability properties of a GMME we consider the asymptotic optimality problem

$$\min_{a \in \mathcal{A}} E_{\theta_0} h^\top(X_1; a) W_0 E_{\theta_0} h(X_1; a), \quad (10)$$

corresponding to (5). Its unique solution is assumed to be $a(P_{\theta_0})$ and to be in the interior of \mathcal{A} . The sequence of necessary (functional) equations

$$\left[E_{\theta_n} \frac{\partial h^\top(X_1; \tilde{a}(P_{\theta_n}))}{\partial a} \right] W_n [E_{\theta_n} h(X_1; \tilde{a}(P_{\theta_n}))] = 0, \quad (11)$$

defining the GMME then converges a.s to the implicit (functional) equation:

$$E_{\theta_0} \frac{\partial h^\top(X_1; a(P_{\theta_0}))}{\partial a} W_0 E_{\theta_0} h(X_1; a(P_{\theta_0})) = 0. \quad (12)$$

In order to describe the stability properties of a GMME in a neighborhood of P_{θ_0} we introduce the following well known concept from the theory of robust statistics; cf. also Hampel et al. (1986).

Definition 1 The influence function $IF(\cdot; \tilde{a}, P_{\theta_0})$ of a statistical functional¹ \tilde{a} is given by

$$IF(x; \tilde{a}, P_{\theta_0}) = \lim_{\epsilon \rightarrow 0} \frac{\tilde{a}((1 - \epsilon)P_{\theta_0} + \epsilon\delta_x) - \tilde{a}(P_{\theta_0})}{\epsilon}, \quad (13)$$

for all δ_x such that this limit exists.

¹In the following we will always assume that the domain of the given statistical functional is an open convex subset of \mathcal{P} containing P_{θ_0} and all empirical measures.

As a consequence, the influence function of a statistical functional describes the linearized asymptotic bias of a statistic under single point contaminations δ_x of the assumed model distribution P_{θ_0} . An unbounded IF implies an unbounded asymptotic bias of a statistic under single point contaminations of the model. Therefore, a natural robustness requirement on a statistical functional is the boundedness of its influence function.

The influence function of a GMME is obtained by implicitly differentiating the necessary condition (12) in an arbitrary direction δ_x . Straightforward calculations then yield²:

$$IF(x; \tilde{a}, P_{\theta_0}) = -S_0 E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} W_0 h(x; a(\theta_0)) \quad . \quad (14)$$

Note that in deriving this expression we used condition (4) which is satisfied by assumption at the model P_{θ_0} . As a consequence, we can see that:

- The IF of a GMME is linearly related to the orthogonality function of the model $h(\cdot; a(P_{\theta_0}))$.
- The IF of a GMME is bounded if and only if the orthogonality function of the underlying model is bounded.

Expression (14) covers as special cases well known situations where h is linear, as in Krasker and Welsch (1985), Krasker (1986) and in Lucas, van Dijk and Kloek (1994).

²In the exactly identified case ($k = H$) this expression simplifies to:

$$IF(x; \tilde{a}, P_{\theta_0}) = - \left[E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} \right]^{-1} h(x; a(\theta_0)) \quad ,$$

the standard expression for an M-estimator defined by a score function h ; cf. for instance Huber (1981).

It is well known that many econometric estimators can be interpreted as GMME, see Hansen (1982). Unfortunately, many of these turn out to be non-robust, because the corresponding function h is unbounded in the observations. Well-known examples in the (linear) instrumental variables framework were analyzed for instance in Krasker and Welsch (1985).

In addition to M-estimators that are defined through the roots of an implicit equation (these estimators can in fact be interpreted as particular GMME), there is a broad class of nonlinear GMME where the given nonlinearity is in contrast with the basic robustness principle of a bounded influence function³. Some examples are listed below.

Example 1: Nonlinear instrumental variables estimators (cf. Amemiya (1974)). Let $(X_t)_{t \geq 0} := (X_t^{(1)}, X_t^{(2)}, X_t^{(3)})_{t \geq 0}$ be a data generating process, with (for brevity) $X_t^{(1)}$ a scalar endogenous variable, $X_t^{(2)}$ a scalar exogenous variable and $X_t^{(3)}$ some instrumental variable inducing the orthogonality restrictions:

$$E \left[X_1^{(3)} (X_1^{(1)} - m(X_1^{(2)}; a)) \right] = 0 \quad ,$$

for some given nonlinear function m . Since the function

$$(x^{(1)}, x^{(2)}, x^{(3)}, a) \longmapsto x^{(3)} (x^{(1)} - m(x^{(2)}; a)) \quad ,$$

is unbounded at least in $x^{(1)}$ and $x^{(3)}$ all these estimators have unbounded IF. Moreover, for different nonlinear forms of m the robustness problems of a given instrumental variables estimator can be quite different. For instance, different polynomial forms of m can induce very different biases for the corresponding estimator under a slight single point contamination of the

³This point is even more important for deriving robust GMM testing procedures; cf. Section 4.

underlying model. It is then useful to have a general procedure for bounding this maximal bias independently of the general form of m .

Example 2: GMM estimation of autoregressive models with conditionally heteroscedastic errors. Let $(y_t)_{t \in \mathbb{N}}$ be the ARCH process as defined in (1) with associated orthogonality conditions (2). The function defining these orthogonality conditions is unbounded. Moreover, note that the observation y_{t-1} enters in the last of these four orthogonality conditions as a polynomial of degree four. Therefore, for some choices of the model parameters the influence function of the implied GMM estimator can be steep in some contamination directions. In Section 5 we will apply the RGMM methodology to this particular example by deriving a RGMM testing procedure for conditional heteroscedasticity in some simulation experiments.

Example 3: GMM estimation of nonlinear empirical asset pricing models (cf. for example Bansal, Hsieh and Viswanathan (1993)). Let a nonlinear pricing kernel $(G_t)_{t \in \mathbb{N}}$ be defined by

$$G_t = G(R_{f,t}, R_{M,t}) = \beta_0 + \beta_1 R_{f,t} + \sum_{j=1,3,5} \beta_{j,M} (R_{M,t})^j \quad , \quad (15)$$

where $(R_{f,t})_{t \in \mathbb{N}}$ and $(R_{M,t})_{t \in \mathbb{N}}$ are some corresponding series of yields to maturity of the Treasury bill and of an aggregate equity index, respectively.

Given a set of instrumental variables $(Z_t)_{t \in \mathbb{N}}$ and a set of n contingent claims pay-offs $(x_t^{(1)}, \dots, x_t^{(n)})_{t \in \mathbb{N}}$, the natural orthogonality conditions implied by the given asset pricing equation are:

$$E[G_{t+1}(x_{t+1}^{(i)} - 1)Z_t] = 0 \quad , \quad i = 1, \dots, n \quad . \quad (16)$$

These orthogonality conditions are again unbounded. Moreover, $R_{M,t+1}$ enters in all orthogonality conditions as a polynomial of degree five. Therefore, for some choices of the model parameters the influence function of the implied GMM estimator is again steep in some contamination directions.

3 Robust GMM Estimation

It is not possible to construct robust GMME that are optimal in the sense of Hampel et al. (1986), because a best ML-estimator of $a(P_{\theta_0})$ at the model is not generally available, even when its influence function is not required to be bounded. Specifically, the covariance matrices of the GMME induced by two different non-nested sets of orthogonality conditions are not generally rankable.

Instead, we require the bound on the IF to be satisfied in a norm that is self-standardized with respect to the covariance matrix of the given GMME. This norm measures the influence of the estimator \tilde{a} relative to its variability expressed by its covariance matrix. We will see in Section 4 that this is the appropriate norm for obtaining robust GMM testing procedures.

Formally, we look at GMME with a bounded self-standardized IF, that is satisfying:

$$\|IF(x; \tilde{a}, P_{\theta_0})\|_{\Sigma_0^{-1}} := \|\Sigma_0^{-\frac{1}{2}} IF(x; \tilde{a}, P_{\theta_0})\| \leq c \quad , \quad (17)$$

where c is a given prespecified positive constant. We can satisfy this condition for our RGMM estimator by bounding the self-standardized norm

$$\|h(x; a(\theta_0))\|_{V_0^{-1}} := \|V_0^{-\frac{1}{2}} h(x; a(\theta_0))\| \quad (18)$$

of h .

Indeed, (17) is satisfied when the self-standardized norm of h is bounded by c because⁴:

⁴In the exactly identified case ($H = k$) this inequality becomes an equality. In this situation the bound on the self-standardized IF of a GMM estimator provided by a bounded self-standardized norm of h is exact and the following arguments in this Section and in Section 4 still hold.

$$\begin{aligned}
\|IF(x; \tilde{a}, P_{\theta_0})\|_{\Sigma_0^{-1}}^2 &= h^\top(x; a(\theta_0))W_0E_{\theta_0}\frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} \\
&\quad \times \left[E_{\theta_0}\frac{\partial h^\top(X_1; a(\theta_0))}{\partial a}W_0V_0W_0E_{\theta_0}\frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} \right]^{-1} \\
&\quad \times E_{\theta_0}\frac{\partial h^\top(X_1; a(\theta_0))}{\partial a}W_0h(x; a(\theta_0)) \\
&\leq \|h(x; a(\theta_0))\|_{V_0^{-1}}^2 \quad ,
\end{aligned}$$

by the orthogonal projection property of the matrix:

$$\begin{aligned}
V_0^{\frac{1}{2}}W_0E_{\theta_0}\frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} \left[E_{\theta_0}\frac{\partial h^\top(X_1; a(\theta_0))}{\partial a}W_0V_0W_0E_{\theta_0}\frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} \right]^{-1} \\
\times E_{\theta_0}\frac{\partial h^\top(X_1; a(\theta_0))}{\partial a}W_0V_0^{\frac{1}{2}} \quad .
\end{aligned}$$

To construct a GMME with self-standardized influence function bounded by c , we introduce the Huber function $\mathcal{H}_c : \mathbb{R}^H \rightarrow \mathbb{R}^H; y \mapsto yw_c(y)$, defined by $w_c(y) := \min\left(1, \frac{c}{\|y\|}\right)$ for $y \neq 0$ and $w_c(0) := 1$, and a new mapping $h_c^{A, \tau} : \mathbb{R}^N \times \mathcal{A} \rightarrow \mathbb{R}^H$ given by

$$h_c^{A, \tau}(x; a) := \mathcal{H}_c(A[h(x; a) - \tau]) = A[h(x; a) - \tau]w_c(A[h(x; a) - \tau]) \quad , \quad (19)$$

for $x \in \mathbb{R}^N$ and $a \in \mathcal{A}$. The nonsingular matrix $A \in \mathbb{R}^{H \times H}$ and the vector $\tau \in \mathbb{R}^H$ are determined by the implicit equations⁵:

$$E_{\theta_0}h_c^{A, \tau}(X_1; a(P_{\theta_0})) = 0 \quad , \quad (20)$$

and

$$E_{\theta_0}h_c^{A, \tau}(X_1; a(P_{\theta_0}))h_c^{A, \tau^\top}(X_1; a(P_{\theta_0})) = I \quad . \quad (21)$$

⁵It is important to note that no further model assumptions are needed in order to perform this construction.

