A Closed-Form Estimator for the GARCH(1,1) Model
Author(s): Dennis Kristensen and Oliver Linton
Source: *Econometric Theory*, Vol. 22, No. 2 (Apr., 2006), pp. 323-337
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/4093228

Econometric Theory, **22**, 2006, 323–337. Printed in the United States of America.
DOI: 10.1017/S0266466606060142

NOTES AND PROBLEMS

A CLOSED-FORM ESTIMATOR FOR THE GARCH(1,1) MODEL

DENNIS KRISTENSEN
University of Wisconsin-Madison

OLIVER LINTON
London School of Economics

We propose a closed-form estimator for the linear GARCH(1,1) model. The estimator has the advantage over the often used quasi-maximum likelihood estimator (QMLE) that it can be easily implemented and does not require the use of any numerical optimization procedures or the choice of initial values of the conditional variance process. We derive the asymptotic properties of the estimator, showing $T^{(\kappa-1)/\kappa}$-consistency for some $\kappa \in (1,2)$ when the fourth moment exists and $\sqrt{T}$-asymptotic normality when the eighth moment exists. We demonstrate that a finite number of Newton–Raphson iterations using our estimator as starting point will yield asymptotically the same distribution as the QMLE when the fourth moment exists. A simulation study confirms our theoretical results.

1. INTRODUCTION

The estimation of the Bollerslev (1986) generalized autoregressive conditional heteroskedasticity (GARCH) model is often carried out using the quasi-maximum likelihood estimator (QMLE). The asymptotic properties of this estimator are by now well established (see Jeantheau, 1998; Lee and Hansen, 1994; Lumsdaine, 1996; Ling and McAleer, 2003), and its finite-sample properties have been examined through Monte Carlo studies (Fiorentini, Calzolari, and Panattoni, 1996; Lumsdaine, 1995). Recently, Giraitis and Robinson (2001) have proposed using the Whittle likelihood estimator, which exploits only the second-order properties of the series but does so in an optimal way. Unfortunately, the calculation of both these estimators requires the use of numerical optimization procedures because closed-form expressions are not available.

The first author's research was supported by the Shoemaker Foundation. The second author's research was supported by the Economic and Social Science Research Council of the United Kingdom. Address correspondence to Dennis Kristensen, Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706, USA; e-mail: dkristen@ssc.wisc.edu. Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom; e-mail: lintono@lse.ac.uk.

 0266-4666/06 $12.00

Therefore, the resulting estimator depends on the implementation, with different optimization techniques leading to potentially different estimators. This has been demonstrated in the studies by Brooks, Burke, and Persand (2001) and McCullough and Renfro (1999) where different commercially available software packages were used to estimate GARCH models by quasi-maximum likelihood (QML). Both studies reported markedly different outputs across the various packages, reflecting the different initialization and algorithmic strategies employed.

We propose a simple estimator of the parameters in the GARCH(1,1) model based on the second-order properties of the series like Giraitis and Robinson (2001). We derive an autoregressive moving average (ARMA) representation of the squared GARCH process and use the implied autocovariance function to define closed-form estimators of the parameters of the GARCH model. This strategy has already been used in the literature on estimating standard ARMA models (see Tuan, 1979; Galbraith and Zinde-Walsh, 1994) and in ARFIMA models recently by Mayoral (2004). The ARMA representation of GARCH models was already noted by Bollerslev (1986), whereas Francq and Zakoïan (2000) and Gouriéroux and Jasiak (2001, p. 130) proposed to utilize this to obtain estimators of the parameters. However, the estimators of Francq and Zakoïan (2000) and Gouriéroux and Jasiak (2001) require numerical optimization. Baillie and Chung (2001) took a somewhat similar approach using a minimum distance estimator based on the autocorrelation function of the squared GARCH process; again, numerical optimization is required to obtain the actual estimator.

