Journal of Financial Econometrics Advance Access published August 23, 2006

# Sample and Implied Volatility in GARCH Models

LAJOS HORVÁTH
*University of Utah*

PIOTR KOKOSZKA
*Utah State University*

RIČARDAS ZITIKIS
*University of Western Ontario*

ABSTRACT

The unconditional variance of various GARCH-type models is a function $h(\boldsymbol{\theta})$ of the parameter vector $\boldsymbol{\theta}$ which is estimated by $\widehat{\boldsymbol{\theta}}$. For most models used in practice, closed-form expressions of $h(\cdot)$ have been found. On the contrary, the unconditional variance can be estimated by the sample variance $\hat{\sigma}^2$. This article establishes the asymptotic distributions of the differences $\hat{\sigma}^2 - h(\boldsymbol{\theta})$ and $\hat{\sigma}^2 - h(\widehat{\boldsymbol{\theta}})$ for broad classes of GARCH-type models. Even though both limit distributions are normal, the asymptotic variances are not equal. Potential practical consequences of these results are discussed.

KEYWORDS: GARCH processes, statistical hypothesis test, volatility

The asymptotic theory for various types of GARCH processes has recently been advanced in several directions, see Horváth, Kokoszka, and Teyssière (2001), Ling and McAleer (2002a,b), Li, Ling, and McAleer (2002), Berkes, Horváth, and Kokoszka (2003, 2004), Francq and Zakoian (2004), and Berkes and Horváth (2004), to name just a few references related to the subject of this article. However, some important and easy-to-formulate problems still remain unexplored.

We thank Professor Eric Renault for two extensive letters clarifying some issues discussed in Section 3 and directing us to relevant literature. L.H. was supported by NSF grants DMS-0604670 and INT-0223262 and NATO grant PST.EAP.CLG 980599. P.K. was supported by NSF grants DMS-0413653 and INT-0223262, and by NATO grant PST.EAP.CLG 980599. R.Z. was supported by a Discovery Research Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. Address correspondence to Piotr Kokoszka, Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, or e-mail: piotr.kokoszka@usu.edu.

doi:10.1093/jjfinec/nbl002

To explain the contribution of this article, recall the GARCH($p$, $q$) equations [cf. Bollerslev (1986)]:

$$y_k = \sigma_k \varepsilon_k, \quad \sigma_k^2 = \omega + \sum_{1\le i\le p} \alpha_i y_{k-i}^2 + \sum_{1\le j\le q} \beta_j \sigma_{k-j}^2, \tag{1}$$

where $\{\varepsilon_k, \ -\infty < k < \infty\}$ is a sequence of zero mean independent and identically distributed random variables, and $\omega > 0$, $\alpha_i \ge 0$, and $\beta_j \ge 0$ are parameters. As it is always done in this context, we also assume that

$$\mathbf{E}\left(\varepsilon_k^2\right) = 1. \tag{2}$$

Furthermore, we assume throughout this article that

$$\alpha_i > 0 \quad \text{and} \quad \beta_j > 0. \tag{3}$$

Necessary and sufficient conditions under which the GARCH($p$, $q$) equations have a unique, strictly stationary, and nonanticipative solution were found by Nelson (1991) when $p = 1$ and $q = 1$, and by Bougerol and Picard (1992a,b) for arbitrary $p \ge 1$ and $q \ge 1$. We assume throughout this article that those conditions hold and note here only that they imply $\beta_1 + \cdots + \beta_q < \rho$, for a constant $\rho < 1$.

Denote by $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ the vector of true GARCH($p$, $q$) parameters. A generic element of the parameter space is denoted by $\boldsymbol{u} = (x, s_1, \ldots, s_p, t_1, \ldots, t_q)$. Define the function $h : \mathbf{R}^{1+p+q} \to [0,\infty)$ by

$$h(\boldsymbol{u}) := \frac{x}{1 - \sum_i s_i - \sum_j t_j}.$$

Note that by assumption (2), the variance implied by the GARCH model is $h(\boldsymbol{\theta})$. On the contrary, the sample variance is

$$\widehat{Y}_n := \frac{1}{n} \sum_{1\le k\le n} y_k^2.$$

It can be expected that the difference $\widehat{Y}_n - h(\boldsymbol{\theta})$ tends to zero as $n \to \infty$, but it is not immediately clear at what rate and what the asymptotic distribution is. If we replace the unknown true parameter vector $\boldsymbol{\theta}$ by its estimator $\widehat{\boldsymbol{\theta}}$, we can then ask the same question about the difference $\widehat{Y}_n - h(\widehat{\boldsymbol{\theta}})$. We will show that under fairly general assumptions and for broad classes of models both differences are of the order $n^{-1/2}$.

In view of the above, the following questions arise: Do the statistics

$$T_n(\boldsymbol{\theta}_0) := \sqrt{n}(\widehat{Y}_n - h(\boldsymbol{\theta})) \tag{4}$$

and

$$T_n(\widehat{\boldsymbol{\theta}}) := \sqrt{n}(\widehat{Y}_n - h(\widehat{\boldsymbol{\theta}})) \tag{5}$$

converge to the same distribution? Are the limiting distributions normal? Can the asymptotic variances be effectively computed? Do these statistics lead to useful tests? To answer these questions is the goal of this article.

The remainder of this article is organized as follows. In Section 1, we state the relevant theoretical results for GARCH($p$, $q$) processes. An extension to a broad class of augmented GARCH processes is developed in Section 2. Section 3 discusses some practical consequences of the theoretical results, whereas Section 4 contains their proofs.

## 1 MAIN RESULTS FOR THE GARCH($p$, $q$) PROCESSES

From now on, we denote by $\widehat{\boldsymbol{\theta}}$ the commonly used pseudo maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$, see for example, Section 7.4.1 of Zivot and Wang (2003).

To formulate our results, we must introduce additional notation. Under the assumptions stated in the introduction, we have $\sigma_k^2 = w_k(\boldsymbol{\theta})$, where the function $w_k : \mathbf{R}^{1+p+q} \to [0,\infty)$ is defined by

$$w_k(\boldsymbol{u}) := c_0(\boldsymbol{u}) + \sum_{1 \le i < \infty} c_i(\boldsymbol{u}) y_{k-i}^2. \tag{6}$$

The coefficient-functions $c_i : \mathbf{R}^{1+p+q} \to [0, \infty)$ are deterministic, known, and can be expressed as solutions of certain recursion equation [cf. Berkes, Horváth, and Kokoszka (2003) for details]. For example, for the GARCH(1,1) model, $c_0(u) = x/(1-t)$ and for $i \geq 1$, $c_i(\boldsymbol{u}) = st^{i-1}$, where, as in the introduction, $\boldsymbol{u} = (x, s, t)$ is the parameter vector whose true value is $\boldsymbol{\theta} = (\omega, \alpha, \beta)$.

