Econometric Theory, **23**, 2007, 1136–1161. Printed in the United States of America.
DOI: 10.1017/S0266466607070454

HIGHER ORDER ASYMPTOTIC THEORY WHEN A PARAMETER IS ON A BOUNDARY WITH AN APPLICATION TO GARCH MODELS

EMMA M. IGLESIAS
Michigan State University

OLIVER B. LINTON
London School of Economics

Andrews (1999, *Econometrica* 67, 1341–1383) derived the first-order asymptotic theory for a very general class of estimators when a parameter is on a boundary. We derive the second-order asymptotic theory in this setting in some special cases. We focus on the behavior of the quasi maximum likelihood estimator (QMLE) in stationary and nonstationary generalized autoregressive conditionally heteroskedastic (GARCH) models when constraints are imposed in the maximization procedure. We show how in this case both a first- and a second-order bias appear in the estimator and how the bias can be quite large. We provide two types of bias correction mechanisms for the researcher to choose in practice: either to bias correct only for a first-order bias or for a first- and second-order bias. We show that when some constraints are imposed, it is advisable to bias correct not only for the first-order bias but also for the second-order bias.

1. INTRODUCTION

There are a number of situations in econometrics where parameter values can lie on the boundary of a natural parameter space. Andrews (1999) presented the first-order asymptotic theory for a very general class of estimators when a parameter is on a boundary. Kim, Stone, and White (2000) applied the theory of Andrews (1999) to the case of a Sharpe "style regression," which is widely used in the investment industry. The methods they propose for conducting inference are quite complex.

We consider the issue of estimating parameters subject to inequality restrictions. We work with a special case where the inequality restricted estimators can be expressed in terms of a number of equality restricted estimators and so

We thank Bruce Hansen and two referees for helpful comments. The first author gratefully acknowledges financial support from the MSU Intramural Research Grants Program. The second author gratefully acknowledges financial support from the ESRC. Address correspondence to Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom; e-mail: o.linton@lse.ac.uk.

 © 2007 Cambridge University Press 0266-4666/07 $15.00

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

consequently inference is much easier. We develop the higher order theory for this case. We are specially interested in the consequences of being in the boundary region on the behavior of the bias and higher order moments of the estimator. We apply our results to the case of generalized autoregressive conditionally heteroskedastic (GARCH) specifications, where boundary values are quite problematic and are the subject of much recent research.

Engle (1982) and Bollerslev (1986) made very popular the use of ARCH (autoregressive conditionally heteroskedastic) and GARCH models. One of the most widely used estimation techniques for GARCH models is the QML (quasi maximum likelihood) procedure of Bollerslev and Wooldridge (1992). The asymptotic properties of this unrestricted estimator have already been shown under different assumptions in the literature (see, e.g., Weiss, 1984, 1986; Bollerslev and Wooldridge, 1992; Lee and Hansen, 1994; Lumsdaine, 1996; Newey and Steigerwald, 1997; Jeantheau, 1998; Comte and Lieberman, 2003; Ling and McAleer, 2003). The most recent contribution is by Jensen and Rahbek (2004a, 2004b), where they show the asymptotic normality of the QML estimator (QMLE) in both the stationary and the nonstationary regions. However, they still require positiveness in the parameter space. Forecasting is another reason to impose constraints in GARCH models. Hansen and Lunde (2005) present the GARCH(1,1) model as one of the best models available nowadays for prediction; however, Chong, Ahmad, and Abdullah (1999) also show the pitfalls of forecasting with GARCH models when the unconditional variance of the process does not exist. Also, in the special case of the GARCH(1,1) model, the positiveness of the parameter space seems to be not only a sufficient condition for a positive conditional variance but also a necessary requirement (see, e.g., Van Dijk, Franses, and Lucas, 1999), although for higher order GARCH models, more relaxed restrictions have been proposed (see, e.g., Doornik and Ooms, 2003). Those are the reasons why there are many papers nowadays that estimate ARCH and GARCH models with QML, some of them subject to constraints. Mainly, these are constraints to impose the stationarity condition of the process and/or that the conditional variance cannot reach negative values. Some recent examples of these papers where constraints are included both in univariate and multivariate frameworks are Tse (1998), Giot and Laurent (2003), and Fiorentini, Sentana, and Calzolari (2003). The study of finite-sample properties of the QMLE has been mainly developed so far for unconstrained estimation (e.g., Linton, 1997; Iglesias and Phillips, 2005). However, these papers do not show what happens when the researcher decides to estimate a GARCH model subject to constraints. Except for Andrews (1999), all the asymptotic theory that has been developed so far for GARCH models captures only the consequences of unconstrained estimation.

In this paper, we investigate constrained estimators. We present the higher order asymptotic theory that appears when GARCH models are estimated subject to constraints. We use this theory to provide advice and a bias correction mechanism to those researchers that follow this option in practice. The results

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

in this paper can be extended straightforwardly when exogenous variables are included in the mean equation (see Iglesias and Phillips, 2005).[1]

The plan of the paper is as follows. In Section 2 we start by presenting the general approach to computing higher order distributional approximations when some parameters are on the boundary. In Section 3 we discuss the application of the general theory to the two leading examples we consider, the ARCH(1) and the GARCH(1,1) models. In Section 4 we discuss the application of our expansions to inference and bias correction. Section 5 contains some simulation evidence. Section 6 concludes. We present some of the cumulant computations in the Appendix.

2. EDGEWORTH EXPANSIONS UNDER INEQUALITY RESTRICTIONS

Suppose that $\theta = (\theta_1, \theta_2, \ldots, \theta_n) \in \mathbb{R}^n$ are the unknown parameters that are subject to some inequality constraints. We write the constraints as

$$g_1(\theta) \geq 0, \ldots, g_p(\theta) \geq 0 \tag{1}$$

for some $p \leq n$, where the functions $g_j(.)$ are known. In vector notation we can write $g(\theta) \geq 0$. There may also be strict inequality restrictions and equality restrictions, but we shall not separately formalize these for notational simplicity. Each restriction g_j implies a subset Θ_j of $\mathbb{R}^n$, and collectively the parameter space becomes $\Theta = \cap_{j=1}^{p} \Theta_j$. A leading example is where $\theta_j \geq 0, j = 1, \ldots, p$, in which case $\Theta_j = \mathbb{R}_+ \times \cdots \times \mathbb{R}_+ \times \mathbb{R}^{n-p}$. Other examples such as $\sum_{j=1}^{n} \theta_j \leq 1$ are more complicated but fall under the general case laid out previously.

We suppose that there is an objective function $l_T(\theta)$ that depends on the data. Maximization of $l_T(\theta)$ over the (restricted) parameter space yields the estimator $\theta^* = (\theta_1^*, \theta_2^*, \ldots, \theta_n^*)$ that we analyze here. Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$ be the unrestricted estimator that maximizes $l_T(\theta)$ over $\mathbb{R}^n$. The estimator objective function can be the log-likelihood, the quasi log-likelihood, a least squares criterion function, a generalized method of moments (GMM) objective function, a minimum distance objective function, an objective function that depends on finite- or infinite-dimensional preliminary estimators, etc. (for more details, see Andrews, 1999). By definition, the extremum estimators θ^*, $\hat{\theta}$ satisfy

$$l_T(\theta^*) = \sup_{\theta \in \Theta} l_T(\theta); \qquad l_T(\hat{\theta}) = \sup_{\theta \in \mathbb{R}^n} l_T(\theta).$$

Our approach to solving the inequality constrained optimization problem is to look at all the subcases where one or more of the restrictions are binding; see Otten and Bams (2001). Suppose that $g_{j_1}(\theta) = 0, \ldots, g_{j_r}(\theta) = 0$ for some r and some $\{j_1, \ldots, j_r\}$; then we have a standard equality constrained optimization problem

$$\sup_{\theta : g_{j_1}(\theta) = 0, \ldots, g_{j_r}(\theta) = 0} l_T(\theta),$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

which can be solved by standard methods. We search over all such subcases and then find the parameter value that satisfies all inequality restrictions and that maximizes the criterion. In general there are $(2^p - 1)$ possible restricted problems to consider, which means that we can get $\tilde{\theta}^1, \ldots, \tilde{\theta}^{(2^p-1)}$ possible estimators subject to binding restrictions. Then the estimator over the set Θ, denoted $\theta^* = (\theta_1^*, \theta_2^*, \ldots, \theta_n^*)$, is whichever of $\hat{\theta}, \tilde{\theta}^1, \tilde{\theta}^2, \ldots, \tilde{\theta}^{(2^p-1)}$ satisfies the estimation criterion (the objective function is optimized) and the restrictions. There is a natural specific to general strategy that should be adopted here: start with the unrestricted problem; if the solution value satisfies all the inequality restrictions then stop. Then look at each of the $r = 1$ subproblems and check whether any of their solution values satisfies all the inequality restrictions. If at least one does, then one can stop and take whichever solution value maximizes $l_T(\theta)$. One proceeds through $r = 1, \ldots, p$ until a solution is found.[2] The estimator θ^* is thus a function of a general vector Z_T of standardized estimators $(\tilde{\theta}^1, \ldots, \tilde{\theta}^{(2^p-1)}, l_T(\tilde{\theta}^1), \ldots, l_T(\tilde{\theta}^{(2^p-1)}))$, i.e., $\theta^* = g(Z_T)$ for some g. Therefore, if the distribution of Z_T is known or can be approximated by some signed measure and if the mapping g is well behaved, then the distribution of θ^* or an approximation to it can be obtained by some manipulations. We will deal with cases where both of these conditions are met.

