



HAL
open science

Bootstrap refinements for QML estimators of the GARCH(1,1) parameters

Valentina Corradi, Emma M. Iglesias

► **To cite this version:**

Valentina Corradi, Emma M. Iglesias. Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *Econometrics*, 2008, 144 (2), pp.500. 10.1016/j.jeconom.2008.03.003 . hal-00495781

HAL Id: hal-00495781

<https://hal.science/hal-00495781>

Submitted on 29 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Bootstrap refinements for QML estimators of the GARCH(1,1) parameters

Valentina Corradi, Emma M. Iglesias

PII: S0304-4076(08)00043-2
DOI: 10.1016/j.jeconom.2008.03.003
Reference: ECONOM 3023

To appear in: *Journal of Econometrics*

Received date: 18 April 2007
Revised date: 26 February 2008
Accepted date: 3 March 2008

Please cite this article as: Corradi, V., Iglesias, E.M., Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *Journal of Econometrics* (2008), doi:10.1016/j.jeconom.2008.03.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Bootstrap Refinements for QML Estimators of the GARCH(1,1) Parameters

VALENTINA CORRADI ^{*‡}

University of Warwick

EMMA M. IGLESIAS ^{†‡}

Michigan State University

This draft: February 2008

Abstract

This paper reconsiders a block bootstrap procedure for Quasi Maximum Likelihood estimation of GARCH models, based on the resampling of the likelihood function, as proposed by Gonçalves and White (2004). First, we provide necessary conditions and sufficient conditions, in terms of moments of the innovation process, for the existence of the Edgeworth expansion of the GARCH(1,1) estimator, up to the k -th term. Second, we provide sufficient conditions for higher order refinements for equally tailed and symmetric test statistics. In particular, the bootstrap estimator based on resampling the likelihood has the same higher order improvements in terms of error in the rejection probabilities as those in Andrews (2002).

Keywords: block bootstrap, Edgeworth expansion, higher order refinements, Quasi Maximum Likelihood, GARCH.

^{*}Valentina Corradi, Department of Economics, University of Warwick, Coventry CV4 7AL, UK. e-mail: vcorradi@warwick.ac.uk.

[†]Emma Iglesias, Department of Economics, Michigan State University, 101 Marshall-Adams Hall, East Lansing, MI 48824-1038, USA. e-mail: iglesia5@msu.edu.

[‡]We wish to thank a co-editor, Ron Gallant, an Associate Editor, two anonymous referees, Silvia Gonçalves, Lutz Kilian, Oliver Linton, Nour Meddahi, Paolo Zaffaroni, Tim Vogelsang and the participants at the Financial Econometrics London-Oxford group at Imperial College for very useful comments and suggestions. Both authors thank the financial support from an ESRC grant (Award number: T026 27 1238) and Intramural Research Grants Program (IRGP) at Michigan State University. Corradi also gratefully acknowledge financial support from ESRC grant code RES-062-23-0311.

1 Introduction

It is well known that the quasi likelihood function of a GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, introduced by Bollerslev (1986), depends on the entire past history of the observables. In this case, resampling blocks of observations is not equivalent to resampling blocks of the likelihood function. This point has been lucidly pointed out by Gonçalves and White (2004), who indeed suggest to construct bootstrap estimators of GARCH models based on resampling blocks of the likelihood function. We go a step further, and we investigate the higher order properties of such estimators. This is accomplished in two steps. First, we establish the necessary conditions for the existence of an Edgeworth expansion up to the k -th term. This collapses to the existence of a minimum number of moments of the innovation process. Second, we provide sufficient conditions for higher order improvements of equally tailed and symmetric t-tests based on Quasi Maximum Likelihood Estimators (QMLE) of GARCH(1,1) parameters. This is done by providing the sufficient number of moments of the innovation process needed for the moments of the actual and bootstrap statistics to approach each other at an appropriate rate. Broadly speaking, this allows to control the rate at which the difference between the Edgeworth expansion of the actual and bootstrap statistic approaches zero. Linton (1997) calculated the Edgeworth-B distribution function for the GARCH(1,1), and we extend his setting to the Edgeworth expansion in the context of the existence of refinements of the bootstrap. If we were willing to assume that all the moments of the innovation process exist, then all the assumptions in Andrews (2002) would be satisfied and higher order refinements would then follow from his Theorem 2. On the other hand, for the GARCH case we can prove refinements for QML estimators under somewhat weaker assumptions on the existence of moments of the score and Hessian derivatives, than those required in Andrews (2002).

In nutshell, we need (i) conditions on the parameter space of the process in order to ensure that the GARCH process is exponentially β -mixing (see e.g. Carrasco and Chen (2002)), (ii) smoothness of the density of the innovation process (iii) the existence of a given number of moments of the innovation process. First, we show that under (i)-(iii) the Götze and Hipp (1994) conditions for the existence of the Edgeworth expansion for weakly dependent observations are satisfied. Then, we show that (i)-(iii) suffice to obtain the same higher order refinements of Andrews (2002). Thus, the block bootstrap, based on resampling the likelihood, leads to an error in rejection probability and confidence interval coverage probability of smaller order than $T^{-1/2-\xi}$ for equal-tailed t-tests for GARCH parameters and of order smaller than $T^{-1-\xi}$ for symmetric t-tests, for $\xi > 0$, and such that $\xi < \gamma$ and $\xi + \gamma < 1/2$, where γ is the parameter controlling the lengths of the block l , i.e. $l \simeq T^\gamma$.

Needless to say, an advantage of using bootstrap estimators is that of obtaining more accurate inference on the GARCH(1,1) parameters. Gonçalves and Kilian (2004, 2007) focused on bootstrap

inference for conditional mean that is robust to GARCH or other unspecified forms of conditional heteroskedasticity. However, there are many cases in practice where we are interested on the GARCH parameters themselves. From a more empirical perspective, one of the most popular application of bootstrap estimation of GARCH parameters is in the context of risk management, and more precisely in the evaluation of Value at Risk and Expected Shortfall. For example, Christoffersen and Gonçalves (2005) rely on bootstrapping GARCH parameters in order to take into proper consideration the contribution of parameter estimation error when evaluating Value at Risk; Mancini and Trojani (2005) address the same issue using an estimation and bootstrap procedure robust to the presence of outliers. Both papers rely on the residual based bootstrap approach outlined by Pascual, Romo and Ruiz (2006).

If one knew the data generating process (DGP), then a residual based bootstrap approach, which makes direct use of the structure of the model, seems to be more natural than a nonparametric bootstrap approach, such as the block bootstrap. Hidalgo and Zaffaroni (2007) show the first order validity of the residual-based bootstrap for $ARCH(\infty)$ process, which indeed include finite $GARCH(1, 1)$. Hence, it is interesting to investigate whether the residual-based bootstrap provides higher order improvements over asymptotic normality, and whether these improvements are sharper than those provided by the block bootstrap. This is a quite challenging task, and we leave it for future research. Nevertheless, we outline how the arguments used by Hidalgo and Zaffaroni to show first order validity cannot be directly extended to show the higher order properties of the residual-based bootstrap.

In the case of possible nonlinear Markov processes, recent papers by Andrews (2005) and Horowitz (2003) have established that higher-order refinements, very close to those attainable in the *iid* case, can be achieved via the use of the Markov bootstrap, even in the case the underlying transition density is unknown, and replaced by a nonparametric estimator. Though a GARCH process is not Markov, under certain conditions is approximate Markov, as in the definition of Horowitz. Thus, we have analyzed the applicability of the Markov process for the estimation of the conditional variance parameters. If the marginal density of the innovation process is known, then the results of Andrews (2005) apply and the Markov bootstrap provides refinements arbitrarily close to those attainable in the *iid* case. On the other hand, whenever the density of the innovation is unknown and has unbounded support, as in the gaussian case for example, then some of the assumptions for the existence of higher order refinements are violated. This point has been already outlined in Horowitz (2003).

The rest of this paper is organized as follows. Section 2 describes the implementation of the bootstrap procedure based on the resampling of the likelihood function. Section 3 establishes the higher order improvement of the bootstrap approximation of the coverage error of confidence intervals for GARCH parameters, and summarizes the main theoretical results. Section 4 outlines the

difficulties involved in the establishment of possible higher order refinements for the residual-based bootstrap. Section 5 briefly outlines the use of the Markov bootstrap, whenever the marginal density of the innovation is unknown and has unbounded support. Section 6 reports Monte Carlo simulation results, which provide some evidence of the improved accuracy of bootstrap, for the error in the coverage probability. All the proofs are collected in the Appendix.

2 The Block Bootstrap: Set-Up

Suppose y_t is generated by the GARCH(1,1) process,

$$(1) \quad y_t = \sqrt{h_t} \eta_t, \quad h_t = \theta_1^\dagger + \theta_2^\dagger y_{t-1}^2 + \theta_3^\dagger h_{t-1}, \quad t = 1, \dots, T$$

where η_t is *iid* and $\theta^\dagger = (\theta_1^\dagger, \theta_2^\dagger, \theta_3^\dagger)$ is the true parameter vector.

We define the log-likelihood function as if $\frac{y_t}{\sqrt{h_t}}$ were normally distributed, thus using a Gaussian likelihood. The quasi maximum likelihood estimator, QMLE, is then defined as:

$$(2) \quad \hat{\theta}_T = \arg \max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T L_t(\theta)$$

where

$$\frac{1}{T} \sum_{t=1}^T L_t(\theta) = -\frac{1}{2} \sum_{t=1}^T \ln h_t(\theta) - \frac{1}{2} \sum_{t=1}^T \frac{y_t^2}{h_t(\theta)}.$$

Note that after a few manipulations, and imposing $h_0 = \eta_0 = 0$, we have that¹

$$(3) \quad \begin{aligned} h_t &= \theta_1 + (\theta_2 \eta_{t-1}^2 + \theta_3) h_{t-1} \\ &= \theta_1 \left[1 + \sum_{j=1}^{t-1} \prod_{i=1}^j (\theta_2 \eta_{t-i}^2 + \theta_3) \right]. \end{aligned}$$

From (3) it is immediate to see that resampling blocks of the observable series y_t is not equivalent to resampling blocks of the log-likelihood.

Thus, we need to resample b blocks of length l from the loglikelihood L_t , setting $bl = T$. Hereafter, let I_i , $i = 1, \dots, b$ denote identical and independent draws from a discrete uniform on $0, 1, \dots, T-l$. Thus, for each $i = 1, \dots, b$ $I_i = j$, with $j = 0, 1, \dots, T-l$, with equal probability $1/(T-l+1)$. Now, for all $\theta \in \Theta$, define $L_1^*(\theta), L_2^*(\theta), \dots, L_{T-l+1}^*(\theta), \dots, L_T^*(\theta)$ to be equal to $L_{I_1+1}(\theta), L_{I_1+2}(\theta), \dots, L_{I_b+1}(\theta), \dots, L_{I_b+l}(\theta)$. Note, that we use the same random draws I_i for any $\theta \in \Theta$. In general, resampling the data or resampling the likelihood is equivalent. However, in the GARCH case, the (quasi) likelihood function

¹Note that if either h_0 or η_0 are different from zero, then the expression in (3) holds up to a term converging to zero exponentially as $t \rightarrow \infty$.

depends on the entire past history of the observables. In this case, resampling blocks of observations is not equivalent to resampling blocks of the likelihood function. This point has been lucidly pointed out by Gonçalves and White (2004), who indeed suggest to construct bootstrap estimators of GARCH model based on resampling blocks of the likelihood function. We use the moving blocks bootstrap (MBB) of Künsch (1989) as in Gonçalves and White (2004).

