Title

Consistent Estimation for Aggregated GARCH

Author
Komunjer, Ivana

Publication Date
2001-05-01

2001-08

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

CONSISTENT ESTIMATION FOR AGGREGATED GARCH PROCESSES

BY

IVANA KOMUNJER

DISCUSSION PAPER 2001-08
MAY 2001

Consistent Estimation for Aggregated GARCH Processes

Ivana Komunjer*

May 2001

Abstract

We study the properties of a quasi-maximum likelihood (QML) for the parameters of a "weak" *GARCH* process obtained by contemporaneous aggregation of two independent "strong" *GARCH* processes. The inconsistency of the Gaussian quasi-likelihood estimator (QMLE) has already been reported by Nijman & Sentana (1996) but has not yet been solved. In this paper we identify the causes of inconsistency of QMLE in the "weak" GARCH case and compare the performance of QMLE when the innovations are assumed to have Gaussian, Laplace (double exponential) or α-stable distribution.

Keywords: aggregation, GARCH, estimation, quasi-maximum likelihood, consistency.

JEL codes: C13, C15, C32, C51

*The author is PhD candidate at HEC - School of Management, Department of Finance, 1 Rue de la Libération, 78351 Jouy-en-Josas, France. This paper was written while the author was visiting the UCSD Department of Economics, whose hospitality is gratefully acknowledged. I am indebted to Clive Granger, James Hamilton, Graham Elliott, Hal White, Allan Timmermann, Robert Engle, Peter Hansen and all the participants of the Econometrics Department Seminar for their precious comments. I am also thankful to my advisor, Michael Rockinger, for his interest in this research.

1 Introduction

It is well known that $GARCH$ type models give a parsimonious representation of the conditional heteroskedasticity exhibited by the financial time series such as exchange rates and stock prices. One serious drawback of this type of models, however, is their internal inconsistency by aggregation. In fact, the class of "strong" $GARCH$ processes, as defined by Engle [10] and Bollerslev [4], is not closed under contemporaneous aggregation. In general, the sum of two independent "strong" $GARCH$ processes cannot be described as a "strong" $GARCH$ whose parameters are functions of the parameters of the two underlying processes. In order to overcome this aggregation problem, Nijman & Sentana [23] define a new class of processes called "weak" $GARCH$. The latter are closed under contemporaneous aggregation and therefore seem to provide a solution to the aggregation problem. However, when trying to estimate the parameters of such a model by Gaussian quasi-likelihood estimator, the authors note that the latter is "approximately consistent in some cases and clearly inconsistent in others".

Table 4 : weak $GARCH$ parameters estimated plim of Gaussian QMLE

β_1	α_1	β_2	α_2	$\frac{\psi_1}{\psi_2}$	β	α	σ_η^2	$p\lim\widehat{\beta}$	$p\lim\widehat{\alpha}$
0.5	0.35	0.8	0.05	1	0.569	0.281	8.282	0.690	0.125
0.5	0.35	0.8	0.05	4	0.516	0.334	112.158	0.581	0.234
0.5	0.35	0.8	0.05	1/4	0.705	0.145	26.167	0.775	0.062

To illustrate their conclusion, we report some of the results shown by Table 4 in [23].

In this paper we identify the causes of inconsistency of the QMLE in the "weak" $GARCH$ models and compare the performance of QMLE under different density assumptions, namely Gaussian, Laplace and α-stable.

2 Problem Identification

To describe the problem we consider, suppose that the data consists of observations $y_t, t = 1, \dots, T$ generated by a univariate $GARCH(1,1)$ process, as first defined by Engle [10] and

Bollerslev [4]. Let $\mathcal{F}_t \equiv (y_{t-1}, y_{t-2}, \ldots)$ denote the information set at period-t and let $\{\xi_t\}$ be a sequence of innovations that is independent and identically distributed (iid). The process $\{y_t\}$ to be estimated can then be described as

$$y_t = \sigma_t \xi_t, \quad (1)$$

$$\sigma_t^2 = \psi + \beta\sigma_{t-1}^2 + \alpha y_{t-1}^2. \quad (2)$$

Throughout, we assume that the period-t innovation is centered, $E[\xi_t] = 0$, and reduced, $E[\xi_t^2] = 1$. Many $GARCH$ models of the form (1) are estimated by assuming a particular form for the innovation density function. For example, the families of density functions used are Gaussian, t density, Gamma, etc.

In order to estimate the parameters of such a process we use a QMLE. If the likelihood is assumed to be Gaussian, then the QMLE is the value of $\theta \equiv (\psi, \beta, \alpha)'$ that maximizes $L_T(\theta) \equiv T^{-1}\sum_{t=1}^{T} l_t(\theta)$ where the period-t conditional log-likelihood of y_t given $\mathcal{F}_t$, $l_t(\theta)$, is defined as

$$l_t(\theta) \equiv -\ln\sigma_t - (\frac{y_t}{\sigma_t})^2. \quad (3)$$

The QMLE obtained by maximizing the Gaussian log-likelihood is known to be consistent and asymptotically normal, as shown by Bollerslev and Wooldridge [7], Lumsdaine [19], Weiss [28] and Lee & Hansen [18], among others.

The asymptotic results on the behavior of the QMLE in the classical "strong" $GARCH$ case (1) however, no longer hold for "weak" $GARCH$ processes, as defined by Nijman & Sentana [23]. The "weak" $GARCH$ processes are typically obtained by contemporaneous aggregation of two univariate "strong" $GARCH$, as in (1). Thus if $y_{1,t}$ and $y_{2,t}$ are two independent "strong" $GARCH(1,1)$ with

$$y_{i,t} = \sigma_{i,t}\xi_{i,t},$$

$$\sigma_{i,t}^2 = \psi_i + \beta_i\sigma_{i,t-1}^2 + \alpha_i y_{i,t-1}^2,$$

for $i = 1, 2$, their sum $y_t \equiv y_{1,t} + y_{2,t}$ is a "weak" $GARCH(1,1)$, provided that their persistence parameters are the same, i.e. $\beta_1 + \alpha_1 = \beta_2 + \alpha_2$. One way to describe a "weak"

$GARCH$ model that is estimated is

$$y_t^2 = \sigma_t^2 + \eta_t, \tag{4}$$
$$\sigma_t^2 = \psi + \beta\sigma_{t-1}^2 + \alpha y_{t-1}^2, \tag{5}$$

where σ_t^2 is the period-t linear projection of y_t^2 on $\mathcal{F}_t' \equiv (\mathbf{1}, y_{t-1}^2, y_{t-2}^2, \ldots)$ and η_t is the period-t innovation. It can easily be shown that η_t is a white noise, $\{\eta_t\} \sim WN(0, \sigma_\eta^2)$.[1]

To understand why identification differs for "strong" and "weak" $GARCH$ processes, we focus on three fundamental differences between the two specifications.

The first notable difference between the models (1) and (4) is the characterization of σ_t^2: the "strong" $GARCH$ σ_t^2 is the period-t conditional expectation of y_t^2 given $\mathcal{F}_t$, i.e. $\sigma_t^2 = E[y_t^2|\mathcal{F}_t]$, whereas the "weak" $GARCH$ σ_t^2 is the period-t linear projection of y_t^2 on $\mathcal{F}_t' \equiv (\mathbf{1}, y_{t-1}^2, y_{t-2}^2, \ldots)$, i.e. $\sigma_t^2 = \widehat{E}[y_t^2|\mathcal{F}_t']$.[2] We will give the estimation related implications of this characterization of σ_t^2 in the following sections.

The second important difference between the two processes lies in the nature of scaled variable, $\frac{y_t}{\sigma_t}$, the ratio of the observed variable to its conditional standard deviation. In the "strong" $GARCH$ case, the scaled observations are independent and identically distributed conditional on $\mathcal{F}_t$, i.e. $\left\{\frac{y_t}{\sigma_t}|\mathcal{F}_t\right\} \sim iid(0,1)$ whereas in the "weak" $GARCH$ case, their conditional distribution has no longer the iid property. Thus the conditional moments of the "weak" $GARCH$ scaled observations $\frac{y_t}{\sigma_t}$, $E[(\frac{y_t}{\sigma_t})^r|\mathcal{F}_t]$ will no longer be independent of $\mathcal{F}_t$.[3]

[1] Throughout the paper we use the following definition of white noise: the process $\{Z_t\}$ is said to be white noise with mean zero and variance σ^2, $\{Z_t\} \sim WN(0, \sigma^2)$, iff $\{Z_t\}$ has zero mean and covariance function

$$\gamma(h) = \begin{cases} \sigma^2 & \text{if } h = 0, \\ 0 & \text{if } h \neq 0. \end{cases}$$

[2] We will use the symbol $\widehat{E}$ to denote a linear projection on the set $\mathcal{F}_t^{'}$ composed of past realizations of y_t^2 along with a constant term $\mathbf{1}$.

[3] Consider for example the conditional second and fourth moments of $\frac{y_t}{\sigma_t}$:

$$\begin{aligned} E[\frac{y_t^2}{\sigma_t^2}|\mathcal{F}_t] &= E[\frac{(y_{1,t} + y_{2,t})^2}{\sigma_t^2}|\mathcal{F}_t] \\ &= \sigma_t^{-2}\left(E[y_{1,t}^2|\mathcal{F}_t] + 2E[y_{1,t}y_{2,t}|\mathcal{F}_t] + E[y_{2,t}^2|\mathcal{F}_t]\right) \\ &= \sigma_t^{-2}\left(\sigma_{1,t}^2 + \sigma_{2,t}^2\right) \end{aligned}$$

The third somewhat less visible difference between the two processes lies in the nature of the parameter θ to estimate. In the strong $GARCH$ case (1) and assuming that the period-t innovation ξ_t is Gaussian, $\xi_t \sim \mathcal{N}(0,1)$, we need to estimate $\theta \equiv (\psi, \beta, \alpha)'$ where different components of θ are mutually independent. In the weak $GARCH$ case however, we need to estimate $\theta \equiv (\psi, \beta, \alpha, \sigma_\eta^2)'$ where σ_η^2, the variance of the period-t innovation η_t, depends on the other components of θ, as will be shown in the next sections.

As before, we are interested in estimating the parameters of a "weak" $GARCH$ process. In order to estimate the parameters of such a process let us follow the original approach of Nijman & Sentana [23] and use a QMLE. Thus, we first need to make a choice of a particular family of distributions for the scaled variable $\frac{y_t}{\sigma_t}$. Since $\frac{y_t}{\sigma_t}$ is symmetrically distributed we adopt their Gaussian assumption $\frac{y_t}{\sigma_t} \sim \mathcal{N}(0, s^2)$ where $s^2 \equiv E[(\frac{y_t}{\sigma_t})^2]$.[4] The Gaussian QMLE is then the value of $\theta = (\psi, \beta, \alpha, s^2)'$ that maximizes $L_T(\theta) \equiv T^{-1} \sum_{t=1}^T l_t(\theta)$ with

$$l_t(\theta) \equiv -\ln s - \ln \sigma_t - s^{-2} (\frac{y_t}{\sigma_t})^2. \tag{6}$$

Unfortunately, the Gaussian QMLE thus obtained is "approximately consistent in some cases and clearly inconsistent in others" as noted by Nijman & Sentana [23]. The question then is wether the previous results on the consistency and asymptotic normality of a Gaussian QMLE in the classical "strong" $GARCH$ case still hold in the "weak" $GARCH$ one.

For example, Weiss [28] proves consistency and asymptotic normality of MLE in the ARCH models, under the assumptions that $\frac{y_t}{\sigma_t}$ are *iid* and the fourth moment of y_t finite. By using slightly different approaches Bollerslev & Wooldridge [7] and Lumsdaine [19] derive the consistency and asymptotic normality conditions in $GARCH(1,1)$ and $IGARCH(1,1)$ models. Lumsdaine [19] assumes that the scaled variable $\frac{y_t}{\sigma_t}$ is *iid* and drawn from a symmet-

and

$$\begin{aligned} E[\frac{y_t^4}{\sigma_t^4}|\mathcal{F}_t] &= E[\frac{(y_{1,t}+y_{2,t})^4}{\sigma_t^4}|\mathcal{F}_t] \\ &= \sigma_t^{-4}\left(E[y_{1,t}^4|\mathcal{F}_t] + 4E[y_{1,t}^3 y_{2,t}|\mathcal{F}_t] + 6E[y_{1,t}^2 y_{2,t}^2|\mathcal{F}_t] + 4E[y_{1,t} y_{2,t}^3|\mathcal{F}_t] + E[y_{2,t}^4|\mathcal{F}_t]\right) \\ &= \sigma_t^{-4}\left(\sigma_{1,t}^4 E[\xi_{1,t}^4] + 6\sigma_{1,t}^2\sigma_{2,t}^2 + \sigma_{2,t}^4 E[\xi_{2,t}^4]\right), \end{aligned}$$

which both depend on the information set.

[4] In fact, $\sigma^2 \equiv E[(\frac{y_t}{\sigma_t})^2]$ is no longer 1 since $\sigma^2 = E[E[(\frac{y_t}{\sigma_t})^2|\mathcal{F}_t]] = E[\sigma_t^{-2}\left(\sigma_{1,t}^2 + \sigma_{2,t}^2\right)]$.

ric unimodal density with 32nd moment finite. On the other hand, Bollerslev & Wooldridge [7] provide quite general but somewhat abstract regularity conditions, which ensure asymptotic normality of the QMLE for a wide class of models.[5] Finally, Lee & Hansen [18] require the scaled variable $\frac{y_t}{\sigma_t}$ to be strictly stationary and ergodic. They prove the consistency of the QMLE under the condition that the conditional $2+\delta$ moment of $\frac{y_t}{\sigma_t}$ is uniformly bounded, for some $\delta > 0$. The asymptotic normality is proved by adding the assumption that the conditional fourth moment of $\frac{y_t}{\sigma_t}$ is uniformly bounded. In addition, all these authors assume that the conditional mean and the conditional variance of y_t are correctly specified.