An important ingredient in our approximation is going to be the approximation to the distribution of Z_T. Consider a general vector $Z_T \in \mathbb{R}^d$ of standardized estimators that is asymptotically normal with variance matrix Ω and satisfies a joint Edgeworth expansion. Suppose that the first three mixed cumulants of Z_T satisfy

$$cum(Z_{Ti}) = \frac{c_i}{\sqrt{T}} + o(T^{-1/2}), \tag{2}$$

$$cum(Z_{Ti}, Z_{Tj}) = c_{ij} + o(T^{-1/2}), \tag{3}$$

$$cum(Z_{Ti}, Z_{Tj}, Z_{Tk}) = \frac{c_{ijk}}{\sqrt{T}} + o(T^{-1/2}) \tag{4}$$

for arrays of constants $\{c_i\}_{i=1}^d$, $\{c_{ij}\}_{i,j=1}^d$, $\{c_{ijk}\}_{i,j,k=1}^d$. Then for any Borel set B,

$$\Pr[Z_T \in B] = \int_B \phi_{0,\Omega}(y) \left\{ 1 + \sum_{i=1}^{d} \frac{c_i}{\sqrt{T}} H_i(y) + \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} \frac{c_{ijk}}{6\sqrt{T}} H_{ijk}(y) \right\} dy + o(T^{-1/2}), \tag{5}$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

where

$$\phi_{0,\Omega}(y) = \frac{1}{(2\pi)^{d/2}\det(\Omega)} \exp\left(-\frac{1}{2} y^{\top}\Omega^{-1} y\right),$$

$$H_i(y) = \frac{-1}{\phi_{0,\Omega}(y)} \frac{\partial}{\partial y_i} \phi_{0,\Omega}(y),$$

$$H_{ijk}(y) = \frac{-1}{\phi_{0,\Omega}(y)} \frac{\partial^3}{\partial y_i \partial y_j \partial y_k} \phi_{0,\Omega}(y).$$

The quantities $H_i(y)$ and $H_{ijk}(y)$ are the multivariate Hermite polynomials of first and third degree. See Taniguchi and Kakizawa (2000, p. 170) and Barndorff-Nielsen and Cox (1989, p. 174).[3] In the Appendix we give some discussion on how to compute the cumulant constants. Finally, to approximate $\Pr[\theta^* \in B^*]$ for Borel set B^* one computes the approximation to $\Pr[Z_T \in B]$ for the implied set B associated with the function g. The explanation of this step is best left to our examples that follow.

In some special cases the calculations involved simplify considerably. Specifically, when the restrictions are "nested" one can dispense with the likelihood functions in Z_T. Suppose that for each $r = 1, \ldots, p$ there is only one case; then one does not need to know the likelihood function value to conclude which value to take as there is a natural ordering just based on r. A trivial example is when there is only one restriction, such as the ARCH(1) process, but the GARCH(1,1) model is also of this type. In Section 3 we discuss these two examples in detail.

3. EXAMPLES

3.1. ARCH(1) Process

Suppose that

$$y_t = \varepsilon_t \sigma_t,$$

$$\sigma_t^2 = \theta_1 + \theta_3 y_{t-1}^2,$$

where ε_t is independent and identically distributed (i.i.d.) with mean zero and variance one, and let $\theta = (\theta_1, \theta_3) \in \mathbb{R}^2$ be the unknown parameters. We suppose that $\theta_1 > 0$ and $\theta_3 \geq 0$ (nonnegativity constraint).[4] The constraint that $\theta_1 > 0$ precludes the possibility that $\theta_1 = 0$ and so can never be binding. Let

$$\Theta = \{\theta : \theta_1 > 0, \theta_3 \geq 0\}$$

be the restricted parameter space. Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_3)$ be the unrestricted QMLE, i.e., the maximizer over $\mathbb{R}^2$ of the quasi log-likelihood. Let $\tilde{\theta} = (\tilde{\theta}_1, 0)$ be the

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

maximizer subject to the binding restriction that $\theta_3 = 0$. Then the maximizer over the set Θ, denoted θ^*, is whichever of $\hat{\theta}, \tilde{\theta}$ makes the likelihood bigger and satisfies the restrictions. This is a nested case because $l_T(\hat{\theta}) \geq l_T(\tilde{\theta})$, and so all that is required is to check whether $\hat{\theta}$ satisfies the restrictions. Thus θ^* is a mixture of the two parameters

$$\theta_3^* = \max\{\hat{\theta}_3, 0\} = \hat{\theta}_3 1(\hat{\theta}_3 \geq 0), \tag{6}$$

$$\theta_1^* = \hat{\theta}_1 1(\hat{\theta}_3 \geq 0) + \tilde{\theta}_1 1(\hat{\theta}_3 < 0). \tag{7}$$

These equations represent the inequality restricted estimator θ^* in terms of two equality restricted estimators $\tilde{\theta}, \hat{\theta}$. Thus the joint distribution of $\tilde{\theta}, \hat{\theta}$ determines the distribution of θ^*. In particular, by the law of total probability (conditioning on the event $\{\hat{\theta}_3 \geq 0\}$ and its complement),

$$\Pr[\theta_3^* \leq x] = \begin{cases} \Pr[\hat{\theta}_3 \leq x] & \text{if } x \geq 0 \\ 0 & \text{if } x < 0, \end{cases}$$

$$\Pr[\theta_1^* \leq x] = \begin{cases} \Pr[\hat{\theta}_1 \leq x, \hat{\theta}_3 \geq 0] + \Pr[\tilde{\theta}_1 \leq x, \hat{\theta}_3 < 0] & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases} \tag{8}$$

To approximate $\Pr[\theta_3^* \leq x]$ we only need the marginal Edgeworth expansion for the estimator $\hat{\theta}_3$, and this calculation is standard; see Linton (1997).[5] If θ is in the strict interior of the parameter space the probabilities like $\Pr[\hat{\theta}_3 \geq 0]$ increase exponentially to one, and the standard expansion applies to $\Pr[\theta_1^* \leq x]$. But when, e.g., $\theta_3 = 0$ this probability converges at a more modest pace to one-half. That expansion is no longer valid, and we must compute the more complicated expression given in (8). Therefore, in general to approximate $\Pr[\theta_1^* \leq x]$ we need to approximate $\Pr[\hat{\theta}_1 \leq x, \hat{\theta}_3 \geq 0]$ and $\Pr[\tilde{\theta}_1 \leq x, \hat{\theta}_3 < 0]$, which require the joint asymptotic expansions of $(\hat{\theta}_1, \hat{\theta}_3)$ and of $(\tilde{\theta}_1, \hat{\theta}_3)$. The calculations involved in obtaining the joint asymptotic expansion are quite similar to those for the marginal expansions and are discussed further in the Appendix.

Suppose that we have approximations Ψ_{Tj}, $j = 0, 1, 2$ (as described in (5)), such that

$$\sup_{x, y \in \mathbb{R}} |\Pr[\hat{\theta}_1 \leq x, \hat{\theta}_3 \leq y] - \Psi_{T1}(x, y)| = o(T^{-1}),$$

$$\sup_{x, y \in \mathbb{R}} |\Pr[\tilde{\theta}_1 \leq x, \hat{\theta}_3 \leq y] - \Psi_{T2}(x, y)| = o(T^{-1}),$$

$$\sup_{x \in \mathbb{R}} |\Pr[\hat{\theta}_3 \leq x] - \Psi_{T0}(x)| = o(T^{-1}).$$

Then let $\Psi_{T3}^*(x) = \Psi_{T0}(x)$ and

$$\Psi_{T1}^*(x) = 1 - \Psi_{T1}(x, 0) + \Psi_{T2}(x, 0) \tag{9}$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

for $x \geq 0$ and $\Psi^*_{T1}(x), \Psi^*_{T3}(x) = 0$ otherwise. It follows that $\Psi^*_{Tj}(x)$ are valid approximations to $\Pr[\theta^*_j \leq x]$ in the sense that

$$\sup_{x \in \mathbb{R}} |\Pr[\theta^*_j \leq x] - \Psi^*_{Tj}(x)| = o(T^{-1}). \tag{10}$$

This is true for all values of θ including the boundary ones.

To compute the approximate bias or skewness one can use and integrate the approximating measure. For example, an approximation to the bias of θ^*_1 is given by $\int_0^\infty x d\Psi^*_T(x) - \theta^*_1$; see Section 3.2.

To apply this general formula one just has to take different vectors Z_T and Borel sets B in (5). To compute $\Pr[\theta^*_1 \leq x]$ we need to apply (5) to the cases $Z_T = (\sqrt{T}(\hat{\theta}_1 - \theta_1), \sqrt{T}(\hat{\theta}_3 - \theta_3))^\top \in \mathbb{R}^2$ and $Z_T = (\sqrt{T}(\tilde{\theta}_1 - \theta_1), \sqrt{T}(\hat{\theta}_3 - \theta_3))^\top \in \mathbb{R}^2$. We also need to compute the probabilities $\Pr[\hat{\theta}_3 \geq 0]$ using the univariate Edgeworth expansion.

3.2. GARCH(1,1)

Suppose that

$$y_t = \varepsilon_t \sigma_t,$$

$$\sigma_t^2 = \theta_1 + \theta_2 \sigma_{t-1}^2 + \theta_3 y_{t-1}^2,$$

where ε_t is i.i.d. with mean zero and variance one. Let $\theta = (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3$ be the unknown parameters where $\theta_1 > 0$ and $\theta_2 \geq 0$, $\theta_3 \geq 0$ (but if $\theta_2 = 0$, then $\theta_3 = 0$). Because of this last qualification this is a nested case, and the restrictions are (we ignore the restriction that $\theta_1 > 0$)

$$g_1(\theta) = \theta_2 \geq 0, \qquad g_2(\theta) = \theta_3 1(\theta_2 > 0) \geq 0.$$