It should be pointed out that, if we were just interested in first order validity, then we could have set $l = 1$ and relied on the *iid* nonparametric bootstrap. In fact, as the score is a martingale difference sequence, bootstrap samples based on *iid* resampling of the log-likelihood would have ensured that the first two bootstrap moments properly mimic the correspondent sample moments. However, as for refinements, we need to match higher moments, and in this case the fact that the score is a martingale difference sequence does not help.

We now use the resample log-likelihood in order to construct the bootstrap estimator $\widehat{\theta}_T^*$, that is:

$$\widehat{\theta}_T^* = \arg \max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \left(L_t^*(\theta) - \left(E^* \left(\nabla L_t^*(\widehat{\theta}_T) \right) \right)' \theta \right).$$

Note that the recentering term, $\left(E^* \left(\nabla L_t^*(\widehat{\theta}_T) \right) \right)' \theta$, ensure that the score, evaluated at $\widehat{\theta}_T$, has zero mean.

Define the Hessian and the variance of the score as B^\dagger and A^\dagger respectively, where

$$B^\dagger = E \left(-\nabla_{\theta}^2 L_t(\theta^\dagger) \right)$$

and

$$A^\dagger = Var \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{\theta} L_t(\theta^\dagger) \right) = E \left(\nabla_{\theta} L_t(\theta^\dagger) \nabla_{\theta} L_t(\theta^\dagger)' \right).$$

Also, define the sample analogs of B^\dagger and A^\dagger , that is

$$\begin{aligned} \widehat{B}_T &= -\frac{1}{T} \sum_{t=1}^T \nabla_{\theta}^2 L_t(\widehat{\theta}_T) \\ \widehat{A}_T &= \frac{1}{T} \sum_{t=1}^T \nabla_{\theta} L_t(\widehat{\theta}_T) \nabla_{\theta} L_t(\widehat{\theta}_T)'. \end{aligned} \tag{4}$$

Our objective is to provide higher order refinements of the approximation of the coverage error for the t-statistic

$$t_{\theta_i, T} = \sqrt{T} \left(\widehat{B}^{-1} \widehat{A} \widehat{B}^{-1} \right)_{ii}^{-1/2} \left(\widehat{\theta}_{i, T} - \theta_i^\dagger \right), \tag{5}$$

for $i = 1, 2, 3$.

Define the bootstrap analogs of \widehat{B}_T and \widehat{A}_T as \widehat{B}_T^* and \widehat{A}_T^* , where

$$\widehat{B}_T^* = \frac{1}{T} \sum_{t=1}^T \nabla \left(\nabla L_t^*(\widehat{\theta}_T^*) - E^* \left(\nabla L_t^*(\widehat{\theta}_T^*) \right) \right)$$

and

$$(6) \quad \widehat{A}_T^* = \frac{1}{T} \sum_{t=1}^T \left(\nabla L_t^*(\widehat{\theta}_T^*) - E^* \left(\nabla L_t^*(\widehat{\theta}_T^*) \right) \right) \left(\nabla L_t^*(\widehat{\theta}_T^*) - E^* \left(\nabla L_t^*(\widehat{\theta}_T^*) \right) \right)'$$

The bootstrap analog of $t_{\theta_i,T}$ is then given by:

$$(7) \quad \widetilde{t}_{\theta_i,T} = \sqrt{T} \left(\widehat{B}_T^{*-1} \widehat{A}_T^* \widehat{B}_T^{*-1} \right)_{ii}^{-1/2} \left(\widehat{\theta}_{i,T}^* - \widehat{\theta}_{i,T} \right).$$

As $\nabla L_t(\theta^\dagger)$ is a martingale difference sequence, \widehat{A}_T is a consistent estimator of $Var \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_\theta L_t(\theta^\dagger) \right)$. However, as within each block, $\nabla L_t^*(\widehat{\theta}_T)$ is correlated with its neighbours, \widehat{A}_T^* cannot be a consistent estimator for $Var^* \left(\frac{1}{T} \sum_{t=1}^T \nabla L_t^*(\widehat{\theta}_T) \right)$. Hence $\widetilde{t}_{\theta_i,T}$ cannot have a standard normal limiting distribution, under the bootstrap probability law. A consistent estimator of $Var^* \left(\frac{1}{T} \sum_{t=1}^T \nabla L_t^*(\widehat{\theta}_T) \right)$ is instead given by:

$$(8) \quad \begin{aligned} \widetilde{A}_T^* &= \frac{b}{T(T-l+1)} \sum_{i=0}^{T-l} \sum_{j=1}^l \sum_{m=1}^l \left(\left(\nabla L_{i+j}(\widehat{\theta}_T) - E^* \left(\nabla L_{i+j}(\widehat{\theta}_T) \right) \right) \right. \\ &\quad \left. \times \left(\nabla_\theta L_{i+m}(\widehat{\theta}_T) - E^* \left(\nabla L_{i+m}(\widehat{\theta}_T) \right) \right) \right)'. \end{aligned}$$

Now, let $\widehat{\sigma}_{T,ii} = \left(\widehat{B}_T^{-1} \widehat{A}_T \widehat{B}_T^{-1} \right)_{ii}^{1/2}$, $\widetilde{\sigma}_{T,ii}^* = \left(\widehat{B}_T^{*-1} \widetilde{A}_T^* \widehat{B}_T^{*-1} \right)_{ii}^{1/2}$ and following Hall and Horowitz (1996), define the adjustment term as

$$\tau_{i,T}^* = \widehat{\sigma}_{T,ii} / \widetilde{\sigma}_{T,ii}^*,$$

and the adjusted t-statistics as:

$$(9) \quad t_{\theta_i,T}^* = \sqrt{T} \left(\widehat{B}_T^{*-1} \widehat{A}_T^* \widehat{B}_T^{*-1} \right)_{ii}^{-1/2} \left(\widehat{\theta}_{i,T}^* - \widehat{\theta}_{i,T} \right) \tau_{i,T}^*.$$

3 The Block Bootstrap: Refinements

Define the vectors containing the centered score, the centered outer product and all their derivatives up to order $d_{1,k}$ as:

$$(10) \quad f_{i,t,d_{1,k}}(\theta) = \left(\overline{\nabla_{\theta_i} L_t(\theta)}, \dots, \nabla_{\theta_i}^{d_{1,k}} \left(\overline{\nabla_\theta L_t(\theta)} \right), \overline{\nabla_{\theta_i} L_t(\theta) \nabla_{\theta_i} L_t(\theta)}, \dots, \nabla_{\theta_i}^{d_{1,k}} \left(\overline{\nabla_\theta L_t(\theta) \nabla_\theta L_t(\theta)'} \right) \right),$$

where $\overline{\nabla_{\theta_i} L_t(\theta)} = \nabla_{\theta_i} L_t(\theta) - E(\nabla_{\theta_i} L_t(\theta))$,

$\overline{\nabla_{\theta_i} L_t(\theta) \nabla_{\theta_i} L_t(\theta)'} = \nabla_{\theta_i} L_t(\theta) \nabla_{\theta_i} L_t(\theta)' - E(\nabla_{\theta_i} L_t(\theta) \nabla_{\theta_i} L_t(\theta)')$, and $d_{1,k}$ depends on k , which is the number of terms in the Edgeworth expansion, i.e. $d_{1,k} \geq k + 3$.

In what follows, we rely on the following assumption:

Assumption A: (i) The parameter space Θ is a compact subset of \mathcal{R}^{3+} such that $\Theta = \{\theta_1, \theta_2, \theta_3; \theta_1, \theta_2, \theta_3 > 0, \theta_2 + \theta_3 < 1\}$. (ii) $\theta^\dagger = (\theta_1^\dagger, \theta_2^\dagger, \theta_3^\dagger)$, $\hat{\theta}_T = (\hat{\theta}_{1T}, \hat{\theta}_{2T}, \hat{\theta}_{3T})$ are in the interior of Θ (iii) η_t is *iid*($0, \sigma_\eta^2$) and it has a positive continuous density.

Assumption A ensures that y_t and $h_t(\theta^\dagger)$ are strictly and covariance stationary β -mixing processes, with exponentially decaying mixing coefficients (see e.g. Carrasco and Chen (2002)). Let N^\dagger be a neighbourhood of θ^\dagger , then for any $\theta \in N^\dagger$, $f_{i,t,d_{1,k}}(\theta)$ is also β -mixing process with exponential decay.

In Lemma A1 in the Appendix, we show that Assumption A ensures that the Götze and Hipp (1994) Cramer type conditions hold. Hence, in order to prove the existence of the k -th term of the Edgeworth expansion for the statistic in (5) we need to show the existence of enough moments for the components of $f_{i,t,d_{1,k}}(\theta)$, for $i = 1, 2, 3$. This is accomplished in Theorem 1, where we provide necessary and sufficient conditions for the existence of the k -th term of the Edgeworth expansion for the t-statistic in (5).

Hereafter, let $\left(1 + \sum_{i=1}^k T^{-i/2} \pi_i(\partial/\partial z)\right) \Phi(z)$ denote the Edgeworth expansion of t_{θ_i} up to order k (see e.g. p.142 in Andrews 2001). We have:

Theorem 1: Let Assumptions A hold. Then

$$(11) \quad \lim_{T \rightarrow \infty} T^{k/2} \sup_{z \in \mathcal{R}} \left| P(t_{\theta_i} \leq z) - \left(1 + \sum_{i=1}^k T^{-i/2} \pi_i(\partial/\partial z)\right) \Phi(z) \right| = 0$$

(i) only if $E\left(\eta_t^{2(k+2)(k+4)}\right) < \infty$

(ii) and if $E\left(\eta_t^{2(k+3)(k+5)}\right) < \infty$.

Theorem 1 provides a set of necessary and a set of sufficient conditions for the existence of the k -th term of the Edgeworth expansion of the statistics. Needless to say, for $k = 2, 3$ (i) is also a necessary condition for obtaining higher order refinements for equally-tailed and symmetric t-tests respectively.

Hereafter, $z_{t_{\theta_i,T}, \alpha/2}^*$ and $z_{|t_{\theta_i,T}|, \alpha}^*$ denote the $(1 - \alpha)/2$ bootstrap quantile for $t_{\theta_i,T}^*$ and the $(1 - \alpha)$ quantile for $|t_{\theta_i,T}^*|$ respectively. As usual, $z_{t_{\theta_i,T}, \alpha/2}^*$ and $z_{|t_{\theta_i,T}|, \alpha}^*$ are constructed from the empirical distribution of the bootstrap statistic $t_{\theta_i,T}^*$.

We have:

Theorem 2: Let Assumption A hold. Let $\xi + \gamma < 1/2$ and $\xi < \gamma$, where γ is the parameter controlling the block size, i.e. $l \simeq T^\gamma$. If $E\left(\eta_t^{2(d_{1,k} + d_{2,k} + 2)(k+2)p}\right) < \infty$, where $d_{1,k} \geq k + 3$, $d_{2,k} \geq k + 1$, and

$p > \frac{k/2}{2(1/2-\xi-\gamma)}$, then:

(i) for $k = 2$,

$$P\left(t_{\theta_i,T} < -z_{t_{\theta_i,T},\alpha/2}^* \text{ or } t_{\theta_i,T} > z_{t_{\theta_i,T},1-\alpha/2}^*\right) = \alpha + o\left(T^{-(1/2+\xi)}\right)$$

(ii) for $k = 3$,

$$P\left(|t_{\theta_i,T}| < z_{|t_{\theta_i,T}|,\alpha}^*\right) = \alpha + o\left(T^{-(1+\xi)}\right).$$

From Theorem 2, it is immediate to say that we obtain the same higher order refinements for the error in rejection probability (ERP), as in Andrews (2002). As an immediate corollary, we also have that the bootstrap error in the coverage probability (ECP) is of order smaller than $T^{-(1/2+\xi)}$ for the equal-tailed test and of $T^{-(1+\xi)}$ for the symmetric test, with $\xi < 1/4$. It is interesting to note that, in the exponentially mixing GARCH(1,1) case, the only conditions we need for higher order refinements are in terms of smoothness of the density and of existence of a sufficient number of moments of the innovation process.