Our estimator can readily be implemented without using numerical optimization methods. Furthermore, the estimator does not require one to choose (arbitrary) initial values for the conditional variance process. In that sense, the proposed estimator is more robust compared to the aforementioned estimators. On the other hand, for the estimator to be consistent and asymptotically normally distributed, relatively strong assumptions about the moments of the GARCH process have to be made. Although consistency and asymptotic normality of the QMLE can be shown under virtually no moment restrictions of the GARCH process, we require fourth moments to obtain consistency and $T^{(\kappa-1)/\kappa}$-convergence toward a so-called α-stable distribution with index $\kappa \in (1,2)$, and eighth moments for $\sqrt{T}$-asymptotic normality.

The $\sqrt{T}$-asymptotic normality result when the eighth moment exists follows from standard central limit theorems because the asymptotic variance in this case is well defined. When the eighth moment does not exist, we are still able to show that a limiting distribution exists, but it does not have second moment, and the convergence rate is slowed down. The latter result is based on recent results by Mikosch and Stărică (2000). By combining our estimator with a finite-order Newton–Raphson (NR) procedure one can achieve $\sqrt{T}$-asymptotic normality, and even full efficiency under Gaussianity of the rescaled errors, based on fourth moments.

2. AN ARMA REPRESENTATION OF THE GARCH(1,1) MODEL

We consider the GARCH(1,1) process given by

$$y_t = \sigma_t z_t, \tag{1}$$

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha y_{t-1}^2. \tag{2}$$

We assume that $E[z_t|\mathcal{F}_{t-1}] = 0$ and $E[z_t^2|\mathcal{F}_{t-1}] = 1$, where $\mathcal{F}_t$ is the σ-field generated by $\{z_t, z_{t-1}, \ldots\}$. We can write $x_t \equiv y_t^2$ as

$$x_t = \omega + \phi x_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1},$$

where $\varepsilon_t = x_t - \sigma_t^2$ is a martingale difference sequence with respect to $\mathcal{F}_t$, $\phi = \alpha + \beta > 0$, and $\theta = -\beta < 0$. From this expression, we see that x_t is a (heteroskedastic) ARMA(1,1) process with parameters ϕ and θ. We shall throughout assume that $\phi < 1$, implying that $E[x_t] < \infty$ (cf. Bollerslev, 1988). We introduce the covariance function of the process

$$\gamma(k) = E[(x_{t+k} - E[x_t])(x_t - E[x_t])].$$

Assuming that $\{x_t\}$ is stationary with second moment, $\gamma(\cdot)$ is well defined. Using standard results, we then have that the autocorrelation function, $\rho(k) = \gamma(k)/\gamma(0)$, solves the following set of Yule–Walker equations:

$$\rho(k) = \phi\rho(k-1), \qquad k = 2,3,\ldots, \tag{3}$$

$$\rho(1) = \frac{(1+\phi\theta)(\phi+\theta)}{1+\theta^2+2\phi\theta} \tag{4}$$

(cf. Harvey, 1993, Ch. 2, eqns. (4.13a) and (4.13b); these equations were also derived in Bollerslev, 1988, and He and Teräsvirta, 1999). We can express (4) as a quadratic equation in θ,

$$\theta^2 + b\theta + 1 = 0, \qquad b \equiv \frac{\phi^2 + 1 - 2\rho(1)\phi}{\phi - \rho(1)}.$$

Observe that b is only well defined if $\phi \neq \rho(1)$. It is easily checked that $\rho(1) \geq \phi$ with equality if and only if $\phi^2 = 1$ or $\theta = 0$. The first case is ruled out by our assumption that $\phi < 1$, whereas in the following discussion we assume $\beta > 0$.[1] Under this assumption, $b > 2$ is well defined, and a solution to the quadratic equation is given by

$$\theta = \frac{-b + \sqrt{b^2 - 4}}{2}. \tag{5}$$

There is a second root that is reciprocal to the one stated here; however this has $|\theta| = \beta > 1$, which is ruled out by $\phi < 1$.[2] Finally, we observe that

$$\omega = \sigma^2(1-\phi), \qquad \sigma^2 \equiv E(y_t^2). \tag{6}$$

This last expression is utilized by Engle and Sheppard (2001) to profile out ω in their estimation procedure.