Clearly, the main assumptions for the consistency and asymptotic normality of a Gaussian QMLE are: (1) correct specification of the conditional mean and variance of y_t, (2) strict stationarity of the scaled variable $\frac{y_t}{\sigma_t}$, and (3) some additional moment conditions on either $\frac{y_t}{\sigma_t}$ or y_t. Before applying the known asymptotic results, we must therefore ensure that these requirements are met in the "weak" $GARCH$ case.

The main goal of this paper is therefore to provide a consistent method for estimating the parameters of aggregated $GARCH$ processes, also known as "weak" $GARCH$. We proceed by carefully examining the four possible sources of bias in the Gaussian QMLE of "weak" $GARCH$ processes, namely: (1) new characterization of the conditional variance σ_t^2, (2) non-iid structure of the scaled variable $\frac{y_t}{\sigma_t}$, (3) dependence in the components of the parameter to estimate, and (4) possible non-existence of higher moments of the residuals.

3 Properties of aggregated $GARCH$ processes

In this section we recall some of the existing results on the contemporaneous aggregation of two $GARCH(1,1)$ processes, as first derived by Nijman & Sentana [23] in 1996. The properties of the "weak" $GARCH$ obtained by aggregation are then compared to the properties of the classical "strong" $GARCH$ some of which we recall hereafter.

[5] Typically, the authors prove consistency by using the uniform law of large numbers. The proof of asymptotic normality of the QMLE relies on the fact that the score $s_t(\theta)$ of the conditional log-likelihood, evaluated at the true parameter θ_0, is a vector martingale difference sequence with respect to $\mathcal{F}_t$. They then apply a martingale central limit theorem to $\{s_t(\theta_0)\}$.

3.1 "Strong" $GARCH$ properties

Let $\{y_{1,t}\}$ and $\{y_{2,t}\}$ be two independent univariate $GARCH(1,1)$ processes defined by

$$\begin{aligned} y_{i,t} &= \sigma_{i,t}\xi_{i,t}, & (7)\\ \sigma_{i,t}^2 &= \psi_i + \beta_i\sigma_{i,t-1}^2 + \alpha_i y_{i,t-1}^2 & (8)\end{aligned}$$

for $i = 1, 2$. As before we assume that the period-t innovations are centered, $E[\xi_{i,t}] = 0$, and reduced, $E[\xi_{i,t}^2] = 1$. Throughout the paper we do not assume any particular form for the innovation density function but, instead, give necessary conditions for the processes $\{y_{i,t}\}, i = 1, 2$ to be well defined. Let us in that view transform (8) and consider the $ARMA(1,1)$ representation for the time series of squared returns

$$y_{i,t}^2 = \psi_i + (\beta_i + \alpha_i)\, y_{i,t-1}^2 + \eta_{i,t} - \beta_i\eta_{i,t-1} \tag{9}$$

where the period-t innovation $\eta_{i,t}$ is defined as $\eta_{i,t} \equiv y_{i,t}^2 - \sigma_{i,t}^2$ for $i = 1, 2$. In order for (9) to be causal and invertible $ARMA(1,1)$ we need to impose the following conditions on the parameters ψ_i, β_i and α_i, $i = 1, 2$,

$$\psi_i > 0\ ,\ \beta_i \geqslant 0 \text{ and } \alpha_j \geqslant 0 \tag{10}$$

and

$$1 - (\beta_i + \alpha_i)z \neq 0 \text{ for all } z \in C \text{ such that } |z| \leqslant 1. \tag{11}$$

Moreover, the period-t innovation $\eta_{i,t}$ needs to be a white noise. This last property is ensured by the following condition.

Proposition 1 *Under conditions* (10) *and* (11) *and provided that*

$$E\left[\xi_{i,t}^4\right] < 1 + \alpha_i^{-2}\left[1 - (\beta_i + \alpha_i)^2\right],$$

the sequence $\{\eta_{i,t}\}$ *of period-t innovations* $\eta_{i,t} \equiv y_{i,t}^2 - \sigma_{i,t}^2$ *satisfies*

$$\{\eta_{i,t}\} \sim WN(0, \sigma_{\eta,i}^2),$$

where

$$\sigma_{\eta,i}^2 = \sigma^4(E\left[\xi_{i,t}^4\right] - 1)\frac{1 - (\beta_i + \alpha_i)^2}{1 - (\beta_i + \alpha_i)^2 + (1 - E\left[\xi_{i,t}^4\right])\alpha_i^2}$$

and

$$\sigma^2 \equiv \frac{\psi}{1-(\beta_i+\alpha_i)}.$$

The fourth moment condition on $\xi_{i,t}, i=1,2$, ensures the positivity of the variance $\sigma^2_{\eta,i}$ of period-t innovation $\eta_{i,t}$ and imposes an additional constraint on the parameters β_i and $\alpha_i, i=1,2$, which also have to satisfy (10) and (11). It is interesting to note that the moment condition $E\left[y^4_{i,t}\right]<\infty$ results in the same restrictions on β_i and $\alpha_i, i=1,2$. Indeed, $E\left[y^4_{i,t}\right]$ and $E\left[\xi^4_{i,t}\right]$ are related through

$$k_{\xi,i} = k_i \frac{1-(\beta_i+\alpha_i)^2+\alpha_i^2}{1-(\beta_i+\alpha_i)^2+k_i\alpha_i^2} \tag{12}$$

where k_i is the kurtosis of $y_{i,t}$, $k_i \equiv E\left[y^4_{i,t}\right](E\left[y^2_{i,t}\right])^{-2}$, and $k_{\xi,i}$ the kurtosis of $\xi_{i,t}$, $k_{\xi,i} \equiv E\left[\xi^4_{i,t}\right]$.[6] It is interesting to note that (12) can also be used in the expression of the variance $\sigma^2_{\eta,i}$ of period-t innovation $\eta_{i,t}$,

$$\sigma^2_{\eta,i} = \sigma^4(k_i-1)\frac{1-(\beta_i+\alpha_i)^2}{1-(\beta_i+\alpha_i)^2+\alpha_i^2} \tag{13}$$

Thinking in terms of $ARMA(1,1)$ representation (9) of $GARCH(1,1)$ process $\{y_{i,t}\}$, we can see that the vector parameter that fully determines (9) is $(\psi_i, \beta_i, \alpha_i, \sigma^2_{\eta,i})$. However, the last component of the parameter vector is not independent from the other components, as shown by (13). We can solve this dependency in parameter components problem by using a different parametrization, i.e. by considering a vector parameter $(\psi_i, \beta_i, \alpha_i, k_{\xi,i})$ whose components are mutually independent.

The results of Proposition 1 can be generalized to higher order $GARCH$ processes. Consider the case where $\{y_{i,t}\}, i=1,2$ are generated by $GARCH(p,q)$

$$y_{i,t} = \sigma_{i,t}\xi_{i,t} \tag{14}$$

$$\sigma^2_{i,t} = \psi_i + \sum_{k=1}^{p}\beta_{i,k}\sigma^2_{i,t-k} + \sum_{l=1}^{q}\alpha_{i,l}y^2_{i,t-l} \tag{15}$$

where the parameters ψ_i , $\beta_{i,k}, 1 \leqslant k \leqslant p$ and $\alpha_{i,l}, 1 \leqslant l \leqslant q$ satisfy the causality and invertibility conditions:

$$\psi_i > 0 \ , \ \beta_{i,k} \geqslant 0 \text{ for all } k: 1 \leqslant k \leqslant p \ , \text{ and } \alpha_{i,l} \geqslant 0 \text{ for all } l: 1 \leqslant l \leqslant q \ , \tag{16}$$

[6] Equivalently, we can say that $k_i = k_{\xi,i}\frac{1-(\beta_i+\alpha_i)^2}{1-(\beta_i+\alpha_i)^2-(k_{\xi,i}-1)\alpha_i^2}$.

and

$$1-(\beta_{i,1}+\alpha_{i,1})z-...-(\beta_{i,r}+\alpha_{i,r})z^{r}\neq 0 \text{ for all } z\in C \text{ such that } |z|\leqslant 1\ , \qquad (17)$$

where $r=\max(p,q)$ and we define $\beta_{i,k}\equiv 0$ for $k>p$ and $\alpha_{i,l}\equiv 0$ for $l>q$. The conditions (16) and (17) are mere generalizations of (10) and (11). The following proposition gives the necessary conditions for the period-t innovation $\eta_{i,t}$ to be a white noise.

Proposition 2 *Under conditions* (16) *and* (17) *and provided that*

$$E\left[\xi_{i,t}^{4}\right]<1+[1-(\sum_{k=1}^{p}\beta_{i,k}+\sum_{l=1}^{q}\alpha_{i,l})^{2}][\sum_{l=1}^{q}\alpha_{i,l}^{2}]^{-1},$$

the sequence $\left\{\eta_{i,t}\right\}$ *of period-t innovations* $\eta_{i,t}\equiv y_{i,t}^{2}-\sigma_{i,t}^{2}$ *satisfies*

$$\left\{\eta_{i,t}\right\}\sim WN(0,\sigma_{\eta,i}^{2}),$$

where

$$\sigma_{i,\eta}^{2}=\sigma_{i}^{4}\left(E\left[\xi_{i,t}^{4}\right]-1\right)\frac{1-(\sum_{k=1}^{p}\beta_{i,k}+\sum_{l=1}^{q}\alpha_{i,l})^{2}}{1-(\sum_{k=1}^{p}\beta_{i,k}+\sum_{l=1}^{q}\alpha_{i,l})^{2}+(1-E\left[\xi_{i,t}^{4}\right])\sum_{l=1}^{q}\alpha_{i,l}^{2}}$$

and

$$\sigma_{i}^{2}\equiv\frac{\psi_{i}}{1-\sum_{k=1}^{p}\beta_{i,k}-\sum_{l=1}^{q}\alpha_{i,l}}.$$

The results of Proposition 2 can be viewed as a generalization of the finite fourth moment conditions derived by both Weiss [28] in the case of univariate ARCH models, and Bollerslev [4] in the special case of a $GARCH(1,1)$ process. We therefore report them in this paper together with a proof.

Proof of Proposition 2. see the Appendix 6. ■

Before considering the aggregation of univariate $GARCH$ processes, let us give two more results on the higher-order moments of $\left\{\eta_{i,t}\right\}$ in the special case where $\{y_{i,t}\}, i=1,2$ are $GARCH(1,1)$ with Gaussian innovations, i.e. $\{\xi_{i,t}\}\sim\mathcal{N}(0,1)$.

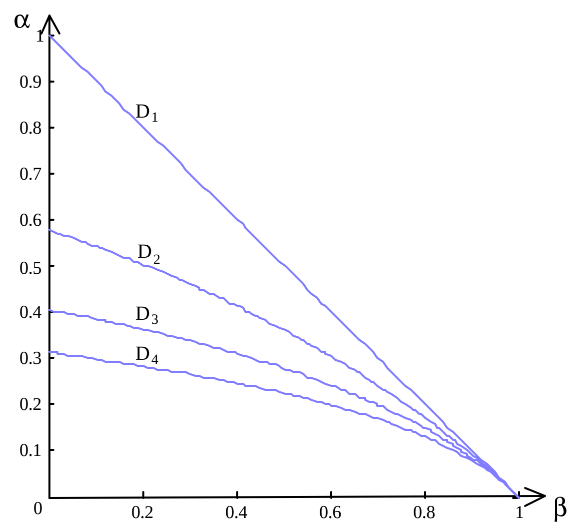

Figure 1: Domains of definition D_r, $D_r \equiv \{(\beta_i, \alpha_i) \in [0,1]^2 : E[\eta_{i,t}^r] < \infty\}$ for $r = 1, \ldots, 4$, of different moments of $\{\eta_{i,t}\}$ in the "strong" $GARCH(1,1)$ case.

Proposition 3 *Under conditions* (10) *and* (11) *and provided that*

$$1 - \beta_i^3 - 3\beta_i^2\alpha_i - 9\beta_i\alpha_i^2 - 15\alpha_i^3 > 0$$

the sequence $\{\eta_{i,t}\}$ *of period-t innovations is third order stationary and* $E[\eta_{i,t}^3]/\sigma_\eta < \infty$.

In the same manner, we have the following result.

Proposition 4 *Under conditions* (10) *and* (11) *and provided that*

$$1 - 4\beta_i^3\alpha_i - 18\beta_i^2\alpha_i^2 - 60\beta_i\alpha_i^3 - \beta_i^4 - 105\alpha_i^4 > 0$$

the sequence $\{\eta_{i,t}\}$ *of period-t innovations is fourth order stationary and* $E[\eta_{i,t}^4]/\sigma_\eta^2 < \infty$.

The Figure 1 represents the domains of definition D_r, $D_r \equiv \{(\beta_i, \alpha_i) \in [0,1]^2 : E[\eta_{i,t}^r] < \infty\}$ for $r = 1, \ldots, 4$, of different moments of $\{\eta_{i,t}\}$ in the "strong" $GARCH(1,1)$ case.

3.2 "Weak" $GARCH$ properties

Let us at present consider the sum $\{y_t\}$ of the two independent univariate $GARCH(1,1)$ processes $\{y_{i,t}\}, i = 1,2$

$$y_t \equiv y_{1,t} + y_{2,t}. \tag{18}$$

As in the case of "strong" $GARCH$ processes $\{y_{i,t}\}, i = 1,2$, we are considering the $ARMA$ representation of the process $\{y_t^2\}$ of squared aggregated returns. In order to determine the parameters of aggregated process we follow the approach of Nijman & Sentana [23] and report their results in Proposition 5.