4 The Residual-Based Bootstrap

The residual-based bootstrap is based on *iid* resampling of the centered residuals. Let $\hat{\eta}_t = y_t/h_t^{1/2}(\hat{\theta}_T)$, we resample, with replacement, from the empirical distribution of $\hat{\eta}_t$, and obtain the sequence η_t^* , where η_1^* is equal to $\hat{\eta}_t - T^{-1} \sum_{t=1}^T \hat{\eta}_t$, $t = 1, \dots, T$ with equal probability $1/T$. $\eta_1^*, \eta_2^*, \dots, \eta_T^*$ are defined analogously, so that η_t^* is *iid* conditional on sample. We then proceed as follows:

$$\begin{aligned} \text{Fix } y_0, h_0, & \implies \hat{h}_1^* = \hat{\theta}_{1,T} + \hat{\theta}_{2,T}y_0^2 + \hat{\theta}_{3,T}h_0 \\ y_1^* &= \hat{h}_1^{*1/2}\eta_1^* \\ \hat{h}_2^* &= \hat{\theta}_{1,T} + \hat{\theta}_{2,T}y_1^{*2} + \hat{\theta}_{3,T}\hat{h}_1^* \\ y_2^* &= \hat{h}_2^{*1/2}\eta_2^* \end{aligned}$$

and so on, until sequences y_t^*, h_t^* , $t = 1, \dots, T$ are formed.

Now, define the bootstrap QML estimator as:

$$\begin{aligned} \tilde{\theta}_T^* &= \arg \max_{\theta \in \Theta} \sum_{t=1}^T L_t^{*(RS)}(\theta) \\ &= \arg \max_{\theta \in \Theta} \left(-\frac{1}{2} \sum_{t=1}^T \ln h_t^*(\theta) - \frac{1}{2} \sum_{t=1}^T \frac{y_t^{*2}}{\hat{h}_t^*(\theta)} \right), \end{aligned}$$

where the superscript (*RS*) denotes the fact that the bootstrap likelihood is based on the residual bootstrap approach.

This residual based bootstrap procedure has been recently suggested by Pascual, Romo and Ruiz (2006). See also Christoffersen and Gonçalves (2005), for an application to Value at Risk evaluation with GARCH models.

In the case of correctly specified ARMA models, the asymptotic first order validity, as well as refinements, of the residual-based bootstrap have been already established (see e.g. Inoue and Kilian (2002), Kreiss and Franke (1992), and Bose (1988)).

Robinson and Zaffaroni (2006) establish conditions for strong consistency and asymptotic normality of QML estimators of ARCH(∞) models. Now ARCH(∞) models encompass GARCH(1,1) models. In particular, the ARCH(∞) model allows for a higher degree of memory, including also long-memory behavior. Hidalgo and Zaffaroni (2007; hereafter HZ) take a step further and establish the first order validity of their residual-based bootstrap analogs.

Below, we outline the key steps in the HZ proof for the residual-based bootstrap first order validity, and sketch the further steps one should follow in order to establish higher order refinements. In fact, the interesting question is whether the residual based-bootstrap may provide sharper refinements than the block-bootstrap does. This is a rather involved issue, and we leave it for further research.

The key point is that we can express h_t , the score, and the hessian, as well as their bootstrap analogs, as a function on entire past of squared innovations. For notational simplicity we just focus on h_t . Recalling equation (3), it's bootstrap analog is given by:

$$\begin{aligned}
 h_t(\theta) &= \theta_1 + (\theta_2 \eta_{t-1}^2 + \theta_3) h_{t-1} \\
 &= \theta_1 \left[1 + \sum_{j=1}^{t-1} \prod_{i=1}^j (\theta_2 \eta_{t-i}^2 + \theta_3) \right] \\
 h_t^*(\theta) &= \theta_1 + (\theta_2 \eta_{t-1}^{*2} + \theta_3) h_{t-1}^* \\
 (12) \quad &= \theta_1 \left[1 + \sum_{j=1}^{t-1} \prod_{i=1}^j (\theta_2 \eta_{t-i}^{*2} + \theta_3) \right]
 \end{aligned}$$

Thus one can write $h_t^*(\theta^\dagger)$, as $h_t^*(\theta^\dagger) = h(\eta_{t-1}^{*2}, \eta_{t-2}^{*2}, \dots, \eta_1^{*2}; \theta^\dagger)$, Now,

$$\begin{aligned}
 &E^* \left(\frac{1}{T} \sum_{t=1}^T h_t^*(\theta^\dagger) \right) \\
 &\simeq \frac{1}{T^T} \sum_{i_1=1}^T \sum_{i_2=1}^T \dots \sum_{i_T=1}^T h(\widehat{\eta}_{i_1}^2, \widehat{\eta}_{i_2}^2, \dots, \widehat{\eta}_{i_T}^2),
 \end{aligned}$$

as any permutation and combination can occur with equal probability $1/T^2$

²Note the abuse of notation. As we are resampling from centered residuals, more correctly we should have $(\widehat{\eta}_{i_1} - T^{-1} \sum_t \widehat{\eta}_t)^2$, though for notational simplicity we simply write $\widehat{\eta}_{i_1}^2$.

Let $h_{t,M}^*(\theta^\dagger)$ be a truncated version of $h_t^*(\theta^\dagger)$, i.e. $h_{t,M}^*(\theta^\dagger) = h(\eta_{t-1}^{*2}, \eta_{t-2}^{*2}, \dots, \eta_{t-M}^{*2}, 0, \dots, 0)$, HZ proof (see Lemmas 6.13-6.15) for first order asymptotic validity is based on three arguments. The first argument is that the process can be truncated, and the truncation error approaches zero fast enough, i.e.

$$E^* \left(\frac{1}{T} \sum_{t=1}^T (h_t^*(\theta^\dagger) - h_{t,M}^*(\theta^\dagger)) \right) = O_p(M^{-d}),$$

where $d > 1$ depends on the degree of memory of the process. Note that,

$$E^* \left(\frac{1}{T} \sum_{t=1}^T h_{t,M}^*(\theta^\dagger) \right) = \frac{1}{T^M} \sum_{i_1=1}^T \sum_{i_2=1}^T \dots \sum_{i_M=1}^T h(\widehat{\eta}_{i_1}^2, \widehat{\eta}_{i_2}^2, \dots, \widehat{\eta}_{i_M}^2) = V_{T,M},$$

where $V_{T,M}$ defines a V -statistic. Define, the U -statistic $U_{T,M}$ as

$$U_{T,M} = \binom{T}{M}^{-1} \sum_{1 \leq i_1 \leq \dots \leq i_M \leq T} h(\widehat{\eta}_{i_1}^2, \widehat{\eta}_{i_2}^2, \dots, \widehat{\eta}_{i_M}^2).$$

The second argument is that V -statistics are very accurately approximated by U -statistics, in fact by e.g. Theorem 5.1 in Grams and Serfling (1973),

$$E(|V_{T,M} - U_{T,M}|^r) = O(T^{-r}), \text{ for } r > 0.$$

Finally, the third argument is that, by the law of large numbers for U -statistics,

$$E \left(U_{T,M} - \frac{1}{T} \sum_{t=1}^T h_{t,M}(\theta^\dagger) \right) = O\left(\frac{M}{T}\right).$$

The same applies if we replace θ^\dagger with $\widehat{\theta}_T$. The point is that, as for first order validity, it suffices to match the Hessian and the variance of the score. As the score is a martingale difference sequence, it all collapses to match sample and bootstrap sample first moments. In fact, as we outline below, things get more complicated once we move to higher moments.

The question is whether we can also obtain refinements for the residual-based bootstrap and whether they are sharper than those obtained via the block-bootstrap.

A crucial condition for bootstrap refinements is that the "distance" between the Edgeworth expansion of the statistic and that of its bootstrap counterpart approaches zero at fast enough rate, which in turn defines the "sharpness" of refinements. This collapses to check the speed at which the distance between higher moments of the statistic and its bootstrap counterpart approaches zero. Hereafter, let $f_{i,t,d_1,k}^* \left(\widehat{\theta}_T \right)$ be the bootstrap analog of $f_{i,t,d_1,k}(\theta^\dagger)$ evaluated at $\widehat{\theta}_T$, defined as in Section 3. Also, with an abuse of notation, let $f_t^* \left(\widehat{\theta}_T \right)$ be a generic element of $f_{i,t,d_1,k}^* \left(\widehat{\theta}_T \right)$.

In a nutshell, to have an improvement in the error in coverage probability (ECP) of $T^{-\zeta}$, we need that (e.g. Andrews 2002):

$$(13) \quad T^a P \left(\left| T^{\alpha(m)} E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_t^* (\hat{\theta}_T) \right)^m - T^{\alpha(m)} E \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_t (\theta^\dagger) \right)^m \right| > T^{-\zeta} \right) \rightarrow 0,$$

where $\alpha(m) = 0$ if m is even and $\alpha(m) = 1/2$ if m is odd, $a = 3/2$ and $m = 4$ for equally tailed test, and $a = 2$, $m = 5$ for symmetric tests. Needless to say, the residual-based block bootstrap will provide sharper refinements if $\zeta > \xi$, where ξ is defined in Theorem 2 above. Now, as convergence in r -mean implies converges in probability, the statement in (13) is implied by

$$(14) \quad T^a E \left(T^\zeta \left| T^{\alpha(m)} E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_t^* (\hat{\theta}_T) \right)^m - T^{\alpha(m)} E \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_t (\theta^\dagger) \right)^m \right|^r \right) = o(1), \text{ for some } r > 0.$$

Now, $f_t^* (\hat{\theta}_T)$ writes as

$$f_t^* (\hat{\theta}_T) = f \left(\eta_{t-1}^{*2}, \eta_{t-2}^{*2}, \dots, \eta_1^{*2}; \hat{\theta}_T \right),$$

and define its truncated version as

$$f_{t,M}^* (\hat{\theta}_T) = f \left(\eta_{t-1}^{*2}, \eta_{t-2}^{*2}, \dots, \eta_{t-M}^{*2}; \hat{\theta}_T \right).$$

The left hand side of (14) is majorized by $I_T^* + II_T^* + III_T^*$, where

$$\begin{aligned} I_T^* &= T^a E \left(T^\zeta \left| T^{\alpha(m)} E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_t^* (\hat{\theta}_T) \right)^m - T^{\alpha(m)} E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M}^* (\hat{\theta}_T) \right)^m \right|^r \right) \\ II_T^* &= T^a E \left(T^\zeta \left| T^{\alpha(m)} E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M}^* (\hat{\theta}_T) \right)^m - T^{\alpha(m)} E \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M} (\theta^\dagger) \right)^m \right|^r \right) \\ III_T^* &= T^a \left(T^\zeta \left| T^{\alpha(m)} E \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M} (\theta^\dagger) \right)^m - T^{\alpha(m)} E \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_t (\theta^\dagger) \right)^m \right|^r \right). \end{aligned}$$

It is immediate to see that the case of $m = 1$ can be treated along the same lines of HZ. Now, given (3) and (10), for $0 < \rho < 1$,

$$I_T^* = T^a O \left(T^{\alpha(m)(r \times m)} T^\zeta \rho^{-M} \right),$$

so that it is enough to set $M = T^\varepsilon$, for $\varepsilon > 0$ arbitrarily small to ensure that $I_T^* = o(1)$. The same order of magnitude applies to III_T^* . The difficult part is II_T^* .