$a(\theta_0)$ can be estimated by the sequence of fixed-points of the algorithm described by (19), (20) and (21). Note that the bound imposed on the self-standardized influence of our GMME cannot be chosen arbitrarily small. Indeed, $c \geq \sqrt{H}$, cf. Hampel et al. (1986), p. 228.

In some robust applications in the iid framework, the functional form implied by (20) for the dependence of τ on a and A can be determined explicitly. For instance, in linear regression models with normal errors, symmetry implies $\tau = 0$; cf. Hampel et al. (1986), ch. 6.

To apply the algorithm to a general GMM situation we propose to estimate A via the sequence of solutions to the empirical version of (21) and to determine τ as the solution of (20) under the model probability P_{θ_0} . In some models - as for instance in the RGMM application presented in Section 5 - this will require a simulation procedure.

Specifically, for a given bound $c > \sqrt{H}$ on the self-standardized influence of \tilde{a} , the computation of the robust GMME can be performed by the following four steps.

- Fix a starting value a_0 for $a(\theta_0)$ and initial values $\tau_0 := 0$ and A_0 such that

$$A_0^\top A_0 = [E_{\theta_n}(h(X_1; a_0)h^\top(X_1; a_0))]^{-1} .$$

- Compute new values τ_1 and A_1 for τ and A defined by

$$\tau_1 := \frac{E_{\theta_0}[h(X_1; a_0)w_c(A_0(h(X_1; a_0) - \tau_0))]}{E_{\theta_0}w_c(A_0(h(X_1; a_0) - \tau_0))} , \quad (22)$$

and

$$(A_1^\top A_1)^{-1} := E_{\theta_n} \left[(h(X_1; a_0) - \tau_0)(h(X_1; a_0) - \tau_0)^\top \times w_c^2(A_0(h(X_1; a_0) - \tau_0)) \right] . \quad (23)$$

- Compute the optimal GMME a_1 associated⁶ to the orthogonality function $h_c^{A_1, \tau_1}$.
- Replace τ_0 and A_0 by τ_1 and A_1 , respectively, and iterate the second and third step described above until convergence of the sequence of optimal GMME $(a_n)_{n \in \mathbb{N}}$ associated with the sequence $(h_c^{A_n, \tau_n})_{n \in \mathbb{N}}$ of bounded orthogonality functions.

The robust GMME obtained in this way can be interpreted as the GMME induced by the truncated orthogonality conditions $h_c^{A, \tau}$ when satisfying the orthogonality condition (20) for τ and when simultaneously estimating A through the empirical version of (21).

$h_c^{A, \tau}$ is a truncated version of h . Because of the truncation, h must be shifted by τ in order to satisfy the orthogonality condition (20). Moreover, (21) ensures that c is an upper bound on the self-standardized influence function of the corresponding GMME, because – by construction – the self-standardized norm of $h_c^{A, \tau}$ is equal to its euclidean norm which itself is bounded by the constant c .

Existence and uniqueness of a solution $(a(P_{\theta_0}), \tau(P_{\theta_0}), A(P_{\theta_0}))$ are implied by the implicit function theorem and the Fréchet differentiability of the equation system defining the GMME of $(a(P_{\theta_0}), \tau(P_{\theta_0}), A(P_{\theta_0}))$ in a neighborhood of P_{θ_0} , which itself is implied by the boundedness of the function $h_c^{A, \tau}$; cf. for example Clarke (1986), Bednarski (1993). More specific conditions for a special model can be found in Krasker and Welsch (1985). Regularity conditions for consistency and asymptotic normality of a GMM estimator at the model P_{θ_0} are provided in Hansen (1982).

⁶Note that (by construction) the optimal asymptotic weighting matrix associated to this particular GMME is the identity matrix.

Whereas the original moment conditions h are usually dictated by economic theory, the truncated version $h_c^{A,\tau}$ takes into account the realistic case that only the "majority of the data" can reasonably fit the original moment conditions. The weights $w_c(A[h(x; a) - \tau])$ assigned to each observation x can be used to detect outlying points. The tuning constant $c \in (\sqrt{H}, \infty)$ controls the degree of robustness imposed on the procedure. It can be chosen by the analyst as a trade-off between her theoretical moment conditions and those supported by the data. Some objective guidelines for the choice of c are presented in the next section where we directly focus on RGMM hypothesis testing. There we show that for a given amount of model contamination the constant c can be determined so that the maximal bias in the level or the power of a GMM test remains below a given bound.

As pointed out by a referee, one possible disadvantage of the RGMM estimator defined above could be the well-known poor performance in small samples of GMM estimators when the asymptotic covariance matrix of the given orthogonality function is estimated; cf. for instance Koenker, Machado, Skeels, Welsh (1994). Some protection in this respect should be supplied by the fact that the orthogonality function $h_c^{A,\tau}$ behind our RGMM estimator is bounded; nevertheless we expect the issue of a covariance matrix estimation in our RGMM framework to be particularly important when the number of orthogonality conditions is "high" (for example of an order higher than $n^{\frac{1}{3}}$; cf. Koenker and Machado (1999)).

One possibility to improve the small sample performance of our RGMM estimator is to use an empirical likelihood version of GMM as proposed by Imbens, Spady and Johnson (1998). Specifically, one can use as estimating equations for the empirical likelihood those given by formula (5), p. 337 or (9) p. 339 in their paper with (in their notation) $\psi(z, \theta) = h_c^{A,\tau}(z, \theta)$. The

boundedness of the function $h_c^{A,\tau}$ will preserve the robustness properties of the estimator while the empirical likelihood version should improve the finite sample performance.

4 Robust Inference with GMM Estimators

This section is devoted to the robustness properties of GMM based test statistics. The key idea in deriving RGMM procedures is to construct GMM estimators based on a bounded self-standardized norm of the given orthogonality function, as for instance in the case of the RGMM estimator defined in the last section. For simplicity of notation we will derive all results for the case of an optimal GMME (that is a weighting matrix $W_0 = V_0^{-1}$) based on a bounded orthogonality function. Modifications to the general case are straightforward⁷.

Several tests derived from a GMME can be constructed, for testing some misspecification of the model or some set of parameter restrictions on a . The GMM specification test proposed by Hansen (1982) is a test of the overidentifying restrictions implied by the null hypothesis given by (4), for the case where $H > k$.

The asymptotic distribution of the statistic defining Hansen's test with respect to a sequence of local misspecifications is a noncentral $\mathcal{X}^2(H - k, \beta)$ distribution. At the model $\beta = 0$; see Newey (1985).

The GMM versions of the classical ML-tests are used to test a null hy-

⁷Only the results obtained for the likelihood ratio statistic are not available in the non-optimal GMM case since the corresponding statistic is then no longer asymptotically equivalent to a symmetric functional form (see below).

pothesis

$$g(a(\theta_0)) = 0 \quad , \quad (24)$$

for a smooth function $g : \mathcal{A} \rightarrow \mathbb{R}^r$ such that $\frac{\partial g^\top}{\partial a}(a(\theta))$ is of full column rank for all $\theta \in \Theta$.

The Wald, score and likelihood-ratio statistics induced from a best GMME are all asymptotically equivalent under the null hypothesis (24) and with respect to a sequence of local alternatives to $a(\theta_0)$. They are asymptotically noncentral $\mathcal{X}^2(r; \beta)$ distributed, with a noncentrality parameter $\beta = 0$ at the model under the null hypothesis (24).

We restrict our attention to GMM test statistics that can be written (at least asymptotically) as simple quadratic forms of a functional U . Specifically, we consider functionals ξ defined asymptotically by a symmetric form

$$n\xi(P_{\theta_n}) = nU(P_{\theta_n})^\top U(P_{\theta_n}) \quad ; \quad n \in \mathbb{N} \quad , \quad (25)$$

and consider the following test statistics.

Hansen's test: Hansen's statistic (ξ^M) is of the symmetric form (25) with a functional U defined by

$$U^M(P_{\theta_n}) := W_0^{\frac{1}{2}} E_{\theta_n} h(X_1; \tilde{a}(P_{\theta_n})) \quad . \quad (26)$$

A consistent estimator of W_0 is given by the sequence \mathcal{W} of positive definite estimators.

Wald-type test: The statistic of a GMM based Wald-type test (ξ^W) is of the form (25) with a functional U defined by

$$U^W(P_{\theta_n}) := \left[\frac{\partial g}{\partial a^\top}(a(\theta_0)) \Sigma_0 \frac{\partial g^\top}{\partial a}(a(\theta_0)) \right]^{-\frac{1}{2}} g(\tilde{a}(P_{\theta_n})) \quad . \quad (27)$$

Practically,

$$\frac{\partial g}{\partial a^\top}(a(\theta_0)) \Sigma_0^{-1} \frac{\partial g^\top}{\partial a}(a(\theta_0)) \quad ,$$

is approximated by estimating Σ_0 with $\Sigma_{\theta_n}(W_n)$ and $a(\theta_0)$ with $\tilde{a}(P_{\theta_n})$.