The expressions (4), (5), and (6) can now be used to obtain estimators of the parameters α, β, and ω. First, we can estimate ϕ by $\hat{\phi} = \hat{\rho}(2)/\hat{\rho}(1)$, where $\hat{\rho}(\cdot)$ is the sample autocorrelation function of x_t, $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$ with

$$\hat{\gamma}(k) = \frac{1}{T-k}\sum_{t=1}^{T-k}(x_{t+k}-\hat{\sigma}^2)(x_t-\hat{\sigma}^2), \tag{7}$$

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} x_t. \tag{8}$$

Substituting the estimator of ϕ into (5), we obtain an estimator of θ,

$$\hat{\theta} = \frac{-\hat{b}+\sqrt{\hat{b}^2-4}}{2}, \qquad \hat{b} \equiv \frac{\hat{\phi}^2+1-2\hat{\rho}(1)\hat{\phi}}{\hat{\phi}-\hat{\rho}(1)},$$

assuming that $\hat{b} \geq 2$. This leads to the following estimators of $\lambda = (\alpha, \beta, \omega)^\top$:

$$\hat{\alpha} = \hat{\theta} + \hat{\phi}, \qquad \hat{\beta} = -\hat{\theta}, \qquad \hat{\omega} = \hat{\sigma}^2(1-\hat{\phi}). \tag{9}$$

In practice, this method may lead to $\hat{\phi} < 0$ or $\hat{\phi} > 1$. To deal with this problem, the estimator can be Winsorized (censored) at zero and one or at ϵ and $1-\epsilon$ for small positive ϵ.

Note that

$$\phi = \sum_{j=1}^{\infty} w_j \frac{\rho(j+1)}{\rho(j)} \tag{10}$$

for any w_j sequence with $\sum_{j=1}^{\infty} w_j = 1$ so that a more general class of estimators can be defined based on this relationship. It can be expected that for a sufficiently general class of weights one can obtain the same efficiency as the Whittle estimator of Giraitis and Robinson (2001). We do not pursue this approach to achieving efficiency because a more standard approach is available; see the discussion that follows. Nevertheless, some smoothing of the ratio of autocorrelations based on (10) may be desirable in practice.

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

We derive the asymptotic properties of our estimator $\hat{\lambda} \equiv (\hat{\alpha}, \hat{\beta}, \hat{\omega})^\top$ under the following set of assumptions:

A.1. The error process $\{z_t\}$ is independent and identically distributed (i.i.d.) with marginal distribution given by a lower semicontinuous density with support $\mathbb{R}$. It satisfies $E[z_t] = 0$ and $E[z_t^2] = 1$.

A.2. The observed process $\{y_t\}$ is strictly stationary with $E[(\beta + \alpha z_t^2)^2] < 1$.

A.3. $E[(\beta + \alpha z_t^2)^4] < 1$.

Under (A.1), the moment condition in (A.2) is necessary and sufficient for the GARCH model to have a strictly stationary solution with a fourth moment. This solution will be β-mixing with geometrically decreasing mixing coefficients (cf. Meitz and Saikkonen, 2004, Thm. 1). We then assume that this is the one we have observed. Assumption (A.3) is a strengthening of (A.2) implying that also the eighth moment of y_t exists.

In some of the results stated later, the i.i.d. assumption in (A.1) can be weakened to $\{z_t\}$ being a stationary martingale difference sequence; this is more realistic, allowing for dependence in the rescaled errors. In (A.2) we assume that we have observed the stationary version of the process. This is merely for technical convenience and can most likely be removed.

Under (A.1)–(A.2), we prove consistency of the estimator and derive its asymptotic distribution (toward which it converges with rate slower than $T^{-1/2}$). If additionally (A.3) holds, the estimator is shown to be $\sqrt{T}$-asymptotically normally distributed.

We first state a lemma that shows that the basic building blocks of our estimator are consistent and gives their asymptotic distribution. Under (A.2), $E[y_t^4] < \infty$, but the eighth moment does not necessarily exist. One can show for ARMA processes with homoskedastic errors that the sample autocorrelation function is asymptotically normally distributed with only a second moment of the errors. In our case however, the errors, ε_t, are heteroskedastic, and to obtain asymptotic normality the fourth moment, $E[\varepsilon_t^4] < \infty$, seems to be needed; see, for example, Hannan and Heyde (1972) for results in both the homoskedastic and heteroskedastic cases. The fourth moment of ε_t translates into the eighth moment of y_t.