**Theorem 1** *If the expectations* $\mathbf{E}(\varepsilon_0^4)$ *and* $\mathbf{E}(\sigma_0^4)$ *are finite, then*

$$\sqrt{n}\,(\widehat{Y}_n - h(\boldsymbol{\theta})) \to_d \mathcal{N}\left(0, \tau_0^2\right) \tag{7}$$

*with the asymptotic variance*

$$\tau_0^2 := \frac{h^2(\boldsymbol{\theta})}{c_0^2(\boldsymbol{\theta})}\left(\mathbf{E}\left[\varepsilon_0^4\right] - 1\right)\mathbf{E}\left[\sigma_0^4\right].$$

The asymptotic variance $\tau_0^2$ in Theorem 1 is a known function of the parameter vector $\boldsymbol{\theta}$ and the first four moments of $\varepsilon_0 :\ h(\boldsymbol{\theta})$ and $c_0(\boldsymbol{\theta})$ are rational functions of the components of $\boldsymbol{\theta}$, and a closed-form expression for $\mathbf{E}(\sigma_0^4)$ was derived by He and Teräsvirta (1999). Further details are presented in Section 3.

To establish our next result, we must assume that the estimator $\widehat{\boldsymbol{\theta}}$ admits the expansion

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{2}(1-\varepsilon_k^2)\frac{w_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})}\mathbf{B}_0^{-1} + o_{\mathbf{P}}(n^{-1/2}), \tag{8}$$

where, on denoting by $\mathcal{D}$ the gradient operator with respect to the coordinates of $\boldsymbol{u}$,

$$\mathbf{B}_0 := \mathbf{E}[\mathcal{D}^2\ell_0(\boldsymbol{\theta})] \quad \text{with} \quad \ell_0(\boldsymbol{\theta}) := -\frac{1}{2}\left[\log w_0(\boldsymbol{\theta}) + \frac{y_k^2}{w_0(\boldsymbol{\theta})}\right].$$

Representation (8) was established by Berkes, Horváth, and Kokoszka (2003) and Francq and Zakoian (2004) under mild conditions.

We also introduce the random variable

$$\eta_0 := \frac{h(\boldsymbol{\theta})}{c_0(\boldsymbol{\theta})}\sigma_0^2 + \frac{1}{2}\left\langle \mathcal{D}h(\boldsymbol{\theta}), \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\mathbf{B}_0^{-1}\right\rangle, \tag{9}$$

with $\langle\cdot,\cdot\rangle$ denoting the inner product between two vectors.

**Theorem 2** *If the assumptions of Theorem 1 and Equation* (8) *hold, then*

$$\sqrt{n}\,(\widehat{Y}_n - h(\widehat{\boldsymbol{\theta}})) \to_d \mathcal{N}(0,\tau^2) \tag{10}$$

*with the asymptotic variance*

$$\tau^2 := \left(\mathbf{E}[\varepsilon_0^4] - 1\right)\mathbf{E}[\eta_0^2].$$

It is not clear whether the asymptotic variance $\tau^2$ in Theorem 2 can be computed explicitly, even if the vector $\boldsymbol{\theta}$ is known.

We note that all estimators of $\boldsymbol{\theta}$ proposed to date satisfy an asymptotic linearity condition similar to (8). Hence, with slight modifications, Theorem 2 holds for the estimators studied in Peng and Yao (2003) and Berkes and Horváth (2004).

Theorems 1 and 2 show that statistics (4) and (5) are both asymptotically normal, but their asymptotic variances are different. The estimation of $\boldsymbol{\theta}$ introduces an additional term, the second term on the right-hand side of (9).

## 2 EXTENSION TO AUGMENTED GARCH PROCESSES

In the previous section, we formulated two results for the GARCH($p$, $q$) process in Theorems 1 and 2. In this section, we extend Theorem 1 to the so-called augmented GARCH process [cf. Duan (1997)] defined below. An extension of Theorem 2 might also be possible, but we do not pursue it here because, as will be seen in Section 3, Theorem 2 is not as useful as Theorem 1. We comment further on this issue at the end of this section.

The augmented GARCH process is defined by

$$y_k = \sigma_k \varepsilon_k, \quad \Lambda(\sigma_k^2) = \Lambda(\sigma_{k-1}^2)c(\varepsilon_{k-1}) + d(\varepsilon_{k-1}), \tag{11}$$

where $\Lambda(x)$ is a completely specified function, whereas $c(x)$ and $d(x)$ are functions of known functional forms but possibly depending on some unknown parameters. Numerous examples in the literature show that the functions $c(x)$ and $d(x)$ can typically be represented as inner products between multidimensional parameters and completely specified functions, taking values in a space of the same dimension as the parameters. That is, we have the representations $c(x) = \langle \boldsymbol{\theta}_c, \boldsymbol{c}(x) \rangle$ and $d(x) = \langle \boldsymbol{\theta}_d, \boldsymbol{d}(x) \rangle$ upon which we shall *not* rely in our considerations but will clearly see when considering examples below.

We assume throughout that Equations in (11) have a unique and strictly stationary solution. For necessary and sufficient conditions when this happens, we refer to Aue, Berkes, and Horváth (2006).

To derive an inferential theory for augmented GARCH processes which extends the theory developed above for the GARCH(1, 1) process, we need to assume additional structure for the function $\Lambda(x)$. A glimpse at various examples shows that many augmented GARCH processes are of either the polynomial or the exponential type, meaning that $\Lambda(x)$ is either $x^\delta$ for some $\delta > 0$ or $\log x$. These two cases suggest imposing the following structural property on the function $\Lambda(x)$:

$$\Lambda(xy) = \Lambda(x)a(y) + b(y), \tag{12}$$

where $a(y)$ and $b(y)$ are some functions. Note that assumption (12) is, indeed, satisfied for both $\Lambda(x) = x^\delta$ and $\Lambda(x) = \log x$. Namely, in the first case we have Equation (12) with $a(y) = y^\delta$ and $b(y) = 0$, whereas in the second case the equation holds with $a(y) = 1$ and $b(y) = \log y$. In fact, these are essentially all the nontrivial solutions (in $\Lambda$, $a$, and $b$) of the generalized Pexider's equation (3.2), as seen from Theorem 4.9 on p. 65 of Castillo, Iglesias, and Ruiz-Cobo (2005).

To formulate the main result of this section, Theorem 3 below, we need to introduce additional notation. First, let

$$\widehat{Y}_n(\Lambda) := \frac{1}{n} \sum_{1 \le k \le n} \Lambda(y_k^2).$$

Note that when $\Lambda(x) = x$, then $\widehat{Y}_n(\Lambda)$ becomes the sample second moment $\widehat{Y}_n$ of the returns.

We now must compute a counterpart of $\widehat{Y}_n(\Lambda)$ implied by the augmented GARCH model, that is, $\mathbf{E}[\Lambda(y_0^2)]$. For this, denote by $\boldsymbol{\theta}$ the parameter vector on which the functions $c(\cdot)$ and $d(\cdot)$ depend. Using Equations (11) and (12), it is easy to verify that

$$\mathbf{E}[\Lambda(y_0^2)] = H(\boldsymbol{\theta}),$$

where

$$H(\boldsymbol{\theta}) := \mathbf{E}\left[b(\varepsilon_0^2)\right] + \Delta_0 \mathbf{E}[d(\varepsilon_0)],$$

and where

$$\Delta_0 := \frac{\mathbf{E}\left[a(\varepsilon_0^2)\right]}{1 - \mathbf{E}[c(\varepsilon_0)]}.$$

We assume of course that all the moments on the right-hand side of the above equations are finite, and that $\mathbf{E}[c(\varepsilon_0)] < 1$.