This is exactly the situation described in Jensen and Rahbek (2004a, 2004b), where the authors need the parameters to be positive. Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ be the unrestricted QMLE (other estimators may be analyzed using the same methodology), i.e., the maximizer over $\mathbb{R}^3$ of the quasi log-likelihood. Let $\tilde{\theta}^1 = (\tilde{\theta}^1_1, 0, \tilde{\theta}^1_3)$ be the maximizer subject to the binding restriction that $\theta_2 = 0$ and let $\tilde{\theta}^2 = (\tilde{\theta}^2_1, 0, 0)$ be the maximizer subject to the binding restrictions that $\theta_2 = 0$ and $\theta_3 = 0$. Then the maximizer over the set Θ, denoted θ^*, is whichever of $\hat{\theta}, \tilde{\theta}^1, \tilde{\theta}^2$ makes the likelihood bigger and satisfies the restrictions. Clearly, $l_T(\hat{\theta}) \geq l_T(\tilde{\theta}^1) \geq l_T(\tilde{\theta}^2)$, and so it amounts to just checking first whether $\hat{\theta}$ satisfies the restrictions, if not then checking whether $\tilde{\theta}^1$ satisfies the remaining restrictions, etc. Thus θ^* is a mixture of the three parameters. In fact

$$\theta^*_2 = \max\{\hat{\theta}_2, 0\} = \hat{\theta}_2 1(\hat{\theta}_2 \geq 0), \tag{11}$$

$$\theta^*_3 = \max\{\hat{\theta}_3, 0\} 1(\hat{\theta}_2 \geq 0) + \max\{\tilde{\theta}^1_3, 0\} 1(\hat{\theta}_2 < 0), \tag{12}$$

$$\theta^*_1 = \hat{\theta}_1 1(\hat{\theta}_2 \geq 0, \hat{\theta}_3 \geq 0) + \tilde{\theta}^1_1 1(\hat{\theta}_2 < 0, \tilde{\theta}^1_3 \geq 0) + \tilde{\theta}^2_1 1(\hat{\theta}_2 < 0, \tilde{\theta}^1_3 < 0). \tag{13}$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

A consequence of (11)–(13) is that

$$\Pr[\theta_2^* \le x] = \begin{cases} \Pr[\hat{\theta}_2 \le x] & \text{if } x \ge 0 \\ 0 & \text{if } x < 0, \end{cases}$$

$$\Pr[\theta_3^* \le x] = \begin{cases} \Pr[\hat{\theta}_3 \le x, \hat{\theta}_2 \ge 0] + \Pr[\tilde{\theta}_3^1 \le x, \hat{\theta}_2 < 0] & \text{if } x \ge 0 \\ 0 & \text{if } x < 0, \end{cases}$$

$$\Pr[\theta_1^* \le x] = \begin{cases} \Pr[\hat{\theta}_1 \le x, \hat{\theta}_2 \ge 0, \hat{\theta}_3 \ge 0] \\ \quad + \Pr[\tilde{\theta}_1^1 \le x, \hat{\theta}_2 < 0, \tilde{\theta}_3^1 \ge 0] \\ \quad + \Pr[\tilde{\theta}_1^2 \le x, \hat{\theta}_2 < 0, \tilde{\theta}_3^1 < 0] & \text{if } x \ge 0 \\ 0 & \text{if } x < 0. \end{cases}$$

We apply the same methodology as in the ARCH case.

4. INFERENCE AND BIAS CORRECTION

The preceding approximate distributions can be used when the parameters θ are known, say, under a null hypothesis. We now discuss how our expansions can be used for inference. The main issue here is that our distributional approximations are discontinuous in the parameters, e.g., $\Psi_{T3}^*(x;\theta)$ is not (uniformly) continuous in θ. This is of no consequence if the parameters are known, but when the parameters are unknown this complicates matters. For example, the natural estimator of the distribution $\Psi_{Tj}^*(x;\hat{\theta})$ does not well approximate $\Psi_{Tj}^*(x;\theta)$ when θ is on the boundary of the parameter space. This is why the standard inference methods do not apply. It follows that the obvious bias correction methods also do not work as usual. We argue that by using an alternative estimator of θ in $\Psi_{Tj}^*(x;\theta)$ one can obtain consistency. We focus on the bias correction issue for simplicity and present the main ideas for the case of the parameter θ_3 in the ARCH model.

From (5), the density of $\hat{\theta}_3$ is approximately (to $o(T^{-1})$)

$$f_T(y) = \sqrt{T}\phi_{0,\omega}\big(\sqrt{T}(y-\theta_3)\big) \times \left[1 + \frac{c_3}{\sqrt{T}} h_3\big(\sqrt{T}(y-\theta_3)\big) + \frac{c_{333}}{6\sqrt{T}} h_{333}\big(\sqrt{T}(y-\theta_3)\big)\right], \tag{14}$$

where ω is the asymptotic variance of $\sqrt{T}(\hat{\theta}_3 - \theta_3)$, whereas h_3, h_{333} are the one order higher Hermite polynomials corresponding to H_3, H_{333}; see Rothenberg (1984, (3.3)). Actually, $h_3(x) = 1$ and $h_{333}(x) = x^3 - 3x$. Therefore,

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

$$
\begin{aligned}
E[\theta_3^*] \simeq & \int_0^\infty y f_T(y)\,dy = \int_0^\infty y\sqrt{T}\,\phi_{0,\omega}\big(\sqrt{T}(y-\theta_3)\big) \\
& \times\left[1+\frac{c_3}{\sqrt{T}}\,h_3\big(\sqrt{T}(y-\theta_3)\big)+\frac{c_{333}}{6\sqrt{T}}\,h_{333}\big(\sqrt{T}(y-\theta_3)\big)\right]dy \\
= & \int_{-\theta_3\sqrt{T}}^\infty\left(\theta_3+\frac{1}{\sqrt{T}}\,y\right)\phi_{0,\omega}(y)\left[1+\frac{c_3}{\sqrt{T}}\,h_3(y)+\frac{c_{333}}{6\sqrt{T}}\,h_{333}(y)\right]dy \\
= & \,\theta_3\int_{-\theta_3\sqrt{T}}^\infty\phi_{0,\omega}(y)\left[1+\frac{c_3}{\sqrt{T}}\,h_3(y)+\frac{c_{333}}{6\sqrt{T}}\,h_{333}(y)\right]dy \\
& +\frac{1}{\sqrt{T}}\int_{-\theta_3\sqrt{T}}^\infty y\phi_{0,\omega}(y)\left[1+\frac{c_3}{\sqrt{T}}\,h_3(y)+\frac{c_{333}}{6\sqrt{T}}\,h_{333}(y)\right]dy.
\end{aligned}
$$

This approximation is valid to order T^{-1}, and the left-hand side is interpreted as an asymptotic moment because the actual moment may not exist. See Srinivasan (1970) for related discussion. We next specialize this general formula in the two cases.

When $\theta_3 = 0$, we have

$$
\begin{aligned}
E[\theta_3^*] = & \frac{1}{\sqrt{T}}\int_0^\infty y\phi_{0,\omega}(y)\left[1+\frac{c_3}{\sqrt{T}}\,h_3(y)+\frac{c_{333}}{6\sqrt{T}}\,h_{333}(y)\right]dy+o(T^{-1}) \\
= & \frac{1}{\sqrt{T}}\int_0^\infty y\phi_{0,\omega}(y)\,dy+\frac{c_3}{T}\int_0^\infty y\phi_{0,\omega}(y)h_3(y)\,dy \\
& +\frac{c_{333}}{6T}\int_0^\infty y\phi_{0,\omega}(y)h_{333}(y)\,dy+o(T^{-1}) \\
= & \frac{1}{\sqrt{T}}\int_0^\infty y\phi_{0,\omega}(y)\,dy+\frac{c_3}{T}\int_0^\infty y\phi_{0,\omega}(y)h_3(y)\,dy+o(T^{-1})
\end{aligned}
$$

because $\int_0^\infty y\phi_{0,\omega}(y)h_{333}(y)\,dy = \int_{-\infty}^\infty y\phi_{0,\omega}(y)h_{333}(y)\,dy/2 = 0$. When $\theta_3 > 0$

$$
\begin{aligned}
E[\theta_3^*] = & \,\theta_3\int_{-\theta_3\sqrt{T}}^\infty\phi_{0,\omega}(y)\left[1+\frac{c_3}{\sqrt{T}}\,h_3(y)+\frac{c_{333}}{6\sqrt{T}}\,h_{333}(y)\right]dy \\
& +\frac{1}{\sqrt{T}}\int_{-\theta_3\sqrt{T}}^\infty y\phi_{0,\omega}(y)\left[1+\frac{c_3}{\sqrt{T}}\,h_3(y)\right. \\
& \qquad\left.+\frac{c_{333}}{6\sqrt{T}}\,h_{333}(y)\right]dy+o(T^{-1}) \\
\simeq & \,\theta_3+\frac{c_3}{T}\int_{-\theta_3\sqrt{T}}^\infty y\phi_{0,\omega}(y)h_3(y)\,dy+o(T^{-1}).
\end{aligned}
$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

Thus when $\theta_3 > 0$ the bias is of order T^{-1}, whereas when $\theta_3 = 0$, the bias is larger and of order $T^{-1/2}$ with a second-order term of order T^{-1}.

To use this for bias correction one should replace the bias $b_T(\theta)$ by the estimated bias

$$b_T(\theta^*) = \theta_3^* \int_{-\infty}^{-\theta_3^*\sqrt{T}} \phi_{0,\omega^*}(y)\,dy + \theta_3^* \int_{-\theta_3^*\sqrt{T}}^{\infty} \phi_{0,\omega^*}(y) \left[\frac{c_3^*}{\sqrt{T}}\, h_3(y) + \frac{c_{333}^*}{6\sqrt{T}}\, h_{333}(y) \right] dy + \frac{1}{\sqrt{T}} \int_{-\theta_3^*\sqrt{T}}^{\infty} y\phi_{0,\omega^*}(y) \left[1 + \frac{c_3^*}{\sqrt{T}}\, h_3(y) + \frac{c_{333}^*}{6\sqrt{T}}\, h_{333}(y) \right] dy, \quad \textbf{(15)}$$

where $\omega^* = \omega(\theta^*)$, $c_3^* = c_3(\theta^*)$, and $c_{333}^* = c_{333}(\theta^*)$. Unfortunately, this does not work when the parameter is on the boundary; specifically, $\sqrt{T}\,b_T(\theta^*)$ converges to a random limit in that case (because $\theta_3^*\sqrt{T}$ has a nondegenerate limiting distribution). Instead we shall follow Andrews (2000) and work with a modified estimator that avoids this problem.[6] Let η_T be a sequence of numbers that satisfies $\eta_T \to 0$ and $\liminf_{T\to\infty} \eta_T\sqrt{T/(2\ln\ln T)} > 1$ and define