Proposition 5 *Let $\{y_{i,t}\}, i = 1,2$ be two independent univariate $GARCH(1,1)$ processes defined by (7) and (8) and such that $\beta_1 + \alpha_1 = \beta_2 + \alpha_2$. Then the aggregated process $y_t \equiv y_{1,t} + y_{2,t}$ satisfies*

$$\begin{aligned} y_t^2 &= \sigma_t^2 + \eta_t, \\ \sigma_t^2 &= \psi + \beta\sigma_{t-1}^2 + \alpha y_{t-1}^2, \end{aligned}$$

where σ_t^2 is the period-t linear projection of y_t^2 on $\mathcal{F}_t' \equiv (\mathbf{1}, y_{t-1}^2, y_{t-2}^2, \ldots)$ and η_t is the period-t innovation. The parameters ψ, β and α are functions of the parameters of the two processes $\{y_{i,t}\}, i = 1,2$ and are determined by

$$\begin{aligned} \psi &= \psi_1 + \psi_2, \\ \beta(1+\beta^2)^{-1} &= \frac{\beta_1\sigma_{\eta,1}^2 + 4(\beta_1+\alpha_1)\sigma_1^2\sigma_2^2 + \beta_2\sigma_{\eta,2}^2}{(1+\beta_1^2)\sigma_{\eta,1}^2 + 4[1+(\beta_1+\alpha_1)^2]\sigma_1^2\sigma_2^2 + (1+\beta_2^2)\sigma_{\eta,2}^2}, \\ \alpha &= (\beta_1+\alpha_1) - \beta, \end{aligned}$$

and η_t is a white noise, $\{\eta_t\} \sim WN(0, \sigma_\eta^2)$, with

$$\sigma_\eta^2 = \sigma^4(k-1)\frac{1-(\beta+\alpha)^2}{1-(\beta+\alpha)^2+\alpha^2}$$

where k is the kurtosis of the aggregated process $\{y_t\}$, $k \equiv E[y_t^4](E[y_t^2])^{-2} = \sigma^{-4}E[y_t^4]$. Thus defined, the aggregated process $\{y_t\}$ is said to be "weak" $GARCH(1,1)$.

There are two results of this Proposition that need to be proven: the functional forms of the parameters ψ, β and α, and the fact that $\{\eta_t\} \sim WN(0, \sigma_\eta^2)$. The expression of ψ, β

and α as functions of the parameters of the two processes $\{y_{i,t}\}, i = 1, 2$ has already been derived by Nijman & Sentana [23] so we refer to their results. We therefore only provide the proof of the second result of Proposition 5, which has not yet been reported in the literature.

Proof of Proposition 5. see the Appendix 7. ∎

The first important result of Proposition 5 is the characterization of σ_t^2: the "weak" $GARCH$ σ_t^2 is the period-t linear projection of y_t^2 on $\mathcal{F}_t' \equiv (\mathbf{1}, y_{t-1}^2, y_{t-2}^2, \ldots)$, i.e. $\sigma_t^2 = \widehat{E}[y_t^2|\mathcal{F}_t']$ and no longer the period-t conditional expectation of y_t^2 given $\mathcal{F}_t$, i.e. $\sigma_t^2 = E[y_t^2|\mathcal{F}_t]$, as in the "strong" $GARCH$ case. Therefore, the conditional variance of y_t is no longer correctly specified. Indeed, we have

$$E[y_t^2|\mathcal{F}_t] = E[(y_{1,t} + y_{2,t})^2|\mathcal{F}_t] = \sigma_{1,t}^2 + \sigma_{2,t}^2 \neq \sigma_t^2. \tag{19}$$

Even though the conditional mean of y_t remains correctly specified, i.e. $E[y_t|\mathcal{F}_t] = 0$, there is a misspecification of the conditional variance of y_t in the "weak" $GARCH$ models. Thus, one of the main assumptions for the consistency and asymptotic normality of the Gaussian QMLE no longer holds in the "weak" $GARCH$ case. Moreover, if we define s_t^2 as the period-t conditional expectation of y_t^2 given $\mathcal{F}_t$, $s_t^2 \equiv E[y_t^2|\mathcal{F}_t]$, then s_t^2 will no longer be an "ARMA(1,1)" as in the "strong" $GARCH$ case. We can thus say that the aggregated process is no longer $GARCH(1,1)$ for the conditional expectation of y_t^2.[7]

The second important result of Proposition 5 is that the scaled variable $\frac{y_t}{\sigma_t}$ is no longer iid conditional on the information set $\mathcal{F}_t$. Indeed, let us consider the conditional moments of $\frac{y_t}{\sigma_t}$, $E[(\frac{y_t}{\sigma_t})^r|\mathcal{F}_t]$ with $r > 0$. For example, the conditional second and fourth moments of $\frac{y_t}{\sigma_t}$ are

$$E[\frac{y_t^2}{\sigma_t^2}|\mathcal{F}_t] = \sigma_t^{-2}\left(\sigma_{1,t}^2 + \sigma_{2,t}^2\right)$$

and

$$E[\frac{y_t^4}{\sigma_t^4}|\mathcal{F}_t] = \sigma_t^{-4}\left(\sigma_{1,t}^4 E[\xi_{1,t}^4] + 6\sigma_{1,t}^2\sigma_{2,t}^2 + \sigma_{2,t}^4 E[\xi_{2,t}^4]\right).$$

Therefore, both second and fourth conditional moments of $\frac{y_t}{\sigma_t}$ depend on the information set $\mathcal{F}_t$, which is in contradiction with the conditional iid property.

[7] I am thankful to Clive Granger for pointing out this property.

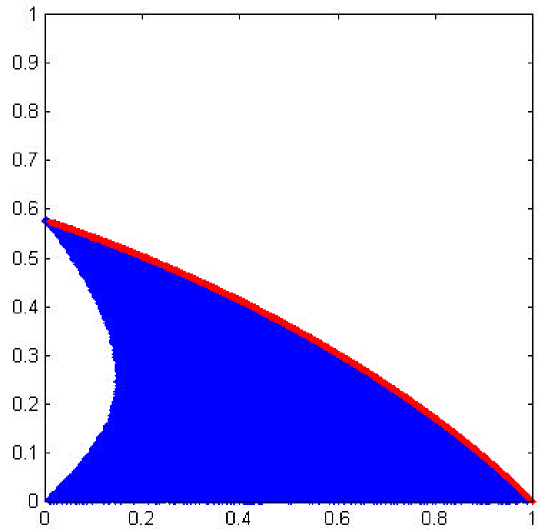

Figure 2: Domain of definition D'_2 of the second moment of "weak" GARCH(1,1) innovations $\{\eta_t\}$, $D'_2 \equiv \{(\beta,\alpha) \in [0,1]^2 : E[\eta_t^2] < \infty\}$, and the boundary of D_2 for "strong" GARCH(1,1) innovations $\{\eta_{i,t}\}$.

The third but somewhat indirect result of Proposition 5 deals with the higher-order moments existence of the period-t innovation η_t. Whereas in the "strong" $GARCH$ case we can easily derive the analytical expressions for the boundaries of the domains of definition of higher-order moments of $\{\eta_t\}$ (see Propositions 3 and 4), we can only determine them numerically in the "weak" $GARCH$ case. Figures 2, 3 and 4 represent the domains D'_r, $D'_r \equiv \{(\beta,\alpha) \in [0,1]^2 : E[\eta_t^r] < \infty\}$ for $r = 2,3,4$, where the higher-order moments of $\{\eta_t\}$ exist in the "weak" $GARCH(1,1)$ model. As we can see from different plots, we have $D'_r \subset D_r$, i.e. D'_r is strictly included in D_r.

In order to illustrate the practical implications of the previous results, let us consider the following example and compute different moments of $\{\eta_t\}$ in the "weak" $GARCH(1,1)$ models originally studied by Nijman & Sentana [23].

Example: "weak" $GARCH(1,1)$ parameters obtained by aggregation of two independent strong $GARCH(1,1)$ with parameters $(\psi_i, \beta_i, \alpha_i)$

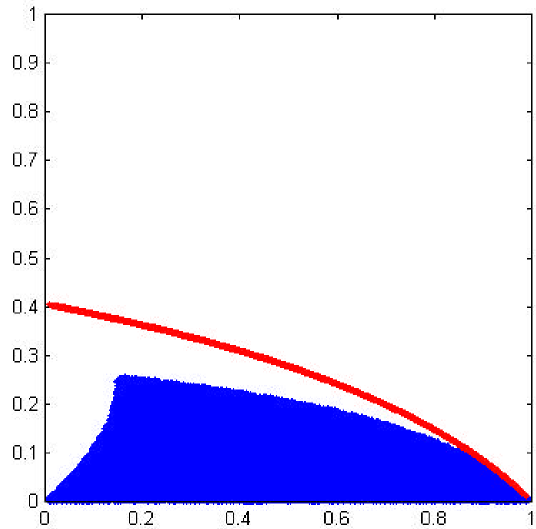

Figure 3: Domain of definition D'_3 of the third moment of "weak" GARCH(1,1) innovations $\{\eta_t\}$, $D'_3 \equiv \{(\beta, \alpha) \in [0,1]^2 : E[\eta_t^3] < \infty\}$, and the boundary of D_3 for "strong" GARCH(1,1) innovations $\{\eta_{i,t}\}$.

β_1	α_1	β_2	α_2	$\frac{\psi_1}{\psi_2}$	β	α	σ_η^2	s_η	k_η
0.5	0.35	0.8	0.05	1	0.569	0.281	8.282	∞	∞
0.5	0.35	0.8	0.05	4	0.516	0.334	112.158	∞	∞
0.5	0.35	0.8	0.05	1/4	0.705	0.145	26.167	$< \infty$	∞

The parameters s_η and k_η denote the skewness and the kurtosis of the "weak" $GARCH$ innovations $\{\eta_t\}$. Their existence depends on the values of the "weak" $GARCH$ β and α. Thus for $\beta = 0.569$ and $\alpha = 0.281$ for example, we have $E[\eta_t^3] = \infty$ and $E[\eta_t^4] = \infty$. This fact will be particularly important for the construction of the Gaussian QMLE, which we study in the next section.

The fact that σ_t^2 is a linear projection of y_t^2 on $\mathcal{F}_t$, and no longer its conditional expectation, together with the non-iid property of the scaled variable $\frac{y_t}{\sigma_t}$ and a possible non-existence of higher order moments of the innovations $\{\eta_t\}$, are important reasons why an aggregated, "weak" $GARCH$, process is fundamentally different from the classical, "strong" $GARCH$

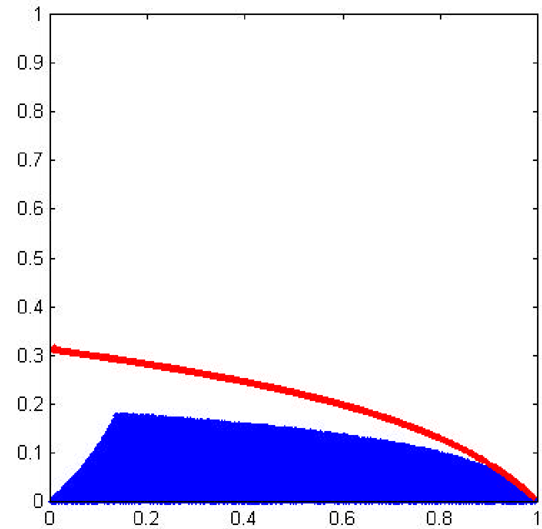

Figure 4: Domain of definition D'_4 of the fourth moment of "weak" $GARCH(1,1)$ innovations $\{\eta_t\}$, $D'_4 \equiv \{(\beta, \alpha) \in [0,1]^2 : E[\eta_t^4] < \infty\}$, and the boundary of D_4 for "strong" GARCH innovations $\{\eta_{i,t}\}$.

one. Moreover, these facts invalidate the applicability of the classical results on the consistency and asymptotic normality of the Gaussian QMLE to "weak" $GARCH$ models.

In the next paragraph we further explore the differences between the two types of processes.

Similar to the "strong" $GARCH$ case, let us consider the $ARMA(1,1)$ representation for the series of squares $\{y_t^2\}$ of aggregated process. It is given by

$$y_t^2 = \psi + (\beta + \alpha) y_{t-1}^2 + \eta_t - \beta \eta_{t-1}, \{\eta_t\} \sim WN(0, \sigma_\eta^2). \tag{20}$$

Thinking in terms of $ARMA(1,1)$ representation (20) of the "weak" $GARCH(1,1)$ process $\{y_t\}$, we can see that the vector parameter that fully determines (20) is $(\psi, \beta, \alpha, \sigma_\eta^2)$. However, as noted before, the last component of this parameter vector is not independent from the other components since

$$\sigma_\eta^2 = \sigma^4 (k-1) \frac{1 - (\beta + \alpha)^2}{1 - (\beta + \alpha)^2 + \alpha^2}. \tag{21}$$

Unlike in the "strong" $GARCH$ case, we can no longer solve this dependency in parameter components problem by using a different parametrization. Indeed the kurtosis k of the aggregated process given by

$$k = \frac{k_1\sigma_1^4 + 6\sigma_1^2\sigma_2^2 + k_2\sigma_2^4}{(\sigma_1^2 + \sigma_2^2)^2} \tag{22}$$

is dependent on the parameters of the two processes $\{y_{i,t}\}, i = 1, 2$. Consequently, k is not independent of (ψ, β, α). As in the "weak" $GARCH$ case we can no longer specify y_t as $y_t = \sigma_t \xi_t$ where $\{\xi_t\}$ is some iid process, such that $E[\xi_t] = 0$ and $E[\xi_t^2] = 1$, we can no longer re-parametrize our problem and use k_ξ instead of σ_η^2. When estimating the parameters of (20) we will therefore have to account for the dependence between the components of the parameter vector $(\psi, \beta, \alpha, \sigma_\eta^2)$.

In conclusion to this section let us resume the fundamental differences between "strong" and "weak" $GARCH$ processes and derive their implications on the estimation methods used for parameter identification in "weak" $GARCH$ processes.