Score Type Tests: The statistic of a GMM based score-type test (ξ^S) is of the form (25) with a functional U defined by

$$U^S(P_{\theta_n}) := \hat{\Sigma}_0^{\frac{1}{2}} E_{\theta_n} \frac{\partial h^\top(X_1; \hat{a}(P_{\theta_n}))}{\partial a} W_n E_{\theta_n} h(X_1; \hat{a}(P_{\theta_n})) \quad , \quad (28)$$

where $\hat{a}(P_{\theta_n})$ is a solution to a constrained GMM minimization problem:

$$\hat{a}(P_{\theta_n}) = \arg \min_{a \in \mathcal{A}, g(a)=0} E_{\theta_n} h^\top(X_1; a) W_n E_{\theta_n} h(X_1; a) \quad , \quad (29)$$

and $\hat{\Sigma}_0$ is the covariance matrix (9) evaluated at $a = \hat{a}(P_{\theta_0})$. In applications $\hat{\Sigma}_0$ can be consistently estimated by $\hat{\Sigma}_{\theta_n}(W_n)$.

Likelihood ratio-type test: The GMM likelihood ratio type-test is constructed with a statistic ξ^R that can be written asymptotically as a symmetric form. It is defined by

$$\begin{aligned} \xi^R(P_{\theta_n}) := & E_{\theta_n} h^\top(X_1; \hat{a}(P_{\theta_n})) W_0 E_{\theta_n} h(X_1; \hat{a}(P_{\theta_n})) \\ & - E_{\theta_n} h^\top(X_1; \tilde{a}(P_{\theta_n})) W_0 E_{\theta_n} h(X_1; \tilde{a}(P_{\theta_n})) \quad . \end{aligned}$$

Asymptotically one has

$$\xi^R(P_{\theta_n}) = U^H(P_{\theta_n})^\top U^H(P_{\theta_n}) + o_p(1) \quad ,$$

with a functional U^H that is explicitly given by (47) in the proof of Theorem 1.

As mentioned in the introduction, the general goal of robust testing procedures is to control the maximal bias on the level and the power of a test that can arise because of a slight distributional misspecification of a null or an alternative hypothesis. This is called robustness of validity and efficiency, respectively.

To analyze the asymptotic local stability properties of these tests we follow the general approach proposed by Heritier and Ronchetti (1994). In

order to apply this methodology to the GMM setting we can assume the following uniform convergence to normality of a robust GMME.

Property 1 Let a bounded influence GMME \tilde{a} of $a(P_{\theta_0})$ be given. It then follows:

$$\sqrt{n}(\tilde{a}(P_{\theta_n}) - \tilde{a}(P_{\epsilon,n,Q}^0)) \rightarrow \mathcal{N}(0, \Sigma_0) \quad , \quad n \rightarrow \infty \quad , \quad (30)$$

in distribution, uniformly over the sequence $(\mathcal{U}_{\epsilon,n}(P_{\theta_0}))_{n \in \mathbb{N}}$ of (ϵ, n) -neighborhoods of P_{θ_0} defined by

$$\mathcal{U}_{\epsilon,n}(P_{\theta_0}) := \left\{ P_{\epsilon,n,Q}^0 := \left(1 - \frac{\epsilon}{\sqrt{n}}\right) P_{\theta_0} + \frac{\epsilon}{\sqrt{n}} Q \quad \mid \quad Q \in \text{dom}(\tilde{a}) \right\} \quad , \quad (31)$$

where the assumptions on $\text{dom}(\tilde{a})$ are given in Footnote 1.

The neighborhood defined by (31) is probably the simplest way to formalize local perturbations of the model P_{θ_0} . Note that $d_K(P_{\epsilon,n,Q}^0, P_{\theta_0}) \leq \epsilon$ for all $n \in \mathbb{N}$ and $Q \in \text{dom}(\tilde{a})$, where d_K denotes the Kolmogoroff distance. Alternatively, one could use more involved notions of distance between distributions.

Property 1 is stronger than the requirement of the existence of the influence function. Generally, one needs a stronger smoothness condition like Fréchet differentiability in order to obtain uniform convergence; cf. Clarke (1986) and Bednarski (1993). However, under appropriate regularity conditions (cf. Clarke (1986) and Heritier and Ronchetti (1994)), bounded influence statistical functionals can be shown to be Fréchet differentiable. As a particular case, the robust RGMM estimator proposed in the last section is Fréchet differentiable.

The next theorem provides a maximal asymptotic bias of the level of a GMM test⁸.

⁸The proofs of all theorems are given in the Appendix.

Theorem 1 Let \tilde{a} be a GMME induced by a bounded orthogonality function h and denote by α the level functional of the tests based on ξ^M , ξ^W , ξ^S and ξ^R , respectively. Let further $(P_{\epsilon,n,Q}^0)_{n \in \mathbb{N}}$ be a sequence of (ϵ, n, Q) -contaminations of the underlying null distribution P_{θ_0} , each of them belonging to a corresponding neighborhood $\mathcal{U}_{\epsilon,n}(P_{\theta_0})$, as defined in (31).

Then:

$$\lim_{n \rightarrow \infty} \alpha(P_{\epsilon,n,Q}^0) = \alpha_0 + \epsilon^2 \mu \left\| \int_{\mathbb{R}^N} IF(x; U, P_{\theta_0}) dQ(x) \right\|^2 + o(\epsilon^2), \quad \epsilon \rightarrow 0, \quad (32)$$

for all $Q \in \text{dom}(\alpha)$, where $U(\cdot)$ is the U -functional corresponding to each test,

$$\mu = -\frac{\partial}{\partial \beta} H_r(\eta_{1-\alpha_0}; \beta)|_{\beta=0} = \frac{(1-\alpha_0)}{2} - \frac{1}{2} H_{r+2}(\eta_{1-\alpha_0}; 0) \quad ,$$

$H_r(\cdot; \beta)$ is the cumulative distribution function of a noncentral $\mathcal{X}^2(r; \beta)$ distribution with r degrees of freedom and noncentrality parameter $\beta \geq 0$, $\eta_{1-\alpha_0}$ is the $1 - \alpha_0$ quantile of a $\mathcal{X}^2(r; 0)$ distribution and $\alpha_0 = \alpha(P_{\theta_0})$ is the nominal level of the test. Moreover, the bias of $\alpha(P_{\epsilon,n,Q}^0)$ is uniformly bounded by the inequality:

$$\lim_{n \rightarrow \infty} |\alpha(P_{\epsilon,n,Q}^0) - \alpha_0| \leq \epsilon^2 \cdot \mu \cdot \sup_x \|h(x; a(\theta_0))\|_{W_0}^2 + o(\epsilon^2) \quad .$$

As a consequence of the theorem, the maximal asymptotic bias of the level of a GMM test that is derived from the robust GMME of the last section can be bounded by the inequality

$$\lim_{n \rightarrow \infty} |\alpha(P_{\epsilon,n,Q}^0) - \alpha_0| \leq \mu \cdot (\epsilon c)^2 + o(\epsilon^2) \quad . \quad (33)$$

For robust testing purposes the asymptotic bound (33) can be used to choose c depending on both the maximal amount of contamination (ϵ) expected by

the researcher – given some prior information on the nature of the data – and the maximal bias for the level (*maxbias*) he or she is willing to accept:

$$c = \frac{1}{\epsilon} \cdot \sqrt{\frac{\text{maxbias}}{\mu}} \quad . \quad (34)$$

The following table presents the implied c values for $\epsilon = 5\%$, $\text{maxbias} = +/- 0.5\%$ and $\alpha_0 = 5\%$.

Insert Table 1 about here

By regressing $\log \mu$ vs. $\log r$ and for the case of a nominal level $\alpha_0 = 5\%$ at the model, one can obtain the simple approximation:

$$c \approx \frac{3}{\epsilon} r^{0.3} (\text{maxbias})^{1/2} \quad . \quad (35)$$

We now come back to the robustness of efficiency properties of a GMM test and first investigate the case of a GMM specification test.

Let

$$(P_{\eta,n}^{alt})_{n \in \mathbb{N}} := \left(\left(1 - \frac{\eta}{\sqrt{n}} \right) P_{\theta_0} + \frac{\eta}{\sqrt{n}} P_{\theta_1} \right)_{n \in \mathbb{N}} \quad (36)$$

be a sequence of local alternatives to P_{θ_0} and

$$\mathcal{U}_{\epsilon,n}(P_{\eta,n}^{alt}) := \left\{ P_{\epsilon,n,Q}^1 := \left(1 - \frac{\epsilon}{\sqrt{n}} \right) P_{\eta,n}^{alt} + \frac{\epsilon}{\sqrt{n}} Q \mid Q \in \text{dom}(\tilde{\alpha}) \right\} \quad , \quad (37)$$

be the corresponding asymptotic neighborhood of $P_{\eta,n}^{alt}$, for given n .

A natural restriction on the magnitude of the contamination is $|\epsilon| < |\eta|$. This allows us to distinguish the neighborhood $\mathcal{U}_{\epsilon,n}(P_{\eta,n}^{alt})$ of the local alternative from the given null hypothesis. On the other side, one could of course compare a given neighborhood $\mathcal{U}_{\epsilon_0,n}(P_{\theta_0})$ of the null hypothesis with

a neighborhood $\mathcal{U}_{\epsilon,n}(P_{\eta,n}^{alt})$ of the local alternative. In this case a natural restriction will be $|\epsilon_0| + |\epsilon| < |\eta|$.

The next theorem is the "power" counterpart of Theorem 1 for the GMM specification test. Similarly to the case of the level, the theorem yields an explicit asymptotic bound by which the maximal asymptotic bias of the power can be bounded.