We need to compute $E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M}^* (\hat{\theta}_T) \right)^m$. For sake of simplicity, we begin by considering the case of $m = 2$. By noting that

$$\begin{aligned} & E^* \left(f \left(\eta_{t-1}^{*2}, \eta_{t-2}^{*2}, \dots, \eta_{t-M}^{*2}; \hat{\theta}_T \right) f \left(\eta_{t-\tau-1}^{*2}, \eta_{t-\tau-2}^{*2}, \dots, \eta_{t-\tau-M}^{*2}; \hat{\theta}_T \right) \right) \\ &= 0 \text{ for } |\tau| > M, \end{aligned}$$

$$\begin{aligned}
& E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M}^* (\hat{\theta}_T) \right)^2 \\
& \simeq \frac{1}{T} \sum_{\tau=-M}^M \sum_{t=2M}^{T-M} E^* \left(f \left(\eta_{t-1}^{*2}, \eta_{t-2}^{*2}, \dots, \eta_{t-M}^{*2}; \hat{\theta}_T \right) f \left(\eta_{t-\tau-1}^{*2}, \eta_{t-\tau-2}^{*2}, \dots, \eta_{t-\tau-M}^{*2}; \hat{\theta}_T \right) \right) \\
& = \frac{1}{T} \sum_{\tau=-M}^M \sum_{t=2M}^{T-M} \frac{1}{T^\tau} \frac{1}{T^M} \sum_{i_1}^{T-M} \dots \sum_{i_\tau}^{T-M} \dots \sum_{i_M}^{T-M} \dots \sum_{i_{\tau+M}}^{T-M} f \left(\hat{\eta}_{i_1}^2, \dots, \hat{\eta}_{i_\tau}^2, \dots, \hat{\eta}_{i_M}^2; \hat{\theta}_T \right) \\
& \quad \times f \left(\hat{\eta}_{i_\tau}^2, \dots, \hat{\eta}_{i_M}^2, \dots, \hat{\eta}_{i_{M+\tau}}^2; \hat{\theta}_T \right).
\end{aligned}$$

Hereafter, let $\tau^{(1)} \leq \min_{i \leq M} |\tau_i|$, $\tau^{(m)} \leq \max_{i \leq M} |\tau_i|$, $\tau^{(1)} \leq \tau^{(2)} \leq \dots \leq \tau^{(m)}$. Thus, for a generic m ,

$$\begin{aligned}
& E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M}^* (\hat{\theta}_T) \right)^m \\
& \simeq \frac{1}{T^{m/2-1}} \sum_{\tau_1=-M}^M \dots \sum_{\tau_m=-M}^M E^* \left(f \left(\eta_{2M-\tau_1}^{*2}, \dots, \eta_{M-\tau_1}^{*2}; \hat{\theta}_T \right) \right. \\
& \quad \times \dots \times f \left(\eta_{2M-\tau_m}^{*2}, \dots, \eta_{M-\tau_m}^{*2}; \hat{\theta}_T \right) \left. \mathbb{1} \left\{ \max_{i \leq m} \tau_i - \min_{i \leq m} \tau_i \leq M \right\} \right) \\
& = \frac{1}{T^{m/2-1}} \sum_{\tau_1=-M}^0 \dots \sum_{\tau_m=-M}^0 \frac{1}{T^{M+\tau^{(m)}}} \sum_{i_1=2M}^{T-M} \dots \sum_{i_{\tau^{(1)}}=2M}^{T-M} \dots \sum_{i_{\tau^{(m)}}=2M}^{T-M} \dots \\
& \quad \dots \sum_{i_M=2M}^{T-M} \dots \sum_{i_{M+\tau^{(1)}}=2M}^{T-M} \dots \sum_{i_{M+\tau^{(m)}}=2M}^{T-M} \left(f \left(\hat{\eta}_{i_1}^2, \dots, \hat{\eta}_{i_{\tau^{(1)}}}^2, \dots, \hat{\eta}_{i_{\tau^{(m)}}}^2, \dots, \hat{\eta}_{i_M}^2; \hat{\theta}_T \right) \right. \\
& \quad \times \dots \times f \left(\hat{\eta}_{i_{\tau^{(m)}}}^2, \dots, \hat{\eta}_{i_M}^2, \dots, \hat{\eta}_{i_{\tau^{(1)}+M}}^2, \dots, \hat{\eta}_{i_{\tau^{(m)}+M}}^2; \hat{\theta}_T \right) \left. \right) \\
& \quad + \frac{1}{T^{m/2-1}} \sum_{\tau_1=1}^M \dots \sum_{\tau_m=1}^M \frac{1}{T^{M+\tau^{(m)}}} \sum_{j_1=2M}^{T-M} \dots \sum_{j_{\tau^{(1)}}=2M}^{T-M} \dots \sum_{j_{\tau^{(m)}}=2M}^{T-M} \dots \\
& \quad \dots \sum_{j_M=2M}^{T-M} \dots \sum_{j_{M+\tau^{(1)}}=2M}^{T-M} \dots \sum_{j_{M+\tau^{(m)}}=2M}^{T-M} \left(f \left(\hat{\eta}_{j_1}^2, \dots, \hat{\eta}_{j_{\tau^{(1)}}}^2, \dots, \hat{\eta}_{j_{\tau^{(m)}}}^2, \dots, \hat{\eta}_{j_M}^2; \hat{\theta}_T \right) \right. \\
& \quad \times \dots \times f \left(\hat{\eta}_{j_{\tau^{(m)}}}^2, \dots, \hat{\eta}_{j_M}^2, \dots, \hat{\eta}_{j_{\tau^{(1)}+M}}^2, \dots, \hat{\eta}_{j_{\tau^{(m)}+M}}^2; \hat{\theta}_T \right) \left. \right)
\end{aligned}$$

It is immediate to see that, contrary to the case of $m = 1$, $E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M}^* (\hat{\theta}_T) \right)^m$ cannot be expressed as a V -statistic, and thus the arguments used by HZ do no longer directly apply.

In fact, we can write

$$\begin{aligned}
& E^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^T f_{t,M}^* (\hat{\theta}_T) \right)^m \\
& = \frac{1}{T^{m/2-1}} \sum_{\tau_1=-M}^0 \dots \sum_{\tau_m=-M}^0 \tilde{V}_{1,T,M+\tau^{(m)}} + \frac{1}{T^{m/2-1}} \sum_{\tau_1=1}^M \dots \sum_{\tau_m=1}^M \tilde{V}_{2,T,M+\tau^{(m)}},
\end{aligned}$$

where for $i = 1, 2$ $\tilde{V}_{i,T,M+\tau^{(m)}}$ is not a V -statistic with $M + \tau^{(m)}$ elements, as different indexes appear with different frequency, as for example $\hat{\eta}_{i_{\tau^{(m)}}}^2, \dots, \hat{\eta}_{i_M}^2$ appears m -times, while $\hat{\eta}_{i_1}^2, \dots, \hat{\eta}_{i_{\tau^{(1)}}}^2$ appears only once. Hence, we cannot directly rely on the same argument used by Hidalgo and Zaffaroni for showing first order validity.

As m is finite, we believe it is possible to find a sharp enough bound for II_T^* ; however this is a rather challenging task and we leave it for future research.

5 The Markov Bootstrap

For nonlinear, finite order markov processes it has been already established that the Markov bootstrap can provide higher order refinements very close to those available for *iid* observations. Andrews (2005) has considered the case in which the transition density is known in closed form, so that the Markov bootstrap is indeed a parametric bootstrap, while Horowitz (2003) has considered the case in which the transition density is unknown. More precisely, Andrews (2005) suggests to recursively resample the data from the likelihood evaluated at the estimated parameters. Thus, the bootstrap sample is generated by the same conditional distribution as the original sample, but with the "true" parameters replaced by the estimated parameters. However, this approach is not directly applicable in our context, as it is well known that y_t , as defined in (1) is non markovian, though (y_t, σ_t^2) are jointly markovian. Nevertheless, if we knew the marginal density of the η , say $\phi_\eta(\cdot)$, then we could draw from it T iid observations, and use them to recursively construct $h_t(\hat{\theta}_T)$ and y_t in the same manner outlined in the previous section. In this case, the only difference between the DGP and the bootstrap DGP is the latter is generated using $\hat{\theta}_T$ instead of θ^\dagger . As a consequence, we could obtain higher order refinements along the same lines as in Andrews. Needless to say, this is not particularly useful, as in general we do not know the marginal density of the errors.

For the case in which the transition density of a Markov process is unknown, Horowitz (2003) has suggested to draw the observations from a kernel density estimator. For one-dimensional Markov processes of order q , Horowitz shows that the error in the bootstrap estimate of one-sided and symmetrical probabilities are $O_P(T^{-1+\varepsilon})$ and $O_P(T^{-3/2+\varepsilon})$, respectively, where $\varepsilon > 0$ can be set arbitrarily small.

The key point in Horowitz result, stated in his Lemma 14, is that the cumulants of the original and bootstrap statistics differ only of a term of order $T^{-1/2+\varepsilon}$, which reflects the uniform rate at which the estimated conditional density converges to the true one, provided one uses a "enough" higher order kernel. In fact, because of the markov property, all moments, as well as smooth functions of them, can be computed via the transition density. Thus it is enough to control the error between the estimated and the true conditional density, in a uniform manner.

It is easy to see that β -mixing GARCH processes, with exponentially decaying mixing coefficients, are approximate Markov, according to the definition of Horowitz (2003, Section 4.3).

Though, one of the assumption in Horowitz, i.e. Assumption 4, requires that y_t has bounded support. In the GARCH context, if the innovation has unbounded support, as in the case of normal innovation e.g., then also y_t has unbounded support and Assumption 4 is violated. Needless say, the boundedness of the support is part of a set of sufficient conditions, and therefore we cannot claim it is indeed necessary. We leave the study of possible higher order refinements via the Markov bootstrap for future research.

6 Simulation results

Table 2 in Gonçalves and White (2004, p. 209) reports the coverage rates of nominal 95% symmetric percentile intervals for an ARCH(1) process, both using asymptotic theory and the block bootstrap. They consider sample sizes T equal to 200 and 500 and, for $\bar{\rho} = 0$ (in their notation, $\bar{\rho}$ implies that η_t in (1) follows a $iidN(0, 1)$ distribution). We extend their simulation experiment to the case of a GARCH(1,1) model as in (1), where the true parameter vector is given by $(\theta_1^\dagger, \theta_2^\dagger, \theta_3^\dagger)' = (0.8, 0.1, 0.8)'$, and η_t is either $iidN(0, 1)$ or a t -distribution with 5 degrees of freedom (t_5). We use the MBB of Künsch (1989). Since our model is a GARCH(1,1), a more complicated process than the simulated ARCH(1) of Gonçalves and White (2004, page 209, Table 2), we report results for sample sizes 300 and 500. The number of Monte Carlo replications is set equal to 4000. Table 1 reports the results of coverage rates of nominal 95% symmetric percentile-t intervals both using the QML asymptotic approximation (ASYM) and our block bootstrap procedure (Bboot). We report the confidence intervals (C. I.) for θ_2 and θ_3 . For the block bootstrap procedure (Bboot), we have used 999 bootstrap replications. We provide the results for block sizes $l = 1, 5, 10, 25$ and 50. When the block size is equal to 1, the block bootstrap only ensures first order validity; and then the difference in the empirical coverage rate will illustrate the degree to which higher-order improvements help³. Pascual, Romo and Ruiz (2006, Tables 1 and 2, p. 2299-2300) provide simulation results for a normal and a t_5 innovation process in the GARCH(1,1) context. Their simulation results provide evidence of the good performance of the residual based bootstrap procedure, although as far as we know, there are no theoretical results available in the literature that support the existence of refinements with this type of bootstrap.