$$\theta_3^{**} = \hat{\theta}_3 1(\hat{\theta}_3 > \eta_T).$$

Then define the estimated bias $b_T(\theta^{**})$ as in (15). It can be shown that

$$b_T(\theta^{**}) - b_T(\theta) = o_p(T^{-1}) \quad \textbf{(16)}$$

for all $\theta_3 \geq 0$. The reason for this is that when $\theta_3 = 0$, $T^r\hat{\theta}_3 1(\hat{\theta}_3 > \eta_T) = o_p(1)$ for any $r > 0$ because $\Pr[\hat{\theta}_3 > \eta_T] \to 0$. Therefore

$$T\theta_3^{**} \int_{-\infty}^{-\theta_3^{**}\sqrt{T}} \phi_{0,\omega^*}(y)\,dy = o_p(1)$$

$$\sqrt{T} \int_{-\theta_3^*\sqrt{T}}^{0} y\phi_{0,\omega^*}(y)\,dy \simeq T\theta_3^*\,\phi_{0,\omega^*}\left(-\theta_3^*\sqrt{T}\right) = o_p(1)$$

and similarly for the other terms. When $\theta_3 > 0$, $b_T(\theta) = O(T^{-1})$, and the estimated version $b_T(\theta^{**})$ is of smaller order in probability. It follows that the additive and two alternative multiplicative bias corrected estimators $\theta_{abc}^* = \theta^* - b_T(\theta^{**})$, $\theta_{mbc1}^* = \theta^*/(1 + b_T(\theta^{**})/\theta^*)$, and $\theta_{mbc2}^* = \theta^* \exp(-b_T(\theta^{**})/\theta^*)$ (where $0/0 = 0$) have bias of order $o(T^{-1})$. The exponential bias corrected estimator may be preferred because it is always nonnegative, whereas the additive bias corrected estimator can be negative and hence violates the restrictions we seek to impose.

The generalization of the bias correction mechanism to the multidimensional case implies retrieving the bias function $b_T(.)$ from (5) and again using a mod-

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

ified estimator to apply in (16). For example, in the GARCH(1,1) case where we impose the positiveness constraint in all the parameters and again using the notation of Section 3, we define the bias correction using the shrinkage estimates: $\theta_2^{**} = \hat{\theta}_2 1(\hat{\theta}_2 > \eta_T)$, $\theta_3^{**} = \hat{\theta}_3 1(\hat{\theta}_3 > \eta_T)1(\hat{\theta}_2 > \eta_T) + \tilde{\theta}_3 1(\tilde{\theta}_3 > \eta_T)1(\hat{\theta}_2 < \eta_T)$, and $\theta_1^{**} = \hat{\theta}_1 1(\hat{\theta}_2 > \eta_T, \hat{\theta}_3 > \eta_T) + \tilde{\theta}_1^1 1(\hat{\theta}_2 < \eta_T, \tilde{\theta}_3^1 > \eta_T) + \tilde{\theta}_1^2 1(\hat{\theta}_2 < \eta_T, \tilde{\theta}_3^1 < \eta_T)$.

In conclusion, we have shown how to correct the constrained estimator for bias, first and second order. Our method is based on analytical calculations and use of the shrinkage estimator θ^{**}. An alternative would be the bootstrap methods proposed in Andrews (2000). Our approach works also for distributional approximation, so that $\Psi_{Tj}^*(x;\theta^{**})$ approximates $\Psi_{Tj}^*(x;\theta)$ to the required order uniformly over x. Note however that as in Andrews (2000) the result (16) is a pointwise result and does not extend to uniformity, e.g., to the sequence of true parameters converging to zero.

5. SIMULATIONS

We now present simulation results to analyze the consequences of imposing restrictions in the GARCH(1,1) model and later also in the ARCH(1) model. Table 1 shows the results for 40,000 replications with a sample size of 500 observations, and it shows the bias of the estimates both when we restrict and when we do not restrict the parameter space to be in the positive region in the GARCH(1,1) model; $\varepsilon_t \sim N(0,1)$. We show the bias results without and with the positiveness restrictions when estimating by QML for the (θ_2, θ_3) parameter values equal to $(0.4, 0.5)$, $(0.1, 0.5)$, $(0.1, 0.1)$, $(0.8, 0.1)$, $(0.0, 0.0)$. The biases are free of the value of θ_1, which has been set to 0.81 in all simulations.[7] As is

TABLE 1. Bias results for the GARCH(1,1) model

	No restriction bias	$\theta_1 > 0, \theta_2 \geq 0, \theta_3 \geq 0$ bias
$\theta_2 = 0.4$	−0.011	−0.011
$\theta_3 = 0.5$	−0.003	−0.002
$\theta_2 = 0.1$	0.015	0.017
$\theta_3 = 0.5$	−0.056	−0.057
$\theta_2 = 0.1$	0.148	0.150
$\theta_3 = 0.1$	−0.010	−0.009
$\theta_2 = 0.8$	−0.095	−0.090
$\theta_3 = 0.1$	−0.009	−0.008
$\theta_2 = 0.0$	0.316	0.326
$\theta_3 = 0.0$	0.015	0.016

Note: Sample size = 500; 40,000 replications.

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

already stated in Linton (1997) (he only analyzed the $O(T^{-1})$ bias), the bias increases significantly when θ_3 is small. In practical applications of GARCH models, it is very common that the $\hat{\theta}_2$ value is large, *while* $\hat{\theta}_3$ *is small*, and so it is important that the researcher be aware that mainly this situation is the one that is very likely to generate relatively large biases. The biases are especially large when both θ_2 and θ_3 are small. The restriction of positiveness slightly increases the bias in general, although less than happens for the ARCH(1) model (as we will see in the next section). The results clearly justify that, both when restrictions are imposed and when they are not, the estimates need a bias correction mechanism in practice. When the true parameter value is on the boundary, the bias increases significantly mainly for $\hat{\theta}_2$.

If we are only interested in correcting the first-order bias, e.g., if the restriction is the positiveness of the parameter space, the first-order bias term (in a standardized case) is simply given by $E(\max(Z,0))$ for $Z \sim N(0,1)$ (according to the theory of Andrews, 1999), whereas if the restriction is to impose that the parameter space is less than 1, then the first-order bias is given by $E(\min(Z,1))$ for $Z \sim N(0,1)$. However, to consider the second-order bias, we need to apply the theory developed in Section 2 to the GARCH(1,1) and the ARCH(1) models. This is done in the Appendix.

We present in Table 2 simulation results to find out about the consequences of imposing restrictions in the estimation of an ARCH(1) model. Table 2 presents results for the bias; with an asterisk, we also provide the standard errors of the estimators, and with a double asterisk we show the skewness. We present the results in five different scenarios: when we do not restrict the estimator, constraining $\theta_3 \geq 0$, constraining $\theta_3 \leq 3.562$ (strict-stationarity condition in the ARCH(1) model), constraining $\theta_3 \leq 1$ (weak stationarity), and $0 \leq \theta_3 \leq 1$. We show the results for different combinations of true values: $(\theta_1, \theta_3) = (1, 0.9)$, $(\theta_1, \theta_3) = (1,1)$, $(\theta_1, \theta_3) = (1,2)$, $(\theta_1, \theta_3) = (1,3)$, and $(\theta_1, \theta_3) = (1,0)$. In this way we analyze the behavior of the ARCH(1) model in both the stationary and nonstationary regions (see Jensen and Rahbek, 2004a, 2004b, where they prove that the QMLE is still consistent and asymptotically normal in the nonstationary area) and also when a restriction is binding or not.

If the true parameter value is 0.9, the fact of having a bias of -0.17 (when we constrain $0 \leq \theta_3 \leq 1$) is an important problem (when the researcher estimates with that constraint in practice), and therefore, to bias correct is quite relevant. In the same way, when $\theta_3 = 2$, to impose the weak stationarity constraint wrongly can have very important consequences. It is true that under restrictions, the standard error is reduced (as is shown in Table 2), although we have to judge that together with the fact that the bias in the restricted estimator is amazingly large. In summary, restrictions can affect significantly the estimates in this context. We also show in bold the behavior of the skewness, the standard errors, and the biases when the restrictions are binding. In Table 2, we show that the biases can increase very significantly: e.g., when the true $\theta_3 = 1$ and we impose the constraint of $0 \leq \theta_3 \leq 1$, the bias reaches the value -0.216

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

TABLE 2. Simulations of bias, standard errors*, and skewness** for the ARCH(1) model

True values	No constraints	$\theta_3 \geq 0$	$\theta_3 \leq 3.562$	$\theta_3 \leq 1$	$0 \leq \theta_3 \leq 1$
$\theta_1 = 1$	0.095	0.088	0.090	0.092	0.087
	0.44*, 1.27**	0.43*, 1.18**	0.43*, 1.32**	0.44*, 0.99**	0.43*, 1.19**
$\theta_3 = 0.9$	−0.094	−0.086	−0.093	−0.165	−0.166
	0.39*, 0.09**	0.38*, 0.15**	0.39*, 0.03**	0.29*, −1.00*	0.28*, −0.89**
$\theta_1 = 1$	0.106	0.098	0.101	0.099	0.101
	0.45*, 1.33**	0.45*, 1.34**	0.45*, 1.65**	0.47*, 1.29**	0.45*, 1.13**
$\theta_3 = 1$	−0.096	−0.090	−0.099	**−0.215**	**−0.216**
	0.41*, 0.08**	0.41*, 0.08**	0.41*, 0.12**	**0.27***, **2.26****	0.27*, **−1.19****
$\theta_1 = 1$	0.239	0.215	0.262	0.261	0.287
	1.21*, 16.18**	1.31*, 17.32**	2.17*, 12.18**	1.50*, 42.62**	2.02*, 29.91**
$\theta_3 = 2$	−0.110	−0.108	−0.112	**−1.018**	**−1.019**
	0.58*, −0.01**	0.58*, −0.004**	0.59*, −0.08**	**0.09***, **−6.90****	0.10*, **−6.55****
$\theta_1 = 1$	1.364	1.223	1.634	1.025	1.711
	16.17*, 37.11**	14.77*, 63.58**	38.62*, 54.28**	7.28*, 64.53**	54.02*, 64.95**
$\theta_3 = 3$	−0.127	−0.109	−0.200	**−2.003**	**−2.003**
	0.76*, 0.06**	0.003*, −49.56**	0.63*, −0.68**	**0.05***, **−20.65****	0.04*, **−23.43****
$\theta_1 = 1$	0.014	0.015	0.010	0.014	0.009
	0.26*, 0.42**	0.26*, 0.36**	0.26*, 0.39**	0.26*, 0.41**	0.25*, 0.42**
$\theta_3 = 0$	−0.008	**0.058**	−0.007	−0.008	**0.055**
	0.16*, 1.330**	**0.11*, 2.94****	0.16*, 1.54**	0.15*, 1.22**	0.11*, 2.91**

Note: Sample size = 50; 10,000 replications.