Theorem 2 Let \tilde{a} be a GMME induced by a bounded orthogonality function h and denote by π the power functional of the test based on ξ^M . Let further $(P_{\epsilon,n,Q}^1)_{n \in \mathbb{N}}$ be a sequence of (ϵ, n, Q) -contaminations of the underlying local alternatives $P_{\eta,n}^{alt}$, each of them belonging to a corresponding neighborhood $\mathcal{U}_{\epsilon,n}(P_{\eta,n}^{alt})$, as defined in (37).

Then:

$$\begin{aligned} \lim_{n \rightarrow \infty} |\pi(P_{\epsilon,n,Q}^1) - \pi(P_{\eta,n}^{alt})| &= 2\mu\epsilon\eta \int_{\mathbb{R}^N} IF^\top(x; U^M, P_{\eta,n}^{alt}) dQ(x) \\ &\quad \times \int_{\mathbb{R}^N} IF(x; U^M, P_{\theta_0}) dP_{\theta_1}(x) + o(\eta), \end{aligned} \quad (38)$$

with μ defined as in Theorem 1. Moreover, the bias of the asymptotic power functional π is uniformly bounded by the inequality:

$$\lim_{n \rightarrow \infty} |\pi(P_{\epsilon,n,Q}^1) - \pi(P_{\eta,n}^{alt})| \leq 2\mu\epsilon\eta \max_{\{P_{\eta,n}^{alt}, P_{\theta_0}\}} \sup_x \|h(x; \tilde{a}(\cdot))\|_{W_0}^2 + o(\eta). \quad (39)$$

Similarly to the case for the level, the maximal asymptotic bias of the power of a GMM specification test derived from the RGMM estimator of the last section can be estimated by the inequality:

$$\lim_{n \rightarrow \infty} |\pi(P_{\epsilon,n,Q}^1) - \pi(P_{\eta,n}^{alt})| \leq 2\mu\epsilon\eta c^2 + o(\eta) \quad . \quad (40)$$

As in the case of the level (see (33)) this inequality can be used to relate the tuning constant c of our RGMM estimator to the maximal bias in the

power of the GMM specification test, given a nominal level α_0 at the model. For instance, assuming $H - k = 1$ and $\epsilon = 5\%$, a bound of 0.5% on the bias from a nominal level of 5% implies $c = 4.18$ and $\mu = 0.1145$ (cf. Table 1). This yields an absolute maximal bias in the power of a corresponding RGMM specification test given by 0.20η . For example for $\eta = 15\%$ the implied maximal bias in the power is approximatively 3%.

Theorem 2 illustrates the trade off existing between power and robustness of a GMM specification test. Indeed, for a given maximal bias over the contaminated neighborhood $\mathcal{U}_{\epsilon,n}(P_{\eta,n}^{alt})$ one cannot impose stronger robustness requirements on a RGMM estimator (that is a lower constant c) without simultaneously looking at local alternatives that are more distant in the given direction P_{θ_1} (that is with a higher constant η). On the other side, imposing stronger robustness requirements by a lower constant c reduces the maximal bias from the power of the given local alternative $P_{\eta,n}^{alt}$. However, for near local alternatives (and therefore low values of η) this will correspond to a low power of the RGMM specification test over the full contaminated neighborhood $\mathcal{U}_{\epsilon,n}(P_{\eta,n}^{alt})$.

We conclude this section by discussing the robustness of efficiency properties of the GMM based Wald, score, and likelihood ratio tests. Consider again the neighborhood defined by (31). For $P_{\epsilon,n,Q}^0 \in \mathcal{U}_{\epsilon,n}(P_{\theta_0})$ we define a sequence of parametric local alternatives to (24) by

$$g\left(a(\theta_0) + \frac{\Delta}{\sqrt{n}}\right) = 0 \quad , \quad (41)$$

with a non-zero vector $\Delta \in \mathbb{R}^k$.

Similarly to the case of the GMM specification test, a natural restriction on the magnitude of the contamination is $|\epsilon| < |\Delta|$. This allows us to distinguish the neighborhood $\mathcal{U}_{\epsilon,n}(P_{\theta_0})$ of the local alternative from the given null hypothesis. The next theorem is the power counterpart of Theorem 1 for the

maximum-likelihood-type GMM tests. Similarly to Theorem 2 an explicit asymptotic bound for the maximal bias in the power of a parametric GMM test is provided.

Theorem 3 Let \tilde{a} be a GMME induced by a bounded orthogonality function h and denote by π the power functional of the test based on ξ^W , ξ^S and ξ^R , respectively.

Then, the bias of π is uniformly bounded by the inequality:

$$\lim_{n \rightarrow \infty} |\pi(P_{\epsilon, n, Q}^0) - \pi(P_{\theta_0})| \leq 2\mu\epsilon \|\Delta\|_{\Sigma_0^{-1}} \sup_x \|h(x; a(\theta_0))\|_{W_0} + o(\Delta) \quad , \quad (42)$$

where μ is defined as in Theorem 1.

Similarly to the case for the power of a GMM specification test, the maximal asymptotic bias of the power of a parametric GMM test derived from the RGMM estimator of the last section can be estimated by the inequality:

$$\lim_{n \rightarrow \infty} |\pi(P_{\epsilon, n, Q}^0) - \pi(P_{\theta_0})| \leq 2\mu\epsilon \|\Delta\|_{\Sigma_0^{-1}} c + o(\Delta) \quad . \quad (43)$$

This bound can be used to relate the choice of c to the maximal bias in the power of a parametric GMM test, given a nominal value α_0 at the model.

5 An Application to RGMM Testing for Conditional Heteroscedasticity

In this section we consider a simple application of our RGMM methodology to a test for ARCH structures in the errors of a linear autoregressive model. The goal is not to perform a full analysis of the robustness properties of ARCH testing procedures but to outline the performance of the RGMM in

a simple application as well as the algorithm used to compute the RGMM estimator of Section 3.

Let $(y_t)_{t \in \mathbb{N}}$ be the autoregressive process (1) with ARCH(1) error terms presented in Example 2. Moreover, consider the orthogonality conditions given by (2) and (3).

A test for a constant conditional variance specification of ϵ_t could be a Hansen specification test for the overidentifying orthogonality conditions implied by the null hypothesis $\alpha_1 = 0$ against an alternative hypothesis $\alpha_1 > 0$.

Note that in the present formulation we treat all parameters that have to be estimated under the null hypothesis symmetrically. Of course, one could easily develop a two-stage RGMM testing procedure if (β_0, β_1) is treated as a vector of nuisance parameters.

To construct a GMM test for conditional heteroscedasticity behaving satisfactorily under local deviations from normality, we consider as a reference model for y_t an autoregressive model with normally distributed errors u_t and compute the RGMM estimator presented earlier for a given choice of the tuning constant c (see below). By construction, the above RGMM test of induced by this RGMM estimator maintains "good" level and power properties under local deviations from the given reference model in a way that is formalized through the inequalities obtained in Theorem 1 to 3.

To compare the performance of the given GMM and RGMM tests for ARCH we simulate the following distributions "near" the normal distribution as candidate models of a possible data generating process for u_t .

1. *Standard normal*

In this experiment we compare the efficiency of the RGMM and the classical GMM testing procedures at the given reference model.

2. *Contaminated normal $CN(\epsilon, K^2)$*

$$F(x) = (1 - \epsilon)\Phi(x) + \epsilon\Phi\left(\frac{x}{K}\right) \quad , \quad x \in \mathbb{R} \quad , \quad (44)$$

where Φ is the cumulative distribution function of a standard normal random variable. Here, we investigate the performance of the classical GMM and the RGMM under a known maximum distance ϵ from the standard normal model and a given degree of contaminating variance K^2 . We simulate this case for a distance $\epsilon = 0.05$ and a very high contaminating variance $K^2 = 100$. This choice is quite extreme. However, it allows us to compare the performances of the RGMM and the GMM under dramatic symmetric deviations from normality that could occur over a short time period in real data.

3. *Student t_ν with ν degrees of freedom*

We consider the cases $\nu = 5, 9$ that allow for the existence of the fourth and the eighth conditional moments of u_t , respectively. Note that the t_9 and t_5 distributions are already very near to the normal. As a consequence, in this example we can compare the numerical performance of the robust and the classical GMM when very small deviations from normality are present. Moreover, in the t_5 case we can investigate the impact of the non-existence of some theoretical conditional moments of u_t (assumed finite by the GMM).

4. *Double exponential DE*

This distribution has a symmetric convex density. It is therefore qualitatively different from the normal already in the center of the distribution. Furthermore, it displays fat tails somewhere between the t_5 and the $CN(0.05, 10^2)$ distribution.

All simulated error distributions were scaled in order to have variance 1. This small simulation design covers a good spectrum of tail behaviours for distributions of u_t that have heavier tails than the normal and still satisfy minimal moment requirements. Indeed, the tail indices (cf. Gasko and Rosenberger (1983) p. 322) of these distributions are 1 for the standard normal distribution, 1.16 and 1.34 for the Student t_9 and t_5 distributions respectively, 1.63 for the double exponential, and 3.42 for the contaminated normal. For comparison, a standard Cauchy distribution has a tail index of 9.22.

We simulate (1) for the parameter choice $(\beta_0, \beta_1, \alpha_0) = (0.4, 0.3, 0.25)$ and for different values of α_1 , ranging from 0 to 0.3, under the different distributions for u_t presented above and for sample sizes $T = 250, 500, 1000$. Note that for $\alpha_1 > \frac{1}{3}$ the fourth unconditional moments of u_t do not exist even under normality of u_t .

As an illustration, some QQ plots⁹ of the unconditional distribution of a process (y_t) without and with ARCH effects ($\alpha_1 = 0$ and $\alpha_1 = 0.2$ respectively) for some of the distributions considered above are presented in Figure 1 and 2.

Insert Figure 1 and 2 about here

From these graphs one can see that the effects on the unconditional distribution of y_t of a "slight" modification of the conditional distribution of u_t can be quite important, when ARCH effects are present. This is particularly true for the tails of the induced distributions. As expected, a given tail index of a conditional distribution for u_t induce fatter tails in the unconditional distributions of y_t when ARCH structures are present.