³We are very grateful to one of the referees for pointing this out.

Table 1: Empirical coverage rates of nominal 95% symmetric percentile-t intervals for GARCH(1,1)

η_t	T	C. I. for θ	ASYM	Bboot				
				$l = 1$	$l = 5$	$l = 10$	$l = 25$	$l = 50$
$iidN(0, 1)$	300	θ_2	85.7	86.1	89.7	90.4	89.3	90.2
		θ_3	81.1	83.3	90.2	91.1	87.4	89.2
		θ_2	92.8	91.1	92.7	92.6	93.2	91.4
	500	θ_2	92.8	91.1	92.7	92.6	93.2	91.4
		θ_3	87.2	88.1	89.5	93.7	93.1	92.2
		θ_2	81.3	82.1	84.1	85.3	86.1	85.4
t_5	300	θ_2	81.3	82.1	84.1	85.3	86.1	85.4
		θ_3	76.1	78.3	79.2	78.1	79.1	79.4
		θ_2	88.1	89.3	90.1	89.7	90.3	91.2
	500	θ_2	88.1	89.3	90.1	89.7	90.3	91.2
		θ_3	83.1	84.2	85.1	87.1	86.1	87.4
		θ_3	83.1	84.2	85.1	87.1	86.1	87.4

From Table 1, it is immediate to see that when η_t is $iidN(0, 1)$, the empirical coverage of ASYM is substantially below the nominal one, specially for θ_3 , even with samples of 500 observations. Also, for sample of 300 observations, the empirical coverage of ASYM is substantially below the nominal one for both θ_2 and θ_3 .

The bootstrap empirical coverages for θ_2 and θ_3 seem to be rather robust to the choice of the block size, provided $l \geq 10$. The improvements in coverage of the block bootstrap relative to first order asymptotics are noticeable, specially for θ_3 for 500 observations, and for both θ_2 and θ_3 in the case of 300 observations. This shows the usefulness of the block bootstrap in this context. The simulation results of Table 1 when η_t is drawn from $iidN(0, 1)$ innovations, all the moments of η_t exist, and thus conditions in Theorem 2 are satisfied. Moreover, when $l = 1$, the bootstrap only ensures first order validity. Indeed, as expected, the improvements over asymptotic normality, are smaller than in the case of $l \geq 10$. This demonstrates once again the usefulness of the block-bootstrap in providing refinements when the block size is carefully chosen.

Finally, Table 1 also reports the results for the case in which the innovations are drawn from t-distribution with 5 degrees of freedom. Hence, the conditions in Theorems 1 and 2 about the existence of a sufficient number of moments, are clearly violated. Even if the block bootstrap performs much worse than in the case of normal innovations, it still provides some improvement over asymptotic normality.

7 Appendix

The proof of Theorem 1 and Theorem 2 below require the following Lemma, which ensures that the Götze and Hipp conditions (1994) for the existence of the Edgeworth expansion hold for the GARCH case.

Lemma A1:

Let Assumption A hold and assume that $E\left(\eta_t^{4(d_{1,k}+2)}\right) < \infty$. Then there exist constants $K_1 < \infty$ and $\delta > 0$ such that for arbitrarily large $\varsigma > 1$ and all integers $m \in (\delta^{-1}, T)$ and $\tau \in R^{\dim(f_{t,d_{1,k}})}$ with $\delta < \|\tau\| < T^\varsigma$:

$$(15) \quad E \left| E \left(\exp \left(i\tau' \sum_{t=1}^{2m+1} f_{t,d_{1,k}}(\theta^\dagger) \right) \mid \{\eta_j : |j-m| > K_1\} \right) \right| \leq \exp(-\delta),$$

where η_j is the innovation process in (1), $i = \sqrt{-1}$ and $f_{t,d_{1,k}}(\theta^\dagger) = (f_{1,t,d_{1,k}}, f_{2,t,d_{1,k}}, f_{3,t,d_{1,k}})$ and for $\iota = 1, 2, 3$ $f_{\iota,t,d_{1,k}}$ is the vector containing the unique components of the score and the outerproduct and their derivatives through order $d_{1,k}$, evaluated at θ^\dagger and defined in (10).

Proof of Lemma A1:

We proceed in two steps. First we show that we can express $f_{t,d_{1,k}}(\theta^\dagger)$ as a function of all the past innovation, i.e., we can write

$$(16) \quad f_{t,d_{1,k}}(\theta^\dagger) := \Gamma(\eta_{i-j} : j > 0; \theta^\dagger), \quad i \in \mathbb{Z}.$$

Then, we show that $f_{t,d_{1,k}}\theta^\dagger$ satisfies conditions (i)-(iii) in Lemma 2.3 in Götze and Hipp (1994), and hence the statement follows.

We begin by showing (16). Recall that the element of $f_{t,d_{1,k}}(\theta^\dagger)$ are the derivatives of the centered score and Hessian up to order $d_{1,k}$. Now, straightforward calculations give:

$$\begin{aligned} \nabla_\theta L_t(\theta^\dagger) &= \frac{\nabla_\theta h_t(\theta^\dagger)}{h_t(\theta^\dagger)} (\eta_t^2 - 1), \\ \nabla_\theta^2 L_t(\theta^\dagger) &= \frac{1}{2}(\eta_t^2 - 1) \left(\frac{\nabla_\theta^2 h_t(\theta^\dagger)}{h_t(\theta^\dagger)} - \frac{\nabla_\theta h_t(\theta^\dagger) \nabla_\theta h_t(\theta^\dagger)'}{h_t^2(\theta^\dagger)} \right) - \frac{1}{2}\eta_t^2 \frac{\nabla_\theta h_t(\theta^\dagger) \nabla_\theta h_t(\theta^\dagger)'}{h_t^2(\theta^\dagger)} \\ \nabla_\theta^3 L_t(\theta^\dagger) &= \frac{1}{2}(\eta_t^2 - 1) \left(\frac{\nabla_\theta^3 h_t(\theta^\dagger)}{h_t(\theta^\dagger)} - \frac{\nabla_\theta h_t(\theta^\dagger)' \otimes \nabla_\theta^2 h_t(\theta^\dagger)'}{h_t^2(\theta^\dagger)} + \frac{2(\nabla_\theta h_t(\theta^\dagger)' \otimes \nabla_\theta h_t(\theta^\dagger) \otimes \nabla_\theta h_t(\theta^\dagger)')}{h_t^3(\theta^\dagger)} \right) \\ &\quad + \frac{1}{2}\eta_t^2 \left(\frac{(\nabla_\theta h_t(\theta^\dagger)' \otimes \nabla_\theta h_t(\theta^\dagger) \otimes \nabla_\theta h_t(\theta^\dagger)')}{h_t^3(\theta^\dagger)} - \frac{\nabla_\theta h_t(\theta^\dagger)' \otimes \nabla_\theta^2 h_t(\theta^\dagger)'}{h_t^2(\theta^\dagger)} \right) \end{aligned}$$

It now becomes apparent, that in the computation of $\nabla_{\theta}^3 L_t(\theta^\dagger)$, by the law of iterated expectations

$$\begin{aligned} & E \left[\eta_t^2 \left(\frac{(\nabla_{\theta} h_t(\theta^\dagger))' \otimes \nabla_{\theta} h_t(\theta^\dagger) \otimes \nabla_{\theta} h_t(\theta^\dagger)}{h_t^3(\theta^\dagger)} \right) \right] \\ &= E \left[\frac{(\nabla_{\theta} h_t(\theta^\dagger))' \otimes \nabla_{\theta} h_t(\theta^\dagger) \otimes \nabla_{\theta} h_t(\theta^\dagger)}{h_t^3(\theta^\dagger)} \right] \end{aligned}$$

and the term of higher order for generic k is

$$E \left(\frac{(\nabla_{\theta} h_t(\theta^\dagger))' \otimes \nabla_{\theta} h_t(\theta^\dagger) \otimes \dots \otimes \nabla_{\theta} h_t(\theta^\dagger)}{h_t^k(\theta^\dagger)} \right).$$

From Lemma 1 in Lumsdaine (1996, equations A1.1-A1.4), it follows that for $i = 1, 2, 3$

$E \left(\left(\frac{\nabla_{\theta_i} h_t(\theta^\dagger)}{h_t(\theta^\dagger)} \right)^k \right)$ is of the same order of magnitude as $E \left(\frac{y_{t-j}^{2k}}{h_t^k(\theta^\dagger)} \right) = E \left(\frac{h_{t-j}^k(\theta^\dagger)}{h_t^k(\theta^\dagger)} \eta_{t-j}^{2k} \right)$ for any

$j < t$. Therefore, a necessary condition for $E \left(\nabla_{\theta_{j_1 \dots j_k}} L_T(\theta^\dagger) \right) < \infty$ is that $E \left(\eta_{t-j}^{2k} \right) < \infty$.

Indeed, this is also a sufficient condition. In fact, Ling and McAleer (2003, p.304) show that a sufficient condition for the existence of $E \left(\frac{(\nabla_{\theta} h_t(\theta^\dagger))' \otimes \nabla_{\theta} h_t(\theta^\dagger)}{h_t^2(\theta^\dagger)} \right)$ is that $E(\eta_i^4) < \infty$. The same argument

can be used to show a sufficient condition for $E \left(\nabla_{\theta_{j_1 \dots j_k}} L_T(\theta^\dagger) \right) < \infty$ is that $E \left(\eta_{t-j}^{2k} \right) < \infty$.

From the expression for $\nabla_{\theta} L_t(\theta)$, $\nabla_{\theta}^2 L_t(\theta)$, ..., $\nabla_{\theta}^k L_t(\theta)$ we see that they can be expressed as ratios of products of derivatives of $h_t(\theta^\dagger)$ over power of $h_t(\theta^\dagger)$. Recalling that,

$$h_t(\theta) = \theta_1 \left[1 + \sum_{j=1}^{t-1} \prod_{i=1}^j (\theta_2 \eta_{t-i}^2 + \theta_3) \right]$$

we have that,

$$\nabla_{\theta_1} h_t = 1 + \sum_{j=1}^{t-1} \prod_{i=1}^j (\theta_2 \eta_{t-i}^2 + \theta_3)$$

and that

$$\begin{aligned} \nabla_{\theta_2} h_t &= \theta_1 \sum_{k=0}^{t-2} \theta_3^k \sum_{j_1=0}^{t-2-k} \theta_3^{j_1} \eta_{t-j_1-1}^2 \\ &+ 2\theta_1 \theta_2 \sum_{k=0}^{t-3} \theta_3^k \sum_{j_2=1}^{t-2-k} \sum_{j_1 < j_2} \theta_3^{j_2-1} \eta_{t-j_1-1}^2 \eta_{t-j_2-1}^2 \\ &+ 3\theta_1 \theta_2^2 \sum_{k=0}^{t-4} \theta_3^k \sum_{j_3=2}^{t-2-k} \sum_{j_2 < j_3} \sum_{j_1 < j_2} \theta_3^{j_3-2} \eta_{t-j_1-1}^2 \eta_{t-j_2-1}^2 \eta_{t-j_3-1}^2 \\ &+ \dots \\ &+ l\theta_1 \theta_2^{l-1} \sum_{k=0}^{t-l-1} \theta_3^k \sum_{j_l=l-1}^{t-2-k} \sum_{j_{l-1} < j_l} \dots \sum_{j_1 < j_2} \theta_3^{j_l-(l-1)} \eta_{t-j_1-1}^2 \dots \eta_{t-j_l-1}^2 \\ &+ \dots + \\ &(t-1) \theta_1 \theta_2^{t-2} \sum_{k=0}^0 \theta_3^k \sum_{j_{t-1}=t-2}^{t-2-k} \sum_{j_{t-2} < j_{t-1}} \dots \sum_{j_1 < j_2} \theta_3^{j_{t-1}-(t-2)} \eta_{t-j_1-1}^2 \dots \eta_{t-j_{t-1}-1}^2 \end{aligned}$$