(it is a bias of around 20% of the true parameter value). Also, the skewness increases very significantly when the true parameter touches the boundary of the restriction. Finally, we also show how the variance of the θ_1 estimator increases significantly when restrictions are imposed in the estimation of θ_3 and when we are in the nonstationary area.

Now that we know from Tables 1 and 2 how the restrictions can affect the bias, standard error, and skewness of the estimators, we want to find out how our theory is able to correct for the problem. In Table 3, we show the simulated second-order bias of the estimator with one asterisk, and with a double asterisk we show the value that our theory predicts and that should match the second-order bias of the estimator. Note that for the bias of $\hat{\theta}_1$ and $\hat{\theta}_3$ we use the expressions of the bias given in the Appendix in equation (A.2). Because the bias expressions in the ARCH(1) model involve summations from 1 up to the sample size, we follow the recommendation in Iglesias and Phillips (2005), where it is shown that if we truncate those summations in 8, they already provide a very good approximation of the true bias. Our theory explains and predicts the movement of the true second-order bias, and it accounts for an important portion of its amount (see Table 3) mainly for $\hat{\theta}_3$. We consider again the estimates under no constraints and with the positiveness constraint under (θ_1, θ_3) parameter values equal to $(1, 0.9)$, $(1, 1)$, $(1, 2)$, $(1, 3)$. We show that it is especially

TABLE 3. Simulated bias* and $O(T^{-1})$ bias approximations** for the ARCH(1) model

True values	No constraints	$\theta_3 \geq 0$
$\theta_1 = 1$	0.095*	0.088*
	0.072**	0.059**
$\theta_3 = 0.9$	−0.094*	−0.086*
	−0.080**	−0.064**
$\theta_1 = 1$	0.106*	0.098*
	0.072**	0.060**
$\theta_3 = 1$	−0.096*	−0.090*
	−0.082**	−0.079**
$\theta_1 = 1$	0.239*	0.215*
	0.174**	0.126**
$\theta_3 = 2$	−0.110*	−0.108*
	−0.083**	−0.061**
$\theta_1 = 1$	1.364*	1.223*
	0.332**	0.286**
$\theta_3 = 3$	−0.127*	−0.109*
	−0.084**	−0.054**

Note: Sample size = 50; 10,000 replications.

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

TABLE 4. First-order bias, total simulated bias*, first-order skewness**, and total simulated skewness*** for the ARCH(1) model

First case		
True values	No constraints	Restriction: $\theta_3 \geq 0$
$\theta_1 = 1$	0, 0.014*	0, 0.015*
	0**, 0.417***	0**, 0.360***
$\theta_3 = 0$	0, −0.008*	0.055, 0.058*
	0**, 1.330***	0.440**, 2.940***
Second case		
True values	No constraints	Restriction: $\theta_3 \leq 1$
$\theta_1 = 1$	0, 0.106*	0, 0.099*
	0**, 1.330***	0**, 1.290***
$\theta_3 = 1$	0, −0.096*	−0.032, −0.215*
	0**, 0.081***	−0.203**, 2.260***

Note: Sample size = 50; 10,000 replications.

interesting how when the true value of $\theta_3 = 1$, without the bias correction mechanism (because of the heavy downward bias of -0.09) we would conclude that the disturbance is weakly stationary, whereas in fact it contains a unit root. That means that using only asymptotic theory can have dangerous consequences in the interpretation of the estimates in some circumstances.

Finally, in Table 4 we show the simulated first-order bias of the estimator (what the theory of Andrews, 1999, predicts) and the total bias when the parameter is on a boundary. The design of the experiment includes the (θ_1, θ_3) parameter values equal to $(1,0)$, $(1,1)$. We find that for some restrictions, the Andrews (1999) theory provides a very good approximation (when the restriction is $\theta_3 \geq 0$), whereas for another restriction (for imposing weak stationarity such that $\theta_3 \leq 1$), the first-order approximation again in a standardized case (coming from $E(\min(Z,1))$ where $Z \sim N(0,1)$) is not very accurate, and higher order refinements are very advisable. We also show the total simulated skewness and the first-order approximation for the skewness that the Andrews (1999) theory would predict. The skewness term comes in the first case of Table 4 from the term $E[(\max(Z,0) - E(\max(Z,0)))^3]$; and for the second restriction from $E[(\min(Z,1) - E(\min(Z,1)))^3]$. We see that the first-order asymptotic theory for the skewness does not offer a good approximation in the case of the first restriction or for the second one.

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

5.1. Inference and Bias Correction

We proceed now to show in practice how our procedure of Section 4 corrects the constrained estimator for bias, first and second order. As we have pointed out, our method is based on analytical calculations and use of the shrinkage estimator θ^{**}. We consider the setting of an ARCH(1) process in (6), with a sample size of 50 observations and 10,000 replications; $\varepsilon_t \sim N(0,1)$. We show the results of our uncorrected and bias corrected estimators using robust estimates in terms of median and interquartile range (because the approximations are valid in probability). We also consider two restrictions that the applied researcher may consider in practice: (1) a positiveness constraint and (2) a constraint to impose a finite unconditional variance.

If we are in a setting where $\theta_1 = 1$ and $\theta_3 = 0$, and we estimate subject to the constraint

$$\theta_3^* = \max\{\hat{\theta}_3, 0\} = \hat{\theta}_3 1(\hat{\theta}_3 \geq 0),$$

then the uncorrected estimator θ_3^* has a bias (see Table 2) of 0.058 and a standard error of 0.111. Therefore, we consider a shrinkage estimator, θ_3^{**}, that we use for bias correction of θ_3^*:

$$\theta_3^{**} = \hat{\theta}_3 1(\hat{\theta}_3 > \eta_T),$$

where, for η_T, we have used a rate of $T^{-1/4}$ and scaled by the square root of the variance of the unrestricted estimator. Therefore, following our methodology in Section 4, we construct three types of bias corrected estimators of θ_3^*: $\theta_{3,abc}^*$, $\theta_{3,mbc1}^*$, and $\theta_{3,mbc2}^*$.

We repeated the same experiment with the second constraint where now $\theta_1 = \theta_3 = 1$ and

$$\theta_3^* = \min\{\hat{\theta}_3, 1\}.$$

The bias and standard error of θ_3^* (see Table 2) are -0.215 and 0.269, respectively. We construct again a shrinkage estimator of the type

$$\theta_3^{**} = \begin{Bmatrix} \hat{\theta}_3 & \text{if } \hat{\theta}_3 < 1 + \eta_T \\ 1 & \text{if } \hat{\theta}_3 > 1 + \eta_T \end{Bmatrix},$$

and we obtain the three bias corrected estimators. Table 5 shows the median and interquartile range of θ_3^* (the uncorrected estimator) and the three bias corrected estimators under the two constraints. We can observe how under constraint 1, the additive bias corrected estimator can reach negative values and therefore it distorts its median and interquartile range in relation to the uncorrected estimator θ_3^*. Out of the three bias corrected estimators, $\theta_{3,mbc2}^*$ is the one that offers the best interquartile range at the same time that is median unbiased. The estimator $\theta_{3,mbc2}^*$ is clearly preferred to the uncorrected estimator θ_3^*. Under constraint 2, the three bias corrected estimators are median un-

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

TABLE 5. Median and interquartile range of θ_3^*, $\theta_{3,abc}^*$, $\theta_{3,mbc1}^*$, and $\theta_{3,mbc2}^*$

Estimating subject to constraint 1: $\theta_3^* = \hat{\theta}_3 1(\hat{\theta}_3 \geq 0)$				
	$\theta_3 = 0$		$\theta_3 = 0.5$	
Estimator	Median	Interquartile range	Median	Interquartile range
θ_3^*	0	0.074	0.417	0.412
$\theta_{3,abc}^*$	−0.045	0.120	0.482	0.474
$\theta_{3,mbc1}^*$	0	0.052	0.507	0.511
$\theta_{3,mbc2}^*$	0	0.051	0.494	0.490
Estimating subject to constraint 2: $\theta_3^* = \min\{\hat{\theta}_3, 1\}$				
	$\theta_3 = 1$		$\theta_3 = 0.5$	
Estimator	Median	Interquartile range	Median	Interquartile range
θ_3^*	0.908	0.361	0.423	0.409
$\theta_{3,abc}^*$	0.983	0.364	0.503	0.409
$\theta_{3,mbc1}^*$	0.987	0.371	0.538	0.370
$\theta_{3,mbc2}^*$	0.996	0.356	0.522	0.399

Note: The conclusions from the simulation results when we set $\theta_3 = 0.05$ are very similar to those of $\theta_3 = 0$.

biased, whereas the uncorrected estimator θ_3^* is not. Again, $\theta_{3,mbc2}^*$ is clearly the preferred estimator in terms of median and interquartile range. Therefore, from our simulations, we advise using θ_{mbc2}^* in practice.

To find out about the behavior of our procedure when the true parameter is outside the boundary, we have run simulations when the true $\theta_1 = 1$ and $\theta_3 = 0.5$. Under the positiveness constraint, θ_3^* presents a negative bias of -0.053 and a standard error of 0.294. When we check the performance of our bias corrected estimators, again $\theta_{3,mbc2}^*$ is the estimator that presents the best median, even though it increases the interquartile range a bit in relation to θ_3^*. Finally, when we check the performance when $\theta_3 = 0.5$ under the second of the constraints, the bias of θ_3^* is -0.074 with a standard error of 0.288. Again, our bias corrected estimators offer an important improvement in the median, with $\theta_{3,mbc2}^*$ being the one that presents both the best median and interquartile range. Therefore, overall, from our simulations, we advise again using θ_{mbc2}^* in practice.