Each model is simulated 1000 times. The corresponding empirical rejection frequency for the RGMM and the GMM Hansen's test is calculated for

⁹All QQ plots are based on simulated samples of 5000 observations.

a fixed nominal level of 5%. The estimated standard error of the empirical rejection frequency \hat{p} is given by (using the binomial distribution) $\left(\frac{\hat{p}(1-\hat{p})}{1000}\right)^{\frac{1}{2}}$. It is 0.7%, 1.0% and 1.5% for $\hat{p} = 5\%$, 20%, 50%, respectively.

The tuning constant for the RGMME was set at $c = 2.09$. This allows to obtain a maximal bias of $\pm 0.5\%$ in the level of the RGMM test also for contaminations $\epsilon = 10\%$ (cf. Table 1 above) of the unconditional distribution of y_t . We imposed such a strong robustness restriction on our RGMME because the unconditional distribution of y_t shows even fatter tails than the conditional distribution of u_t when ARCH effects are present, a fact that can make the distance between the induced unconditional distributions of y_t larger than the distance between the assumed conditional distributions of u_t .

For each simulation run we used $a_0 = (0.2, 0.2, 0.2)$ as a starting point for the algorithm and always obtained convergence.

In the first step of the algorithm we set $\tau = 0$ and updated the matrix A after having estimated the covariance matrix of $h(X_1, a_0)$ with a Newey West (1987b) covariance matrix estimator. In the second step of the algorithm we simulated an ARCH process corresponding to the parameter choice a_0 and computed the expectations needed to solve (22) and thereby obtain τ_1 . A_1 is obtained after having estimated the covariance matrix of $(h(X_1, a_0) - \tau_0) \cdot w_c(A_0(h(X_1, a_0) - \tau_0))$ with a Newey-West (1987b) covariance matrix estimator. Note that at this stage of the algorithm an autocorrelation robust covariance matrix estimator is necessary even when $(h(X_t, a_0))_{t \in \mathbb{N}}$ is conditionally uncorrelated because this does not generally imply that $(h_c^{A_0, \tau_0}(X_t, a_0))_{t \in \mathbb{N}}$ is uncorrelated. In the third step we computed the GMME a_1 associated to the orthogonality function $h_c^{A_1, \tau_1}$. The second and third step above are then iterated until convergence of the sequence $(a_n)_{n \in \mathbb{N}}$ of GMME associated to the sequence $(h_c^{A_n, \tau_n})_{n \in \mathbb{N}^*}$ of bounded

orthogonality functions.

The results are presented in Table 2 to 6. Although the goal of this simulation is not to perform a full analysis of the robustness properties of ARCH testing procedures, some of the features obtained are worth noting.

Insert Table 2 to 6 about here

First of all, the RGMM test yields empirical sizes that are very stable across all simulated distributions (generally between 0.02 and 0.03). On the other side, the empirical sizes of the classical GMM tests are far less stable ranging between 0.05 (for the normal distribution and for a sample size $T = 1000$) and 0.11 (in some experiments with the DE and t_5 distributions) for distributions that are not "too far" from the normal. In the case of heavier tails (the contaminated normal case) the classical GMM test breaks down. Although this simulation is too limited to draw final conclusions, we observe empirical sizes reflecting a rather "conservative" behaviour of the RGMM test and a drastic liberal behaviour of the classical GMM test. Indeed, already under normality the empirical sizes of the classical GMM test are often higher than the given nominal level of 5% (for sample sizes $T = 250, 500$).

Secondly, the RGMM test yields empirical power curves that are fairly stable for almost all the simulated distributions. In particular, for distributions that are not "too far" from the normal (the DE , the t_9 and the t_5 distributions) the empirical asymptotic power when $T = 1000$ deviates from that obtained under normality by no more than ± 0.09 (with a maximal absolute deviation of 0.09 obtained for the DE case when $\alpha_1 = 0.2$). In the contaminated normal case differences are larger. However, note that in this case the classical GMM test does not even produce a monotonically increasing power curve.

The stability of the RGMM is paid for through a loss in power under normality, compared to the classical GMM test. For instance, when $\alpha_1 = 0.1$ the power of the robust GMM test under normality is half that of the classical one. Somehow surprisingly however, this clear power advantage is already lost for very small deviations from normality. Indeed, for the different sample sizes the classical and robust power curves in the t_9 experiment are quite comparable, with some small advantages for the GMM (RGMM) for large (small) sample sizes. For larger deviations, (the t_5 , the DE and the CN case) the power of the RGMM is clearly higher than that of the classical test. This suggests that in real data applications already very small contaminations of the underlying model could affect the efficiency of classical GMM testing procedures. On the other side, a RGMM procedure could then be helpful in maintaining this efficiency loss below a given bound.

6 Conclusions

We derived a RGMM estimator that generates robust tests for a broad class of GMM test statistics. Special cases are Hansen's specification test and likelihood-type GMM tests like the Wald, the score and the likelihood ratio test.

We presented an algorithm to compute our RGMM estimator, in which the degree of robustness required by a researcher can be controlled through the choice of an appropriate tuning constant c .

We explicitly related the choice of this tuning constant to two key variables: the amount of contamination that one can realistically assume with respect to a given data set and to the available data information, and the maximal bias of level and (or) power of a GMM test that one is ready to

admit for the given test.

In some simulated experiments we presented evidence that the optimal performance of a GMM test at the model can be strongly worsened even when small deviations are present. In these experiments the RGMM testing procedure behaves well in controlling for small distributional deviations from the assumptions. Moreover, the efficiency loss at the model of the RGMM procedure seems to be reasonable when considering its performance under small model misspecifications.

Further research on RGMM testing includes the study of its performance under more general model structures and model deviations (for instance asymmetric deviations) than those presented above. Applications to more complex macroeconomic and financial models where a reference model for the data distribution can be assumed could produce interesting robust results that can be compared with those obtained with classical methodologies. Finally, a further issue is the small sample behaviour of RGMM statistics.

Acknowledgments

The authors thank the Co-Editor, the Associate Editor and three referees for very valuable comments that improved the presentation of the paper.

Appendix

Proof of Theorem 1: We prove the statement of the theorem only for the score and likelihood ratio statistics. Those for ξ^M and ξ^W can be proved by similar arguments. As noted after Property 1 \tilde{a} is Fréchet differentiable¹⁰. This implies the Fréchet differentiability of U^S . A first order von Mises (1947) expansion of U^S then gives up to terms of order $o(\epsilon)$

$$\sqrt{n}(U^S(P_{\theta_n}) - U^S(P_{\epsilon,n,Q}^0)) \rightarrow \mathcal{N}(0, I_r) \quad , \quad n \rightarrow \infty \quad , \quad (45)$$

in distribution uniformly for all $Q \in \text{dom}(U^S)$, using (30). As shown by Heritier and Ronchetti (1994) the asymptotic level under contamination of the corresponding symmetric test functional induced by U^S can be then approximated by the second order expansion given by (32) with $U = U^S$. Note that the equality for μ in the statement of the theorem is obtained by a result of Johnson and Kotz (1991), ch. 28, p. 132, formula 1. Then, by (4) and using the hypothesis $g(a(P_{\theta_0})) = 0$, we obtain:

$$IF(x; U^S, P_{\theta_0}) = \Sigma_0^{\frac{1}{2}} E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} W_0 \times \left(E_{\theta_0} \frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} IF(x; \hat{a}, P_{\theta_0}) + h(x; a(\theta_0)) \right) .$$

The constrained GMM estimator $(\hat{a}(P_{\theta_n}), \hat{\lambda}(P_{\theta_n}))$ is defined by the system of first order conditions

$$E_{\theta_n} \frac{\partial h^\top(X_1; \hat{a}(P_{\theta_n}))}{\partial a} W_n E_{\theta_n} h(X_1; \hat{a}(P_{\theta_n})) - \frac{\partial g^\top}{\partial a}(\hat{a}(P_{\theta_n})) \hat{\lambda}(P_{\theta_n}) = 0$$

$$g(\hat{a}(P_{\theta_n})) = 0 \quad ,$$

where $\hat{\lambda} : \text{dom}(\hat{\lambda}) \rightarrow \mathbb{R}^r$ is the corresponding statistical functional of Lagrange multipliers. Differentiating implicitly the limit version of these necessary conditions in direction $\delta_x \in \text{dom}(\hat{a})$ – while imposing (4) and $\hat{\lambda}(P_{\theta_0}) = 0$ – and solving the corresponding system of implicit equations gives

$$IF(x; \hat{a}, P_{\theta_0}) = (I - \Sigma_0 M_0) IF(x; \tilde{a}, P_{\theta_0}) \quad , \quad (46)$$

¹⁰See Clarke (1986), Bednarski (1993) and Heritier and Ronchetti (1994).

where

$$M_0 = \frac{\partial g^\top}{\partial a}(a(\theta_0)) \left(\frac{\partial g}{\partial a^\top}(a(\theta_0)) \Sigma_0 \frac{\partial g^\top}{\partial a}(a(\theta_0)) \right)^{-1} \frac{\partial g}{\partial a^\top}(a(\theta_0)) \quad .$$

Inserting this result in (46) and using (14) with $W_0 = V_0^{-1}$ yields

$$\|IF(x; U^S, P_{\theta_0})\|^2 = \left\| M_0 \Sigma_0 E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} W_0 h(x; a(\theta_0)) \right\|_{\Sigma_0}^2 \quad .$$

Moreover, by the orthogonal projection property of $M_0 \Sigma_0$ (with respect to the scalar product induced by Σ_0):

$$\begin{aligned} \|IF(x; U^S, P_{\theta_0})\|^2 &\leq \left\| E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} W_0 h(x; a(\theta_0)) \right\|_{\Sigma_0}^2 \\ &\leq \|h(x; a(\theta_0))\|_{W_0}^2 \quad . \end{aligned}$$

This proves the theorem for α^S .

