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

February 10, 2006

Abstract

Andrews (1999) 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 behaviour of the QMLE in stationary and nonstationary GARCH models when constraints are imposed in the maximisation procedure. We show how in this case both a first and second-order bias appears in the estimator, and how it 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, 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, 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.

*Department of Economics, Michigan State University, 101 Marshall-Adams Hall, East Lansing, MI 48824-1038, USA. e-mail: iglesia5@msu.edu. Financial support from the MSU Intramural Research Grants Program is gratefully acknowledged.

†Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom. E-mail address: o.linton@lse.ac.uk. Financial support from the ESRC is gratefully acknowledged.

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 consequently inference is a lot 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 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 heteroscedastic) and GARCH (generalized-ARCH) 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) and 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 both in the stationary and the nonstationary region. However, they still require the positiveness in the parameter space. Forecasting is another reason to impose constraints in GARCH models. Hansen and Lunde (2005) present the GARCH(1,1) as one of the best models available nowadays for prediction; however, Chong, Ahmed and Abdullah (1999) show as well 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), 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, and some of them subject to constraints. Mainly, they 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 framework are Tse (1998), Giot and Laurent (2003) and Fiorentini, Sentana and Calzolari (2003). The study of finite sample properties of the QML estimator has been mainly developed so far as well for unconstrained estimation (e.g. Linton (1997) and 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 in this paper can be extended straightforwardly when exogenous variables are included in the mean equation (see Iglesias and Phillips (2005)).

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). In section 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 like $\sum_{j=1}^{n} \theta_j \leq 1$ are more complicated but fall under the general case laid out above.

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 $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_n)$ be the unrestricted estimator that maximizes $l_T(\theta)$ over $\mathbb{R}^n$. The estimator objective function can be the log-likelihood, the quasi-loglikelihood, a least squares criterion function, a GMM objective function, a minimum distance objective function, an objective function that depends on finite or infinite dimensional preliminary estimators, etc. (see Andrews (1999) for more details). By definition, the extremum estimators θ^*, $\widehat{\theta}$ satisfy:

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

Our approach to solving the inequality constrained optimization problem is to look at all the sub-cases 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

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

which can be solved by standard methods. We search over all such sub-cases and then find the parameter value that satisfies all inequality restrictions and which maximizes the criterion. In general there are $(2^p - 1)$ possible restricted problems to consider, which means that we can get $\widetilde{\theta}^1, \ldots, \widetilde{\theta}^{(2^p-1)}$ 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 $\widehat{\theta}, \widetilde{\theta}^1, \widetilde{\theta}^2, \ldots, \widetilde{\theta}^{(2^p-1)}$ 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 $r = 1$ sub-problems 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. The estimator θ^* is thus a function of $Z_T = (\widetilde{\theta}^1, \ldots, \widetilde{\theta}^{(2^p-1)}, l_T(\widetilde{\theta}^1), \ldots, l_T(\widetilde{\theta}^{(2^p-1)}))$, 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 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:

$$\begin{aligned} cum(Z_{Ti}) &= \frac{c_i}{\sqrt{T}} + o(T^{-1/2}) && (2)\\ cum(Z_{Ti}, Z_{Tj}) &= c_{ij} + o(T^{-1/2}) && (3)\\ cum(Z_{Ti}, Z_{Tj}, Z_{Tk}) &= \frac{c_{ijk}}{\sqrt{T}} + o(T^{-1/2}) && (4)\end{aligned}$$

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,

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

where

$$\begin{aligned}
\phi_{0,\Omega}(y) &= \frac{1}{(2\pi)^{d/2}\det(\Omega)}\exp\left(-\frac{1}{2}y^{\top}\Omega^{-1}y\right)\\
H_i(y) &= \frac{-1}{\phi_{0,\Omega}(y)}\frac{\partial}{\partial y_i}\phi_{0,\Omega}(y)\\
H_{ijk}(y) &= \frac{-1}{\phi_{0,\Omega}(y)}\frac{\partial^3}{\partial y_i\partial y_j\partial y_k}\phi_{0,\Omega}(y).
\end{aligned}$$

The quantities $H_i(y)$ and $H_{ijk}(y)$ are the multivariate Hermite polynomials of first and third degree. See Taniguchi and Kakizawa (2000, p170), Barndorff-Nielsen and Cox (1989, p 174).[1] 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 below.