$$\nabla_{\theta_1} h_t / h_t^* = 1 / \theta_1.$$

and

$$\begin{aligned} \nabla_{\theta_3} h_t &= \theta_1 \theta_2 \left[\sum_{k=0}^{t-2} k \theta_3^{k-1} \sum_{j_1=0}^{t-2-k} \theta_3^{j_1} \eta_{t-j_1-1}^{2*} + \sum_{k=0}^{t-2} \theta_3^k \sum_{j_1=0}^{t-2-k} j_1 \theta_3^{j_1-1} \eta_{t-j_1-1}^{2*} \right] \\ &+ \theta_1 \theta_2^2 \left[\sum_{k=0}^{t-3} k \theta_3^{k-1} \sum_{j_3=1}^{t-2-k} \sum_{j_1 < j_2} \theta_3^{j_2-1} \eta_{t-j_1-1}^{2*} \eta_{t-j_2-1}^{2*} \right] \end{aligned}$$

$$\begin{aligned}
 & + \sum_{k=0}^{t-3} \theta_3^k \sum_{j_2=1}^{t-2-k} \sum_{j_1 < j_2} (j_2 - 1) \theta_3^{j_2-2} \eta_{t-j_1-1}^{2*} \eta_{t-j_2-1}^{2*} \Big] \\
 & + \theta_1 \theta_2^3 \left[\sum_{k=0}^{t-4} k \theta_3^{k-1} \sum_{j_3=2}^{t-2-k} \sum_{j_2 < j_3} \sum_{j_1 < j_2} \theta_3^{j_3-2} \eta_{t-j_1-1}^{2*} \eta_{t-j_2-1}^{2*} \eta_{t-j_3-1}^{2*} \right. \\
 & + \sum_{k=0}^{t-4} \theta_3^k \sum_{j_3=2}^{t-2-k} \sum_{j_2 < j_3} \sum_{j_1 < j_2} (j_3 - 2) \theta_3^{j_3-3} \eta_{t-j_1-1}^{2*} \eta_{t-j_2-1}^{2*} \eta_{t-j_3-1}^{2*} \Big] + \\
 & + \dots + \\
 & + \theta_1 \theta_2^l \left[\sum_{k=0}^{t-l-1} k \theta_3^{k-1} \sum_{j_l=l-1}^{t-2-k} \sum_{j_{l-1} < j_l} \dots \sum_{j_1 < j_2} \theta_3^{j_l-(l-1)} \eta_{t-j_1-1}^{2*} \dots \eta_{t-j_l-1}^{2*} \right. \\
 & + \sum_{k=0}^{t-l-1} \theta_3^k \sum_{j_l=l-1}^{t-2-k} \sum_{j_{-1} < j_l} \dots \sum_{j_1 < j_2} (j_l - (l-1)) \theta_3^{j_l-l} \eta_{t-j_1-1}^{2*} \dots \eta_{t-j_l-1}^{2*} \Big] + \\
 & + \dots + \\
 & \theta_1 \theta_2^{t-1} \sum_{j_{t-1}=t-2}^{t-2-k} \sum_{j_{t-2} < j_{t-1}} \dots \sum_{j_1 < j_2} (j_{t-1} - (t-2)) \theta_3^{j_{t-1}-(t-1)} \eta_{t-j_1-1}^{2*} \dots \eta_{t-j_{t-1}-1}^{2*}.
 \end{aligned}$$

Thus, we can write $f_{t,d_1,k}(\theta^\dagger) = (f_{1,t,d_1,k}(\theta^\dagger), f_{2,t,d_1,k}(\theta^\dagger), f_{3,t,d_1,k}(\theta^\dagger))$ as

$$(17) \quad f_{t,d_1,k} := \Gamma(\eta_{t-i} : i \geq 0), \quad t \in \mathbb{Z}$$

where η_t is $iid(0, \sigma_\eta^2)$ with positive and continuous density, Γ is a measurable function $\Gamma : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^{3 \times 2d_1,k}$. For $\mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$ and $x \in \mathbb{R}$ define $(\mathbf{y}, x)^t$ as the sequence with coordinates,

$$\eta_i = \begin{cases} y_i & i < t \\ x & i = t \\ y_{i-1} & i > t \end{cases}$$

Recalling that $E(\eta^{4(d_1,k+2)}) < \infty$, it follows that $E(\|\Gamma((\eta, x_1)^t)\|) < \infty$. Also, given assumption A, $f_{t,d_1,k}$ is β -mixing with mixing coefficients decaying at an exponential rate. Hence, condition (i) in Götze and Hipp (1994, Lemma 2.3) follows, i.e.

There exists $K < \infty$ and $\bar{\alpha} > 0$ such that for $j \in \mathbb{Z}$ and $x_1, x_2 \in \mathbb{R}$

$$E \|\Gamma((\eta, x_1)^t) - \Gamma((\eta, x_2)^t)\| \leq K e^{-\bar{\alpha}|t|} |x_1 - x_2|.$$

As all partial derivatives of the Gaussian likelihood for GARCH(1,1) models are almost surely continuous provided the innovation process has a continuous positive density, then condition (ii) in Götze and Hipp (1994, Lemma 2.3), follows, i.e.

For $t \in \mathbb{Z}$ there exists $\Lambda_t \subset \mathbb{R}$, $P(\Lambda_t) = 1$, such that for all $x_0 \in \Lambda_t$, $\bar{\beta}, \bar{\delta} > 0$, there exists $\bar{\tau} > 0$ satisfying

$$\begin{aligned}
 & P(y \in \mathbb{R}^{\mathbb{Z}} : \forall x \in \mathbb{R}, |x - x_0| < \bar{\tau}, \frac{\partial}{\partial \eta_0} f_{t,d_1,k} \text{ exists at the point } (y, x)^t \text{ and} \\
 & \left| \frac{\partial}{\partial \eta_0} f_{t,d_1,k}((y, x)^t) - \frac{\partial}{\partial \eta_0} f_{t,d_1,k}((y, x_0)^t) \right| \leq \bar{\delta}) \geq 1 - \bar{\beta}.
 \end{aligned}$$

Finally, for a Gaussian likelihood, the partial derivatives, with respect to the same argument, of the various component of $f_{t,d_{1,k}}$ are not linearly dependent; also the partial derivatives of a given component of $f_{t,d_{1,k}}$, with respect to generic $6d_{1,k}$ arguments, are not linearly dependent. Hence, (iii) in Götze and Hipp (1994, Lemma 2.3) follows, i.e.

For some distinct $l_1, \dots, l_{6d_{1,k}} \geq 0$,

$$\det \left(\sum_{t=0}^{\infty} \frac{\partial}{\partial \eta_{l_{\bar{v}}}} f_{t,d_{1,k}} : \bar{v} = 1, \dots, 6d_{1,k} \right) \neq 0.$$

on a set of positive P-probability.

Proof of Theorem 1:

(i) We begin with the "only if" part. As we are interested in necessary conditions it suffices to consider $T^{1/2} (\hat{\theta}_{i,T} - \theta_i^\dagger)$. We first need to establish the existence of a stochastic expansion of $T^{1/2} (\hat{\theta}_{i,T} - \theta_i^\dagger)$ up to a term of order $o(T^{-\frac{k-1}{2}})$. Given A, and assuming that $E(\eta_t^4) < \infty$, it follows that $T^{1/2} (\hat{\theta}_{i,T} - \theta_i^\dagger)$ is asymptotically normal (see e.g. Bollerslev and Wooldridge 1992), thus

$$T^{1/2} (\hat{\theta}_T - \theta^\dagger) = T^{1/2} \delta_T = O_P(1).$$

Let $L_T(\theta) = \frac{1}{T} \sum_{t=1}^T L_t(\theta)$, then the Taylor expansion of $\nabla_\theta L_T(\hat{\theta}_T)$ around θ^\dagger , up to the k order writes as

$$\begin{aligned} 0 &= \nabla_\theta L_T(\theta^\dagger) + \nabla_\theta^2 L_T(\theta^\dagger) \delta_T \\ &+ \frac{1}{2} \left(\begin{array}{l} \sum_{j_1=1}^3 \sum_{j_2=1}^3 \nabla_{\theta_1 \theta_{j_1} \theta_{j_2}} L_T(\theta^\dagger) \delta_{T,j_1} \delta_{T,j_2} \\ \sum_{j_1=1}^3 \sum_{j_2=1}^3 \nabla_{\theta_2 \theta_{j_1} \theta_{j_2}} L_T(\theta^\dagger) \delta_{T,j_1} \delta_{T,j_2} \\ \sum_{j_1=1}^3 \sum_{j_2=1}^3 \nabla_{\theta_3 \theta_{j_1} \theta_{j_2}} L_T(\theta^\dagger) \delta_{T,j_1} \delta_{T,j_2} \end{array} \right) \\ &+ \dots + \frac{1}{k!} \left(\begin{array}{l} \sum_{j_1=1}^3 \dots \sum_{j_k=1}^3 \nabla_{\theta_1 \theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger) \delta_{T,j_1} \times \delta_{T,j_2} \times \dots \times \delta_{T,j_k} \\ \sum_{j_1=1}^3 \dots \sum_{j_k=1}^3 \nabla_{\theta_2 \theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger) \delta_{T,j_1} \times \delta_{T,j_2} \times \dots \times \delta_{T,j_k} \\ \sum_{j_1=1}^3 \dots \sum_{j_k=1}^3 \nabla_{\theta_3 \theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger) \delta_{T,j_1} \times \delta_{T,j_2} \times \dots \times \delta_{T,j_k} \end{array} \right) \\ &+ e_{k+1,T}, \end{aligned}$$

where $\nabla_{\theta_1 \theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger)$ denote the $k+1$ -th derivative with respect to $\theta_1, \theta_{j_1}, \dots, \theta_{j_k}$, and δ_{T,j_k} the

j_k -th component of δ_T . Thus,

$$\begin{aligned}
& T^{1/2} \left(\widehat{\theta}_T - \theta^\dagger \right) \\
&= \left(-\nabla_{\theta}^2 L_T(\theta^\dagger) \right)^{-1} \left(T^{1/2} \nabla_{\theta} L_T(\theta^\dagger) \right) \\
&+ \frac{1}{2} T^{1/2} \left(\begin{array}{l} \sum_{j_1=1}^3 \sum_{j_2=1}^3 \nabla_{\theta_1 \theta_{j_1} \theta_{j_2}} L_T(\theta^\dagger) \delta_{T,j_1} \delta_{T,j_2} \\ \sum_{j_1=1}^3 \sum_{j_2=1}^3 \nabla_{\theta_2 \theta_{j_1} \theta_{j_2}} L_T(\theta^\dagger) \delta_{T,j_1} \delta_{T,j_2} \\ \sum_{j_1=1}^3 \sum_{j_2=1}^3 \nabla_{\theta_3 \theta_{j_1} \theta_{j_2}} L_T(\theta^\dagger) \delta_{T,j_1} \delta_{T,j_2} \end{array} \right) + \dots \\
&+ \frac{1}{k!} T^{1/2} \left(\begin{array}{l} \sum_{j_1=1}^3 \dots \sum_{j_k=1}^3 \nabla_{\theta_1 \theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger) \delta_{T,j_1} \times \delta_{T,j_2} \times \dots \times \delta_{T,j_k} \\ \sum_{j_1=1}^3 \dots \sum_{j_k=1}^3 \nabla_{\theta_2 \theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger) \delta_{T,j_1} \times \delta_{T,j_2} \times \dots \times \delta_{T,j_k} \\ \sum_{j_1=1}^3 \dots \sum_{j_k=1}^3 \nabla_{\theta_3 \theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger) \delta_{T,j_1} \times \delta_{T,j_2} \times \dots \times \delta_{T,j_k} \end{array} \right) \\
&+ T^{1/2} e_{k+1,T}.
\end{aligned}$$

Hereafter, $\nabla_{\theta}^k L_T(\theta^\dagger)$ is the matrix with generic element $\nabla_{\theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger)$, $j_i = 1, 2, 3$ for all i . The expression for $\nabla_{\theta} L_t(\theta^\dagger)$, $\nabla_{\theta}^2 L_t(\theta^\dagger)$, $\nabla_{\theta}^3 L_t(\theta^\dagger)$, etc. have been already provided in the proof of the Lemma.