For the GARCH(1, 1) process, we have

$$a(y) = y,\ b(y) = 0,\ c(y) = \alpha y^2 + \beta,\ d(y) = \omega. \tag{13}$$

Hence, $H(\boldsymbol{\theta})$ equals $\omega/(1-\alpha-\beta)$, which is exactly the quantity $h(\boldsymbol{\theta})$ introduced in the introduction.

We now formulate an extension of Theorem 1 to the augmented GARCH processes defined by (11) with the function $\Lambda(x)$ satisfying (12).

**Theorem 3** *Assume that the random variables $a(\varepsilon_0^2)$, $b(\varepsilon_0^2)$, $c(\varepsilon_0)$, $d(\varepsilon_0)$, and $\Lambda(\sigma_0^2)$ have finite second moments. Furthermore, assume that $\mathbf{E}[c(\varepsilon_0)]<1$. Then*

$$\sqrt{n}\,(\widehat{Y}_n(\Lambda) - H(\boldsymbol{\theta})) \to_d \mathcal{N}\left(0, T_0^2\right), \tag{14}$$

*where the asymptotic variance $T_0^2$ is given by*

$$\begin{aligned} T_0^2 :&= \mathbf{E}\left[\Lambda^2(\sigma_0^2)\right]\mathbf{Var}\left[a(\varepsilon_0^2) + \Delta_0 c(\varepsilon_0)\right] + \mathbf{Var}\left[b(\varepsilon_0^2) + \Delta_0 d(\varepsilon_0)\right] \\ &\quad + 2\mathbf{E}\left[\Lambda(\sigma_0^2)\right]\mathbf{Cov}\left[a(\varepsilon_0^2) + \Delta_0 c(\varepsilon_0), b(\varepsilon_0^2) + \Delta_0 d(\varepsilon_0)\right]. \end{aligned}$$

Theorem 3 is proved in Section 4. In the remainder of this section, we compute the centering constant $H(\boldsymbol{\theta})$ and the asymptotic variance $T_0^2$ for several specific augmented GARCH models.

As our first example, we shall show that, as expected, the equality $T_0^2 = \tau_0^2$ holds for the GARCH(1, 1) process. To this end, we first note that in the

GARCH(1, 1) case the functions $b(y)$ and $d(y)$ are constant. Interestingly, we shall also see in our other examples that we usually have either both $a(y)$ and $c(y)$ constant, or both $b(y)$ and $d(y)$ constant. In the former case we then get

$$T_0^2 = \mathbf{Var}\left[b(\varepsilon_0^2) + \Delta_0 d(\varepsilon_0)\right], \tag{15}$$

whereas in the latter case we have

$$T_0^2 = \mathbf{E}\left[\Lambda^2(\sigma_0^2)\right]\mathbf{Var}\left[a(\varepsilon_0^2) + \Delta_0 c(\varepsilon_0)\right]. \tag{16}$$

Hence, recalling that for the GARCH(1, 1) process we have the expressions in (13) and also $\Lambda(y) = y$, we use formula (16) and immediately obtain the following corollary to Theorem 3.

**Corollary 1** *If the expectations* $\mathbf{E}(\varepsilon_0^4)$ *and* $\mathbf{E}(\sigma_0^4)$ *are finite, then we have for the* GARCH(1, 1) *process that*

$$\sqrt{n}\left(\widehat{Y}_n - \frac{\omega}{1-\alpha-\beta}\right) \to_d \mathcal{N}\left(0, \tau_0^2\right) \tag{17}$$

*with the asymptotic variance*

$$\tau_0^2 := \left(\frac{1-\beta}{1-\alpha-\beta}\right)^2 \left(\mathbf{E}\left[\varepsilon_0^4\right] - 1\right)\mathbf{E}\left[\sigma_0^4\right]. \tag{18}$$

Consider now a few other examples of augmented GARCH(1, 1) processes. We start with the asymmetric model, AGARCH, introduced by Ding, Granger, and Engle (1993) and defined by the equation

$$\sigma_k^2 = \omega + \beta\sigma_{k-1}^2 + \alpha(|y_{k-1}| - \mu y_{k-1})^2.$$

This equation can also be rewritten as

$$\sigma_k^2 = \sigma_{k-1}^2\left[\beta + \alpha(|\varepsilon_{k-1}| - \mu\varepsilon_{k-1})^2\right] + \omega,$$

which shows that the AGARCH process is a special case of the general augmented GARCH process in (11) and (12) when the functions are

$$\Lambda(x) = x, \quad a(x) = x, \quad b(x) = 0, \quad d(x) = \omega, \tag{19}$$

$$c(x) = \beta + \alpha(1+\mu^2)x^2 - 2\alpha\mu x|x|. \tag{20}$$

Because the next process that we shall look at is GJR-GARCH, and thus satisfies Equations (11) and (12) with the same functions $\Lambda(x)$, $a(x)$, $b(x)$, and $d(x)$ are as in (19), but with a different $c(x)$, we now formulate a general corollary to Theorem 3.

**Corollary 2** *Assume that the expectations* $\mathbf{E}(\varepsilon_0^4)$ *and* $\mathbf{E}(\sigma_0^4)$ *are finite. Furthermore, assume that the functions* $\Lambda(x)$, $a(x)$, $b(x)$, *and* $d(x)$ *are as in (19), and the function* $c(x)$ *is generic but such that* $\mathbf{E}[c(\varepsilon_0)]<1$. *Then*

$$\sqrt{n}\left(\widehat{Y}_n - \frac{\omega}{1-\mathbf{E}[c(\varepsilon_0)]}\right) \to_d \mathcal{N}\left(0, T_0^2\right) \tag{21}$$

*with the asymptotic variance*

$$T_0^2 = \mathbf{Var}\left[\varepsilon_0^2 + \frac{c(\varepsilon_0)}{1-\mathbf{E}[c(\varepsilon_0)]}\right]\mathbf{E}\left[\sigma_0^4\right]. \tag{22}$$

Coming now back to the AGARCH process, we have $c(x)$ in (20) and thus the equality

$$\mathbf{E}[c(\varepsilon_0)] = \beta + \alpha(1+\mu^2) - 2\alpha\mu\mathbf{E}[\varepsilon_0|\varepsilon_0|]. \tag{23}$$

Equality (23) is helpful in expressing the centering constant $\omega/(1-\mathbf{E}[c(\varepsilon_0)])$ in statement (21) in terms of unknown parameters. Furthermore, we use this formula as well as the following two ones [cf. Equations (24) and (25)] to express the variance on the right-hand side of Equation (22) in terms of unknown parameters that can later be estimated. The promised two formulas are

$$\mathbf{E}\left[\varepsilon_0^2 c(\varepsilon_0)\right] = \beta + \alpha(1+\mu^2)\mathbf{E}\left[\varepsilon_0^4\right] - 2\alpha\mu\mathbf{E}\left[\varepsilon_0^3|\varepsilon_0|\right] \tag{24}$$

and

$$\begin{aligned}\mathbf{E}\left[c^2(\varepsilon_0)\right] &= \beta^2 + 2\beta\alpha(1+\mu^2) + \alpha^2(1+6\mu^2+\mu^4)\mathbf{E}\left[\varepsilon_0^4\right] \\ &\quad - 4\beta\alpha\mu\mathbf{E}[\varepsilon_0|\varepsilon_0|] - 4\alpha^2\mu(1+\mu^2)\mathbf{E}\left[\varepsilon_0^3|\varepsilon_0|\right].\end{aligned} \tag{25}$$

This finishes our discussion of the AGARCH process.