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. Specifically, 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 like the ARCH(1) process but the GARCH(1,1) model is also of this type. We discuss below these two examples in detail.

3 Examples

3.1 ARCH(1) Process

Suppose that

$$\begin{aligned}
y_t &= \varepsilon_t\sigma_t\\
\sigma_t^2 &= \theta_1 + \theta_3 y_{t-1}^2,
\end{aligned}$$

where ε_t is 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 $\theta_3 \geq 0$ (non-negativity constraint).[2] The constraint that $\theta_1 > 0$ precludes the possibility that $\theta_1 = 0$ and so can never be binding. Let

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

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

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

be the restricted parameter space. Let $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_3)$ be the unrestricted QMLE, i.e., the maximizer over $\mathbb{R}^2$ of the quasi-log-likelihood. Let $\widetilde{\theta} = (\widetilde{\theta}_1, 0)$ be the maximizer subject to the binding restriction that $\theta_3 = 0$. Then the maximizer over the set Θ, denoted θ^*, is whichever of $\widehat{\theta}, \widetilde{\theta}$ makes the likelihood bigger and satisfies the restrictions. This is a nested case because $l_T(\widehat{\theta}) \geq l_T(\widetilde{\theta})$, so all that is required is to check whether $\widehat{\theta}$ satisfies the restrictions. Thus θ^* is a mixture of the two parameters:

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

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

These equations represent the inequality restricted estimator in terms of two equality restricted estimators $\widetilde{\theta}, \widehat{\theta}$. It follows that

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

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

To approximate $\Pr\left[\theta_3^* \leq x\right]$ we only need the marginal Edgeworth expansion for the estimator $\widehat{\theta}_3$, and this calculation is standard, see Linton (1997).[3] If θ is in the strict interior of the parameter space the probabilities like $\Pr[\widehat{\theta}_3 \geq 0]$ increase exponentially to one and the standard expansion applies to $\Pr\left[\theta_1^* \leq x\right]$. But when for example $\theta_3 = 0$ this probability converges at a more modest pace to one half and that expansion is no longer valid and we must compute the more complicated expression given in (8). Therefore, in general to approximate $\Pr\left[\theta_1^* \leq x\right]$ we need to approximate $\Pr[\widehat{\theta}_3 \leq x, \widehat{\theta}_1 \geq 0]$ and $\Pr[\widetilde{\theta}_1 \leq x, \widehat{\theta}_3 < 0]$, which require the joint asymptotic expansions of $(\widehat{\theta}_1, \widehat{\theta}_3)$ and of $(\widetilde{\theta}_1, \widehat{\theta}_3)$. 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

$$\begin{aligned} \sup_{x,y\in\mathbb{R}} \left|\Pr\left[\widehat{\theta}_1 \leq x, \widehat{\theta}_3 \leq y\right] - \Psi_{T1}(x,y)\right| &= o(T^{-1}) \\ \sup_{x,y\in\mathbb{R}} \left|\Pr\left[\widetilde{\theta}_1 \leq x, \widehat{\theta}_3 \leq y\right] - \Psi_{T2}(x,y)\right| &= o(T^{-1}) \\ \sup_{x\in\mathbb{R}} \left|\Pr\left[\widehat{\theta}_3 \leq x\right] - \Psi_{T0}(x)\right| &= o(T^{-1}). \end{aligned}$$

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

$$\Psi_{T1}^*(x) = \left[1 - \Psi_{T1}(x,0)\right]\left[1 - \Psi_{T0}(0)\right] + \Psi_{T2}(x,0)\Psi_{T0}(0) \tag{9}$$

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

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

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

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

To compute the approximate bias or skewness one can use 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 below.