First, when constructing the unobservable series $\{\sigma_t^2\}$ based on the observations $\{y_t\}$ obtained by contemporaneous aggregation of two independent $GARCH$ processes, we need to take into account the fact that σ_t^2 is the linear projection of y_t^2 on $\mathcal{F}_t'$, and no longer its conditional expectation. Consequently, the conditional variance of y_t is misspecified and we can no longer apply the classical results on the consistency and asymptotic normality of the Gaussian QMLE to "weak" $GARCH$ models. Another complication due to this new characterization of σ_t^2 is related to the construction of the latter. In order to construct estimates of σ_t^2, i.e. linear projections of y_t^2, we have to use an appropriate procedure. One of them is the innovations algorithm defined by Brockwell & Davis [8], which we describe in more details in the next section.

Second, once we have the series $\{\sigma_t^2\}$ together with the observations $\{y_t\}$, we know that the scaled variable $\frac{y_t}{\sigma_t}$ is not iid. Together with the previous considerations, this property makes the classical QMLE inapplicable in the "weak" $GARCH$ case. Instead, we propose to estimate the parameters of the "weak" $GARCH$ by considering the $ARMA$ representation for $\{y_t^2\}$. We use QMLE based on the innovations $\{\eta_t\}$.

Finally, when estimating the parameter vector of the $ARMA$ series of squared observations $\{y_t^2\}$, we need to take into account the possible dependence within its components as well as a possible non-existence of higher order moments of $\{\eta_t\}$.

4 QMLE for aggregated $GARCH$ processes

In this section we construct estimators of parameters of a "weak" $GARCH$ process. These are obtained as QMLE of parameters of an $ARMA$ process generated by a white noise sequence which does not have the usual "good" properties, i.e. an *iid* process whose variance is independent of other system parameters' values. Moreover, some of the higher-order moments of the innovations might not exist, depending on the true values of the system parameters β and α. We start by constructing the estimates of the innovations $\{\eta_t\}$ by using the Innovations algorithm. We then compute the QMLE under the Gaussian assumption. Finally, alternative density assumptions, like Laplace or α-stable, are examined.

For the simplicity of computations involved in the construction of estimates, we hereafter consider the $GARCH(1,1)$ case. Our results can however be easily generalized in the case of a higher order $GARCH$.

4.1 Innovations η_t

Recall from the previous section (20) that the process $\{y_t^2\}$ of squared returns $y_t = y_{1,t} + y_{2,t}$ follows

$$y_t^2 = \psi + (\beta + \alpha)\, y_{t-i}^2 + \eta_t - \beta\eta_{t-1}, \{\eta_t\} \sim WN\left(0, \sigma_\eta^2\right),$$

where the parameters ψ, β and α satisfy the conditions (10) and (11) and the variance σ_η^2 of the period-t innovation η_t is given by the (21). Let $\{x_t\}$ be a centered series such that for every $t \geqslant 0$ we have

$$x_t \equiv y_t^2 - E[y_t^2] = y_t^2 - \sigma^2. \tag{23}$$

The sequence $\{x_t\}$ thus defined satisfies the recursion

$$x_t - \phi x_{t-1} = \eta_t + \theta\eta_{t-1}, \{\eta_t\} \sim WN\left(0, \sigma_\eta^2\right) \tag{24}$$

where $\phi \equiv \beta + \alpha$ and $\theta \equiv -\beta$.

Given an observed sequence $\{y_t\}$ we first need to construct the innovations $\{\eta_t\}$ appearing in (24). Recall from Proposition (5) that the period-t innovation η_t is defined as $\eta_t \equiv y_t^2 - \sigma_t^2$ where σ_t^2 is the linear projection y_t^2 on the set $\mathcal{F}'_t$, i.e. $\sigma_t^2 = \widehat{E}[y_t^2|\mathcal{F}'_t]$.[8] Thus $\eta_t = y_t^2 - \widehat{E}[y_t^2|\mathcal{F}'_t]$, or equivalently,

$$\eta_t = x_t - \widehat{E}[x_t|\mathcal{F}'_t]. \tag{25}$$

We compute η_t by using Brockwell & Davis' [8] innovations algorithm, which we describe next.

Let $\widehat{x}_t$ be the best linear predictor of x_t in terms of the constant 1 and the past observations $(x_1, ..., x_{t-1})$ and r_{t-1} its mean squared error, i.e.

$$\widehat{x}_j \equiv \widehat{E}[x_j|x_1, ..., x_{j-1}] \text{ for } 2 \leqslant j \leqslant T \text{ , } \widehat{x}_1 = 0, \tag{26}$$

$$r_j = \sigma_\eta^{-2} E(x_{j+1} - \widehat{x}_{j+1})^2 \text{ for } 1 \leqslant j \leqslant T-1 \text{ .} \tag{27}$$

In the case of a causal-invertible $ARMA(1,1)$ process $\widehat{x}_t$ and r_{t-1} can be computed recursively by using the innovations algorithm proposed by Brockwell & Davis [8].[9] We hereafter recall their main results, the details of which can be found in [8] (ch 5.3).

Proposition 6 (The Innovations Algorithm) *Let $\{xt\}$ be a zero mean sequence generated by $ARMA(1,1)$ process (24) and whose autocovariance function we denote by $\gamma_X(.)$. Let $m = 1$ and by convention $\theta_j = 0$ for $j > 1$. The best linear one-step predictors $\widehat{x}_{n+1}, 0 \leqslant n \leqslant T-1$ and their mean squared errors $r_n, 0 \leqslant n \leqslant T-1$ are then given by*

$$\widehat{x}_{n+1} = \phi x_n + \theta_{n,1}\left(x_n - \widehat{x}_n\right), 1 \leqslant n \leqslant T-1,$$

and

$$E(x_{n+1} - \widehat{x}_{n+1})^2 = \sigma_\eta^2 r_n$$

[8] Recall that $\mathcal{F}'_t \equiv (\mathbf{1}, y_{t-1}^2, y_{t-2}^2, \ldots)$.

[9] The same type of result holds for $ARMA(p,q)$ processes.

where $\theta_{n,j}$ and r_n can be computed recursively from

$$\begin{cases} r_0 = \sigma_\eta^{-2}\gamma_X(0), \\ \theta_{n,n-k} = r_k^{-1}[k(n+1,k+1) - \sum_{j=0}^{k-1}\theta_{k,k-j}\theta_{n,n-j}r_j], \quad k = 0,1,\ldots,n-1, \\ r_n = k(n+1,n+1) - \sum_{j=0}^{n-1}\theta_{n,n-j}^2 r_j, \end{cases}$$

with

$$k(i,j) = \begin{cases} \sigma_\eta^{-2}\gamma_X(0), & i = j = 1, \\ \sigma_\eta^{-2}\left[\gamma_X(1) - \phi\gamma_X(0)\right], & \min(i,j) = 1 \textit{ and } \max(i,j) = 2, \\ \theta^2, & i = j > 1, \\ 0, & \textit{otherwise}. \end{cases}$$

Proof of Proposition 6. See Proposition 5.2.2. and (5.3.1)-(5.3.10) in [8]. ■

Before continuing, let us make several remarks on the above results. The main difference between the representation

$$\widehat{x}_{n+1} = \phi x_n + \theta_{n,1}\left(x_n - \widehat{x}_n\right), 1 \leqslant n \leqslant T-1, \tag{28}$$

and the one obtained through Durbin-Levinson algorithm is that in (28) $\widehat{x}_{n+1}$ depends of only one past observations x_n and only one past innovation $(x_n - \widehat{x}_n)$. Recall that the Durbin-Levinson algorithm gives a recursive formula for computing $\widehat{x}_{n+1}$ as linear combination of n past observations $x_n, \ldots, x_1$, i.e. $\widehat{x}_{n+1} = \phi_{n,1}x_n + \ldots + \phi_{n,n}x_1, 1 \leqslant n \leqslant T-1$. The representation (28) for $\widehat{x}_{n+1}$ is therefore particularly convenient from the practical point of view because it requires the storage of at most 2 past realizations in order to predict x_{n+1}.

It can also be shown that if $\{x_t\}$ is invertible, then $r_n \longrightarrow 1$, $\theta_{n,1} \longrightarrow \theta$, as $n \longrightarrow \infty$.[10]

The quantities $k(i,j)$ defined in Proposition (6) depend on $ARMA(1,1)$ parameters ϕ and θ. Moreover, by (21) we know that the variance σ_η^2 of the innovation process depends on ϕ and θ as well. This implies that the quantities $\theta_{n,1}$ and r_n are not independent of σ_η^2. This result is particularly important for the construction of the QMLE of the parameters in (24), as we show next.

[10] See [8].

4.2 Corrected Gaussian QMLE

We propose to estimate the parameters in (24) by using QMLE under the Gaussian assumption. However, since the components of the vector parameter to estimate $(\phi, \theta, \sigma_\eta^2)$ are no longer independent, we need to define a "corrected" log-likelihood function which takes into account this dependence.

Following [8] the Gaussian log-likelihood $l(\cdot)$ of the vector of observations $\mathtt{X}_T \equiv (x_1, ..., x_T)'$ is

$$l(\phi, \theta, \sigma_\eta^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma_\eta^2) - \frac{1}{2}\sum_{t=1}^{T}\ln(r_{t-1}) - \frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}} \tag{29}$$

where $\widehat{x}_t$ and r_{t-1} are obtained recursively by innovation algorithm defined in Proposition 6. The following Proposition gives the optimality conditions for $l(\phi, \theta, \sigma_\eta^2)$.

Proposition 7 *Let $l(\phi, \theta, \sigma_\eta^2)$ be defined by* (29). *Then the first order conditions for an optimum of $l(\phi, \theta, \sigma_\eta^2)$ are*

$$\sigma_\eta^2 T\left[1 + \delta_T(\phi, \theta, \sigma_\eta^2)\right] = \sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}} \tag{30}$$

$$\frac{\partial}{\partial\phi}\left(\sum_{t=1}^{T}\ln r_{t-1} + T\ln\sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}}\right) = - T\delta_T(\phi, \theta, \sigma_\eta^2)\frac{\partial}{\partial\phi}\ln\sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}} \tag{31}$$

and

$$\frac{\partial}{\partial\theta}\left(\sum_{t=1}^{T}\ln r_{t-1} + T\ln\sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}}\right) = - T\delta_T(\phi, \theta, \sigma_\eta^2)\frac{\partial}{\partial\theta}\ln\sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}} \tag{32}$$

where

$$\delta_T(\phi, \theta, \sigma_\eta^2) \equiv T^{-1}\left[\sigma_\eta^2\frac{\partial}{\partial\sigma_\eta^2}\sum_{t=1}^{T}\ln r_{t-1} + \frac{\partial}{\partial\sigma_\eta^2}\sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}}\right]. \tag{33}$$

Proof of Proposition 7. see the Appendix 9. ■

By considering the special case where the quantities $\widehat{x}_t$ and r_{t-1} do not depend on σ_η^2, i.e. the case where the system parameters $\left(\phi, \theta, \sigma_\eta^2\right)$ are independent, we obtain the usual set of first order conditions

$$\sigma_\eta^2 = T^{-1} \sum_{t=1}^{T} \frac{(x_t - \widehat{x}_t)^2}{r_{t-1}}, \tag{34}$$

$$\frac{\partial}{\partial \phi} \left(\sum_{t=1}^{T} \ln r_{t-1} + T \ln \sum_{t=1}^{T} \frac{(x_t - \widehat{x}_t)^2}{r_{t-1}} \right) = 0, \tag{35}$$

and

$$\frac{\partial}{\partial \theta} \left(\sum_{t=1}^{T} \ln r_{t-1} + T \ln \sum_{t=1}^{T} \frac{(x_t - \widehat{x}_t)^2}{r_{t-1}} \right) = 0. \tag{36}$$

The Gaussian QMLE $(\widehat{\phi}, \widehat{\theta}, \widehat{\sigma}_\eta^2)$ then satisfy

$$\widehat{\sigma}_\eta^2 = T^{-1} S(\widehat{\phi}, \widehat{\theta}), \tag{37}$$

where

$$S(\widehat{\phi}, \widehat{\theta}) = \sum_{t=1}^{T} \frac{(x_t - \widehat{x}_t)^2}{r_{t-1}}, \tag{38}$$

and $\widehat{\phi}, \widehat{\theta}$ are the minimizers of

$$l(\phi, \theta) \equiv T^{-1} \sum_{t=1}^{T} \ln r_{t-1} + \ln(T^{-1} S(\phi, \theta)). \tag{39}$$

(see for example Brockwell & Davis [8] (8.7.7)). If θ is invertible, then $r_t \longrightarrow 1$, and the term $T^{-1} \sum_{t=1}^{T} \ln r_{t-1}$ is negligible compared to $\ln(T^{-1} S(\phi, \theta))$. Therefore, the minimization of $l(\phi, \theta)$ in (39) is equivalent to the minimization of $S(\phi, \theta)$, and the maximum likelihood estimator will have similar asymptotic properties as the least squares estimator.