6. CONCLUSIONS

In this paper we obtained higher order distributional expansions when some parameters may lie on the boundary of the parameter space. We specialized our

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

results to find out about the bias behavior of the QMLE in GARCH and ARCH models when constraints are imposed in the maximization procedure. We show that practitioners that use these models should consider the application of a bias correction mechanism because of the large biases created both when constraints are imposed and when they are not. We provide a very simple mechanism to bias correct for the first-order term and a procedure to bias correct for the first- and second-order bias in case the researcher chooses this last option. In some cases, a first-order bias correction is not enough, and a first- and second-order bias correction is advisable. We note that our methods are based on analytic calculation and use of the Hodges shrinkage estimator. This approach is simple and computationally convenient and may work better than the subsampling or corrected bootstrap approaches suggested in Andrews (2000). On the other hand we are also subject to the critique of Leeb and Pötscher (2003) that our estimators do not perform uniformly well across the parameter space.

NOTES

1. Note that our results are very easy to apply in any model provided that we have analytical expressions for the higher order biases. Linton (1997) provides the higher order bias of the QMLE in the GARCH(1,1) model, and Iglesias and Phillips (2005) generalize that result when we include in the GARCH model any number of exogenous variables in the mean equation. Therefore, we can use their expressions directly in this case. Other specifications (such as asymmetric GARCH or other nonlinear models) can be directly analyzed with our approach provided that first we find analytically the higher order biases.

2. Note that our search method is a sensible algorithm to use when the number of constraints is small. However, if the number of constraints is large, a search over all subcases can be very computationally expensive. More efficient methods can be used in this case (see, e.g., the ones available in the numerical optimization literature such as in Nocedal and Wright, 1999).

3. The theoretical validity of the approximations for smooth functions of sample moments has been established by Sargan (1974, 1976) and Phillips (1977b).

4. The case of weak stationarity where $\theta_3 \leq 1$ can be treated identically.

5. This approximation can be justified by the results of Götze and Hipp (1983) under conditions on the parameters that ensure that the process is stationary and strong mixing with geometric decay.

6. Actually, Andrews (2000) proposes bootstrap and subsampling methods based on this idea rather than analytic correction.

7. As Lumsdaine (1995) showed, the QMLEs of θ_2 and θ_3 are invariant to changes in θ_1.

REFERENCES

Andrews, D.W.K. (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 1341–1383.

Andrews, D.W.K. (2000) Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68, 399–405.

Barndorff-Nielsen, O.E. & D.R. Cox (1989) *Asymptotic Techniques for Use in Statistics*. Chapman and Hall.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

Bollerslev, T. & J.M. Wooldridge (1992) Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews* 11, 143–172.

Chong, C.W., M.I. Ahmad, & M.Y. Abdullah (1999) Performance of GARCH models in forecasting stock market volatility. *Journal of Forecasting* 18, 333–343.

Comte, F. & O. Lieberman (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.

Cox, D.R. & P. Hall (2002) Estimation in a simple random effects model with nonnormal distributions. *Biometrika* 89, 831–840.

Doornik, J.A. & M. Ooms (2003) Multimodality in the GARCH regression model. Economics Working papers 2003-W20, Nuffield.

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

Fiorentini, G., E. Sentana, & G. Calzolari (2003) Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student *t* innovations. *Journal of Business & Economic Statistics* 21, 532–546.

Giot, P. & S. Laurent (2003) Value at risk for long and short trading positions. *Journal of Applied Econometrics* 18, 641–664.

Götze, F. & C. Hipp (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64, 311–239.

Hansen, P.R. & A. Lunde (2005) A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20, 873–889.

Iglesias, E.M. & G.D.A. Phillips (2005) Small Sample Estimation Bias in GARCH Models with Any Number of Exogenous Variables in the Mean Equation. Working paper, Michigan State University.

Jeantheau, T. (1998) Strong consistency of estimators of multivariate ARCH models. *Econometric Theory* 14, 70–86.

Jensen, S.T. & A. Rahbek (2004a) Asymptotic normality of the QMLE of ARCH in the nonstationary case. *Econometrica* 72, 641–646.

Jensen, S.T. & A. Rahbek (2004b) Asymptotic inference for nonstationary GARCH. *Econometric Theory* 20, 1203–1226.

Kim, T.-H., D. Stone, & H. White (2000) Asymptotic and Bayesian Confidence Intervals for Sharpe Style Weights. Discussion paper 2000–27, University of California at San Diego.

Lee, S.-W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Leeb, H. & B. Pötscher (2003) Can one estimate the unconditional distribution of post-model-selection estimators? Cowles Foundation Discussion Paper 1444, Yale University.

Ling, S. & M. McAleer (2003) Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 278–308.

Linton, O. (1997) An asymptotic expansion in the GARCH(1,1) model. *Econometric Theory* 13, 558–581.

Lumsdaine, R.L. (1995) Finite sample properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models: A Monte Carlo investigation. *Journal of Business & Economic Statistics* 13, 1–10.

Lumsdaine, R.L. (1996) Asymptotic properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. *Econometrica* 64, 575–596.

Newey, W.K. & D.G. Steigerwald (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65, 587–599.

Nocedal, J. & S.J. Wright (1999) *Numerical Optimization*. Springer-Verlag.

Otten, R. & D. Bams (2001) Statistical tests for return-based style analysis. Paper presented at the EFMA meetings, Lugano.

Phillips, P.C.B. (1977a) Approximations to some finite sample distributions associated with a first-order stochastic difference equation. *Econometrica* 45, 463–485.

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

Phillips, P.C.B. (1977b) A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators. *Econometrica* 45, 1517–1534.
Phillips, P.C.B. (1978) Edgeworth and saddle point approximations in the first-order noncircular autoregression. *Biometrika* 65, 91–98.
Rothenberg, T.J. (1984) Approximating the distributions of econometric estimators and test statistics. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, 881–935. North-Holland.
Sargan, J.D. (1974) On the validity of Nagar's expansion for the moments of econometric estimators. *Econometrica* 42, 169–176.
Sargan, J.D. (1976) Econometric estimators and the Edgeworth expansion. *Econometrica* 44, 421–448; erratum, 45, 272.
Srinivasan, T.N. (1970) Approximations to finite sample moments of estimators whose exact sampling distributions are unknown. *Econometrica* 38, 533–541.
Taniguchi, M. & Y. Kakizawa (2000) *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag.
Tse, Y.K. (1998) The conditional heteroscedasticity of the yen-dollar exchange rate. *Journal of Applied Econometrics* 13, 49–55.
Van Dijk, D., P.H. Franses, & A. Lucas (1999) Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics* 14, 539–562.
Weiss, A.A. (1984) ARMA models with ARCH errors. *Journal of Time Series Analysis* 5, 129–143.
Weiss, A.A. (1986) Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2, 107–131.

APPENDIX

We discuss here the computation of the cumulant constants $\{c_i\}_{i=1}^d$, $\{c_{ij}\}_{i,j=1}^d$, $\{c_{ijk}\}_{i,j,k=1}^d$ in (5). The complete set of calculations was carried out using computer algebra carried out by a program developed by the first author; here we just give some heuristics to indicate some issues.

Suppose that

$$\sqrt{T}(\hat{\theta}_j - \theta_j) = X_j + \frac{A_j B_j}{\sqrt{T}} + o_p(T^{-1/2}), \tag{A.1}$$

where $(X_j, A_j, B_j), j = 1, \ldots, d$ are jointly asymptotically normal or sums of jointly asymptotically normal random variables. The first cumulants $\{c_i\}_{i=1}^d$ are obtained from the univariate expansions. The second cumulants $\{c_{ij}\}_{i,j=1}^d$ are just the elements of the asymptotic covariance matrix of Z_T. To calculate the third cumulants we have two cases. First,

$$cum(Z_{Tj}, Z_{Tj}, Z_{Tj}) = E\left[\left(X_j + \frac{\{A_j B_j - E(A_j B_j)\}}{\sqrt{T}}\right)^3\right]$$

$$\simeq E[X_j^3] + 3\,\frac{E[X_j^2\{A_j B_j - E(A_j B_j)\}]}{\sqrt{T}},$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

which are the standard cumulants from the univariate expansions. However, we also need

$$cum(Z_{Ti}, Z_{Ti}, Z_{Tj}) = E\left[\left(X_i + \frac{\{A_i B_i - E(A_i B_i)\}}{\sqrt{T}}\right)^2 \left(X_j + \frac{\{A_j B_j - E(A_j B_j)\}}{\sqrt{T}}\right)\right]$$

$$\simeq E[X_i^2 X_j] + \frac{E[X_i^2\{A_j B_j - E(A_j B_j)\}]}{\sqrt{T}}$$

$$+ \frac{2E[X_i X_j\{A_i B_i - E(A_i B_i)\}]}{\sqrt{T}},$$

$$cum(Z_{Ti}, Z_{Tj}, Z_{Tk}) = E\left[\prod_{r=i,j,k}\left(X_r + \frac{\{A_r B_r - E(A_r B_r)\}}{\sqrt{T}}\right)\right]$$

$$\simeq E[X_i X_j X_k] + \frac{E[X_j X_k\{A_i B_i - E(A_i B_i)\}]}{\sqrt{T}}$$

$$+ \frac{E[X_i X_k\{A_j B_j - E(A_j B_j)\}]}{\sqrt{T}} + \frac{E[X_i X_j\{A_k B_k - E(A_k B_k)\}]}{\sqrt{T}}.$$

This involves cross-product terms that depend on the joint asymptotic normal distribution of (X, A, B) and on the higher order behavior of X. Note that X_i are sums of (stationary) martingale difference sequences, whereas A, B are not. This means that the asymptotic distribution of A, B involves double sums or long-run variances (see Linton, 1997), whereas the asymptotic distribution of X is much simpler.