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

3.2 GARCH(1,1)

Suppose that

$$\begin{aligned} y_t &= \varepsilon_t \sigma_t \\ \sigma^2_t &= \theta_1 + \theta_2 \sigma^2_{t-1} + \theta_3 y^2_{t-1}, \end{aligned}$$

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 $\theta_1 > 0$)

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

This is exactly the situation described in Jensen and Rahbek (2004a, 2004b) where they need the parameters to be positive. Let $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ 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 $\widetilde{\theta}^1 = (\widetilde{\theta}^1_1, 0, \widetilde{\theta}^1_3)$ be the maximizer subject to the binding restriction that $\theta_2 = 0$, and let $\widetilde{\theta}^2 = (\widetilde{\theta}^2_1, 0, 0)$ 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 $\widehat{\theta}, \widetilde{\theta}^1, \widetilde{\theta}^2$ makes the likelihood bigger and satisfies the restrictions. Clearly, $l_T(\widehat{\theta}) \geq l_T(\widetilde{\theta}^1) \geq l_T(\widetilde{\theta}^2)$ so it amounts to just checking first whether $\widehat{\theta}$ satisfies the restrictions, if not then checking whether $\widetilde{\theta}^1$ satisfies the remaining restrictions etc. Thus θ^* is a mixture of the three parameters. In fact

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

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

$$\theta_1^* = \widehat{\theta}_1 1(\widehat{\theta}_2 \geq 0, \widehat{\theta}_3 \geq 0) + \widetilde{\theta}_1^1 1(\widehat{\theta}_2 < 0, \widetilde{\theta}_3^1 \geq 0) + \widetilde{\theta}_1^2 1(\widehat{\theta}_2 < 0, \widetilde{\theta}_3^1 < 0). \tag{13}$$

A consequence of (11)-(13) is that

$$\Pr\left[\theta_2^* \leq x\right] = \begin{cases} \Pr\left[\widehat{\theta}_2 \leq x\right] & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

$$\Pr\left[\theta_3^* \leq x\right] = \begin{cases} \Pr\left[\widehat{\theta}_3 \leq x, \widehat{\theta}_2 \geq 0\right] \Pr\left[\widehat{\theta}_2 \geq 0\right] + \Pr\left[\widetilde{\theta}_3^1 \leq x, \widehat{\theta}_2 < 0\right] \Pr\left[\widehat{\theta}_2 < 0\right] & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

$\Pr\left[\theta_1^* \leq x\right]$ is similarly defined in terms of some joint probabilities. We apply the same methodology as in the ARCH case.

4 Inference and Bias Correction

The above 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 matter if the parameters are known but when the parameters are unknown this complicates matters. For example, the natural estimator of the distribution $\Psi_{Tj}^*(x;\widehat{\theta})$ 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.

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

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

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

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

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 Srinavasan (1970) for related discussion. We next specialize this general formula in the two cases.

When $\theta_3 = 0$, we have

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

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

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

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

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

where $\omega^* = \omega(\theta^*)$, $c_3^* = c_{333}(\theta^*)$, 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 non-degenerate limiting distribution). Instead we shall follow Andrews (2000) and work with a modified estimator that avoids this problem.[4] 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

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

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

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

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

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

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

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 estimators may be preferred because it is always non-negative 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 involves to retrieve the bias function $b_T(.)$ from (5), and to use again a modified estimator to apply in (16). For example, in the GARCH(1,1) case where we impose the positiveness constraint in all the parameters and using again the notation of Section 3, we define the bias correction using the shrinkage estimates: $\theta_2^{**} = \widehat{\theta}_2 1(\widehat{\theta}_2 > \eta_T)$, $\theta_3^{**} = \widehat{\theta}_3 1(\widehat{\theta}_3 > \eta_T)1(\widehat{\theta}_2 > \eta_T) + \widetilde{\theta}_3 1(\widetilde{\theta}_3 > \eta_T)1(\widehat{\theta}_2 < \eta_T)$, and $\theta_1^{**} = \widehat{\theta}_1 1(\widehat{\theta}_2 > \eta_T, \widehat{\theta}_3 > \eta_T) + \widetilde{\theta}_1^1 1(\widehat{\theta}_2 < \eta_T, \widetilde{\theta}_3^1 > \eta_T) + \widetilde{\theta}_1^2 1(\widehat{\theta}_2 < \eta_T, \widetilde{\theta}_3^1 < \eta_T)$.

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 sequence of true parameters converging to zero.