To apply the approximation (32) to the likelihood-ratio statistic remember that \tilde{a} is Fréchet differentiable. A second order von Mises (1947) expansion of ξ^R under the hypotheses (4) and $g(a(P_{\theta_0})) = 0$ then gives up to terms of order $o(\epsilon^2)$:

$$\begin{aligned} \xi^R(P_{\epsilon, n, Q}^0) &= \frac{\epsilon^2}{n} \left[\int_{\mathbb{R}^N} [E_{\theta_0} \frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} IF(x; \hat{a}, P_{\theta_0}) + h(x; a(\theta_0))]^\top dQ(x) W_0 \times \right. \\ &\quad \int_{\mathbb{R}^N} [E_{\theta_0} \frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} IF(x; \hat{a}, P_{\theta_0}) + h(x; a(\theta_0))] dQ(x) \\ &\quad - \int_{\mathbb{R}^N} [E_{\theta_0} \frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} IF(x; \tilde{a}, P_{\theta_0}) + h(x; a(\theta_0))]^\top dQ(x) W_0 \times \\ &\quad \left. \int_{\mathbb{R}^N} [E_{\theta_0} \frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} IF(x; \tilde{a}, P_{\theta_0}) + h(x; a(\theta_0))] dQ(x) \right] . \end{aligned}$$

This expression can be simplified by using (46), (14) with $W_0 = V_0^{-1}$, and the orthogonal projection property of $M_0 \Sigma_0$ (with respect to the scalar product induced by Σ_0), to obtain:

$$\begin{aligned} \xi^R(P_{\epsilon, n, Q}^0) &= \frac{\epsilon^2}{n} \left[\int_{\mathbb{R}^N} (IF(x; \hat{a}, P_{\theta_0}) - IF(x; \tilde{a}, P_{\theta_0}))^\top dQ(x) \Sigma_0^{-1} \times \right. \\ &\quad \left. \int_{\mathbb{R}^N} (IF(x; \hat{a}, P_{\theta_0}) - IF(x; \tilde{a}, P_{\theta_0})) dQ(x) \right] + o(\epsilon^2) \quad , \quad \epsilon \rightarrow 0 \quad . \end{aligned}$$

The expression on the right hand side of this formula is the second order von Mises expansion under the hypotheses (4) and $g(a(P_{\theta_0})) = 0$ of a Hausman

functional $\xi^H : \text{dom}(\xi^H) \rightarrow \mathbb{R}^+$, defined by the symmetric form $\xi^H(P_{\theta_n}) := U^H(P_{\theta_n})^\top U^H(P_{\theta_n})$, where

$$U^H(P_{\theta_n}) := \Sigma_0^{-\frac{1}{2}}[\hat{a}(P_{\theta_n}) - \tilde{a}(P_{\theta_n})] \quad , \quad (47)$$

cf. Hausman (1978) and Holly (1982). As a consequence the difference between the levels under contamination of the likelihood ratio test and a Hausman test defined by the critical region $\{n\xi^H(P_{\theta_n}) \geq \eta_{1-\alpha_0}\}$, where $\eta_{1-\alpha_0}$ is the $1 - \alpha_0$ quantile of a $\chi^2(r, 0)$ distribution, is of order $o(\epsilon^2)$. Hence, the asymptotic bias under a given $P_{\epsilon, n, Q}^0$ -contamination of the level of the likelihood ratio test can be equivalently investigated by analyzing that of the Hausman test. Similar arguments to those developed above for U^S can be now applied to U^H . The IF of U^H is given by

$$\begin{aligned} IF(x; U^H, P_{\theta_0}) &= -\Sigma_0^{\frac{1}{2}} M_0 IF(x; \tilde{a}, P_{\theta_0}) \\ &= \Sigma_0^{\frac{1}{2}} M_0 \Sigma_0 E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} W_0 h(x; a(\theta_0)), \end{aligned}$$

using (46) and (14). Furthermore, again by the properties of orthogonal projections, this quantity can be bounded by the self-standardized norm of h as follows:

$$\begin{aligned} \|IF(x; U^H, P_{\theta_0})\|^2 &= \|M_0 \Sigma_0 E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} W_0 h(x; a(\theta_0))\|_{\Sigma_0}^2 \\ &\leq \|E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} W_0 h(x; a(\theta_0))\|_{\Sigma_0}^2 \\ &\leq \|h(x; a(\theta_0))\|_{W_0}^2 \quad . \end{aligned}$$

This proves the theorem for the level functional α^R of the likelihood-ratio type test. □

Proof of Theorem 2: By the Fréchet differentiability of \tilde{a} , $n\xi^M(P_{\theta_n})$ is asymptotically uniformly $\chi^2(r, \beta(\epsilon))$ distributed with $\beta(\cdot) = n\|U^M(P_{\cdot, n, Q}^1)\|^2$. Moreover, up to order $O(1/n)$ we have $\pi(P_{\cdot, n, Q}^1) = 1 - H_r(\eta_{1-\alpha_0}; \beta(\cdot))$. A first order Taylor expansion then yields

$$\pi(P_{\epsilon, n, Q}^1) - \pi(P_{\eta, n}^{alt}) = \left. \frac{\partial \pi(P_{\cdot, n, Q}^1)}{\partial \epsilon} \right|_{\epsilon=0} \cdot \epsilon + o(\epsilon) \quad ,$$

up to terms of order $O(1/n)$.

Some calculations then yield:

$$\begin{aligned} \frac{\partial \pi(P_{\cdot,n,Q}^1)}{\partial \epsilon} \Big|_{\epsilon=0} &= \mu \cdot \frac{\partial \beta}{\partial \epsilon} \Big|_{\epsilon=0} = 2\mu n \cdot \frac{\partial U^M(P_{\cdot,n,Q}^1)^\top}{\partial \epsilon} \Big|_{\epsilon=0} U^M(P_{\eta,n}^{alt}) \\ &= 2\mu\sqrt{n} \cdot \left[\int_{\mathbb{R}^N} IF^\top(x; U^M, P_{\eta,n}^{alt}) dQ(x) \right] U^M(P_{\eta,n}^{alt}) \quad . \quad (48) \end{aligned}$$

up to terms of order $O(1/n)$. Writing:

$$U^M(P_{\eta,n}^{alt}) = U^M(P_{\theta_0}) + \frac{\eta}{\sqrt{n}} \int_{\mathbb{R}^N} IF(x; U^M, P_{\theta_0}) dP_{\theta_1}(x) + o(\eta) \quad , \quad (49)$$

we obtain:

$$\begin{aligned} |\pi(P_{\epsilon,n,Q}^1) - \pi(P_{\eta,n}^{alt})| &\leq 2\mu\epsilon\eta \max_{\{P_{\eta,n}^{alt}; P_{\theta_0}\}} \sup_x \|IF(x; U^M, \cdot)\|^2 + o(\eta) \\ &\leq 2\mu\epsilon\eta \max_{\{P_{\eta,n}^{alt}; P_{\theta_0}\}} \sup_x \|h(x, \tilde{a}(\cdot))\|_{W_0}^2 + o(\eta) \quad , \quad (50) \end{aligned}$$

using $U^M(P_{\theta_0}) = 0$. This concludes the proof of the theorem. \square

Proof of Theorem 3: The functional U^W is asymptotically equivalent to U^S and U^R at the model under the local alternatives given by (41); cf. Gourieroux and Monfort (1989). Moreover the Fréchet differentiability of \tilde{a} implies that this equivalence is uniform. It is therefore sufficient to prove the theorem for the functional U^W . The statistic $n\xi^W(P_{\theta_n})$ is asymptotically uniformly $\mathcal{X}^2(r, \beta(\epsilon))$ distributed with $\beta(\cdot) = n\|U^W(P_{\cdot,n,Q}^0)\|^2$. Again, up to order $O(1/n)$ we have $\pi(P_{\cdot,n,Q}^0) = 1 - H_r(\eta_{1-\alpha_0}; \beta(\cdot))$. As in the proof of Theorem 2 a first order Taylor expansion then yields

$$\pi(P_{\epsilon,n,Q}^0) - \pi(P_{\theta_0}) = \frac{\partial \pi(P_{\cdot,n,Q}^0)}{\partial \epsilon} \Big|_{\epsilon=0} \cdot \epsilon + o(\epsilon) \quad , \quad (51)$$

up to terms of order $O(1/n)$.

Similar calculations to those in the proof of Theorem 2 then give:

$$\begin{aligned} \frac{\partial \pi(P_{\cdot,n,Q}^0)}{\partial \epsilon} \Big|_{\epsilon=0} &= \mu \cdot \frac{\partial \beta}{\partial \epsilon} \Big|_{\epsilon=0} = 2\mu n \cdot \frac{\partial U^W(P_{\cdot,n,Q}^0)^\top}{\partial \epsilon} \Big|_{\epsilon=0} U^W(P_{\theta_0}) \\ &= 2\mu\sqrt{n} \cdot \left[\int_{\mathbb{R}^N} IF^\top(x; U^W, P_{\theta_0}) dQ(x) \right] U^W(P_{\theta_0}) \quad , \end{aligned}$$

up to terms of order $O(1/n)$.