Recalling that, $E \left(\nabla_{\theta_1 \theta_{j_1} \dots \theta_{j_k}} L_T(\theta^\dagger) \right) < \infty$ is that $E(\eta_t^{2k}) < \infty$, in order to have an Edgeworth expansion for the LHS of (11), up to the k -th term, we need the existence of the $(k+2)$ moment of the $(k+2)$ derivative of the score and the outerproduct, which in turn requires that $E \left(\eta_t^{2(k+2)(k+4)} \right) < \infty$.

(ii) We now turn to the "if" part. Given Assumption A and recalling (17), we have that there exist a constant K , and $\alpha > 0$, such that for $m \geq 1$,

$$(18) \quad E \left(\left\| \Gamma(\eta_t, \eta_{t-1}, \dots, \eta_{t-m}, \dots, \eta_1, \eta_0) - \Gamma(\eta_t, \eta_{t-1}, \dots, \eta_{t-m}, 0, \dots, 0) \right\| \right) \leq K \exp(-\alpha m).$$

Lemma A1 ensure that condition (15) is satisfied. Now, let $d_{1,k} = k + 3$, where $d_{1,k}$ is defined in Lemma A1. Recalling (10), $E \left(\eta_t^{2(k+3)(k+5)} \right) < \infty$ ensures that the $k + 3$ -th moments of all the component of $f_{t,k+3}$ is finite. Now, the latter condition, plus (15) and (18) ensures that (see e.g. Götze and Hipp (1983, p.216-217) that

$$\lim_{T \rightarrow \infty} T^{k/2} \sup_{z \in \mathcal{R}} \left| P(G(S_{T,k+3}(\theta^\dagger)) \leq z) - \left(1 + \sum_{i=1}^k T^{-1/2} \pi_i(\partial/\partial z) \right) \Phi(z) \right| = 0,$$

where $S_{T,k+3}(\theta^\dagger) = \frac{1}{T} \sum_{t=1}^T f_{t,k+3}(\theta^\dagger)$. Finally, by Proposition 1 in Hall and Horowitz (1996),

$$\lim_{T \rightarrow \infty} T^{k/2} \left(\sup_{z \in \mathcal{R}} \left| P(G(S_{T,k+3}(\theta^\dagger)) \leq z) - P(t_{\theta_i} \leq z) \right| > T^{-k/2} \right) = 0.$$

The statement then follows.

Hereafter, let P^* denoting the probability law governing the block bootstrap, E^* and Var^* the mean and variance operator under P^* , also $\xrightarrow{pr^*}$ and $\xrightarrow{d^*}$ denote convergence in probability and in distribution under P^* , conditionally on the sample.

Proof of Theorem 2:

It is immediate to see that our set-up can be casted into Andrews (2002) set-up, with $L_t(\theta)$ playing the same role as $\rho(X_t; \theta)$ and $\kappa = 1$, as the score is a martingale difference sequence. Now, given (18), Assumption 1 in Andrews (2002) holds, and given Lemma A1 his Assumption 4 is satisfied too. As all the derivatives of the gaussian likelihood are continuous, and recalling that, because of Lemma 1 in Lumsdaine (1996), $E\left(\nabla_{\theta_{j_1 \dots \theta_{j_k}}} L_T(\theta^\dagger)\right) < \infty$ if and only if $E(\eta_t^{2k}) < \infty$, it follows that Assumptions 2, 3 and 5 in Andrews were satisfied if we assumed that the innovation process has all moments finite. On the other hand, we have only required that $E\left(\eta_t^{2(d_{1,k}+d_{2,k}+2)(k+2)p}\right) < \infty$, where $d_{1,k} \geq k+3$, $d_{2,k} \geq k+1$, and $p > \frac{k/2}{2(1/2-\xi-\gamma)}$. The moment conditions on the likelihood derivatives of the elements of $f_{i,t,d_{1,k}}(\theta^\dagger)$ play a crucial role in the proof of Lemma 14 of Andrews. Here, we proceed by showing that the moments conditions imposed on the innovation term ensure that the statement in his Lemma 14 applies also in our context. The rest of the proof comes straightforwardly from Andrews (2002).

Define the bootstrap analog of (10), that is:

$$(19) \quad f_{i,t,d_{1,k}}^*(\theta) = \left(\overline{\nabla_{\theta_i} L_t^*}(\theta), \dots, \nabla_{\theta_i}^{d_{1,k}} \left(\overline{\nabla_{\theta} L_t^*}(\theta) \right), \overline{\nabla_{\theta_i} L_t^* \nabla_{\theta_i} L_t^*}(\theta)', \dots, \nabla_{\theta_i}^{d_{1,k}} \left(\overline{\nabla_{\theta} L_t^* \nabla_{\theta} L_t^*}(\theta)' \right) \right),$$

where $\overline{\nabla_{\theta_i} L_t^*}(\theta) = \nabla_{\theta_i} L_t^*(\theta) - E^*(\nabla_{\theta_i} L_t^*(\theta))$, and

$\overline{\nabla_{\theta_i} L_t^* \nabla_{\theta_i} L_t^*}(\theta)' = \nabla_{\theta_i} L_t^* \nabla_{\theta_i} L_t^*(\theta)' - E^*(\nabla_{\theta_i} L_t^* \nabla_{\theta_i} L_t^*(\theta)')$. Also, define

$$S_{T,i,d_{1,k}}(\theta^\dagger) = \frac{1}{T} \sum_{t=1}^T f_{i,t,d_{1,k}}(\theta^\dagger)$$

$$S_{T,i,d_{1,k}}^*(\hat{\theta}_T) = \frac{1}{T} \sum_{t=1}^T f_{i,t,d_{1,k}}^*(\hat{\theta}_T)$$

Given A , $d_1 \geq k+2$, $\gamma > 0$, there exists an infinite differentiable function $G(\cdot)$, with $G(S_i) = 0$ and $G(S_i^*) = 0$, where $S_i = E(S_{T,i,d_{1,k}}(\theta^\dagger))$, $S_i^* = E^*(S_{T,i,d_{1,k}}^*(\hat{\theta}_T))$, such that, for $k = 2, 3$

$$(20) \quad \lim_{T \rightarrow \infty} \sup_z T^{k/2} \left| \Pr(t_{\theta_i,T} \leq z) - \Pr(T^{1/2} G(S_{T,i,d_{1,k}}(\theta^\dagger)) \leq z) \right| = 0$$

and

$$(21) \quad \lim_{T \rightarrow \infty} \sup_z T^{k/2} \left| \Pr(\tilde{t}_{\theta_i,T} \leq z) - \Pr(T^{1/2} G(S_{T,i,d_{1,k}}^*(\hat{\theta}_T)) \leq z) \right| = 0$$

Let $\Psi_{i,k,T} = T^{1/2}(S_{i,T,d_{1,k}}(\theta^\dagger) - E(S_{i,T,d_{1,k}}(\theta^\dagger)))$ and $\Psi_{i,k,T}^* = T^{1/2}(S_{i,T,d_{1,k}}^*(\hat{\theta}_T) - E^*(S_{i,T,d_{1,k}}^*(\hat{\theta}_T)))$, and let $\Psi_{i,k,T,j}$ be the j -th element of $\Psi_{i,k,T}$. Define

$$(22) \quad v_{i,T,k} = E(T^{\alpha(m)} \prod_{\mu=1}^m \Psi_{i,k,T,j_\mu})$$

and let $v_{i,k} = \lim_{T \rightarrow \infty} v_{i,T,k}$, and

$$(23) \quad \tilde{v}_{i,T,k} = E^* \left(T^{\alpha(m)} \Pi_{\mu=1}^m \Psi_{i,k,T,j_\mu}^* \right)$$

where $\alpha(m) = 1/2$ if m is odd and 0 if m even, and $2 \leq m \leq k+2$. We need to show that

$$(24) \quad \lim_{T \rightarrow \infty} T^{k/2} P \left(\left| T^{1/2} E^* \left(\Pi_{\mu=1}^m \Psi_{i,k,T,\mu}^* \right) - T^{1/2} E \left(\Pi_{\mu=1}^m \Psi_{i,k,T,\mu} \right) \right| > c T^{-\xi} \right) = 0$$

for some $c > 0$, and

$$(25) \quad \lim_{T \rightarrow \infty} T^\xi \left| T^{1/2} E \left(\Pi_{\mu=1}^m \Psi_{i,T,j_\mu} \right) - \lim_{T \rightarrow \infty} T^{1/2} E \left(\Pi_{\mu=1}^m \Psi_{i,T,j_\mu} \right) \right| = 0$$

Hereafter, as the arguments used in the proof are the same across all i , when there is no risk on confusion, we omit the subscript i . Let $f_{i,t,d_{1,k}}(\theta^\dagger)$ and $f_{i,t,d_{1,k}}^*(\widehat{\theta}_T)$ be defined as (10) and (19), respectively. Define $\bar{f}_{t,k} = f_{i,t,d_{1,k}}(\theta^\dagger) - E(f_{i,t,d_{1,k}}(\theta^\dagger))$ and $\bar{f}_{t,k}^* = f_{i,t,d_{1,k}}^*(\widehat{\theta}_T) - E^*(f_{i,t,d_{1,k}}^*(\widehat{\theta}_T))$; also let $Y_j = \sum_{t \in b_j} \bar{f}_{t,k}$, $\tilde{Y}_j = \sum_{t \in b_j} \bar{f}_{t,k}^*$, and $Y_j^* = \sum_{t \in b_j^*} \bar{f}_{t,k}^*$, where for $j = 1, \dots, T_l$, $T_l = T - l + 1$, $b_j = (j, \dots, j + l - 1)$ $j = 1, \dots, b$ and b_j^* is an *iid* discrete uniform random variable which is equal to b_j with equal probability $1/T_l$. As shown in Andrews (2002), the most difficult case is that of $m = 3$. Now,

$$\begin{aligned} T^{1/2} E^* \left(\Pi_{\mu=1}^3 \Psi_{i,T,\mu}^* \right) &= T^{-1} \sum_{j_1=1}^b \sum_{j_2=1}^b \sum_{j_3=1}^b E^* \left(Y_{j_1}^* Y_{j_2}^* Y_{j_3}^* \right) \\ &= T^{-1} b E^* \left(Y_1^{*3} \right) = T^{-1} b T_l^{-1} \sum_{j=1}^{T_l} \tilde{Y}_j^3. \end{aligned}$$