Other examples can be investigated analogously. Take, for example, the GJR-GARCH process of Glosten, Jagannathan, and Runkle (1993) defined by Equations (19) and the function

$$c(x) = \beta + \alpha_1 x^2 \mathbf{1}\{x<0\} + \alpha_2 x^2 \mathbf{1}\{x \geq 0\}. \tag{26}$$

Corollary 2 is applicable. We only need to recalculate formulas (23)–(25), which we do and obtain the following ones:

$$\mathbf{E}[c(\varepsilon_0)] = \beta + \alpha_1 \mathbf{E}\big[\varepsilon_0^2 \mathbf{1}\{\varepsilon_0 < 0\}\big] + \alpha_2 \mathbf{E}\big[\varepsilon_0^2 \mathbf{1}\{\varepsilon_0 \geq 0\}\big], \tag{27}$$

$$\mathbf{E}\big[\varepsilon_0^2 c(\varepsilon_0)\big] = \beta + \alpha_1 \mathbf{E}\big[\varepsilon_0^4 \mathbf{1}\{\varepsilon_0 < 0\}\big] + \alpha_2 \mathbf{E}\big[\varepsilon_0^4 \mathbf{1}\{\varepsilon_0 \geq 0\}\big], \tag{28}$$

and

$$\begin{aligned}\mathbf{E}\big[c^2(\varepsilon_0)\big] = \beta^2 &+ \alpha_1^2 \mathbf{E}\big[\varepsilon_0^4 \mathbf{1}\{\varepsilon_0 < 0\}\big] + \alpha_2^2 \mathbf{E}\big[\varepsilon_0^4 \mathbf{1}\{\varepsilon_0 \geq 0\}\big] \\ &+ 2\beta\alpha_1 \mathbf{E}\big[\varepsilon_0^2 \mathbf{1}\{\varepsilon_0 < 0\}\big] + 2\beta\alpha_2 \mathbf{E}\big[\varepsilon_0^2 \mathbf{1}\{\varepsilon_0 \geq 0\}\big].\end{aligned} \tag{29}$$

Consider yet another example, such that Corollary 2 is not applicable, in which case we appeal to the general Theorem 3. This happens, for example, in the case of the MGARCH process introduced by Geweke (1986), in which case we have the equation:

$$\log \sigma_k^2 = \omega + \beta \log \sigma_{k-1}^2 + \alpha \log y_{k-1}^2.$$

Rewriting the above equation as

$$\log \sigma_k^2 = \big(\log \sigma_{k-1}^2\big)(\alpha + \beta) + \omega + \alpha \log \varepsilon_{k-1}^2.$$

we see that we have the functions:

$$\Lambda(x) = \log x, \quad a(x) = 1, \quad b(x) = \log x, \quad d(x) = \omega + \alpha \log x^2, \tag{30}$$

$$c(x) = \alpha + \beta. \tag{31}$$

Clearly now, the MGARCH process is an example of those with constant functions $a(x)$ and $c(x)$, and so to speed up the calculation of the asymptotic variance, we appeal to formula (15). The following corollary to Theorem 3 becomes an easy exercise.

**Corollary 3** *Assume that the random variables $\log \varepsilon_0^2$ and $\log \sigma_0$ have finite second moments. Furthermore, assume that the functions $\Lambda(x)$, $a(x)$, $b(x)$, $c(x)$, and $d(x)$ are given by formulas (30) and (31), and let $\alpha + \beta < 1$. Then*

$$\sqrt{n}\left(\widehat{Y}_n(\Lambda) - \frac{\omega + (1-\beta)\mathbf{E}\big[\log \varepsilon_0^2\big]}{1-(\alpha+\beta)}\right) \to_d \mathcal{N}\big(0, T_0^2\big), \tag{32}$$

*where the asymptotic variance $T_0^2$ is given by*

$$T_0^2 := \left(\frac{1-\beta}{1-(\alpha+\beta)}\right)^2 \mathbf{Var}\left[\log \varepsilon_0^2\right].$$

To derive an extension of Theorem 2 to augmented GARCH processes, we would need to follow the lines of its proof. This would involve differentiation of the centering constants with respect to parameters and also constructing empirical estimators for the parameters. Although the former problem is an easy, though harrowing, exercise, the latter one is a challenging task. Asymptotic theory for parameter estimation in GARCH-type models which can be imbedded into stochastic recurrence equations is developed in Straumann (2005).

## 3 SOME PRACTICAL CONSEQUENCES

We illustrate the theory developed above using the GARCH(1, 1) model

$$y_k = \sigma_k \varepsilon_k, \quad \sigma_k^2 = \omega + \alpha y_{k-1}^2 + \beta \sigma_{k-1}^2. \tag{33}$$

Extensions to other models involve more complex analogs of formulas (34) and (35) below, which we shall rely on in this section.

The observations presented in this section are not meant as a guide for practitioners, but merely as an illumination of the theoretical results.

Theorem 1 provides the null distribution of a test statistics in the following test:

**Parametric test:**

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \ H_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0,$$

where $\boldsymbol{\theta}_0$ is a vector of parameter values specified by a researcher. (GARCH(1, 1) model is assumed correct.)

For the GARCH(1, 1) model with standard normal errors, we have $\varepsilon_k$

$$h(\boldsymbol{\theta}) = \frac{\omega}{1-\alpha-\beta}, \quad c_0(\boldsymbol{\theta}) = \frac{\omega}{1-\beta}, \tag{34}$$

$$\mathbf{E}\left[\sigma_0^4\right] = \frac{\omega^2 + 2\omega h(\boldsymbol{\theta})(\alpha+\beta)}{1-(3\alpha^2+\beta^2+2\alpha\beta)},$$

and so $\tau_0$ is given by the formula:

$$\tau_0 = \frac{1-\beta}{1-\alpha-\beta}\sqrt{2\mathbf{E}\left[\sigma_0^4\right]}. \tag{35}$$

Under $H_0$, the statistic $\sqrt{n}(\widehat{Y}_n - h(\boldsymbol{\theta}))/\tau_0$ has asymptotic standard normal distribution.

Theorem 2 could, in principle, be used to develop the following test:

**Specification test:**

$H_0$: GARCH($p$, $q$) is the correct model
$H_A$: GARCH($p$, $q$) is not the correct model.

We first evaluate the performance of the parametric test using two parameter vectors:

Model 1: $\omega = 0.5$, $\alpha = 0.1$, $\beta = 0.7$
Model 2: $\omega = 0.1$, $\alpha = 0.2$, $\beta = 0.3$

Although Model 1 is more typical for financial returns, Model 2 is used to evaluate the scope of the findings.