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

5 Simulations

We present now simulation results to analyze the consequences of imposing restrictions in the GARCH(1,1) model and later also in the ARCH(1). Table 1 shows the results for 40000 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). $\varepsilon_t \sim N(0,1)$. 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. As it is already stated in Linton (1997) (although there, it was 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 $\widehat{\theta}_2$ value is large, *while* $\widehat{\theta}_3$ *is small,* so it is important that the researcher can be aware that, specially this situation is the one that is very likely to generate large biases. The biases are especially large when both θ_2 and θ_3 are small. The restriction of positiveness increases slightly the bias, although less than it 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 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 $\widehat{\theta}_2$.

Table 1. Bias results for the GARCH(1,1)

	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

Sample Size=500. 40000 replications

If we are only interested in correcting the first order bias, for example, 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 Andrews (1999) theory); while if the restriction is to impose that the parameter space to be 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) model. This is given 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)), 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 behaviour of the ARCH(1) both in the stationary and non-stationary region (see Jensen and Rahbek (2004a, 2004b), where they prove that the QML estimator is still consistent and asymptotically normal in the nonstationary area), and also when a restriction is binding or not.

Table 2: Simulations of bias, standard errors* and skewness**

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****

Sample Size=50. 10000 replications

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 it 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 behaviour of the skewness, standard errors and the biases when the restrictions are binding. In Table 2, we show that the biases can increase very significantly: for example 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 (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 $\widehat{\theta}_1$ and $\widehat{\theta}_3$, we use the expressions of the bias given in the appendix in (18). Since the bias expressions in the ARCH(1) 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 $\widehat{\theta}_3$. 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 specially, it is interesting how when the true value of $\theta_3 = 1$, without the bias correction mechanism (due to the heavy downward bias of -0.09) we would conclude that the disturbance is weakly stationary, while 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.

Table 3: Simulated bias* and $O\left(T^{-1}\right)$ bias approximations**

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**

Sample Size=50. 10000 replications

Finally, in Table 4 we show the simulated first order bias of the estimator (what Andrews (1999) theory 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, Andrews (1999) theory provides a very good approximation (when the restriction is $\theta_3 \geq 0$), while for another restrictions (for imposing weak stationarity such that $\theta_3 \leq 1$), the first order approximation again in a standardized case (coming from $E\left(\min\left(Z,1\right)\ \right)$ where $Z \sim N\left(0,1\right)$) 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 Andrews (1999) theory would predict. The skewness term comes in the first case of Table 4 from the term $E\left[\left(\max\left(Z,0\right) - E\left(\max\left(Z,0\right)\right)\right)^3\right]$; and for the second restriction from $E\left[\left(\min\left(Z,1\right) - E\left(\min\left(Z,1\right)\right)\right)^3\right]$. We see how the first asymptotic theory for the skewness does not offer a good approximation neither for the first restriction nor for the second one.

Table 4. First order bias, total simulated bias*, first order skewness** and total simulated skewness***

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

Sample Size=50. 10000 replications

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 10000 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 (since 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

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

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^*

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

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

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

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

$$\theta_3^{**} = \left\{ \begin{array}{ll} \widehat{\theta}_3 & \text{if } \widehat{\theta}_3 < 1 + \eta_T \\ 1 & \text{if } \widehat{\theta}_3 > 1 + \eta_T \end{array} \right\}$$

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. $\theta_{3,mbc2}^*$ is clearly preferred to the uncorrected estimator θ_3^*. Under constraint 2, the three bias corrected estimators are median unbiased, while 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 advice to use θ_{mbc2}^* in practice.

In order to find out about the behaviour 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 little 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, being $\theta_{3,mbc2}^*$ the one that presents both the best median and interquartile range. Therefore, overall, from our simulations, we advice again to use θ_{mbc2}^* in practice.

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

Estimating subject to Constraint 1: $\theta_3^* = \widehat{\theta}_3 1(\widehat{\theta}_3 \geq 0)$

	$\theta_3 = 0$		$\theta_3 = 0.5$	
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: $\theta_3^* = \min\{\widehat{\theta}_3, 1\}$

	$\theta_3 = 1$		$\theta_3 = 0.5$	
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

6 Conclusions

In this paper we obtained higher order distributional expansions when some parameters may lie on boundary of the parameter space. We specialized our results to find out about the bias behaviour 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, due to the large biases created, both when constraints are imposed or 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.