The main difference between the two sets of first order conditions, (34) – (36) and (30) – (32), is that the quantity δ_T defined in (33) is no longer zero if the variance σ_η^2 of the innovations depends on the parameters ϕ and θ. We therefore need to correct the scores of the classical log-likelihood function. We propose a correction method in which we replace the right hand sides of (31) and (32) by their estimated values. For example, consider the case where we have a weakly consistent preliminary estimate $(\widetilde{\phi}, \widetilde{\theta}, \widetilde{v}_n)$ of our system parameter $(\phi, \theta, \sigma_\eta^2)$.[11] We can then use $(\widetilde{\phi}, \widetilde{\theta}, \widetilde{v}_n)$ to construct the series of best linear predictors $\{\widetilde{x}_t\}$ of

[11] For details concerning the construction of $(\widetilde{\phi}, \widetilde{\theta}, \widetilde{v}_n)$ see Appendix 8.

the process $\{x_t\}$ as well as their mean squared errors $\{\widetilde{r}_{t-1}\}$. An estimator $\widetilde{\delta_T}$ of the quantity $\delta_T(\phi,\theta,\sigma_\eta^2)$ defined in (33) is then obtained from the first order condition (30)

$$\widetilde{\delta_T} = \left[T^{-1}\sum_{t=1}^{T}\frac{(x_t-\widetilde{x}_t)^2}{\widetilde{r}_{t-1}}\right]/\widetilde{v}_n - 1. \tag{40}$$

By substitution, the two first order conditions (31) and (32) now become

$$\frac{\partial}{\partial\phi}(\sum_{t=1}^{T}\ln r_{t-1} + T\ln\sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}}) = -T\widetilde{\delta_T}\widetilde{s}_\phi \tag{41}$$

and

$$\frac{\partial}{\partial\theta}(\sum_{t=1}^{T}\ln r_{t-1} + T\ln\sum_{t=1}^{T}\frac{\left(X_t-\widehat{X}_t\right)^2}{r_{t-1}}) = -T\widetilde{\delta_T}\widetilde{s}_\theta \tag{42}$$

where the constants $\widetilde{s}_\phi$ and $\widetilde{s}_\theta$ are obtained through

$$\widetilde{s}_\phi = \frac{\partial}{\partial\phi}(\ln\sum_{t=1}^{T}\frac{(x_t-\widetilde{x}_t)^2}{\widetilde{r}_{t-1}}) \text{ and } \widetilde{s}_\theta = \frac{\partial}{\partial\theta}(\ln\sum_{t=1}^{T}\frac{(x_t-\widetilde{x}_t)^2}{\widetilde{r}_{t-1}}). \tag{43}$$

The estimators $\widehat{\phi}$ and $\widehat{\theta}$ of the system parameters are the values of ϕ and θ which minimize

$$cl(\phi,\theta) \equiv T^{-1}\sum_{t=1}^{T}\ln r_{t-1} + \ln\sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}} + \widetilde{\delta_T}(\phi\widetilde{s}_\phi + \theta\widetilde{s}_\theta) \tag{44}$$

quantity that we shall refer to as the "corrected" log-likelihood. In the same manner, we obtain an estimate $\widehat{\sigma}_\eta^2$ of the variance σ_η^2 of the period-t innovation η_t

$$\widehat{\sigma}_\eta^2 = T^{-1}[1+\widetilde{\delta_T}]^{-1}\sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}} \tag{45}$$

where $\widehat{x}_t$ and r_{t-1} are obtained at the optimum, i.e. when $(\phi,\theta) = (\widehat{\phi},\widehat{\theta})$.

The estimator $(\widehat{\phi}_{CG},\widehat{\theta}_{CG},\widehat{\sigma}_{\eta CG}^2)$ obtained by maximizing the "corrected" Gaussian quasi-log-likelihood (44) will be called "corrected" Gaussian QMLE.

4.3 Alternative density assumptions

In order to estimate the parameters of the model (24)

$$x_t - \phi x_{t-1} = \eta_t + \theta\eta_{t-1}, \{\eta_t\} \sim WN\left(0,\sigma_\eta^2\right)$$

we have assumed in the previous paragraph that the innovations $\{\eta_t\}$ had a Gaussian density. We have then maximized the corresponding Gaussian quasi-log-likelihood $l(\cdot)$ of the vector of observations $\mathtt{X}_T \equiv (x_1, ..., x_T)'$

$$l(\phi, \theta, \sigma_\eta^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma_\eta^2) - \frac{1}{2}\sum_{t=1}^{T}\ln(r_{t-1}) - \frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}}$$

with $\widehat{x}_t$ and r_{t-1} being obtained recursively by Innovations algorithm defined in Proposition 6. This approach however does not take into account the heavy-tailedness of the innovations $\{\eta_t\}$, whose third and fourth moment might not exist for a particular set of values (ϕ, θ) (see the discussion in the previous section). The estimates discussed so far are standard procedures developed for the case where the innovations are Gaussian or have finite fourth moment. The estimates of the parameters of the model (24) are in that case consistent and asymptotically normal, as shown by Hannan & Rissanen [16] for example.

In the case where the innovation distribution is not normal, other density assumptions might be more suitable. Thus, Money & al [21], propose a rule, based on the kurtosis of the innovations, for selecting the most appropriate value of p for the family of $\mathcal{L}_p$ norm estimates.[12] It is well known that the $\mathcal{L}_p$ norm estimation is equivalent to the QML estimation under the assumption that the innovations have power distribution whose density is

$$\frac{1}{2\sigma\Gamma(1+1/p)}\exp[-|\frac{z-\mu}{\sigma}|^p].^{13} \tag{46}$$

Special cases include the equivalence between the $\mathcal{L}_2$ norm estimator (LS) and the Gaussian QMLE, or the one between the $\mathcal{L}_1$ norm estimator (LAD) and the Laplace (or double exponential) QMLE. We can therefore apply the results by Money & al [21] for selecting the most appropriate density assumption in the construction of the QMLE. In the case where the sample kurtosis of the innovations $\{\eta_t\}$ is greater than 6, the authors suggest the choice of $p = 1$, which corresponds to the QML under Laplace (or double exponential) assumption.[14]

[12] The $\mathcal{L}_p$ norm estimators minimize the sum of the p-th powers of the absolute deviations of the errors.

[14] In fact, the authors found that a suitable p could be chosen by using the formula

$$p = \frac{9}{K_\eta^2} + 1$$

where K_η is the kurtosis of the innovations $\{\eta_t\}$.

We therefore construct the corresponding Laplace quasi-log-likelihood $l_L(\cdot)$ of the vector of observations $\mathtt{X}_T \equiv (x_1, ..., x_T)'$

$$l_L(\phi, \theta, \sigma_\eta^2) = -\frac{T}{2}\ln(\sigma_\eta^2) - \frac{1}{2}\sum_{t=1}^{T}\ln(r_{t-1}) - \frac{1}{\sigma_\eta}\sum_{t=1}^{T}\frac{|x_t - \widehat{x}_t|}{\sqrt{r_{t-1}}} \tag{47}$$

with $\widehat{x}_t$ and r_{t-1} being obtained recursively by Innovations algorithm defined in Proposition 6. The QMLE obtained by maximizing (47) is denoted by $(\widehat{\phi}_L, \widehat{\theta}_L, \widehat{\sigma}_{\eta L}^2)$.

The latter estimate is suited for the cases where the innovations are leptokurtic, i.e. with finite kurtosis $k_\eta > 3$. In some cases, however, the kurtosis of the "weak" $GARCH$ innovations $\{\eta_t\}$ might not exist. In fact, the "weak" $GARCH(1,1)$ parameters β and α being close to the boundary of D_2', the variance of the innovations $\{\eta_t\}$ increases and tends to infinity.[15] Based on this ground, we make the assumption that the innovations $\{\eta_t\}$ have an α-stable distribution $\mathcal{S}_\alpha(\sigma, \beta, \mu)$ with parameters $(\alpha, \sigma, \beta, \mu)$. The probability densities of α-stable random variables exist and are continuous but, with few exceptions, they are not known in closed form (see [24]). Among the exceptions are (1) the Gaussian distribution $\mathcal{S}_2(\sigma, 0, \mu) = \mathcal{N}(\mu, 2\sigma^2)$ whose density is $\frac{1}{2\sigma\sqrt{\pi}}\exp[-\frac{(z-\mu)^2}{4\sigma^2}]$,and (2) the Cauchy distribution $\mathcal{S}_1(\sigma, 0, \mu)$ whose density is

$$\frac{\sigma}{\pi((z-\mu)^2 + \sigma^2)}.$$

In general, we only know the characteristic function $\Phi_Z(t) \equiv E[\exp(itZ)]$ of a random variable Z that has a stable distribution $\mathcal{S}_\alpha(\sigma, \beta, \mu)$

$$\Phi_Z(t) = \begin{cases} \exp[-\sigma^\alpha |t|^\alpha (1 - i\beta(\text{sign } t)\tan\frac{\pi\alpha}{2}) + i\mu t], & \text{if } \alpha \neq 1, \\ \exp[-\sigma|t|(1 - i\beta\frac{2}{\pi}(\text{sign } t)\ln|t|) + i\mu t], & \text{if } \alpha = 1. \end{cases} \tag{48}$$

The parameter α is the index of stability, $0 < \alpha \leqslant 2$, and

$$\text{sign } t = \begin{cases} 1, & \text{if } t > 0, \\ 0, & \text{if } t = 0, \\ -1, & \text{if } t < 0. \end{cases}$$

The scale parameter σ, $\sigma \geqslant 0$, the skewness parameter β, $-1 \leqslant \beta \leqslant 1$, and the location (or shift) parameter μ, $\mu \in \mathbb{R}$, are unique.[16] Let $f_\alpha(\cdot)$ denote a symmetric stable density

[15] Recall that $D_2' \equiv \{(\beta, \alpha) \in [0,1]^2 : E[\eta_t^2] < \infty\}$.

[16] β is irrelevant when $\alpha = 2$.

with index α and scale parameter 1. We compute the α-stable QMLE by minimizing the quasi-log-likelihood $l_S(\cdot)$ of the vector of observations $\mathtt{X}_T \equiv (x_1, ..., x_T)'$

$$l_S(\phi, \theta, (\alpha, \sigma, \beta, \mu)) = -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{t=1}^{T}\ln(r_{t-1}) - \sum_{t=1}^{T}\ln f_\alpha(\frac{x_t - \widehat{x}_t - \mu}{\sigma\sqrt{r_{t-1}}}) \tag{49}$$

with $\widehat{x}_t$ and r_{t-1} being obtained recursively by Innovations algorithm defined in Proposition 6.[17] The α-stable QMLE could be obtained by simultaneously maximizing $l_S(\cdot)$ over the ARMA parameters (ϕ, θ) and the α-stable distribution parameters $(\alpha, \sigma, \beta, \mu)$. Instead, we use the weakly consistent preliminary estimates $(\widetilde{\phi}, \widetilde{\theta}, \widetilde{v}_n)$ to construct the series of best linear predictors $\{\widetilde{x}_t\}$ of the process $\{x_t\}$, as well as their mean squared errors $\{\widetilde{r}_{t-1}\}$. We then construct consistent quantile estimates $(\widetilde{\alpha}, \widetilde{\sigma}, \widetilde{\beta}, \widetilde{\mu})$ of the α-stable distribution parameters $(\alpha, \sigma, \beta, \mu)$ by using McCulloch's [20] method (see the Appendix 10). These estimates are then used to form a "reduced" $\widetilde{\alpha}$-stable quasi-log-likelihood $l_{\widetilde{S}}(\cdot)$

$$l_{\widetilde{S}}(\phi, \theta) = -\frac{1}{2}\sum_{t=1}^{T}\ln(r_{t-1}) - \sum_{t=1}^{T}\ln f_{\widetilde{\alpha}}(\frac{x_t - \widehat{x}_t - \widetilde{\mu}}{\widetilde{\sigma}\sqrt{r_{t-1}}}). \tag{50}$$

The QMLE obtained by maximizing (50) is denoted by $(\widehat{\phi}_{\widetilde{S}}, \widehat{\theta}_{\widetilde{S}})$.

The asymptotic properties of different estimators obtained, namely: Gaussian QMLE, Laplace QMLE and α-stable QMLE, are further studied by Monte Carlo simulations. As an alternative approach, bootstrapping methods can be employed to approximate the sampling distribution of these QMLE estimates. Unfortunately, the limit of the bootstrap approximation to the sampling distribution in an ARMA setting is still unknown.

4.4 Asymptotic properties of QML estimators

We study the properties of different QML estimators by performing a series of Monte-Carlo simulations. The true data generating process (TDGP) parameters θ_0 used to generate the series are the ones originally studied by Nijman & Sentana [23] and reported in Table 4. The sample means and standard deviations of the QML estimates obtained under different density assumptions are reported in Tables 1-3.

[17] Note that if $Z \sim \mathcal{S}_\alpha(1, 0, 0)$ then $\sigma Z + \mu \sim \mathcal{S}_\alpha(\sigma, 0, \mu)$.

Table 1: "Weak" $GARCH$ with TDGP parameter $(\beta_0, \alpha_0) = (0.569, 0.281)$ and its QMLEs under various density assumptions.

	$\boldsymbol{\theta}_0$	$\widehat{\boldsymbol{\theta}}_{N\&S}$	$\widehat{\boldsymbol{\theta}}_G$	$\widehat{\boldsymbol{\theta}}_L$	$\widehat{\boldsymbol{\theta}}_S$
β	**0.569**	0.690 (–)	0.6572 (0.0857)	0.4293 (0.1700)	**0.5591** (0.1701)
α	**0.281**	0.125 (–)	0.1547 (0.0488)	0.3238 (0.0640)	**0.3130** (0.1822)
N	–	30	36717	509	1403
T	–	80000	5000	5000	5000

By examining the sample means of the QML estimators $\widehat{\boldsymbol{\theta}}_G \equiv (\widehat{\phi}_G, \widehat{\theta}_G)$, $\widehat{\boldsymbol{\theta}}_L \equiv (\widehat{\phi}_L, \widehat{\theta}_L)$ and $\widehat{\boldsymbol{\theta}}_{\tilde{S}} \equiv (\widehat{\phi}_{\tilde{S}}, \widehat{\theta}_{\tilde{S}})$ obtained, respectively, under Gaussian, Laplace and α-stable law assumption, we can see that $\widehat{\boldsymbol{\theta}}_{\tilde{S}}$ performs the best in estimating the MA parameter β in all three "weak" $GARCH(1,1)$ processes studied. The asymptotic bias of $\widehat{\beta}$ obtained by QML under α-stable law assumption is thus smaller than under either Gaussian or Laplace assumption. This result is quite intuitive in the case when both skewness and kurtosis of the aggregated process are infinite, and its variance close to the boundary of D_2', i.e. the case $(\beta_0, \alpha_0) = (0.516, 0.334)$. The assumption that the innovations are α-stable distributed, which implies that they have infinite variance, seem suited for this particular case of figure. We would expect, however, that in the case when skewness and kurtosis of the aggregated process are infinite but its variance quite small, i.e. the case $(\beta_0, \alpha_0) = (0.569, 0.281)$, the Laplace assumption would lead to the best results in terms of asymptotic consistency. This expectation is not confirmed by the empirical findings reported in Table 1. The result obtained in the finite skewness case, i.e. when $(\beta_0, \alpha_0) = (0.705, 0.145)$, is even more interesting with that the respect to the point raised previously. In this particular case, the Laplace QMLE performs very badly in terms of asymptotic consistency, whereas the α-stable QMLE presents small asymptotic bias. These findings would suggest that as long as the kurtosis of the "weak" $GARCH$ innovations is infinite, the best performing QMLE is the one that does not assume the finiteness of their variance. The importance of the assumption generally used when estimating the parameters of a general ARMA model, seems to be confirmed in our case. Unfortunately, there are yet no theoretical results on the asymptotic behavior of QMLE relating the latter to the existence of different moments of the innovations.