If for some $C < \infty$, and for $i, j = 1, 2, 3$ $E \left(\left(\nabla_{\theta}^{d_{1,k} + d_{2,k}} \left(\nabla_{\theta_i} L_T(\theta) \nabla_{\theta_i} L_T(\theta)' \right) \right)^{3p} \right) < C$ uniformly in a neighborhood of θ^\dagger , with $d_{1,k} \geq k + 3$, $d_{2,k} \geq k + 1$, p defined as below and if $\xi + \gamma < 1/2$, then from Andrews (2001),

$$(26) \quad \lim_{T \rightarrow \infty} T^{k/2} P \left(T^{-1} b T_l^{-1} \sum_{j=1}^{T_l} \left| \tilde{Y}_j^m - Y_j^m \right| > T^{-\xi} \right) = 0.$$

Given (26), in order to show (24), one has to show that B_1 and B_2 below approach zero, where

$$B_1 = \lim_{T \rightarrow \infty} T^{k/2} P \left(\left| T^{-1} b T_l^{-1} \sum_{j=1}^{T_l} \left(Y_j^3 - E(Y_j^3) \right) \right| > T^\xi \right)$$

and that

$$B_2 = \lim_{T \rightarrow \infty} T^\xi \left| T^{-1} b E(Y_1^3) - T^{1/2} E \left(\Pi_{\mu=1}^3 \Psi_{i,T,\mu} \right) \right|.$$

By Markov inequality and a double application of Yokoyama-Doukan inequality (Doukhan 1995, p.25-30), there exists a constant C and $\delta > 0$ such that,

$$\begin{aligned}
 B_1 &\leq CT^{k/2+p(\xi-1)} E \left| bT_l^{-1} \sum_{j=1}^{T_l} (Y_j^3 - E(Y_j^3)) \right|^p \\
 &\leq C \lim_{T \rightarrow \infty} T^{k/2+p(\xi-1)} b^{p/2} l^{3p/2} \left(E |\bar{f}_{1,k}|^{3p+4\delta} \right)^{\frac{3p(3+\delta)}{(3+\delta)(3(p+\delta)+\delta)}} \\
 (27) \quad &\leq C \lim_{T \rightarrow \infty} T^{k/2+p(\xi-1)} T^{p/2} T^{p\gamma} \left(E |\bar{f}_{1,k}|^{3p+4\delta} \right)^{\frac{3p(3+\delta)}{(3+\delta)(3(p+\delta)+\delta)}}
 \end{aligned}$$

Thus, $B_1 \rightarrow 0$, provided $p > \frac{k}{2(1/2-\gamma-\xi)}$ and if $E |\bar{f}_{1,k}|^{3p+4\delta} < \infty$.

Also, by the proof of Lemma 14 in Andrews (2001), it follows that for the case of $m = 5$, we need $p > k/(3 - 2\xi - 4\gamma)$, requiring $\xi + 2\gamma < 3/2$, which is implied by $\xi + \gamma < 1/2$ and $E |\bar{f}_{1,k}|^{5p+4\delta} < \infty$. Given (10) and the last line of (27), it suffice that the first $mp + 4\delta$ -th moments of all score derivatives up to order $(d_{1,k} + d_{2,k})$ exist. In the proof of Lemma A1 we have shown that $E \left((\nabla_{\theta_i}^{r_1} L_T(\theta^\dagger))^{r_2} \right) < \infty$ if and only if $E \left(\eta_{t-j}^{2(r_1+r_2)} \right) < \infty$; by the same argument, and as a straightforward consequence of the chain rule $E \left(\left(\nabla_{\theta_i}^{r_1} (L_T(\theta^\dagger) L_T(\theta^\dagger)') \right)^{r_2} \right) < \infty$ if and only if $E \left(\eta_{t-j}^{2(r_1+r_2+1)} \right) < \infty$. Now, recalling that $2 \leq m \leq k + 2$, setting $\delta = 1/4$, it follows that $E \left(|\bar{f}_{1,k}|^{mp+4\delta} \right) < \infty$ and $E \left(\left(\nabla_{\theta}^{d_{1,k}+d_{2,k}} (\nabla_{\theta_i} l_t(\theta) \nabla_{\theta_i} l_t(\theta)') \right)^{mp} \right) < \infty$ if $E \left(\eta_t^{2(d_{1,k}+d_{2,k}+2)(k+2)p} \right) < \infty$, where $d_{1,k} \geq k + 3$, $d_{2,k} \geq k + 1$, and $p > \frac{k/2}{2(1/2-\xi-\gamma)}$. Thus, the left hand side of (27) is approaching zero, provided $E \left(\eta_t^{2(2k+6)(k+2)p} \right) < \infty$.

As for B_2 , along the lines of proof of Lemma 14 in Andrews (2002), recalling that $l = T^\gamma$,

$$\begin{aligned}
 &T^{-1} b E(Y_1^3) - T^{1/2} E(\Pi_{\mu=1}^3 \Psi_{i,T,\mu}) \\
 &= \frac{1}{l} E \left(\sum_{i=1}^l \bar{f}_{i,k} \right)^3 - \frac{1}{T} \sum_{i_1=1}^T \sum_{i_2=1}^T \sum_{i_2=1}^T \bar{f}_{i_1,k} \bar{f}_{i_2,k} \bar{f}_{i_3,k} \\
 &= \sum_{i_1=-l+1}^{l-1} \sum_{i_2=-l+1}^{l-1} \left(1 - \frac{(i_1 + i_2)}{l} \right) E(\bar{f}_{0,k} \bar{f}_{i_1,k} \bar{f}_{i_2,k}) \\
 &\quad - \sum_{i_1=-T+1}^{T-1} \sum_{i_2=-T+1}^{T-1} \left(1 - \frac{(i_1 + i_2)}{T} \right) E(\bar{f}_{0,k} \bar{f}_{i_1,k} \bar{f}_{i_2,k}).
 \end{aligned}$$

Now,

$$T^\xi \left| \sum_{i_1=-l+1}^{l-1} \sum_{i_2=-l+1}^{l-1} E(\bar{f}_{0,k} \bar{f}_{i_1,k} \bar{f}_{i_2,k}) - \sum_{i_1=-T+1}^{T-1} \sum_{i_2=-T+1}^{T-1} E(\bar{f}_{0,k} \bar{f}_{i_1,k} \bar{f}_{i_2,k}) \right| \rightarrow 0$$

by the Doukhan mixing inequality (1995, p.9), provided $\gamma > \xi$. Also,

$$T^{\xi-1} \left| \sum_{i_1=-T+1}^{T-1} \sum_{i_2=-T+1}^{T-1} (i_1 + i_2) E(\bar{f}_{0,k} \bar{f}_{i_1,k} \bar{f}_{i_2,k}) \right| \rightarrow 0$$

by mixing inequality, and

$$T^{\xi-\gamma} \left| \sum_{i_1=-l+1}^{l-1} \sum_{i_2=-l+1}^{l-1} (i_1 + i_2) E(\bar{f}_{0,k} \bar{f}_{i_1,k} \bar{f}_{i_2,k}) \right| \rightarrow 0,$$

provided $\xi < \gamma$. The statement in (24) follows by the same argument used to show that $B_2 \rightarrow 0$.

Finally, from (24) and (25) it follows that

$$\lim_{T \rightarrow \infty} T^{k/2} P(|\tilde{v}_{i,T,k} - v_{i,k}| > T^\xi) = 0.$$

The statement in Lemma 14 in Andrews then follows.

■

8 References

- Andrews, D. W. K. (2001), Higher-Order Improvements of a Computationally Attractive k -Step Bootstrap for Extremum Estimators, *Cowles Foundation Discussion Paper #1230R*.
- Andrews, D. W. K. (2002), Higher-Order Improvements of a Computationally Attractive k -Step Bootstrap for Extremum Estimators, *Econometrica* 70, 119-162.
- Andrews, D. W. K. (2005), High Order Improvements of the Parametric Bootstrap for Markov Processes, in *Identification and Inference in Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg*, ed. by D.W.K. Andrews and J.H. Stock, Cambridge University Press.
- Bhattacharya R. N. (1987), Some Aspects of Edgeworth Expansions in Statistics and Probability, in *New Perspectives in Theoretical and Applied Statistics*, ed. by M.L. Puri, J.P. Vilaploma, and W. Wertz, Wiley.
- Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307-327.
- Bollerslev, T., and J. M. Wooldridge (1992), Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances, *Econometric Reviews* 11, 143-172.
- Bose, A. (1988), Edgeworth Correction by Bootstrap in Autoregressions, *Annals of Statistics* 16, 4, 1709-1722.
- Carrasco, M., and X. Chen (2002), Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, *Econometric Theory* 18, 17-39.
- Chandra, T.K., and J.K. Ghosh (1979), Valid Asymptotic Expansions for the Likelihood Ratio Statistic and Other Perturbed Chi-Square Variables, *Sankhya*, 41, Series A, 22-47.
- Christoffersen, P. and S. Gonçalves (2005), Estimation Risk in Financial Risk Management, *Journal of Risk* 7, 1-29.
- Doukhan, P. (1995), *Mixing Properties and Examples*, Springer and Verlag, New York.
- Gonçalves, S. and L. Kilian (2004), Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form, *Journal of Econometrics* 123, 89-120.
- Gonçalves, S. and L. Kilian (2007), Asymptotic and Bootstrap Inference for $AR(\infty)$ Processes with Conditional Heteroskedasticity, *Econometric Reviews* 26, 6, 609-641.

- Gonçalves, S. and H. White (2004), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics* 119, 199-219.
- Götze, F. and C. Hipp (1983), Asymptotic Expansions for Sums of Weakly Dependent Random Vectors. *Zeitschrift für Wahrscheinlichkeits Theorie und Verwandte Gebiete* 64, 211-239.
- Götze, F. and C. Hipp (1994), Asymptotic Distributions of Statistics in Time Series, *Annals of Statistics* 22, 2062-2088.
- Grams, W.F. and R.J. Serfling (1973), Convergence Rates for U-statistics and Related Statistics, *Annals of Statistics* 1, 153-160.
- Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*. New York: Springer.
- Hall, P., and J. L. Horowitz (1996), Bootstrap Critical Values for Tests based on Generalized-Method-of-Moments Estimators, *Econometrica* 64, 4, 891-916.
- Hidalgo, F.J., and P. Zaffaroni (2007), A Goodness of Fit Test for ARCH(∞), *Journal of Econometrics*, 141, 835-875.
- Horowitz, J.L. (2003) Bootstrap Methods for Markov Processes, *Econometrica*, 71, 1049-1082.
- Inoue A., and L. Kilian (2002), Bootstrapping Autoregressive Processes with Possible Unit Roots, *Econometrica* 70, 1, 377-393.
- Kreiss, J.-P., and J. Franke (1992), Bootstrapping Stationary Autoregressive Moving-Average Models, *Journal of Time Series Analysis* 13, 4, 297-317.
- Künsch, H. R. (1989), The Jackknife and the Bootstrap for General Stationary Observations, *Annals of Statistics* 17, 1217-1241.
- Ling, S.Q., and M. McAleer, (2003), Asymptotic theory for a vector ARMA-GARCH model, *Econometric Theory* 19, 280-310.
- Linton, O. (1997), An Asymptotic Expansion in the GARCH(1,1) Model. *Econometric Theory* 13, 558-581.
- Lumsdaine, R. L. (1996), Asymptotic Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models. *Econometrica* 64, 3, 575-596.

- Mancini, L., and F. Trojani, (2005), Robust Semiparametric Bootstrap Methods for Value at Risk Prediction under GARCH-type Volatility Processes, SSRN Working Paper, University of St. Gallen.
- Pascual L., J. Romo and E. Ruiz (2006), Bootstrap Prediction for Returns and Volatilities in GARCH Models, *Computational Statistics and Data Analysis* 50, 2293-2312.
- Robinson, P.M. and P. Zaffaroni (2006), Pseudo Maximum Likelihood Estimation of ARCH(∞) Models, *Annals of Statistics*, 34, 1049-1074.