7 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 were carried out using computer algebra carried out by a programme developed by the first author; here we just give some heuristics to indicate some issues.

Suppose that

$$\sqrt{T}(\widehat{\theta}_j - \theta_j) = X_j + \frac{A_j B_j}{\sqrt{T}} + o_p(T^{-1/2}), \tag{17}$$

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,

$$cum(Z_{Tj}, Z_{Tj}, Z_{Tj}) = E\left[\left(X_j + \frac{\{A_j B_j - E(A_j B_j)\}}{\sqrt{T}}\right)^3\right] \simeq E\left[X_j^3\right] + 3\frac{E\left[X_j^2\{A_j B_j - E(A_j B_j)\}\right]}{\sqrt{T}}$$

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

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

$$\begin{aligned} cum(Z_{Ti}, Z_{Tj}, Z_{Tk}) &= E\left[\prod_{r=i,j,k}\left(X_r + \frac{\{A_r B_r - E(A_r B_r)\}}{\sqrt{T}}\right)\right] \\ &\simeq E\left[X_i X_j X_k\right] + \frac{E\left[X_j X_k\{A_i B_i - E(A_i B_i)\}\right]}{\sqrt{T}} + \frac{E\left[X_i X_k\{A_j B_j - E(A_j B_j)\}\right]}{\sqrt{T}} \\ &\quad + \frac{E\left[X_i X_j\{A_k B_k - E(A_k B_k)\}\right]}{\sqrt{T}}. \end{aligned}$$

This involves cross-product terms that depend on the joint asymptotic normal distribution of (X, A, B) and on the higher order behaviour of X. Note that X_i are sums of (stationary) martingale difference sequences, while 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

$$\begin{aligned}
E\left[X_i X_j X_k\right] &= E\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t^i \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t^j \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t^k\right] \\
&= \frac{1}{T\sqrt{T}}\sum_{t=1}^{T} E\left[X_t^i X_t^j X_t^k\right] + \frac{1}{T\sqrt{T}}\sum_{\substack{t=1 \\ t,s,r \text{ distinct}}}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T} E\left[X_t^i X_s^j X_r^k\right] \\
&\quad + \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\sum_{\substack{s=1 \\ s\neq t}}^{T} E\left[X_t^i X_t^j X_s^k\right] + E\left[X_t^i X_s^j X_t^k\right] + E\left[X_t^i X_s^j X_s^k\right] \\
&= \frac{1}{\sqrt{T}} E\left[X_t^i X_t^j X_t^k\right] + \frac{2}{T\sqrt{T}}\sum_{t=1}^{T-1}\sum_{h=1}^{T-t} E\left[X_t^i X_{t+h}^j X_{t+h}^k\right]
\end{aligned}$$

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

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{s=1 \\ s\neq t}}^{T} E\left[X_t^i X_t^j X_s^k\right] + E\left[X_t^i X_s^j X_t^k\right] + E\left[X_t^i X_s^j X_s^k\right] \simeq \sum_{h=1}^{\infty}\gamma^{i,jk}(h) + \gamma^{k,ij}(h) + \gamma^{j,ik}(h) < \infty.$$

Therefore:

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

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

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) $Z_T = [\sqrt{T}(\widehat{\theta}_1 - \theta_1), \sqrt{T}(\widehat{\theta}_3 - \theta_3)]^\top$; (2) $Z_T = [\sqrt{T}(\widetilde{\theta}_1 - \theta_1), \sqrt{T}(\widehat{\theta}_3 - \theta_3)]^\top$. The first case just requires the cumulants and cross cumulants of the unrestricted GARCH estimators. The second case requires also 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:

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

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

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

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 $Z_T = [\sqrt{T}(\widehat{\theta}_1 - \theta_1), \sqrt{T}(\widehat{\theta}_3 - \theta_3)]^\top$, the cumulants are computed from the stochastic expansion (17) where

$$\begin{aligned} X_i &= -\tau^{ir} V_r \\ A_i B_i &= \tau^{ir}\tau^{s_1 t_1} \left\{ V_{rs_1} V_{t_1} - E\left(V_{rs_1} V_{t_1}\right)\right\} - \frac{1}{2}\tau^{ir}\tau^{s_1 t_1}\tau^{u_1 v_1}\tau_{rt_1 v_1} \left\{ V_{s_1} V_{u_1} - E\left(V_{s_1} V_{u_1}\right)\right\}/2 \end{aligned}$$

and $\tau_{ir} = T^{-1} E\left(l_{ir}\right)$ where raising pairs of indices signifies components from the matrix inversion, $V_r = T^{-1/2} l_r$, $V_{rs_1} = T^{-1/2}\left(l_{rs_1} - E\left(l_{rs_1}\right)\right)$, $\tau_{rt_1 v_1} = T^{-1} E\left(l_{rt_1 v_1}\right)$. 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 $\widehat{\theta}_j$, $j = 1, 2, 3$ and the own skewnesses c_{iii}. Specifically,

$$c_j = \tau^{ji}\tau^{kl}\left\{\tau_{ik,l} + \tau_{ikl}\left(\kappa_4 + 2\right)/4\right\}, \tag{18}$$

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

$$\begin{aligned} c_{ijk} &= -T^{1/2}\tau^{ir}\tau^{jm}\tau^{kl} E\left[V_r V_m V_l\right] && \text{(19a)} \\ &+ \left[\tau^{jm}\tau^{kl}\tau^{ir}\tau^{s_1 t_1}\left[\tau_{m,t_1}\tau_{rs_1,l} + \tau_{m,rs_1}\tau_{l,t_1} - \frac{\tau^{u_1 v_1}\tau_{rt_1 v_1}}{2}\left(\tau_{m,s_1}\tau_{l,u_1} + \tau_{m,u_1}\tau_{l,s_1}\right)\right]\right] && \text{(19b)} \\ &+ \left[\tau^{ir}\tau^{jm}\tau^{kl}\tau^{s_2 t_2}\left[\tau_{r,t_2}\tau_{ls_2,m} + \tau_{r,ls_2}\tau_{m,t_2} - \frac{\tau^{u_2 v_2}\tau_{rt_2 v_2}}{2}\left(\tau_{r,s_2}\tau_{m,u_2} + \tau_{r,u_2}\tau_{m,s_2}\right)\right]\right] && \text{(19c)} \\ &+ \left[\tau^{ir}\tau^{kl}\tau^{jm}\tau^{s_3 t_3}\left[\tau_{r,t_3}\tau_{ms_3,l} + \tau_{r,ms_3}\tau_{l,t_3} - \frac{\tau^{u_3 v_3}\tau_{rt_3 v_3}}{2}\left(\tau_{r,s_3}\tau_{l,u_3} + \tau_{r,u_3}\tau_{l,s_3}\right)\right]\right] && \text{(19d)} \end{aligned}$$

since $E\left[V_m V_l\left\{V_{rs_1} V_{t_1} - E\left(V_{rs_1} V_{t_1}\right)\right\}\right] = \tau_{m,t_1}\tau_{rs_1,l} + \tau_{m,rs_1}\tau_{l,t_1} + o\left(1\right)$ and $E\left[V_m V_l\left\{V_{s_1} V_{u_1} - E\left(V_{s_1} V_{u_1}\right)\right\}\right] = \tau_{m,s_1}\tau_{l,u_1} + \tau_{m,u_1}\tau_{l,s_1} + o\left(1\right)$ with $\tau_{m,t_1} = T^{-1} E\left(l_m l_{t_1}\right)$ and $\tau_{rs_1,l} = T^{-1} E\left(l_{rs_1} l_l\right)$.