Table 1 summarizes empirical sizes of the test for the two models. The test is seen to have excellent size even for series of length $n = 100$, which is considered small in the context of GARCH modeling. Typically, series of length close to $n = 1000$ are required in procedures that involve parameter estimation; the parametric test does not involve parameter estimation and uses the asymptotic variance $\tau_0$ computed as a function of the parameters postulated under $H_0$.

To evaluate the power of the test, we introduce the following two alternatives:

Model 1A: $\omega = 0.5$, $\alpha = 0.1$, $\beta = 0.75$
Model 2A: $\omega = 0.05$, $\alpha = 0.2$, $\beta = 0.3$

Model 1A is very close to Model 1—the parameter $\beta$ changes from 0.70 to 0.75. Model 2A may appear close to Model 2, but, in fact, the unconditional variance in Model 2A is only half of the unconditional variance in Model 2. These alternatives are thus far apart from the practical point of view. Empirical power is displayed in Table 2.

We now turn to the *potential* specification test. The idea would be to use the difference (5) as a test statistic which should be small if the model is correctly

**Table 1** Empirical size of the test of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ based on 5000 replications.

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| $n$ | 10% | 5% | 10% | 5% |
| 100 | 0.0822 | 0.0472 | 0.0870 | 0.0506 |
| 250 | 0.0848 | 0.0418 | 0.0884 | 0.0450 |
| 500 | 0.0972 | 0.0492 | 0.0888 | 0.0436 |
| 1000 | 0.0946 | 0.0486 | 0.1032 | 0.0530 |

Standard errors are approximately 0.003.

**Table 2** Empirical power of the test of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ based on 5000 replications.

| | Model 1 versus 1A | | Model 2 versus 2A | |
|---|---|---|---|---|
| $n$ | 10% | 5% | 10% | 5% |
| 100 | 0.4054 | 0.3238 | 0.9206 | 0.8248 |
| 250 | 0.6912 | 0.5968 | 0.9984 | 0.9970 |
| 500 | 0.8998 | 0.8488 | 1.0000 | 1.0000 |
| 1000 | 0.9906 | 0.9822 | 1.0000 | 1.0000 |

Under $H_0$ the data follow Model 1, resp. 2, under $H_A$, Model 1A, resp. 2A.

specified and large if it is not, because then the variance implied by the estimated model would be very different from the sample variance. This idea is related to the approach known as *variance targeting*, see Engle and Mezrich (1996) and Ding and Engle (2001). Variance targeting, developed primarily for multivariate GARCH models, involves reparametrizing the model in such a way the the unconditional variance becomes one of the parameters, for example, the second equation in (33) can be written as

$$\sigma_k^2 = (1 - \alpha - \beta)\mathbf{E}\sigma_0^2 + \alpha y_{k-1}^2 + \beta\sigma_{k-1}^2 \tag{36}$$

with parameters $\mathbf{E}\sigma_0^2$, $\alpha$, and $\beta$. A two-step estimation procedure is then applied in which $\mathbf{E}\sigma_0^2$ is first estimated by the sample variance, and then $\alpha$ and $\beta$ by maximum likelihood. This approach reduces the computational complexity and guarantees that the implied variance is equal to the sample variance. On the contrary, it makes it impossible to roughly check whether the selected model is correct by comparing the implied variance to the sample variance because they are automatically identically equal. It is not our objective to explore the latter idea in depth, but it may be helpful to briefly discuss some aspects of this problem.

We first take a look at the finite-sample distribution of the statistic $T_n(\widehat{\boldsymbol{\theta}})$ defined by (5) whose asymptotic distribution is normal with mean zero. Figure 1 shows QQ-plots against standard normal quantiles when the data follow Model 1. For $n = 1000$, the distribution is approximately normal, but the mean is positive. For $n = 250$, the distribution has heavy tails and about two-thirds of its mass is concentrated on the positive half-line. The reason why the asymptotic theory does not provide a reasonable finite-sample approximation appears to lie in the well-known finite-sample biases of the MLEs if one of the parameters $\alpha$ and $\beta$ is greater than approximately 0.3. Using computer-intensive estimation methods, it is possible to reduce these biases, see Gourieroux, Renault, and Touzi (2000), and this might allow our theory to be more readily applied.

This is of relevance as many goodness-of-fit tests proposed in the GARCH context have very poor power. For example, if we test GARCH(1, 1) specification against the alternative GARCH(0, 2) using the Ljung-Box test with squared residuals, see for example, Section 7.4.2. of Zivot and Wang (2003), the power is

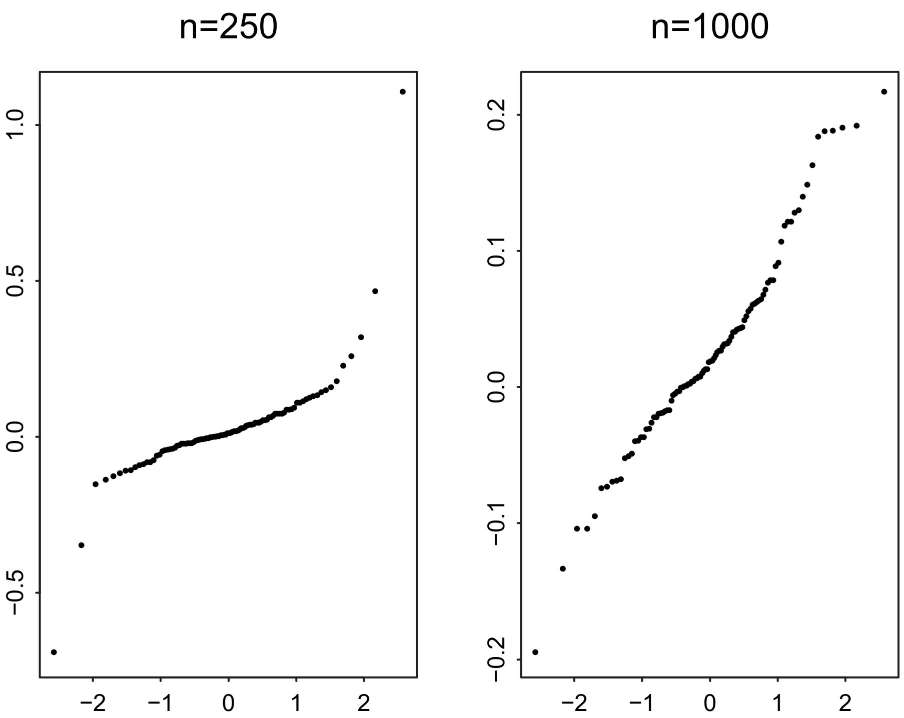


**Figure 1** QQ-plots against standard normal quantiles for the statistic $T_n(\hat{\boldsymbol{\theta}})$ defined by (5) when the data follow Model 1. Each plot is based on 100 replications of the statistic.

often smaller than the nominal size. In an experiment in which the data generating process is GARCH(0, 2) with $\omega = 0.1$, $\beta_1 = 0$, $\beta_2 = 0.5$, and $n = 1000$, the 5% Ljung-Box test using 12 lags has empirical size of 2.3%. Similar findings are reported by Lundbergh and Teräsvirta (2002) and Horváth, Kokoszka, and Teyssière (2004). There are some good directional test, for example, an alternative distribution of residuals can be detected using the tests proposed by Horváth, Kokoszka, and Teyssière (2001, 2004). A bootstrap test proposed by Hidalgo and Zaffaroni (2005) has excellent power against misspecification of the functional form of conditional variance.