Letting $X_i = T^{-1/2}\sum_{t=1}^{T} X_t^i$, we have

$$E[X_i X_j X_k] = E\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t^i \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t^j \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t^k\right]$$

$$= \frac{1}{T\sqrt{T}}\sum_{t=1}^{T} E[X_t^i X_t^j X_t^k] + \frac{1}{T\sqrt{T}}\sum_{\substack{t=1}}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}{}_{t,s,r\ \text{distinct}} E[X_t^i X_s^j X_r^k]$$

$$+ \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\sum_{\substack{s=1\\ s\neq t}}^{T} E[X_t^i X_t^j X_s^k] + E[X_t^i X_s^j X_t^k] + E[X_t^i X_s^j X_s^k]$$

$$= \frac{1}{\sqrt{T}} E[X_t^i X_t^j X_t^k] + \frac{2}{T\sqrt{T}}\sum_{t=1}^{T-1}\sum_{h=1}^{T-t} E[X_t^i X_{t+h}^j X_{t+h}^k]$$

because the terms involving three separate indices are zero. Then defining $\gamma^{i,jk}(h) = E[X_t^i X_{t+h}^j X_{t+h}^k]$, we have

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{s=1\\ s\neq t}}^{T} E[X_t^i X_t^j X_s^k] + E[X_t^i X_s^j X_t^k] + E[X_t^i X_s^j X_s^k]$$

$$\simeq \sum_{h=1}^{\infty} \gamma^{i,jk}(h) + \gamma^{k,ij}(h) + \gamma^{j,ik}(h) < \infty.$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

Therefore

$$E[X_i X_j X_k] = \frac{1}{\sqrt{T}}\left[\gamma^{ijk}(0) + \sum_{h=1}^{\infty} \gamma^{i,jk}(h) + \gamma^{k,ij}(h) + \gamma^{j,ik}(h)\right]$$

$$E[X_i^2 X_j] = \frac{1}{\sqrt{T}}\left[\gamma^{iij}(0) + \sum_{h=1}^{\infty} 2\gamma^{i,ij}(h) + \gamma^{j,ii}(h)\right].$$

We proceed now to find the relevant cumulant constants $\{c_i\}_{i=1}^d$, $\{c_{ij}\}_{i,j=1}^d$, $\{c_{ijk}\}_{i,j,k=1}^d$ for the GARCH(1,1) case. We need to do this for two cases: (1) $Z_T = [\sqrt{T}(\hat{\theta}_1 - \theta_1), \sqrt{T}(\hat{\theta}_3 - \theta_3)]^\top$; (2) $Z_T = [\sqrt{T}(\tilde{\theta}_1 - \theta_1), \sqrt{T}(\hat{\theta}_3 - \theta_3)]^\top$. The first case just requires the cumulants and cross cumulants of the unrestricted GARCH estimators. The second case also requires cross cumulants between GARCH and ARCH estimates.

The unrestricted GARCH(1,1) estimator has an asymptotic expansion in terms of the standardized log-likelihood derivatives:

$$l_i = \frac{1}{2}\sum_{t=1}^{T}(\varepsilon_t^2 - 1)h_{t;i},$$

$$l_{ij} = \frac{1}{2}\sum_{t=1}^{T}(\varepsilon_t^2 - 1)h_{t;ij} - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t^2 h_{t;i}h_{t;j},$$

$$l_{ijk} = \frac{1}{2}\sum_{t=1}^{T}(\varepsilon_t^2 - 1)h_{t;ijk} - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t^2\{h_{t;ij}h_{t;k} + h_{t;ik}h_{t;j} + h_{t;jk}h_{t;i} - h_{t;i}h_{t;j}h_{t;k}\},$$

where $h_t = \ln\sigma_t^2$, $h_{t;i} = \partial h_t/\partial\theta_i$. Here $i,j,k \in \{1,2,3\}$.

In the case that $Z_T = [\sqrt{T}(\hat{\theta}_1 - \theta_1), \sqrt{T}(\hat{\theta}_3 - \theta_3)]^\top$, the cumulants are computed from the stochastic expansion (A.1) where

$$X_i = -\tau^{ir}V_r,$$

$$A_i B_i = \tau^{ir}\tau^{s_1 t_1}\{V_{rs_1}V_{t_1} - E(V_{rs_1}V_{t_1})\} - \frac{1}{2}\tau^{ir}\tau^{s_1 t_1}\tau^{u_1 v_1}\tau_{rt_1 v_1}\{V_{s_1}V_{u_1} - E(V_{s_1}V_{u_1})\}/2,$$

and $\tau_{ir} = T^{-1}E(l_{ir})$ where raising pairs of indices signifies components from the matrix inversion, $V_r = T^{-1/2}l_r$, $V_{rs_1} = T^{-1/2}(l_{rs_1} - E(l_{rs_1}))$, $\tau_{rt_1 v_1} = T^{-1}E(l_{rt_1 v_1})$. We use the summation convention that repeated indices in an expression are to be summed over. Linton (1997) gives an (almost) explicit general expression for c_j, the bias of the estimator $\hat{\theta}_j$, $j = 1,2,3$ and the own skewnesses c_{iii}. Specifically,

$$c_j = \tau^{ji}\tau^{kl}\{\tau_{ik,l} + \tau_{ikl}(\kappa_4 + 2)/4\}, \qquad \textbf{(A.2)}$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

where κ_4 is the fourth cumulant of the innovation, $\tau_{ijk} = T^{-1}E(l_{ijk})$, and $\tau_{ik,l} = T^{-1}E(l_{ik}l_l)$. Furthermore

$$c_{ijk} = -T^{1/2}\tau^{ir}\tau^{jm}\tau^{kl}E[V_rV_mV_l] \tag{A.3a}$$

$$+\left[\tau^{jm}\tau^{kl}\tau^{ir}\tau^{s_1t_1}\left[\tau_{m,t_1}\tau_{rs_1,l}+\tau_{m,rs_1}\tau_{l,t_1}-\frac{\tau^{u_1v_1}\tau_{rt_1v_1}}{2}\left(\tau_{m,s_1}\tau_{l,u_1}+\tau_{m,u_1}\tau_{l,s_1}\right)\right]\right] \tag{A.3b}$$

$$+\left[\tau^{ir}\tau^{jm}\tau^{kl}\tau^{s_2t_2}\left[\tau_{r,t_2}\tau_{ls_2,m}+\tau_{r,ls_2}\tau_{m,t_2}-\frac{\tau^{u_2v_2}\tau_{rt_2v_2}}{2}\left(\tau_{r,s_2}\tau_{m,u_2}+\tau_{r,u_2}\tau_{m,s_2}\right)\right]\right] \tag{A.3c}$$

$$+\left[\tau^{ir}\tau^{kl}\tau^{jm}\tau^{s_3t_3}\left[\tau_{r,t_3}\tau_{ms_3,l}+\tau_{r,ms_3}\tau_{l,t_3}-\frac{\tau^{u_3v_3}\tau_{rt_3v_3}}{2}\left(\tau_{r,s_3}\tau_{l,u_3}+\tau_{r,u_3}\tau_{l,s_3}\right)\right]\right] \tag{A.3d}$$

because $E[V_mV_l\{V_{rs_1}V_{t_1} - E(V_{rs_1}V_{t_1})\}] = \tau_{m,t_1}\tau_{rs_1,l} + \tau_{m,rs_1}\tau_{l,t_1} + o(1)$ and $E[V_mV_l\{V_{s_1}V_{u_1} - E(V_{s_1}V_{u_1})\}] = \tau_{m,s_1}\tau_{l,u_1} + \tau_{m,u_1}\tau_{l,s_1} + o(1)$ with $\tau_{m,t_1} = T^{-1}E(l_ml_{t_1})$ and $\tau_{rs_1,l} = T^{-1}E(l_{rs_1}l_l)$.

In the ARCH case the same expansion applies, only there are only two free parameters. Following the notation of Section 2, for $\tilde{\theta}^1$, the terms c_j (given in Iglesias and Phillips, 2005) and c_{ijk} come from applying (A.2) and (A.3) to the ARCH special case. Specifically,

$$c_1 = H_1^{-1}\left[E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right)^2H_2-\left(E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right)E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)\right)(H_3+2H_4)\right]$$

$$+H_1^{-1}\left[\left(E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2+E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right)E\left(\frac{1}{\sigma_t^4}\right)\right)H_5\right.$$

$$\left.+E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2H_6-E\left(\frac{1}{\sigma_t^4}\right)E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)H_7\right],$$

$$c_3 = H_1^{-1}\left[-E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right)E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)H_2+E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2H_3\right.$$

$$\left.+\left(E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right)E\left(\frac{1}{\sigma_t^4}\right)+E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2\right)H_4\right]$$

$$+H_1^{-1}\left[\left(-2E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)E\left(\frac{1}{\sigma_t^4}\right)\right)H_5-E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)E\left(\frac{1}{\sigma_t^4}\right)H_6+E\left(\frac{1}{\sigma_t^4}\right)^2H_7\right],$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

where

$$H_1 = \left(E\left(\frac{1}{\sigma_t^4}\right)E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right) - E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2\right)^2, \qquad H_2 = \sum_{i=1}^{T} E\left(\frac{1}{\sigma_t^4\sigma_{t-i}^2} - \frac{y_{t-i}^2}{\sigma_t^4\sigma_{t-i}^2}\right),$$

$$H_3 = \sum_{i=1}^{T} E\left(\frac{y_{t-i-1}^2}{\sigma_t^4\sigma_{t-i}^2} - \frac{y_{t-i-1}^2 y_{t-i}^2}{\sigma_t^4\sigma_{t-i}^4}\right), \qquad H_4 = \sum_{i=1}^{T} E\left(\frac{y_{t-1}^2}{\sigma_t^4\sigma_{t-i}^2} - \frac{y_{t-1}^2 y_{t-i}^2}{\sigma_t^4\sigma_{t-i}^4}\right),$$

$$H_5 = \sum_{i=1}^{T} E\left(\frac{y_{t-1}^2 y_{t-i-1}^2}{\sigma_t^4\sigma_{t-i}^2} - \frac{y_{t-1}^2 y_{t-i-1}^2 y_{t-i}^2}{\sigma_t^4\sigma_{t-i}^4}\right), \qquad H_6 = \sum_{i=1}^{T} E\left(\frac{y_{t-1}^4}{\sigma_t^4\sigma_{t-i}^2} - \frac{y_{t-1}^4 y_{t-1}^2}{\sigma_t^4\sigma_{t-i}^4}\right),$$

$$H_7 = \sum_{i=1}^{T} E\left(\frac{y_{t-1}^4 y_{t-1-i}^4}{\sigma_t^4\sigma_{t-i}^2} - \frac{y_{t-1}^4 y_{t-1-i}^4 y_{t-i}^2}{\sigma_t^4\sigma_{t-i}^4}\right).$$

The expressions (as noted in Linton, 1997) are much more complicated than in a linear AR(1) model (see, e.g., Phillips, 1977a, 1978).