In the ARCH case the same expansion applies only there are only two free parameters. Following the notation of Section 2, for $\widetilde{\theta}^1$, the terms c_j (given in Iglesias and Phillips (2005)) and c_{ijk} come from applying (18) and (19) to the ARCH special case. Specifically,

$$\begin{aligned} c_1 &= H_1^{-1}\left[E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right)^2 H_2 - \left(E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right) E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)\right)\left(H_3 + 2H_4\right)\right] \\ &\quad + H_1^{-1}\left[\left(E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2 + E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right) E\left(\frac{1}{\sigma_t^4}\right)\right) H_5 + E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2 H_6 - E\left(\frac{1}{\sigma_t^4}\right) E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right) H_7\right], \end{aligned}$$

$$c_3 = H_1^{-1}\left[-E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right)E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)H_2+E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2H_3+\left(E\left(\frac{y_{t-1}^4}{\sigma_t^4}\right)E\left(\frac{1}{\sigma_t^4}\right)+E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)^2\right)H_4\right]$$
$$+H_1^{-1}\left[\left(-2E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)E\left(\frac{1}{\sigma_t^4}\right)\right)H_5-E\left(\frac{y_{t-1}^2}{\sigma_t^4}\right)E\left(\frac{1}{\sigma_t^4}\right)H_6+E\left(\frac{1}{\sigma_t^4}\right)^2H_7\right],$$

where

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

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

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

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

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 (19), c_{ijk} is made of two components: $c_{ijk}=\Pi^1_{ijk}+\Pi^2_{ijk}$, with $\Pi^1_{ijk}=-T^{1/2}\tau^{ir}\tau^{jm}\tau^{kl}E\left[V_rV_mV_l\right]$.

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

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

where $\Omega=(\{\tau_{i,j}\}_{i,j=1}^n)^{-1}$, $\mu_{i,j}=E\left[\Gamma_{ti}\Gamma_{tj}\right]$, we construct the $n\times1$ vector $\Gamma_t=(\Gamma_{ti})_{i=1}^n$ and e_i is a general $n\times1$ vector of zeros with a unique 1 in position i. For example in the ARCH(1),

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

$(\tau_{11},\tau_{13},\tau_{33})^\top=E[\frac{-1}{2\sigma_t^4}\left(1,y_{t-1}^2,y_{t-1}^4\right)]^\top$ and also, $\left(\mu_{1,1,1},\mu_{1,1,3},\mu_{1,3,3},\mu_{3,3,3}\right)^\top=E[\frac{1}{\sigma_t^6}\left(1,y_{t-1}^2,y_{t-1}^4,y_{t-1}^6\right)]^\top$. 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^2_{ijk} in (19b), (19c) and (19d). The simplest case is given in Linton (1997) where

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

We extend now the previous result to Π_{ijk}^2, where, as we note in (19b), (19c) and (19d), we can decompose it again in three sub-terms: $\Pi_{ijk}^2 = \Pi_{ijk}^{21} + \Pi_{ijk}^{22} + \Pi_{ijk}^{23}$. The first generic sub-term is given by

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

where:

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

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

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

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

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

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

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

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

$$= -\frac{(\kappa_4+2)^2}{8}\left[E\left[2e_j^\top\Omega\Gamma_t\Omega e_k e_i^\top\Omega\Gamma_t - \left(e_i^\top\Omega\Gamma_t\right)\left(e_j^\top\Omega\Gamma_t\right)\left(e_k^\top\Omega\Gamma_t\right)\right]\right] + \frac{(\kappa_4+2)}{8}\left[\sum_{h=1}^{\infty}\phi^{jk,i}(h)\right]$$

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

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

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

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

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

$$\widetilde{\theta}^2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2, \tag{20}$$

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 for example 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

$$\begin{aligned} y_t &= \varepsilon_t \sigma_t \\ \sigma_t^2 &= \theta_1 + \theta_3 y_{t-1}^2, \end{aligned}$$

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 $\theta_3 \geq 0$ (non-negativity constraint), this is simply a special case of the GARCH(1,1) example considered previously.

References

[1] Andrews, D. W. K. (1999), Estimation When a Parameter Is on a Boundary. *Econometrica* 67, 1341-1383.

[2] Andrews, D. W. K. (2000), Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space. *Econometrica* 68, 2, 399-405.

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

[4] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[5] 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.

[6] Chong, C. W., M. I. Ahmad and M. Y. Abdullah (1999), Performance of GARCH Models in Forecasting Stock Market Volatility. *Journal of Forecasting* 18, 333-343.

[7] Comte F. and O. Lieberman (2003), Asymptotic Theory for Multivariate GARCH Processes. *Journal of Multivariate Analysis* 84, 1, 61-84.