Table 2: "Weak" $GARCH$ with TDGP parameter $(\beta_0, \alpha_0) = (0.516, 0.334)$ and its QMLEs under various density assumptions.

	$\boldsymbol{\theta}_0$	$\widehat{\boldsymbol{\theta}}_{N\&S}$	$\widehat{\boldsymbol{\theta}}_G$	$\widehat{\boldsymbol{\theta}}_L$	$\widehat{\boldsymbol{\theta}}_S$
β	**0.516**	0.581 $(-)$	0.5782 (0.0835)	0.4093 (0.1284)	**0.5256** (0.1431)
α	**0.334**	0.234 $(-)$	0.2351 (0.0507)	**0.3427** (0.0490)	0.4116 (0.1524)
N	–	30	32046	417	152
T	–	80000	5000	5000	5000

In terms of the $GARCH(1,1)$ AR parameter α estimates, we can no longer claim that the α-stable QMLE performs the best. As seen from Tables 1-3, $\widehat{\alpha}_{\widetilde{S}}$ presents the smallest asymptotic bias in the case when skewness and kurtosis of the aggregated process are infinite but its variance quite small, i.e. the case $(\beta_0, \alpha_0) = (0.569, 0.281)$. Counter intuitively, it performs worse than $\widehat{\alpha}_L$ in the case $(\beta_0, \alpha_0) = (0.516, 0.334)$, i.e. when both skewness and kurtosis of the aggregated process are infinite, and its variance close to the boundary of D_2'. We however need to point out that this result might be influenced by the fact that we have quite a small sample of $\widehat{\alpha}_{\widetilde{S}}$ obtained by simulations (152 simulations). This is due to the very slow convergence of our optimization algorithm in this particular case of figure. It should also be noted that the primitive parameters estimated by maximizing the quasi-likelihood functions (39), (47) and (50), are in fact $(\phi, \theta) \equiv (\beta + \alpha, -\beta)$. Hence we can complete the results of Table 2 by noting that $\widetilde{\phi}_{\widetilde{S}} = 0.9372$, $\widetilde{\phi}_L = 0.7520$ and $\widetilde{\phi}_G = 0.8133$, knowing that the TDGP value ϕ_0 is always given by $\phi_0 = 0.85$. The Gaussian assumption seems therefore to give the best QML estimate of the AR parameter ϕ. The same applies to the case studied in Table 3, when the skewness of the "weak" $GARCH$ innovations is finite. In the latter case we have $\widetilde{\phi}_{\widetilde{S}} = 0.8050$, $\widetilde{\phi}_L = 0.7018$ and $\widetilde{\phi}_G = 0.8296$. These result would suggest that the QMLE is more robust to the departures from moment finiteness when estimating for the

AR parameter in (20), than for the MA parameter. The latter observation seems to confirm the well know empirical fact that the MA parameter an ARMA(1,1) model, such as (20), is more difficult to estimate than the AR one.[18]

Table 3: "Weak" $GARCH$ with TDGP parameter $(\beta_0, \alpha_0) = (0.705, 0.145)$ and its QMLEs under various density assumptions.

	$\boldsymbol{\theta}_0$	$\widehat{\boldsymbol{\theta}}_{N\&S}$	$\widehat{\boldsymbol{\theta}}_G$	$\widehat{\boldsymbol{\theta}}_L$	$\widehat{\boldsymbol{\theta}}_S$
β	**0.705**	0.775	0.7578	0.4066	**0.7381**
		(−)	(0.0845)	(0.2089)	(0.1465)
α	**0.145**	0.062	**0.0718**	0.2952	0.0669
		(−)	(0.0330)	(0.0879)	(0.0997)
N	–	30	30081	613	821
T	–	80000	5000	5000	5000

5 Conclusion

In conclusion let us resume the fundamental properties of the "weak" $GARCH$ models and stress out some of their implications. First, the scaled variable $\frac{y_t}{\sigma_t}$ looses its iid property by aggregation. Therefore, any testing procedure based on this property becomes invalid by aggregation of the series. Second, the "weak" $GARCH$ models are difficult to estimate by QML because of the misspecification of the conditional second moment of y_t. Different density assumptions will lead to different asymptotic properties of the estimators, which are so far tractable only by numerical simulation. As an alternative approach one could use the generalized method of moments (GMM) estimators. The major inconvenient of this approach however, is the necessity to derive analytic expressions for various moments of the innovations η_t. The latter are not only tedious to derive but also dependent on the parameters of the two original "strong" $GARCH$ processes, which are in general unknown. This brings us to the third important property of the "weak" $GARCH$. The problem of the

[18] In an ARMA setting the MA parameters are the coefficients relatives to the innovation lags and are therefore more sensitive to different specifications of the innovations' distribution.

existence of some higher-order moments of the innovations η_t cannot be assessed directly since we do not have an analytical form of their domains of definition.

The QML estimation of "weak" $GARCH$ models is a nice example of what can happen to the asymptotic properties of the estimators if some important assumptions are relaxed. The speed of convergence of the QMLE might also be affected by the choice of a particular density, as is the case for the MLE. Indeed, it is well know that in the Gaussian case the MLE is $n^{1/2}$-consistent, while in the α-stable case, its rate of convergence is proportional to $n^{1/\alpha}$. Wether this property still holds for QMLE is an open question. A theoretical study of the asymptotical properties of the QMLE under alternative density assumptions, namely Laplace or α-stable, appears to be an important question for the future research, which is however, out of the scope of this paper

The Monte-Carlo simulation study conducted in this paper shows that α-stable assumptions leads to overall better QML estimates of the parameters of an aggregated $GARCH(1,1)$ process. The α-stable QMLE of the MA parameter has the smallest asymptotic bias among the class of estimators studied. The asymptotic bias of an α-stable QMLE for the AR parameter is however larger than for the MA parameter, and is minimal under Gaussian assumption. The latter result would suggest that the QMLE is more robust to the departures from moment finiteness when estimating for the AR parameter than for the MA parameter. We leave a more theoretical study of this empirical conclusion for future research.

References

[1] Akaike, H., (1969), Fitting autoregressive models for prediction, *Annals of the Institute for Statistical Mathematics*, 21, 243-247.

[2] Akaike, H., (1977), On entropy maximization principle, in P.R.Krishnaiah, ed., *Applications of Statistics*, North Holland: Amsterdam, 27-41.

[3] Bera, A.K., and M.L. Higgins, (1992), A class of nonlinear ARCH models, *International Economic Review*, 33(1), 137–158.

[4] Bollerslev, T., (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31(3), 307–328.

[5] Bollerslev, T., Chou, R.Y., and K.F. Kroner, (1992), ARCH modeling in finance, *Journal of Econometrics*, 52(1-2), 5–59.

[6] Bollerslev, T., Engle, R.F. and D.B. Nelson, (1994), ARCH models, in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, vol. 4, Elsevier Science, 2959–3038.

[7] Bollerslev, T., and Wooldridge, J. M. (1992), Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews*, 11, 143-179.

[8] Brockwell, P.J., and Davis, R.A., (1987), *Time Series: Theory and Methods*, Springer-Verlag New York.

[9] Drost, F.C., and Nijman, T.E. (1993), Temporal Aggregation of GARCH Processes, *Econometrica*, 61, 909-927.

[10] Engle, R.F., (1982), Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50(4), 987–1007.

[11] Gourieroux, C., Monfort, A., and Trognon, (1984), Pseudo maximum likelihood methods: theory, *Econometrica*, 52, 681-700.

[12] Hannan, E.J., (1970), *Multiple Time Series*, ed. John Wiley & Sons.

[13] Hannan, E.J., (1975), The estimation of ARMA models, *The Annals of Statistics*, 4, 975-981.

[14] Hannan, E.J., (1980), The estimation of the order of an ARMA process, *The Annals of Statistics*, 5, 1071-1081.

[15] Hannan, E.J., and Quinn, B.G., (1979), The determination of the order of of an autoregression, *Journal of the Royal Statistical Society: Series B*, 41, 190-195.

[16] Hannan, E.J., and Rissanen, R., (1982), Recursive estimation of ARMA order, *Biometrika*, 69, 81-94.

[17] Johnson, N.L., S. Kotz, and N. Balakrishnan, (1994), *Continuous univariate distribution*, vol. 1, John Wiley and Sons, New York, 2nd ed.

[18] Lee, S.W., and Hansen, B.E., (1994), Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator, Econometric Theory, 10, 29-52.

[19] Lumsdaine, R. L., (1996), Consistency and Asymptotic Normality of the Quasi-maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models, *Econometrica*, 64, 575-596.

[20] McCulloch, J.H., (1986), Simple consistent estimators of stable distribution parameters, *Communications in Statistics - Computation and Simulation*, 15, 1109–1136.

[21] Money, A.H., and al., (1982), The linear regression model: $\mathcal{L}_p$ norm estimation and the choice of p, *Communications in Statistics - Computation and Simulation*, 11, 89-109.

[22] Newey, W.K. and Steigerwald, D.G., (1997), Asymptotic Bias for Quasi-Maximum Likelihood Estimators in Conditional Heteroskedasticity Models, *Econometrica*, 65, 587-599.

[23] Nijman, T.E, and Sentana, E., (1996), Marginalization and Contemporaneous Aggregation in Multivariate GARCH processes, *Journal of Econometrics*, 71, 71-87.

[24] Samorodnitsky, G., and Taqqu, M.S., (1994), *Stable non-Gaussian random processes*, Chapman & Hall, New York.

[25] Schwarz, G., (1978), Estimating the dimension of a model, *The Annals of Statistics*, 6, 461-464.

[26] Stout, W.F., (1974), *Almost Sure Convergence*, New York: Academic Press.

[27] Weiss, A. A. (1984), ARMA models with ARCH errors, *Journal of Time Series Analysis*, 5, 129-143.

[28] Weiss, A. A. (1986a), Asymptotic Theory for ARCH Models: Estimation and Testing, *Econometric theory*, 2, 107-131.

[29] White, H., (1982), Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1-25.

6 Appendix

Proof of Proposition 6. In what follows we consider a single univariate $GARCH(p,q)$ process y_t defined by

$$y_t = \sigma_t \xi_t \tag{51}$$

$$\sigma_t^2 = \psi + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j y_{t-j}^2 \tag{52}$$

with $E[\xi_t] = 0$ and $E\left[\xi_t^2\right] = 1$. Moreover define the period-t innovation $\eta_t \equiv y_t^2 - \sigma_t^2$. Let F_t be the σ-algebra generated by the observations $y_1^2, ..., y_t^2$, i.e. $F_t \equiv \overline{sp}\{y_s, s \leqslant t\}$. We then have

$$\begin{aligned} E[\eta_t] &= E\left[y_t^2 - \sigma_t^2\right] \\ &= E\left[\left(\xi_t^2 - 1\right)\sigma_t^2\right] \\ &= E\left[E\left[\left(\xi_t^2 - 1\right)\sigma_t^2 | F_{t-1}\right]\right] \\ &= E\left[\sigma_t^2\left(E\left[\xi_t^2 | F_{t-1}\right] - 1\right)\right] \\ &= 0 \end{aligned}$$

for all t . We can further calculate $E[\eta_t^2]$ for all t by considering

$$\begin{aligned} E\left[\eta_t^2\right] &= E\left[\left(\left(\xi_t^2 - 1\right)\sigma_t^2\right)^2\right] \\ &= E\left[\sigma_t^4\left(\xi_t^4 - 2\xi_t^2 + 1\right)\right] \\ &= E\left[\sigma_t^4 E\left[\xi_t^4 - 2\xi_t^2 + 1 | F_{t-1}\right]\right] \\ &= \left(E\left[\xi_t^4\right] - 1\right) E\left[\sigma_t^4\right]. \end{aligned}$$

From equation (52) we have

$$\begin{aligned}\sigma_t^4 &= (\psi+\sum_{i=1}^{p}\beta_i\sigma_{t-i}^2+\sum_{j=1}^{q}\alpha_j y_{t-j}^2)^2 \\ &= \psi^2+\sum_{i=1}^{p}\beta_i^2\sigma_{t-i}^4+\sum_{j=1}^{q}\alpha_j^2 y_{t-j}^4+2\psi\sum_{i=1}^{p}\beta_i\sigma_{t-i}^2+2\psi\sum_{j=1}^{q}\alpha_j y_{t-j}^2+2\sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_j\beta_i\sigma_{t-i}^2 y_{t-j}^2.\end{aligned}$$

We know from (51) that $E\left[y_t^4\right]=E\left[\sigma_t^4\xi_t^4\right]=E\left[\sigma_t^4\right]E\left[\xi_t^4\right]$. Together with the fact that $E\left[y_t^2\sigma_t^2\right]=E\left[\sigma_t^4\right]$ the above relation becomes

$$E\left[\sigma_t^4\right]=E\left[\sigma_t^4\right](\sum_{i=1}^{p}\beta_i^2+E\left[\xi_t^4\right]\sum_{j=1}^{q}\alpha_j^2+2\sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_j\beta_i)+\psi^2+2\psi(\sum_{i=1}^{p}\beta_i+\sum_{j=1}^{q}\alpha_j)\sigma^2$$

where

$$\sigma^2\equiv E\left[y_t^2\right]=\psi\left(1-\sum_{i=1}^{p}\beta_i-\sum_{j=1}^{q}\alpha_j\right)^{-1}.$$