In (A.3), c_{ijk} is made of two components: $c_{ijk} = \Pi_{ijk}^1 + \Pi_{ijk}^2$, with $\Pi_{ijk}^1 = -T^{1/2}\tau^{ir}\tau^{jm}\tau^{kl}E[V_r V_m V_l]$.

Consider $\Pi_{ijk}^1 = -(\kappa_{23}/8)\tau^{ir}\tau^{jm}\tau^{kl}\mu_{i,j,k}$ where $\tau_{ij} = -\mu_{i,j}/2$ and $\tau_{ijk} = -\{\mu_{ij,k} + \mu_{ik,j} + \mu_{jk,i} - \mu_{i,j,k}\}$. In matrix notation

$$\Pi_{ijk}^1 = -\frac{\kappa_{23}}{8}E[(e_i^\top \Omega\Gamma_t)(e_j^\top \Omega\Gamma_t)(e_k^\top \Omega\Gamma_t)],$$

where $\Omega = (\{\tau_{i,j}\}_{i,j=1}^n)^{-1}$, $\mu_{i,j} = E[\Gamma_{ti}\Gamma_{tj}]$, we construct the $n \times 1$ vector $\Gamma_t = (\Gamma_{ti})_{i=1}^n$, and e_i is a general $n \times 1$ vector of zeros with a unique 1 in position i. For example in the ARCH(1) model,

$$\Omega = \begin{pmatrix} \tau_{11} & \tau_{13} \\ \tau_{31} & \tau_{33} \end{pmatrix}^{-1}, \qquad \Gamma_t = \begin{pmatrix} \dfrac{1}{\sigma_t^2} \\ \dfrac{y_{t-1}^2}{\sigma_t^2} \end{pmatrix}, \qquad e_i = \begin{pmatrix} (1,0)^\top \text{ when } i = 1 \\ (0,1)^\top \text{ when } i = 3 \end{pmatrix},$$

$(\tau_{11}, \tau_{13}, \tau_{33})^\top = E[(-1/2\sigma_t^4)(1, y_{t-1}^2, y_{t-1}^4)]^\top$, and also $(\mu_{1,1,1}, \mu_{1,1,3}, \mu_{1,3,3}, \mu_{3,3,3})^\top = E[(1/\sigma_t^6)(1, y_{t-1}^2, y_{t-1}^4, y_{t-1}^6)]^\top$. The case of $i = j = k$ is given in Linton (1997). In this paper we generalize that expression to allow for the existence of cross products due to the introduction of restrictions in the estimation.

Consider Π_{ijk}^2 in (A.3b)–(A.3d). The simplest case is given in Linton (1997) where

$$\Pi_{iii}^2 = \frac{3(\kappa_4+2)}{4}\left[\frac{(\kappa_4+2)}{2}E[e_i^\top \Omega\Gamma_t \Omega e_i e_i^\top \Omega\Gamma_t] + T^{-1}\sum_{s<t}\sum E[(e_i^\top \Omega\Gamma_t)^2 e_i^\top \Omega\Gamma_s(\varepsilon_s^2 - 1)]\right].$$

We now extend the previous result to Π_{ijk}^2, where, as we note in (A.3b)–(A.3d) we can decompose it again in three subterms: $\Pi_{ijk}^2 = \Pi_{ijk}^{21} + \Pi_{ijk}^{22} + \Pi_{ijk}^{23}$. The first generic subterm is given by

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

$$\Pi_{ijk}^{21} = \left[\tau^{jm}\tau^{kl}\tau^{ir}\tau^{s_1t_1}\left[\tau_{m,t_1}\tau_{rs_1,l} + \tau_{m,rs_1}\tau_{l,t_1} - \frac{\tau^{u_1v_1}\tau_{rt_1v_1}}{2}\left(\tau_{m,s_1}\tau_{l,u_1} + \tau_{m,u_1}\tau_{l,s_1}\right)\right]\right],$$

where

$$\tau^{jm}\tau^{kl}\tau^{ir}\tau^{s_1t_1}\left[-\frac{\tau^{u_1v_1}\tau_{rt_1v_1}}{2}\left(\tau_{m,s_1}\tau_{l,u_1} + \tau_{m,u_1}\tau_{l,s_1}\right)\right]$$

$$= \frac{(\kappa_4+2)^2}{8}\left[\tau^{jt_1}\tau^{ir}\tau^{kv_1}\{\mu_{rt_1,v_1} + \mu_{rv_1,t_1} + \mu_{t_1v_1,r} - \mu_{r,v_1,t_1}\}\right]$$

$$= \frac{(\kappa_4+2)^2}{8}E\left[3e_j^\top \Omega\Gamma_t\Omega e_k e_i^\top \Omega\Gamma_t - (e_i^\top \Omega\Gamma_t)(e_j^\top \Omega\Gamma_t)(e_k^\top \Omega\Gamma_t)\right]$$

$$\times \tau^{jm}\tau^{kl}\tau^{ir}\tau^{s_1t_1}\left[\tau_{m,t_1}\tau_{rs_1,l} + \tau_{m,rs_1}\tau_{l,t_1}\right]$$

$$= -\frac{(\kappa_4+2)^2}{8}\left[E\left[2e_j^\top \Omega\Gamma_t\Omega e_k e_i^\top \Omega\Gamma_t - (e_i^\top \Omega\Gamma_t)(e_j^\top \Omega\Gamma_t)(e_k^\top \Omega\Gamma_t)\right]\right]$$

$$+ \frac{(\kappa_4+2)}{4}\left[T^{-1}\sum_{s<t}\sum E\left[(e_j^\top \Omega\Gamma_t)(e_k^\top \Omega\Gamma_t)e_i^\top \Omega\Gamma_s(\varepsilon_s^2 - 1)\right]\right].$$

Letting $s = t - h$ and $E[(e_j^\top \Omega\Gamma_t)(e_k^\top \Omega\Gamma_t)e_i^\top \Omega\Gamma_{(t-h)}(\varepsilon_{(t-h)}^2 - 1)] = \phi^{jk,i}(h)$, this last term also equals

$$\tau^{jm}\tau^{kl}\tau^{ir}\tau^{s_1t_1}\left[\tau_{m,t_1}\tau_{rs_1,l} + \tau_{m,rs_1}\tau_{l,t_1}\right]$$

$$= -\frac{(\kappa_4+2)^2}{8}\left[E\left[2e_j^\top \Omega\Gamma_t\Omega e_k e_i^\top \Omega\Gamma_t - (e_i^\top \Omega\Gamma_t)(e_j^\top \Omega\Gamma_t)(e_k^\top \Omega\Gamma_t)\right]\right]$$

$$+ \frac{(\kappa_4+2)}{8}\left[\sum_{h=1}^{\infty}\phi^{jk,i}(h)\right].$$

Therefore, finally, adding the two previous terms to retrieve Π_{ijk}^{21} and following the same methodology with Π_{ijk}^{22} and Π_{ijk}^{23}, we obtain

$$\Pi_{ijk}^{2} = \frac{(\kappa_4+2)}{4}\left[\frac{(\kappa_4+2)}{2}E\left[e_j^\top \Omega\Gamma_t\Omega e_k e_i^\top \Omega\Gamma_t\right] + \frac{1}{2}\sum_{h=1}^{\infty}\phi^{jk,i}(h)\right]$$

$$+ \frac{(\kappa_4+2)}{4}\left[\frac{(\kappa_4+2)}{2}E\left[e_i^\top \Omega\Gamma_t\Omega e_j e_k^\top \Omega\Gamma_t\right] + \frac{1}{2}\sum_{h=1}^{\infty}\phi^{ij,k}(h)\right]$$

$$+ \frac{(\kappa_4+2)}{4}\left[\frac{(\kappa_4+2)}{2}E\left[e_i^\top \Omega\Gamma_t\Omega e_k e_j^\top \Omega\Gamma_t\right] + \frac{1}{2}\sum_{h=1}^{\infty}\phi^{ik,j}(h)\right].$$

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press

Finally, for the homoskedastic case of $\tilde{\theta}^2$, we have

$$\tilde{\theta}^2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2, \tag{A.4}$$

which is exactly unbiased. The skewness coefficient is $c_{111} = \theta_1^3 \kappa_{23}$, where $\kappa_{23} = E[(\varepsilon_t^2 - 1)^3]$. ■

From the practical point of view, to use the expressions applied to the QMLE, we need an estimate of the fourth cumulant of the true conditional distribution (κ_4) and κ_{23}. We can use different procedures to get that in practice, although we can recommend, e.g., the methodology proposed in Cox and Hall (2002).

In practice, we can find many different restrictions that can be applied. For example, in the case of an ARCH(1) process

$$y_t = \varepsilon_t \sigma_t,$$

$$\sigma_t^2 = \theta_1 + \theta_3 y_{t-1}^2,$$

and let $\theta = (\theta_1, \theta_3) \in \mathbb{R}^2$ be the unknown parameters. If the researcher restricts in the QML estimation that $\theta_1 > 0$ and $\theta_3 \geq 0$ (nonnegativity constraint), this is simply a special case of the GARCH(1,1) example considered previously.

https://doi.org/10.1017/S0266466607070454 Published online by Cambridge University Press