## 4 PROOFS

**Proof of Theorem 1** We will establish the representation

$$\widehat{Y}_n - h(\boldsymbol{\theta}) = \frac{1}{n} \sum_{1 \le k \le n} \frac{h(\boldsymbol{\theta})}{c_0(\boldsymbol{\theta})} \sigma_k^2 (\varepsilon_k^2 - 1) + o_{\mathbf{P}}(n^{-1/2}). \tag{37}$$

The summands $\sigma_k^2(\varepsilon_k^2 - 1)$ are martingale differences with respect to the $\sigma$-algebras $\mathcal{F}_k$ generated by the random variables $\varepsilon_i$, $-\infty < i \leq k$. Hence, using a martingale convergence theorem [cf. Hall and Heyde (1980)], we obtain that the arithmetic average on the right-hand side of (37) has, when multiplied by $\sqrt{n}$, the same asymptotic distribution as specified in the formulation of Theorem 1.

To establish (37), note that by the first equation of (1),

$$\widehat{Y}_n = \frac{1}{n} \sum_{1 \leq k \leq n} \sigma_k^2(\varepsilon_k^2 - 1) + \frac{1}{n} \sum_{1 \leq k \leq n} \sigma_k^2 \tag{38}$$

Next, using Equation (1), we express the arithmetic average on the right-hand side of equation (38) in terms of $y_k$'s. We start with the equalities

$$\begin{aligned} \frac{1}{n} \sum_{1 \leq k \leq n} \sigma_k^2 &= \omega + \frac{1}{n} \sum_{1 \leq k \leq n} \left( \sum_{1 \leq i \leq p} \alpha_i y_{k-i}^2 + \sum_{1 \leq j \leq q} \beta_j \sigma_{k-j}^2 \right) \\ &= \omega + \frac{1}{n} \sum_{1 \leq i \leq p} \alpha_i \sum_{1-i \leq k \leq n-i} y_k^2 + \frac{1}{n} \sum_{1 \leq j \leq q} \beta_j \sum_{1-j \leq k \leq n-j} \sigma_k^2 \\ &= \omega + \sum_{1 \leq i \leq p} \alpha_i \frac{1}{n} \sum_{1 \leq k \leq n} y_k^2 + \sum_{1 \leq j \leq q} \beta_j \frac{1}{n} \sum_{1 \leq k \leq n} \sigma_k^2 + o_{\mathbf{P}}(n^{-1/2}). \end{aligned} \tag{39}$$

From (39), we obtain the representation

$$\frac{1}{n} \sum_{1 \leq k \leq n} \sigma_k^2 = \frac{\omega}{1 - \sum \beta_j} + \frac{\sum \alpha_i}{1 - \sum \beta_j} \widehat{Y}_n + o_{\mathbf{P}}(n^{-1/2}). \tag{40}$$

Using (40) on the right-hand side of (38), we obtain the equality

$$\widehat{Y}_n = \frac{\omega}{1 - \sum \alpha_i - \sum \beta_j} + \left( \frac{1 - \sum \beta_j}{1 - \sum \alpha_i - \sum \beta_j} \right) \frac{1}{n} \sum_{1 \leq k \leq n} \sigma_k^2(\varepsilon_k^2 - 1) + o_{\mathbf{P}}(n^{-1/2}). \tag{41}$$

We recall that the first ratio on the right-hand side of (41) equals $h(\boldsymbol{\theta})$. Furthermore, using the well-known expression for $c_0(\boldsymbol{u})$ in the definition of $w_k(\boldsymbol{u})$ [cf. Berkes, Horváth, and Kokoszka (2003)], we have that the second ratio (in parentheses) on the right-hand side of (41) equals $h(\boldsymbol{\theta})/c_0(\boldsymbol{\theta})$. We have therefore arrived at representation (37). □

**Proof of Theorem 2** We start with the equality

$$\widehat{Y}_n - h(\widehat{\boldsymbol{\theta}}) = \left(\widehat{Y}_n - h(\boldsymbol{\theta})\right) - \left(h(\widehat{\boldsymbol{\theta}}) - h(\boldsymbol{\theta})\right). \tag{42}$$

In light of (37), our goal is to express the difference $h(\widehat{\boldsymbol{\theta}}) - h(\boldsymbol{\theta})$ as an arithmetic average with a remainder term of the order $o_{\mathbf{P}}(n^{-1/2})$. This can be achieved using representation (8).

By (8), $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| = o_{\mathbf{P}}(n^{-1/2})$. Because $h$ is a rational function well defined in a neighborhood of $\boldsymbol{\theta}$, the Taylor formula implies that

$$h(\widehat{\boldsymbol{\theta}}) - h(\boldsymbol{\theta}) = \langle \mathcal{D}h(\boldsymbol{\theta}), \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \rangle + o_{\mathbf{P}}(n^{-1/2}), \tag{43}$$

where $\mathcal{D}h(\boldsymbol{\theta})$ is the vector of the first partial derivatives of $h(\boldsymbol{\theta})$, and $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors. Hence, we have the desired asymptotic representation

$$h(\widehat{\boldsymbol{\theta}}) - h(\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{2}(1 - \varepsilon_k^2)\left\langle \mathcal{D}h(\boldsymbol{\theta}), \frac{w_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})}\mathbf{B}_0^{-1} \right\rangle + o_{\mathbf{P}}(n^{-1/2}). \tag{44}$$

Using representations (37) and (44) on the right-hand side of (42), we obtain that

$$\widehat{Y}_n - h(\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{1 \le k \le n} (\varepsilon_k^2 - 1)\eta_k + o_{\mathbf{P}}(n^{-1/2}), \tag{45}$$

where

$$\eta_k := \frac{h(\boldsymbol{\theta})}{c_0(\boldsymbol{\theta})}\sigma_k^2 + \frac{1}{2}\left\langle \mathcal{D}h(\boldsymbol{\theta}), \frac{w_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})}\mathbf{B}_0^{-1} \right\rangle.$$

The summands $(\varepsilon_k^2 - 1)\eta_k^2$ are also martingale differences with respect to the $\sigma$-algebras $\mathcal{F}_k$ generated by the $\varepsilon_i$, $-\infty < i \le k$. Hence, (10) follows again from martingale central limit theorem. □