Thus

$$E\left[\sigma_t^4\right](1-\sum_{i=1}^{p}\beta_i^2-E\left[\xi_t^4\right]\sum_{j=1}^{q}\alpha_j^2-2\sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_j\beta_i)=\psi^2+2\psi(\sum_{i=1}^{p}\beta_i+\sum_{j=1}^{q}\alpha_j)\sigma^2 \qquad (53)$$

Condition (17) together with the non-negativity conditions (16) ensures that $\sum_{i=1}^{p}\beta_i+\sum_{j=1}^{q}\alpha_j<1$ so that the process $\{y_t^2\}$ is covariance-stationary. The relation (53) can be rewritten as

$$E\left[\sigma_t^4\right](1-(\sum_{i=1}^{p}\beta_i+\sum_{j=1}^{q}\alpha_j)^2+(1-E\left[\xi_t^4\right])\sum_{j=1}^{q}\alpha_j^2)=\sigma^4(1-(\sum_{i=1}^{p}\beta_i+\sum_{j=1}^{q}\alpha_j)^2).$$

Given the condition

$$1-(\sum_{i=1}^{p}\beta_i+\sum_{j=1}^{q}\alpha_j)^2+(1-E\left[\xi_t^4\right])\sum_{j=1}^{q}\alpha_j^2>0$$

we then have

$$E\left[\eta_t^2\right]=\sigma^4\left(E\left[\xi_t^4\right]-1\right)(1-(\sum_{i=1}^{p}\beta_i+\sum_{j=1}^{q}\alpha_j)^2)(1-(\sum_{i=1}^{p}\beta_i+\sum_{j=1}^{q}\alpha_j)^2+\left(1-E\left[\xi_t^4\right]\right)\sum_{j=1}^{q}\alpha_j^2)^{-1}$$

which proves that $E\left[\eta_t^2\right]=\sigma_\eta^2<\infty$, for all t. Moreover for every pair (t,s) such that $s<t$ we have

$$\begin{aligned}E\left[\eta_t\eta_s\right] &= E\left[\left(y_t^2-\sigma_t^2\right)\left(y_s^2-\sigma_s^2\right)\right] \\ &= E\left[\left(y_s^2-\sigma_s^2\right)\left(E\left[y_t^2|F_s\right]-E\left[\sigma_t^2|F_s\right]\right)\right].\end{aligned}$$

Thus, in order to show that $E\left[\eta_t\eta_s\right]=0$ we need to prove

$$E\left[y_t^2|F_s\right]=E\left[\sigma_t^2|F_s\right] \text{ , for all } s<t. \tag{54}$$

This is easily verified by noting that for all $s<t$ we have $E\left[y_t^2|F_s\right]=E\left[E\left[E\left[y_t^2|F_{t-1}\right]|F_{t-2}\right]...|F_s\right]$ since $F_s\subseteq ...\subseteq F_{t-1}$. Thus

$$\begin{aligned} E\left[y_t^2|F_s\right] &= E\left[E\left[\sigma_t^2|F_{t-2}\right]...|F_s\right] \\ &= E\left[\sigma_t^2|F_s\right] \end{aligned}$$

which establishes (54). Finally, the process $\{\eta_t\}$ satisfies

$$E\left[\eta_t\right]=0 \text{ and } E[\eta_t\eta_s]=\begin{cases} \sigma_\eta^2<\infty \text{ , if } t=s \\ 0 \text{ , if } t\neq s \end{cases},$$

i.e. $\{\eta_t\}\sim WN\left(0,\sigma_\eta^2\right)$ which completes the proof of Proposition 2. ■

7 Appendix

Proof of Proposition 5. Following the developments of Nijman & Sentana [23] we know that

$$\begin{aligned} y_t^2 &= (y_{t,1}+y_{t,2})^2 \\ &= \psi_1+\psi_2+\left(\beta_1+\alpha_1\right)y_{t-1}^2 \\ &\quad +\eta_{1,t}-\beta_1\eta_{1,t-1}+\eta_{2,t}-\beta_2\eta_{2,t-1}+2y_{1,t}y_{2,t}-2\left(\beta_1+\alpha_1\right)y_{1,t-1}y_{2,t-1}. \end{aligned} \tag{55}$$

By introducing the parameter $\psi\equiv\psi_1+\psi_2$ and taking into account that $\beta_1+\alpha_1=\beta_2+\alpha_2$, we rewrite (55) as

$$y_t^2=\psi+(\beta_1+\alpha_1)y_{t-1}^2+u_t,$$

where

$$u_t\equiv\eta_{1,t}-\beta_1\eta_{1,t-1}+\eta_{2,t}-\beta_2\eta_{2,t-1}+2y_{1,t}y_{2,t}-2\gamma y_{1,t-1}y_{2,t-1}. \tag{56}$$

It is easy to show that $\{u_t\}$ is a zero-mean stationary process with autocovariance function $\gamma_u(\cdot)$ such that $\gamma_u(h)=0$ for $|h|>1$ and $\gamma_u(1)\neq 0$. Following the results of Proposition

3.2.1 [8] we know that $\{u_t\}$ is an $MA(1)$ process, i.e. there exists a coefficient β and a white noise process $\{\eta_t\}$ such that

$$u_t = \eta_t - \beta\eta_{t-1}, \{\eta_t\} \sim WN(0, \sigma_\eta^2). \tag{57}$$

Rewriting (57) then yields to (Equation 10, [23])

$$y_t^2 = \psi + (\beta_1 + \alpha_1)y_{t-1}^2 + \eta_t - \beta\eta_{t-1}, \{\eta_t\} \sim WN(0, \sigma_\eta^2). \tag{58}$$

The autoregressive parameter is identified as $\beta + \alpha$, so that (58) becomes

$$y_t^2 = \psi + (\beta + \alpha)y_{t-1}^2 + \eta_t - \beta\eta_{t-1}, \{\eta_t\} \sim WN(0, \sigma_\eta^2).$$

The invertibility condition being satisfied we consider the $MA(\infty)$ representation of $\{y_t^2\}$ given by

$$y_t^2 = \sigma^2 + \eta_t + \sum_{j=1}^{\infty} \alpha(\alpha + \beta)^{j-1}\eta_{t-j} \tag{59}$$

where $\sigma^2 \equiv E[y_t^2] = \dfrac{\psi}{1 - (\beta + \alpha)}$. Using (59) we can easily compute the second moment of $\{y_t^2\}$. Thus

$$\begin{aligned} E\left[y_t^4\right] &= E[(\sigma^2 + \eta_t + \sum_{j=1}^{\infty} \alpha(\beta + \alpha)^{j-1}\eta_{t-j})^2] \\ &= \sigma^4 + E[\eta_t^2] + \sum_{j=1}^{\infty} \alpha^2(\beta + \alpha)^{2(j-1)}E[\eta_{t-j}^2] \\ &= \sigma^4 + \sigma_\eta^2(1 + \alpha^2 \sum_{j=1}^{\infty} (\beta + \alpha)^{2(j-1)}) \\ &= \sigma^4 + \sigma_\eta^2 \frac{1 - (\beta + \alpha)^2 + \alpha^2}{1 - (\beta + \alpha)^2}. \end{aligned}$$

If $1 - (\beta + \alpha)^2 + \alpha^2 > 0$ we have

$$\sigma_\eta^2 = \sigma^4(k - 1)\frac{1 - (\beta + \alpha)^2}{1 - (\beta + \alpha)^2 + \alpha^2} \tag{60}$$

where we define the kurtosis k of the aggregated process $\{y_t\}$ as $k \equiv E\left[y_t^4\right]\left(E\left[y_t^2\right]\right)^{-2} = \sigma^{-4}E\left[y_t^4\right]$. ■

8 Appendix

In this Appendix we construct weakly consistent preliminary estimate $(\widetilde{\phi}, \widetilde{\theta}, \widetilde{v}_n)$ of the parameter vector $(\phi, \theta, \sigma_\eta^2)$ by fitting a high-order MA model to $\{x_t\}$. The series $\{x_t\}$ being a causal zero-mean stationary sequence, we can write its $MA(\infty)$ representation

$$x_t = \sum_{i=0}^{\infty} \pi_i \eta_{t-i} \ , \ \{\eta_t\} \sim WN(0, \sigma_\eta^2), \tag{61}$$

where, by using the $MA(\infty)$ representation of (24), the coefficients π_i satisfy

$$\begin{cases} \pi_0 = 1 \\ \pi_i = \phi^{i-1}(\theta + \phi), \quad i = 1, \ldots \end{cases} \tag{62}$$

We can fit an MA process of order $n < T$ to the data $\mathtt{X}_T = (x_1, ..., x_T)'$ by using the innovation algorithm [8].[19] The fitted $MA(n)$ process is

$$x_t = \sum_{i=0}^{n} \widetilde{\theta}_{n,i} \eta_{t-i} \ , \ \{\eta_t\} \sim WN(0, \widetilde{v}_n) \text{ for } t = n+1, ..., T \tag{63}$$

where $\widetilde{\theta}_n \equiv \left(\widetilde{\theta}_{n,1}, \ldots, \widetilde{\theta}_{n,n}\right)'$ and $\widetilde{v}_n$ are the innovation estimates. We use the innovations algorithm[20] to fit moving average processes of increasing orders $n = 1, 2, ..., N$,with $N = o(T^{1/3})$ to the data and thus determine $\widetilde{\theta}_n$ and $\widetilde{v}_n$ from the recursions

$$\begin{cases} \widetilde{v}_0 = \widehat{\gamma}(0), \\ \widetilde{\theta}_{n,n-k} = \widetilde{v}_k^{-1} \left(\widehat{\gamma}(n-k) - \sum_{j=0}^{k-1} \widetilde{\theta}_{k,k-j} \widetilde{\theta}_{n,n-j} \widetilde{v}_j \right), \quad k = 0, 1, \ldots, n-1, \\ \widetilde{v}_n = \widehat{\gamma}(0) - \sum_{j=0}^{n-1} \widetilde{\theta}_{n,n-j}^2 \widetilde{v}_j, \end{cases} \tag{64}$$

where $\widehat{\gamma}(\cdot)$ is the empirical autocovariance function, $\widehat{\gamma}(h) \equiv T^{-1} \sum_{t=1}^{T-h} (x_t - \overline{x}_T)(x_{t+h} - \overline{x}_T)$, for every $h : 0 \leqslant h \leqslant T-1$, and $\overline{x}_T \equiv T^{-1} \sum_{t=1}^{T} x_t$.

[19] Later on we determine the optimal value of n.

[20] See, for example, Definition 8.3.1, in [8].

We can then compute $f(n) = \ln \widetilde{v}_n + n\frac{\ln T}{T}$, $n = 1, 2, ..., N$ and choose the optimal value n^* such that $n^* \equiv \arg\min_n f(n)$. We then obtain the parameters $\widetilde{\theta}_{n^*}$ and $\widetilde{v}_{n^*}$, which are estimates of the coefficients $\pi_i, 1 \leqslant i \leqslant n^*$ and the innovation variance σ_η^2 as defined in (61). For $i = 1, 2$ we then replace π_i by $\widetilde{\theta}_{n^*,i}$ in (62) and obtain the following system of equations

$$\begin{aligned} \widetilde{\theta}_{n^*,1} &= \theta + \phi, \\ \widetilde{\theta}_{n^*,2} &= \theta(\theta + \phi). \end{aligned} \tag{65}$$

Thus

$$\begin{aligned} \widetilde{\theta} &= \widetilde{\theta}_{n^*,2}/\widetilde{\theta}_{n^*,1}, \\ \widetilde{\phi} &= \widetilde{\theta}_{n^*,1} - \widetilde{\theta}_{n^*,2}/\widetilde{\theta}_{n^*,1}. \end{aligned} \tag{66}$$

The preliminary estimate $\left(\widetilde{\phi}, \widetilde{\theta}, \widetilde{v}_n\right)$ of the system parameter $(\phi, \theta, \sigma_\eta^2)$ obtained by this method is weakly consistent as shown in [8].