[8] Cox, D. R. and P. Hall (2002), Estimation in a Simple Random Effects Model with Nonnormal Distributions. *Biometrika* 89, 4, 831-840.

[9] Engle, R. F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50, 987-1007.

[10] Doornik, J.A. and M. Ooms (2003), Multimodality in the GARCH Regression Model. 2003-W20 in Nuffield Economics Working Papers.

[11] Fiorentini, G., E. Sentana and G. Calzolari (2003), Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroscedastic Dynamic Regression Models with Student t Innovations. *Journal of Business and Economic Statistics 21, 4, 532-546.*

[12] Giot, P. and S. Laurent (2003), Value at Risk for Long and Short Trading Positions. *Journal of Applied Econometrics* 18, 6, 641-664.

[13] Götze, F. and C. Hipp (1983), Asymptotic Expansions for Sums of Weakly Dependent Random Vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64, 311-239.

[14] Hansen, P. R. and A. Lunde (2005), A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?. *Journal of Applied Econometrics* 20, 873-889.

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

[16] Jeantheau, T. (1998), Strong Consistency of Estimators of Multivariate ARCH Models. *Econometric Theory* 14, 70-86.

[17] Jensen, S. T. and A. Rahbek (2004a), Asymptotic Normality of the QMLE of ARCH in the Nonstationary Case. *Econometrica* 72, 2, 641-646.

[18] Jensen, S. T. and A. Rahbek (2004b), Asymptotic Inference for Nonstationary GARCH. *Econometric Theory* 20, 6, 1203-1226.

[19] Kim, T.-H., D. Stone and H. White (2000), Asymptotic and Bayesian Confidence Intervals for Sharpe Style Weights. Discussion Paper 2000-27, UCSD.

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

[21] Leeb, H. and B. Pötscher (2003), Can one estimate the unconditional distribution of post-model-selection estimators?. *Cowles Foundation Discussion Paper* no 1444.

[22] Ling S. and M. McAleer (2003), Asymptotic Theory for a Vector ARMA-GARCH Model. *Econometric Theory* 19, 278-308.

[23] Linton, O. (1997), An Asymptotic Expansion in the GARCH(1,1) Model. *Econometric Theory* 13, 558-581.

[24] 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.

[25] Newey, W. K. and D. G. Steigerwald (1997), Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroskedasticity Models. *Econometrica* 65, 3, 587-599.

[26] Otten, R., and D. Bams (2001), Statistical Tests for Return-Based Style Analysis. EFMA meetings, Lugano.

[27] Phillips, P. C. B. (1977a), Approximations to Some Finite Sample Distributions Associated with a First-Order Stochastic Difference Equation. *Econometrica* 45, 463-485.

[28] 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.

[29] Phillips, P. C. B. (1978), Edgeworth and Saddle Point Approximations in the First-Order Noncircular Autoregression. *Biometrika* 65, 91-98.

[30] Rothenberg, T. J. (1984), Approximating the Distributions of Econometric Estimators and Test Statistics. In Z. Griliches & M. D. Intriligator (eds.), *Handbook of Econometrics*, v. 2, 881-935. Amsterdam: North-Holland.

[31] Sargan, J. D. (1974), On the Validity of Nagar's Expansion for the Moments of Econometric Estimators. *Econometrica* 42, 169-176.

[32] Sargan, J. D. (1976), Econometric Estimators and the Edgeworth Expansion. *Econometrica* 44, 421-448; erratum, 45, 272.

[33] Srinavasan, T.N. (1970), Approximations to Finite Sample Moments of Estimators whose Exact Sampling Distributions are unknown. *Econometrica,* 38, 533-541.

[34] Taniguchi, M., and Y. Kakizawa (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag, Berlin.

[35] Tse, Y. K. (1998), The Conditional Heteroscedasticity of the Yen-Dollar Exchange Rate. *Journal of Applied Econometrics* 13, 49-55.

[36] Van Dijk, D., P. H. Franses and A. Lucas (1999), Testing for ARCH in the Presence of Additive Outliers. *Journal of Applied Econometrics* 14, 539-562.

[37] Weiss, A. A. (1984), ARMA Models with ARCH Errors. *Journal of Time Series Analysis* 5, 2, 129-143.

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