**Proof of Theorem 3** Using the first equation of (11) and also the structural property in (12), we have the following equalities:

$$\begin{aligned}
\widehat{Y}_n(\Lambda) &= \frac{1}{n}\sum_{1 \le k \le n} \left[\Lambda(\sigma_k^2)a(\varepsilon_k^2) + b(\varepsilon_k^2)\right] \\
&= \frac{1}{n}\sum_{1 \le k \le n} \Lambda(\sigma_k^2)\left\{a(\varepsilon_k^2) - \mathbf{E}\left[a(\varepsilon_k^2)\right]\right\} + \frac{1}{n}\sum_{1 \le k \le n} \left\{b(\varepsilon_k^2) - \mathbf{E}\left[b(\varepsilon_k^2)\right]\right\} \\
&\quad + \mathbf{E}\left[a(\varepsilon_0^2)\right]\frac{1}{n}\sum_{1 \le k \le n} \Lambda(\sigma_k^2) + \mathbf{E}\left[b(\varepsilon_0^2)\right].
\end{aligned} \tag{46}$$

Later in our considerations, we shall use the following notations for the first two sums on the right-hand side of Equation (46):

$$\mathcal{S}_n(1) := \frac{1}{\sqrt{n}} \sum_{1\le k\le n} \Lambda(\sigma_k^2)\big\{a(\varepsilon_k^2) - \mathbf{E}\big[a(\varepsilon_k^2)\big]\big\},$$
$$\mathcal{S}_n(2) := \frac{1}{\sqrt{n}} \sum_{1\le k\le n} \big\{b(\varepsilon_k^2) - \mathbf{E}\big[b(\varepsilon_k^2)\big]\big\}.$$

Note that the summands in both $S_n(1)$ and $S_n(2)$ are martingale differences with respect to the $\sigma$-algebras $\mathcal{F}_k$ generated by the random variables $\varepsilon_i,\ -\infty < i \le k$. We shall appeal to this observation later in the proof.

Using now the second equation in (11), we derive an expression for the sample average of $\Lambda(\sigma_k^2)$ on the right-hand side of Equation (46):

$$\begin{aligned}
\frac{1}{n}\sum_{1\le k\le n}\Lambda(\sigma_k^2) &= \frac{1}{n}\sum_{2\le k\le n+1}\Lambda(\sigma_k^2) + \frac{1}{n}\Lambda(\sigma_1^2) - \frac{1}{n}\Lambda(\sigma_{k+1}^2)\\
&= \frac{1}{n}\sum_{2\le k\le n+1}\big[\Lambda(\sigma_{k-1}^2)c(\varepsilon_{k-1}) + d(\varepsilon_{k-1})\big] + o_{\mathbf{P}}(n^{-1/2})\\
&= \frac{1}{n}\sum_{1\le k\le n}\Lambda(\sigma_k^2)c(\varepsilon_k) + \frac{1}{n}\sum_{1\le k\le n} d(\varepsilon_k) + o_{\mathbf{P}}(n^{-1/2})\\
&= \frac{1}{n}\sum_{1\le k\le n}\Lambda(\sigma_k^2)\{c(\varepsilon_k) - \mathbf{E}[c(\varepsilon_k)]\} + \frac{1}{n}\sum_{1\le k\le n}\{d(\varepsilon_k) - \mathbf{E}[d(\varepsilon_k)]\}\\
&\quad + \mathbf{E}[c(\varepsilon_0)]\frac{1}{n}\sum_{1\le k\le n}\Lambda(\sigma_k^2) + \mathbf{E}[d(\varepsilon_0)] + o_{\mathbf{P}}(n^{-1/2}). \qquad (47)
\end{aligned}$$

Before proceeding, we introduce additional notation for the first two sums on the right-hand side of Equation (47):

$$S_n(3) := \frac{1}{\sqrt{n}}\sum_{1\le k\le n}\Lambda(\sigma_k^2)\{c(\varepsilon_k) - \mathbf{E}[c(\varepsilon_k)]\},$$
$$S_n(4) := \frac{1}{\sqrt{n}}\sum_{1\le k\le n}\{d(\varepsilon_k) - \mathbf{E}[d(\varepsilon_k)]\}.$$

The summands in both $S_n(3)$ and $S_n(4)$ are martingale differences with respect to the earlier introduced $\sigma$-algebras $\mathcal{F}_k$. We do not need this observation at the moment but will use it later.

From Equation (47), we obtain the following formula for the sample average of $\Lambda(\sigma_k^2)$:

$$\frac{1}{n}\sum_{1\le k\le n}\Lambda(\sigma_k^2) = \frac{S_n(3) + S_n(4)}{\sqrt{n}\,\{1 - \mathbf{E}[c(\varepsilon_0)]\}} + \frac{\mathbf{E}[d(\varepsilon_0)]}{1 - \mathbf{E}[c(\varepsilon_0)]} + o_{\mathbf{P}}(n^{-1/2}). \qquad (48)$$

Using formula (48) on the right-hand side of Equation (46), we obtain the following one:

$$\widehat{Y}_n(\Lambda) = \frac{S_n(1) + S_n(2)}{\sqrt{n}} + \frac{\mathbf{E}\left[a(\varepsilon_0^2)\right]}{1 - \mathbf{E}[c(\varepsilon_0)]}\frac{S_n(3) + S_n(4)}{\sqrt{n}} + \frac{\mathbf{E}\left[a(\varepsilon_0^2)\right]\mathbf{E}[d(\varepsilon_0)]}{1 - \mathbf{E}[c(\varepsilon_0)]} + \mathbf{E}\left[b(\varepsilon_0^2)\right] + o_{\mathbf{P}}(n^{-1/2}). \tag{49}$$

Note that the third and the fourth summands on the right-hand side of equation (49) give us the earlier introduced $H(\boldsymbol{\theta})$. Hence, with the notation $\Delta_0$ given in the formulation of the theorem, we obtain from Equation (49) that

$$\sqrt{n}\,(\widehat{Y}_n(\Lambda) - H(\boldsymbol{\theta})) = M_n + o_{\mathbf{P}}(1), \tag{50}$$

where

$$M_n := S_n(1) + S_n(2) + \Delta_0 S_n(3) + \Delta_0 S_n(4).$$

We have already noted that each $S_n(i)$ is the sum of martingale differences with respect to the same filtration. Hence, the same can be said about the linear combination $M_n$ of $S_n(i)$'s. Using a martingale convergence theorem [cf. Hall and Heyde (1980)], we obtain that the asymptotic distribution of $M_n$ is centered normal with the variance $\mathbf{E}[M_n^2]$. We need to check that the latter moment equals $T_0^2$. Toward this end, we first calculate the second moment of every summand in the definition of $M_n$ and obtain the four equations:

$$\begin{aligned}
\mathbf{E}\left[S_n^2(1)\right] &= \mathbf{E}\left[\Lambda^2(\sigma_0^2)\right]\mathbf{Var}\left[a(\varepsilon_0^2)\right], \\
\mathbf{E}\left[S_n^2(2)\right] &= \mathbf{Var}\left[b(\varepsilon_0^2)\right], \\
\Delta_0^2\mathbf{E}\left[S_n^2(3)\right] &= \Delta_0^2\mathbf{E}\left[\Lambda^2(\sigma_0^2)\right]\mathbf{Var}[c(\varepsilon_0)], \\
\Delta_0^2\mathbf{E}\left[S_n^2(4)\right] &= \Delta_0^2\mathbf{Var}[d(\varepsilon_0)].
\end{aligned}$$