Expanding $U^W(P_{\theta_0})$ with respect to Δ we have:

$$U^W(P_{\theta_0}) = - \left[\frac{\partial g}{\partial a^\top}(a(\theta_0))_{\Sigma_0} \frac{\partial g^\top}{\partial a}(a(\theta_0)) \right]^{-\frac{1}{2}} \frac{\partial g}{\partial a^\top}(a(\theta_0)) \cdot \frac{\Delta}{\sqrt{n}} + o\left(\frac{\Delta}{\sqrt{n}}\right). \quad (52)$$

Since

$$\begin{aligned} IF(x; U^W, P_{\theta_0}) &= - \left[\frac{\partial g}{\partial a^\top}(a(\theta_0))_{\Sigma_0} \frac{\partial g^\top}{\partial a}(a(\theta_0)) \right]^{-\frac{1}{2}} \\ &\quad \times \frac{\partial g}{\partial a^\top}(a(\theta_0))_{\Sigma_0} E_{\theta_0} \frac{\partial h^\top(X_1; a(\theta_0))}{\partial a} W_0 h(x; a(\theta_0)), \end{aligned} \quad (53)$$

we obtain up to terms of order $o(\Delta)$:

$$\begin{aligned} \frac{\partial \pi(P_{\cdot, n, Q}^0)}{\partial \epsilon} \Big|_{\epsilon=0} &= 2\mu \cdot \left[\int_{\mathbb{R}^N} h^\top(x; a(\theta_0)) dQ(x) \right] W_0 E_{\theta_0} \frac{\partial h(X_1; a(\theta_0))}{\partial a^\top} \Big|_{\Sigma_0} \\ &\quad \times \frac{\partial g^\top}{\partial a}(a(\theta_0)) \left[\frac{\partial g}{\partial a^\top}(a(\theta_0))_{\Sigma_0} \frac{\partial g^\top}{\partial a}(a(\theta_0)) \right]^{-1} \frac{\partial g}{\partial a^\top}(a(\theta_0)) \Delta. \end{aligned}$$

The Cauchy-Schwarz inequality then gives:

$$\begin{aligned} \left| \left(\frac{\partial \pi(P_{\cdot, n, Q}^0)}{\partial \epsilon} \Big|_{\epsilon=0} \right) \right| &\leq 2\mu \cdot \left\| \int_{\mathbb{R}^N} h(x; a(\theta_0)) dQ(x) \right\|_{W_0} \\ &\quad \times \left\| \frac{\partial g^\top}{\partial a}(a(\theta_0)) \left[\frac{\partial g}{\partial a^\top}(a(\theta_0))_{\Sigma_0} \frac{\partial g^\top}{\partial a}(a(\theta_0)) \right]^{-1} \right. \\ &\quad \left. \frac{\partial g}{\partial a^\top}(a(\theta_0)) \Delta \right\|_{\Sigma_0} + o(\Delta). \end{aligned}$$

Using again the properties of orthogonal projections we obtain:

$$\left| \left(\frac{\partial \pi(P_{\cdot, n, Q}^0)}{\partial \epsilon} \Big|_{\epsilon=0} \right) \right| \leq 2\mu \cdot \sup_x \|h(x; a(\theta_0))\|_{W_0} \|\Delta\|_{\Sigma_0^{-1}} + o(\Delta). \quad (54)$$

Finally, by inserting this expression in the Taylor expansion (51) we get:

$$|\pi(P_{\epsilon, n, Q}^0) - \pi(P_{\theta_0})| \leq 2\mu\epsilon \|\Delta\|_{\Sigma_0^{-1}} \sup_x \|h(x; a(\theta_0))\|_{W_0} + o(\Delta). \quad (55)$$

This concludes the proof of the Theorem. \square

References

- AMEMIYA, T. (1974): "The Nonlinear Two-Stages Least Squares Estimator," *Journal of Econometrics*, 2, 105–110.
- BANSAL, R., D. HSIEH, AND S. VISWANATHAN (1993): "No Arbitrage and Arbitrage Pricing: a New Approach," *Journal of Finance*, 48, 1719–1747.
- BEDNARSKI, T. (1993): "Fréchet Differentiability of Statistical Functionals and Implications to Robust Statistics," in: MORGENTHALER, S., E. RONCHETTI, AND W. STAHEL (eds.): *New Directions in Statistical Data Analysis and Robustness*, Birkhäuser, Basel, 26–34.
- CLARKE, B. R. (1986): "Nonsmooth Analysis and Fréchet Differentiability of M-Functionals," *Probability Theory and Related Fields*, 73, 197–209.
- ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.
- GASKO, M., AND J. L. ROSENBERGER (1983): "Comparing Location Estimators: Means, Medians, and Trimean," in: HOAGLIN, D. C., F. MOSTELLER, AND J. W. TUKEY (eds.): *Understanding Robust and Exploratory Data Analysis*, Wiley, New York, 297–338.
- GOURIEROUX, C., AND A. MONFORT (1989): "A General Framework for Testing a Null Hypothesis in a Mixed Form," *Econometric Theory*, 5, 63–82.
- HAMPEL, F. R. (1968): "Contribution to the Theory of Robust Estimation," Ph.D. thesis, University of California, Berkeley.
- (1974): "The Influence Curve and its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383–393.
- HAMPEL F. R., E. M. RONCHETTI, P. J. ROUSSEEUW, AND W. A. STAHEL (1986): *Robust Statistics: The Approach Based on Influence Functions*. Wiley,

New York.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HAUSMAN, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1272.

HERITIER, S., AND E. RONCHETTI (1994): "Robust Bounded-Influence Tests in General Parametric Models," *Journal of the American Statistical Association*, 89, 897–904.

HOLLY, A. (1982): "A Remark on Hausman's Specification Test," *Econometrica*, 50, 749–759.

HUBER, P. (1981): *Robust Statistics*. Wiley, New York.

IMBENS, G. W., SPADY, R. H., AND P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

JOHNSON, N., AND S. KOTZ (1991): *Continuous and Discrete Distributions*, vol. 2. Wiley, New York.

KOENKER, R. W. (1982): "Robust Methods in Econometrics," *Econometric Review*, 1, 213–255.

KOENKER, R. W., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

KOENKER, R. W., MACHADO, J. A. F., SKEELS, C. L., AND A. H. WELSH (1994): "Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimation," *Econometric Theory*, 10, 172–197.

KOENKER, R. W., AND J. A. F. MACHADO (1999): "GMM Inference when the

- Number of Orthogonality Conditions is Large,” *Journal of Econometrics*, 93, 327–344.
- KRASKER, W. S. (1986): “Two-Stage Bounded-Influence Estimators for Simultaneous Equations Models,” *Journal of Business and Economic Statistics*, 4, 437–444.
- KRASKER, W. S., AND R. E. WELSCH (1985): “Resistant Estimation for Simultaneous-Equations Models Using Weighted Instrumental Variables,” *Econometrica*, 53, 1475–1488.
- KRISHNAKUMAR, J., AND E. RONCHETTI (1997): “Robust Estimators for Simultaneous Equations Models,” *Journal of Econometrics*, 78, 295–314.
- KÜNSCH, H. (1984): “Infinitesimal Robustness for Autoregressive Processes,” *Annals of Statistics*, 12, 843–863.
- LUCAS, A., R. VAN DIJK, AND T. KLOEK (1994): “Outlier Robust GMM Estimation of Leverage Determinants,” *Tinbergen Institute Discussion Paper 94-132*.
- MARKATOU, M., AND E. RONCHETTI (1997): “Robust Inference: The Approach Based on Influence Functions,” in: MADDALA, G.S., AND C.R. RAO (eds.): *Handbook of Statistics*, vol. 15, North Holland, 49–75.
- MARTIN, R. D., AND V. J. YOHAI (1986): “Influence Functionals for Times Series,” *Annals of Statistics*, 14, 781–818.
- NEWKEY, W. K. (1985): “Generalized Method of Moments Specification Testing,” *Journal of Econometrics*, 29, 229–256.
- NEWKEY, W. K., AND K. D. WEST (1987a): “Hypothesis Testing with Efficient Method of Moments Estimation,” *International Economic Review*, 28, 777–787.
- (1987b): “A Simple Positive-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimator,” *Econometrica*, 55, 703–708.

PERACCHI, F. (1990): “Robust M-Estimators,” *Econometric Review*, 9, 1–30.

——— (1991): “Robust M-Tests,” *Econometric Theory*, 7, 69–84.

ROUSSEEUW, P. J., AND A. LEROY (1987): *Robust-Regression and Outlier Detection*. Wiley, New York.

VON MISES, R. (1947): “On the Asymptotic Distribution of Differentiable Statistical Functions,” *Annals of Mathematical Statistics*, 18, 309–348.

Tables

Table 1: Values of the Tuning Constant c for Bounding the Maximal Bias of the Level of a GMM test

The values of the tuning constant c are for a nominal level 5% at the model, for a maximal bias given by $maxbias = +/- 0.5\%$ and for a model contamination $\epsilon = 5\%$. r is the number of degrees of freedom implied by the χ^2 -test under scrutiny.

r	$\eta_{1-\alpha_0}$	μ	c
1	3.84	0.1145	4.18
2	5.99	0.0749	5.17
3	7.81	0.0584	5.85
4	9.94	0.0490	6.39
5	11.07	0.0428	6.83
6	12.59	0.0383	7.22
7	14.07	0.0350	7.56
8	15.51	0.0323	7.87
9	16.92	0.0301	8.15
10	18.31	0.0283	8.41
∞		0	∞

Table 2: GMM and RGMM Simulation Results under $u_t \sim \mathcal{N}(0, 1)$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the χ^2 test. The constant c for the RGMM test was set to $c = 2.09$.

	Rejection frequency GMM			Rejection frequency RGMM		
α_1	$T = 250$	$T = 500$	$T = 1000$	$T = 250$	$T = 500$	$T = 1000$
0.00	0.08	0.08	0.05	0.02	0.02	0.02
0.05	0.05	0.09	0.19	0.02	0.06	0.07
0.10	0.09	0.28	0.62	0.06	0.14	0.29
0.15	0.20	0.52	0.90	0.12	0.31	0.62
0.20	0.32	0.74	0.97	0.21	0.51	0.87
0.25	0.45	0.84	0.98	0.35	0.71	0.95
0.30	0.56	0.89	0.98	0.49	0.86	0.99

Table 3: GMM and RGMM Simulation Results under $u_t \sim DE$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the χ^2 test. The constant c for the RGMM test was set to $c = 2.09$.

	Rejection frequency GMM			Rejection frequency RGMM		
α_1	$T = 250$	$T = 500$	$T = 1000$	$T = 250$	$T = 500$	$T = 1000$
0.00	0.11	0.10	0.09	0.03	0.03	0.03
0.05	0.04	0.04	0.09	0.03	0.06	0.12
0.10	0.04	0.12	0.31	0.07	0.14	0.32
0.15	0.06	0.23	0.54	0.11	0.26	0.58
0.20	0.10	0.33	0.71	0.18	0.41	0.78
0.25	0.16	0.43	0.78	0.26	0.57	0.91
0.30	0.21	0.50	0.79	0.34	0.70	0.96

Table 4: GMM and RGMM Simulation Results under $u_t \sim t_9$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the χ^2 test. The constant c for the RGMM test was set to $c = 2.09$.