9 Appendix

Proof of Proposition 7. In this proof we derive the first order conditions for an optimum of

$$l(\phi, \theta, \sigma_\eta^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma_\eta^2) - \frac{1}{2}\sum_{t=1}^{T}\ln\left(r_{t-1}\right) - \frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}\frac{(x_t - \widehat{x}_t)^2}{r_{t-1}}.$$

Differentiating $l(\cdot)$ gives

$$dl(\phi, \theta, \sigma_\eta^2) = \frac{\partial l(\phi, \theta, \sigma_\eta^2)}{\partial \phi}d\phi + \frac{\partial l(\phi, \theta, \sigma_\eta^2)}{\partial \theta}d\theta + \frac{\partial l(\phi, \theta, \sigma_\eta^2)}{\partial \sigma_\eta^2}d\sigma_\eta^2. \tag{67}$$

Recall that $\phi \equiv \beta + \alpha$, $\theta = -\beta$ and the variance σ_η^2 of the innovation process is given by (21). Thus we have

$$\begin{aligned} d\phi &= d\alpha + d\beta, \\ d\theta &= -d\beta, \\ d\sigma_\eta^2 &= \frac{\partial \sigma_\eta^2}{\partial \psi}d\psi + \frac{\partial \sigma_\eta^2}{\partial \beta}d\beta + \frac{\partial \sigma_\eta^2}{\partial \alpha}d\alpha. \end{aligned} \tag{68}$$

Taking into account (68) the expression (67) becomes

$$
\begin{aligned}
dl(\phi,\theta,\sigma_\eta^2) &= [\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \phi} - \frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \theta} + \frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \sigma_\eta^2}\frac{\partial \sigma_\eta^2}{\partial \beta}]d\beta \\
&\quad +[\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \phi} + \frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \sigma_\eta^2}\frac{\partial \sigma_\eta^2}{\partial \alpha}]d\alpha \\
&\quad +\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \sigma_\eta^2}\frac{\partial \sigma_\eta^2}{\partial \psi}d\psi.
\end{aligned} \tag{69}
$$

The first order conditions corresponding to the maximum of the Gaussian log-likelihood function $l(\cdot)$ are then

$$\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \sigma_\eta^2}\frac{\partial \sigma_\eta^2}{\partial \psi} = 0, \tag{70}$$

$$\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \phi} + \frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \sigma_\eta^2}\frac{\partial \sigma_\eta^2}{\partial \alpha} = 0, \tag{71}$$

$$\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \phi} - \frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \theta} + \frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \sigma_\eta^2}\frac{\partial \sigma_\eta^2}{\partial \beta} = 0. \tag{72}$$

Since $\frac{\partial \sigma_\eta^2}{\partial \psi} \neq 0, \frac{\partial \sigma_\eta^2}{\partial \alpha} \neq 0$ and $\frac{\partial \sigma_\eta^2}{\partial \beta} \neq 0$ by substitution we have

$$\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \sigma_\eta^2} = 0, \tag{73}$$

$$\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \phi} = 0, \tag{74}$$

$$\frac{\partial l(\phi,\theta,\sigma_\eta^2)}{\partial \theta} = 0. \tag{75}$$

We can at present use the expressions for $\widehat{x}_t$ and r_{t-1} given by Proposition 6 and differentiate the log-likelihood $l(\cdot)$ with respect to the system parameters $(\phi,\theta,\sigma_\eta^2)$. Noting that $\widehat{x}_t$ and r_{t-1} are not independent of σ_η^2 we derive the following optimality conditions

$$\sigma_\eta^2 T\left[1+\delta_T(\phi,\theta,\sigma_\eta^2)\right] = \sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}} \tag{76}$$

$$
\begin{aligned}
\frac{\partial}{\partial \phi}(\sum_{t=1}^{T}\ln r_{t-1} + T\ln\sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}}) = \\
-T\delta_T(\phi,\theta,\sigma_\eta^2)\frac{\partial}{\partial \phi}\ln\sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}}
\end{aligned} \tag{77}
$$

and

$$\frac{\partial}{\partial\theta}(\sum_{t=1}^{T}\ln r_{t-1}+T\ln\sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}}) = \qquad (78)$$

$$-T\delta_T(\phi,\theta,\sigma_\eta^2)\frac{\partial}{\partial\theta}\ln\sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}}$$

where

$$\delta_T(\phi,\theta,\sigma_\eta^2)\equiv T^{-1}[\sigma_\eta^2\frac{\partial}{\partial\sigma_\eta^2}\sum_{t=1}^{T}\ln r_{t-1}+\frac{\partial}{\partial\sigma_\eta^2}\sum_{t=1}^{T}\frac{(x_t-\widehat{x}_t)^2}{r_{t-1}}]. \qquad (79)$$

which completes the proof of Proposition 7. ■

10 Appendix

In this Appendix we describe McCulloch's [20] quantile estimator for the parameters of an α-stable distribution $\mathcal{S}_\alpha(\sigma,\beta,\mu)$. Let $Z\sim\mathcal{S}_\alpha(\sigma,\beta,\mu)$. We denote by Z_p the p-th quantile of the distribution of Z, i.e. $\Pr\{z\leqslant Z_p\}=p$. In order to estimate the index parameter α, $0.6\leqslant\alpha\leqslant 2$, and the skewness parameter β, $-1\leqslant\beta\leqslant 1$, McCulloch considers the following quantities

$$\Phi_1(\alpha,\beta) = \frac{Z_{.95}-Z_{.05}}{Z_{.75}-Z_{.25}}, \qquad (80)$$

$$\Phi_2(\alpha,\beta) = \frac{Z_{.95}+Z_{.05}-2Z_{.50}}{Z_{.95}-Z_{.05}}. \qquad (81)$$

These functions can be inverted so to obtain

$$\alpha = \Psi_1(\Phi_1,\Phi_2),$$

$$\beta = \Psi_2(\Phi_1,\Phi_2),$$

where the functions $\Psi_1$and Ψ_2 have been tabulated by McCulloch [20] for different values of Φ_1 and Φ_2. In order to form an estimator of α and β, we sample from the empirical distribution of Z, and form sample quantiles Z_p^*. We then use (80) and (80) to compute Φ_1^* and Φ_2^*. Since Z_p^* is consistent for Z_p, the obtained quantities Φ_1^* and Φ_2^* are consistent for

Φ_1 and Φ_2. Let

$$\begin{aligned} \alpha^* &= \Psi_1(\Phi_1^*, \Phi_2^*), \\ \beta^* &= \Psi_2(\Phi_1^*, \Phi_2^*). \end{aligned}$$

By using the tables, we can compute α^* and β^* which are consistent estimators of α and β. McCulloch uses a similar procedure to estimate the values of the scale parameter σ and the location parameter μ.

11 Figures

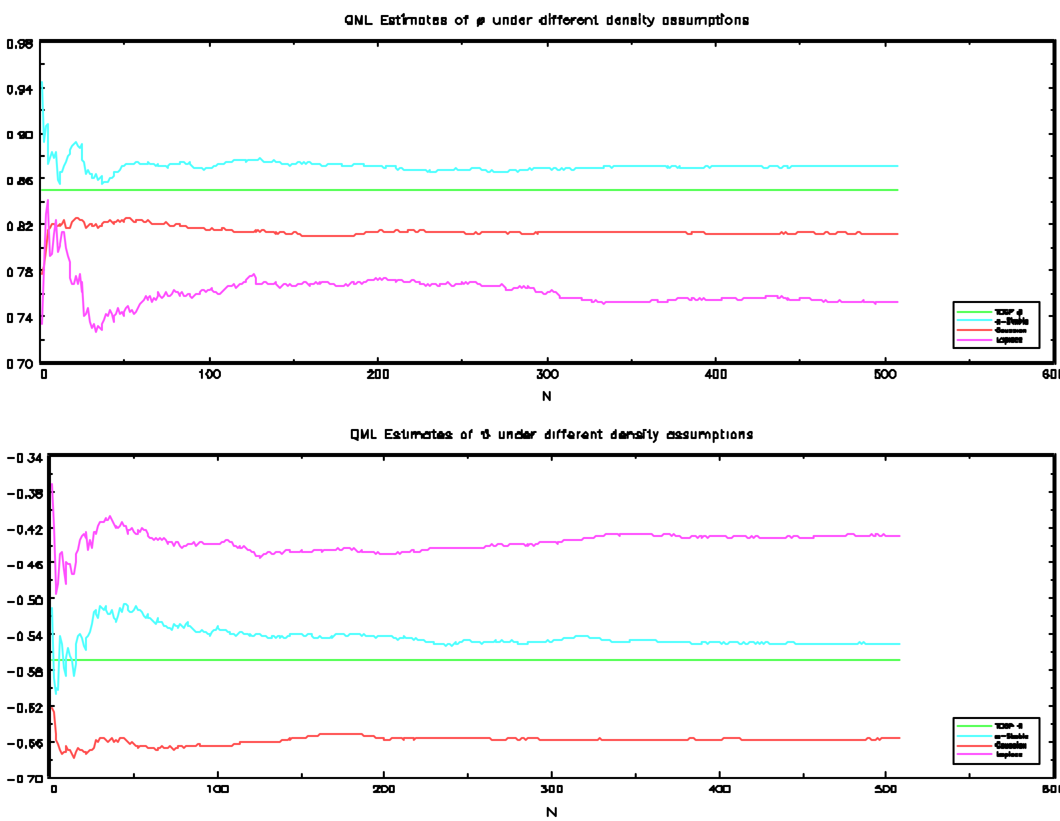

Figure 5: Convergence of the sample mean of the QMLEs of the "weak" *GARCH* with TDGP parameters $(\beta_0, \alpha_0) = (0.569, 0.281)$.

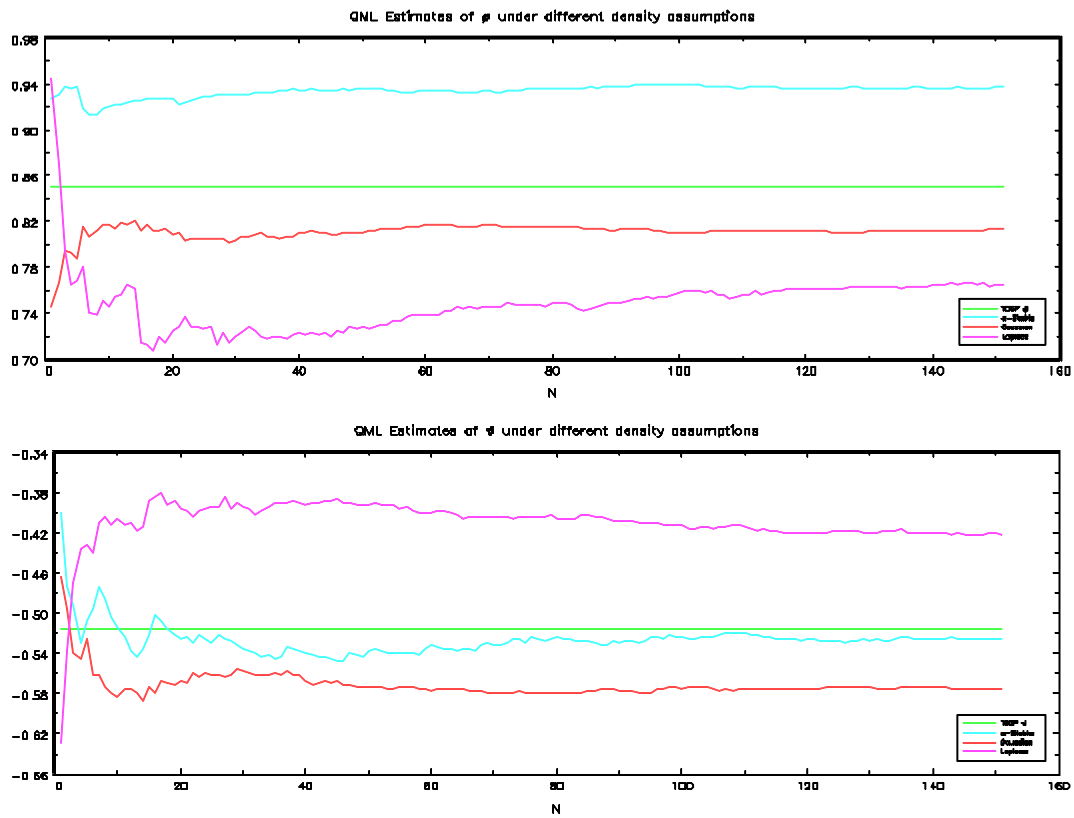

Figure 6: Convergence of the sample mean of the QMLEs of the "weak" *GARCH* with TDGP parameters $(\beta_0, \alpha_0) = (0.516, 0.334)$.

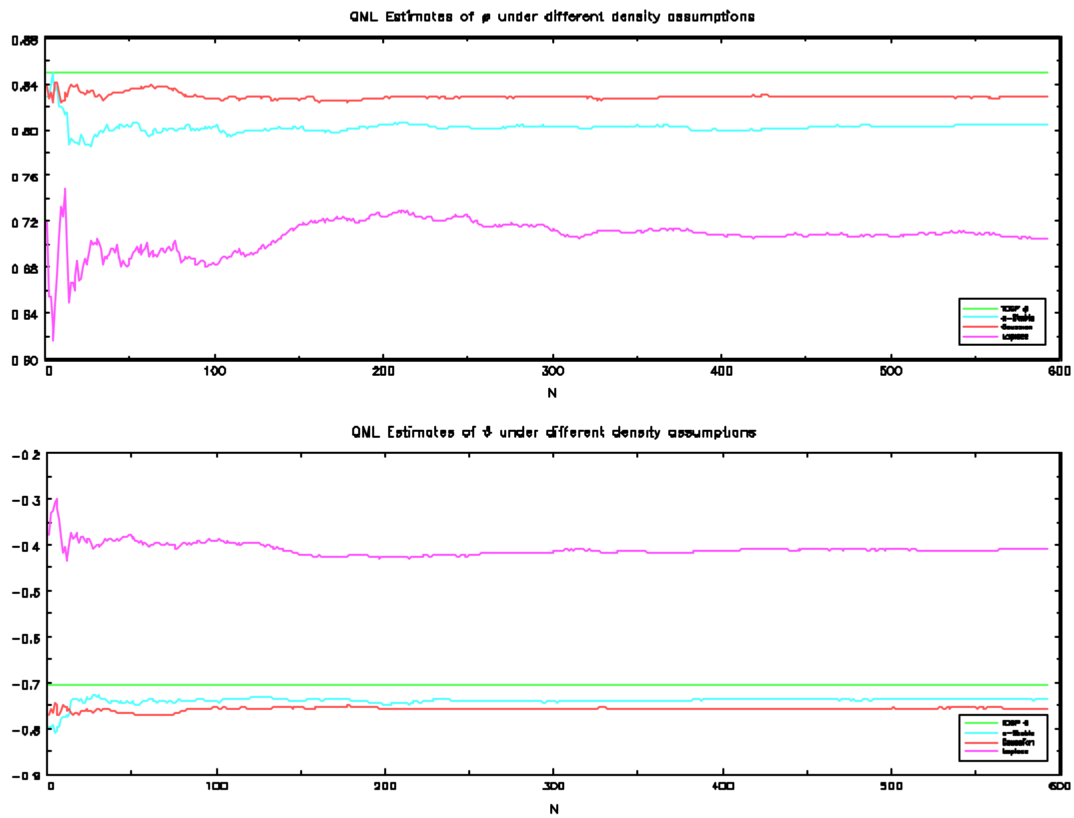

Figure 7: Convergence of the sample mean of the QMLEs of the "weak" $GARCH$ with TDGP parameters $(\beta_0, \alpha_0) = (0.705, 0.145)$.