Next, we calculate the six mixed moments of the four summands in the definition of $M_n$:

$$\begin{aligned}
2\mathbf{E}[S_n(1)S_n(2)] &= 2\mathbf{E}\left[\Lambda(\sigma_0^2)\right]\mathbf{Cov}\left[a(\varepsilon_0^2), b(\varepsilon_0^2)\right], \\
2\Delta_0\mathbf{E}[S_n(1)S_n(3)] &= 2\Delta_0\mathbf{E}\left[\Lambda^2(\sigma_0^2)\right]\mathbf{Cov}\left[a(\varepsilon_0^2), c(\varepsilon_0)\right], \\
2\Delta_0\mathbf{E}[S_n(1)S_n(4)] &= 2\Delta_0\mathbf{E}\left[\Lambda(\sigma_0^2)\right]\mathbf{Cov}\left[a(\varepsilon_0^2), d(\varepsilon_0)\right], \\
2\Delta_0\mathbf{E}[S_n(2)S_n(3)] &= 2\Delta_0\mathbf{E}\left[\Lambda(\sigma_0^2)\right]\mathbf{Cov}\left[b(\varepsilon_0^2), c(\varepsilon_0)\right], \\
2\Delta_0\mathbf{E}[S_n(2)S_n(4)] &= 2\Delta_0\mathbf{Cov}\left[b(\varepsilon_0^2), d(\varepsilon_0)\right], \\
2\Delta_0^2\mathbf{E}[S_n(3)S_n(4)] &= 2\Delta_0^2\mathbf{E}\left[\Lambda(\sigma_0^2)\right]\mathbf{Cov}[c(\varepsilon_0), d(\varepsilon_0)].
\end{aligned}$$

Summing up the right-hand sides of the 10 equations above, we obtain the formula of $T_0^2$ given in Theorem 3. This finishes the proof of the theorem. □

*Received July 12, 2005; revised July 17, 2006; accepted July 20, 2006*

## REFERENCES

Aue, A., I. Berkes, and L. Horváth. (2006). "Strong Approximation for the Sums of Augmented GARCH Sequences." *Bernoulli 12(4)*, 583–608.

Berkes, I., and L. Horváth. (2004). "The Efficiency of the Estimators of the Parameters in GARCH Processes." *The Annals of Statistics 32*, 633–665.

Berkes, I., L. Horváth, and P. S. Kokoszka. (2003). "GARCH Processes: Structure and Estimation." *Bernoulli 9*, 201–227.

Berkes, I., L. Horváth, and P. S. Kokoszka. (2004). "Probabilistic and Statistical Properties of GARCH Processes." In L. Horváth and B. Szyszkowicz (eds.), *Asymptotic Methods in Stochastics: Festschrift for Miklós Csörgö*. Fields Institute Communications, vol. 44, pp. 409–430. Providence, RI: American Mathematical Society.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics 31*, 307–327.

Bougerol, P., and N. Picard. (1992a). "Strict Stationarity of Generalized autoregressive processes." *Annals of Probability 20*, 1714–1730.

Bougerol, P., and N. Picard. (1992b). "Stationarity of GARCH Processes and of Some Nonnegative Time Series." *Journal of Econometrics 52*, 115–127.

Castillo, E., A. Iglesias and R. Ruiz-Cobo. (2005). *Functional Equations in Applied Sciences*. Amsterdam: Elsevier.

Ding, Z., and R. Engle. (2001). "Large Scale Conditional Covariance Modeling, Estimation and Testing." *Academia Economic Papers 29*, 157–184.

Ding, Z., C. W. J. Granger, and R. Engle. (1993). "A Long Memory Property of Stock Marked Returns and a New Model." *Journal of Empirical Finance 1*, 83–106.

Duan, J. C. (1997). "Augmented GARCH($p$, $q$) Process and its Diffusion Limit." *Journal of Econometrics 79*, 97–127.

Engle, R. F., and J. Mezrich. (1996). "GARCH for Groups." *Risk 9*, 36–40.

Francq, C., and J-M. Zakoian. (2004). "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes." *Bernoulli 10*, 605–637.

Geweke, J. (1986). "Modeling the Persistence of Conditional Variances: A Comment." *Econometric Reviews 5*, 57–61.

Glosten, L. R., R. Jagannathan, and D. Runkle. (1993). "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance 48*, 1779–1801.

Gourieroux, C., E. Renault, and N. Touzi. (2000). "Calibration and Simulation for Small Sample Bias Correction." In R. Mariano, T. Schuermann, and M. J. Weeks (eds.), *Simulation-Based Inference in Econometrics*. Cambridge, England: Cambridge University Press.

Hall, P., and C. C. Heyde. (1980). *Martingale Limit Theory and its Applications*. New York: Academic Press.

He, C., and T. Teräsvirta. (1999). "Fourth Moment Structure of the GARCH($p$, $q$) Process." *Econometric Theory 15*, 824–846.

Hidalgo, J., and P. Zaffaroni. (2005). "A Goodness of Fit Test for ARCH($\infty$) Models." Technical Report. Imperial College. Available at http://www3.imperial.ac.uk/people/p.zaffaroni.

Horváth, L., P. S. Kokoszka, and G. Teyssière. (2001). "Empirical Process of the Squared Residuals of an ARCH Sequence." *The Annals of Statistics 29*, 445–469.

Horváth, L., P. S. Kokoszka, and G. Teyssière. (2004). "Bootstrap Misspecification Tests for ARCH Based on the Empirical Process of Squared Residuals." *Journal of Statistical Computation and Simulation 74*, 469–485.

Kokoszka, P., G. Teyssière, and A. Zhang. (2004). "Confidence Intervals for the Autocorrelations of the Squares of GARCH Sequences." In M. Bubak, G. Van Albada, P. Sloot and J. Dongarra (eds.), *Computational Science–ICCS 2004*, Lecture Notes in Computer Science, 3039:837–844. Berlin: Springer Verlag.

Li, W. K., S. Ling, and M. McAleer. (2002). "A Survey of Recent Theoretical Results for Time Series Models with GARCH Errors." *Journal of Economic Surveys 16*, 245–269.

Ling, S., and M. McAleer. (2002a). "Necessary and Sufficient Moment Conditions for the GARCH(p, q) and Asymmetric Power GARCH(p, q) Models." *Econometric Theory 18*, 722–729.

Ling, S., and M. McAleer. (2002b). "Stationarity and the Existence of Moments of a Family of GARCH Processes." *Journal of Econometrics 106*, 109–117.

Lundbergh, S., and T. Teräsvirta. (2002). "Evaluating GARCH Models." *Journal of Econometrics 110*, 417–435.

Nelson, D. B. (1990). "Stationarity and persistence in the GARCH(1; 1) model." *Econometric Theory 6*, 318–334.

Peng, L., and Q. Yao. (2003). "Least Absolute Deviations Estimation for ARCH and GARCH Models." *Biometrika 90*, 967–975.

Straumann, D. (2005). *Estimation in Conditionally Heteroskedastic Time Series Models*. Lecture Notes in Statistics, vol. 181. Berlin: Springer.

Zivot, E., and J. Wang. (2003). *Modeling Financial Time Series with S-Plus*. New York: Springer.