	Rejection frequency GMM			Rejection frequency RGMM		
α_1	$T = 250$	$T = 500$	$T = 1000$	$T = 250$	$T = 500$	$T = 1000$
0.00	0.09	0.09	0.07	0.02	0.02	0.02
0.05	0.05	0.05	0.11	0.04	0.06	0.10
0.10	0.05	0.16	0.42	0.09	0.14	0.30
0.15	0.12	0.35	0.69	0.13	0.31	0.62
0.20	0.21	0.54	0.83	0.23	0.50	0.84
0.25	0.30	0.65	0.87	0.35	0.83	0.95
0.30	0.38	0.73	0.88	0.46	0.83	0.99

Table 5: GMM and RGMM Simulation Results under $u_t \sim t_5$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the χ^2 test. The constant c for the RGMM test was set to $c = 2.09$.

	Rejection frequency GMM			Rejection frequency RGMM		
α_1	$T = 250$	$T = 500$	$T = 1000$	$T = 250$	$T = 500$	$T = 1000$
0.00	0.10	0.11	0.11	0.02	0.02	0.03
0.05	0.05	0.05	0.06	0.03	0.07	0.11
0.10	0.06	0.10	0.24	0.05	0.14	0.33
0.15	0.11	0.18	0.43	0.11	0.28	0.61
0.20	0.15	0.29	0.59	0.17	0.46	0.82
0.25	0.21	0.40	0.67	0.29	0.64	0.93
0.30	0.27	0.48	0.71	0.40	0.78	0.97

Table 6: GMM and RGMM Simulation Results under

$$u_t \sim CN(0.05, 100)$$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the χ^2 test. The constant c for the RGMM test was set to $c = 2.09$.

	Rejection frequency GMM			Rejection frequency RGMM		
α_1	$T = 250$	$T = 500$	$T = 1000$	$T = 250$	$T = 500$	$T = 1000$
0.00	0.35	0.51	0.48	0.02	0.01	0.02
0.05	0.16	0.19	0.17	0.02	0.03	0.06
0.10	0.09	0.08	0.05	0.03	0.06	0.14
0.15	0.06	0.04	0.02	0.06	0.11	0.24
0.20	0.04	0.03	0.03	0.07	0.16	0.36
0.25	0.04	0.03	0.06	0.10	0.22	0.48
0.30	0.04	0.04	0.11	0.13	0.28	0.60

Figures

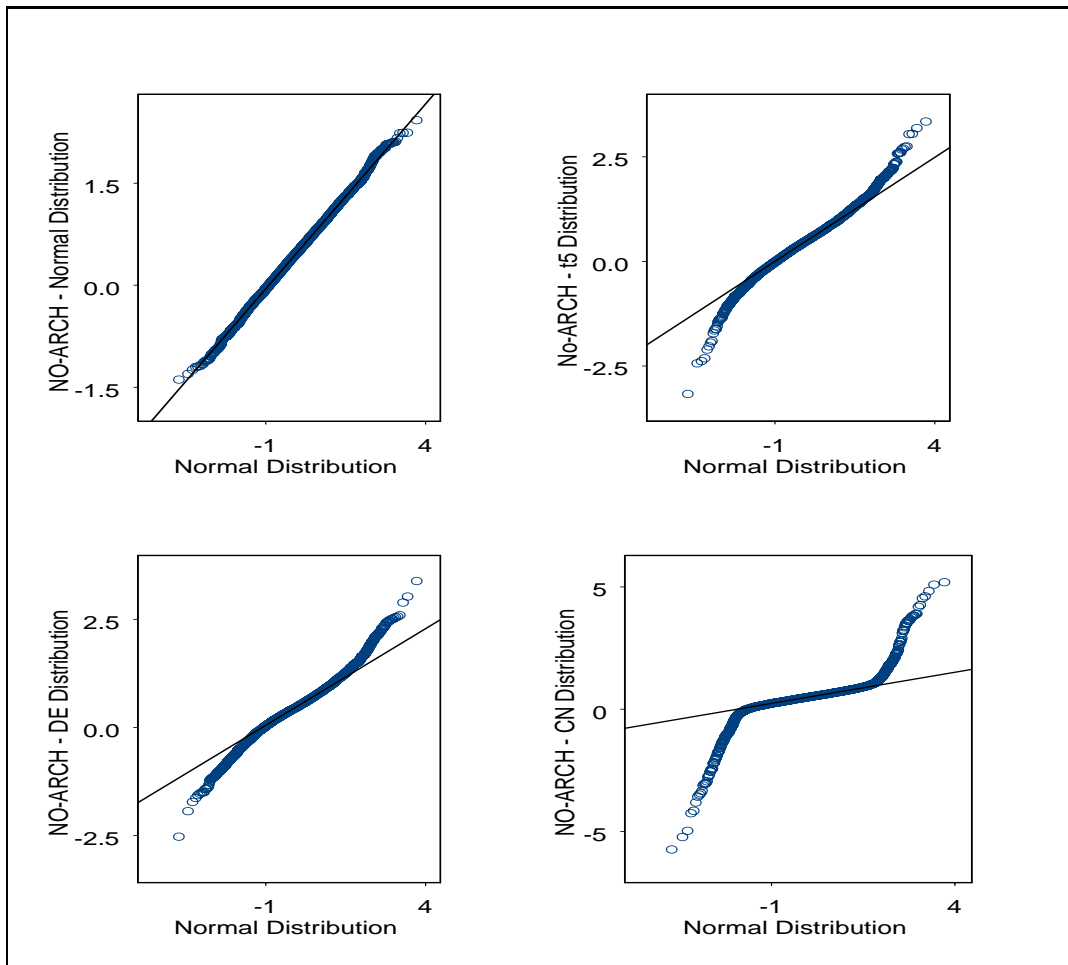


Figure 1: QQ Plot of the unconditional distribution of (y_t) (sample size 5'000 observations) under standard normal, t_5 , double exponential and contaminated normal ($\epsilon = 0.05$, $K = 10$) errors. The ARCH parameter was set to $\alpha_1 = 0$.

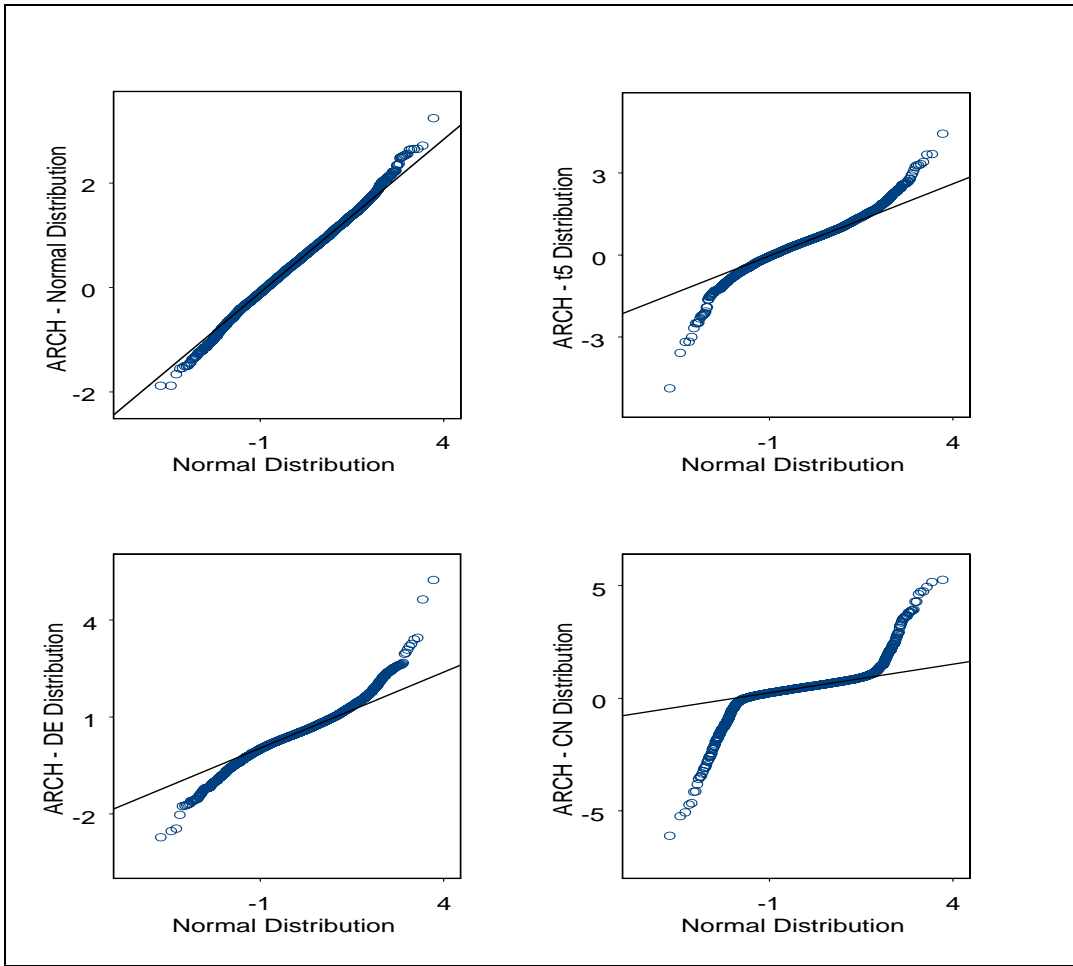


Figure 2: QQ Plot of the unconditional distribution of (y_t) (sample size 5'000 observations) under standard normal, t_5 , double exponential and contaminated normal ($\epsilon = 0.05$, $K = 10$) errors. The ARCH parameter was set to $\alpha_1 = 0.2$.