The estimator $\hat{\gamma}(k)$ has a well-defined asymptotic distribution without requiring the eighth moment of the GARCH process to exist however, but the limit will not be Gaussian. This result has been established in Mikosch and Stărică (2000), see also Basrak, Davis, and Mikosch (2002) and Davis and Mikosch (1998). They show that $\hat{\gamma}(k)$ converges in distribution toward a so-called stable, regularly varying distribution with index $\kappa \in (1,2)$; see Samorodnitsky and Taqqu (1994) and Resnick (1987) for an introduction. The index κ is shown to be the solution to $E[(\alpha z_t^2 + \beta)^\kappa] = 1$. The convergence takes place with rate Ta_T^{-1} where $a_T = T^{1/\kappa} l(T)$ and $l(T)$ is a slowly varying function. The sequence a_T satisfies $TP(y_t^4 > a_T) \to 1$, such that the index κ is a measure of the tail thickness.

LEMMA 1. *Under (A.1)–(A.2), the estimators* $\hat{R}(m) = (\hat{\rho}(k))_{k=1,\dots,m}$ *and* $\hat{\sigma}^2$ *in (7) and (8) satisfy* $\hat{R}(m) \xrightarrow{p} R(m) = (\rho(k))_{k=1,\dots,m}$, $m \geq 1$, *and* $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. *Furthermore,*

$$\sqrt{T}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, V_{\sigma^2}), \tag{11}$$

where $V_{\sigma^2} = \gamma(0) + 2\sum_{k=1}^{\infty}\gamma(k)$ *and*

$$T^{(\kappa-1)/\kappa}(\hat{R}(m) - R(m)) \xrightarrow{d} \gamma^{-1}(0)W(m), \tag{12}$$

for some $\kappa \in (1,2)$, *where* $W(m) = (Z_i - \rho(i)Z_0)_{i=1,\dots,m}$ *and* $(Z_i)_{i=0,\dots,m}$ *has a* κ*-stable distribution. If additionally (A.3) holds, then*

$$\sqrt{T}\begin{bmatrix} \hat{\sigma}^2 - \sigma^2 \\ \hat{R}(m) - R(m) \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} V_{\sigma^2} & O_{1\times m} \\ O_{m\times 1} & V_\rho(m) \end{bmatrix}\right), \tag{13}$$

where $V_\rho(m)$ *is the* $m \times m$ *matrix given by*

$$V_\rho(m) = \operatorname{var}(Y_t) + \sum_{k=1}^{\infty} \operatorname{cov}(Y_t, Y_{t+k}),$$

where Y_t *is an* $m \times 1$ *vector with* $Y_{t,i} = (y_t^2 - \sigma^2)(y_{t+i}^2 - \sigma^2)$, $i = 1,\dots,m$.

Mikosch and Stărică (2000) establish a weak convergence result without the fourth moment condition (A.2). But in this case, one does not have any consistency result because the law of large numbers does not hold. The preceding weak convergence result of $\hat{\gamma}(k)$ under (A.1)–(A.2) has the flaw that the limit distribution is not in explicit form, which makes it difficult in practice to carry out any inference.

The consistency results and the weak convergence results in (11) and (13) in the preceding lemma can be proved without the i.i.d. assumption on $\{z_t\}$ in (A.1) by using martingale limit theory. This can be done by utilizing the ARMA structure of $\{x_t\}$ and applying the results of Hannan and Heyde (1972). It is not clear whether the more general weak convergence result in (12) can be extended to a non-i.i.d. setting however.

Observe that our estimator of λ can be expressed in terms of $\hat{\sigma}^2$ and $\hat{R}(2)$. A simple application of the continuous mapping theorem therefore yields the following result.

THEOREM 2. *Under (A.1)–(A.2), the estimator* $\hat{\lambda} = (\hat{\alpha}, \hat{\beta}, \hat{\omega})^\top$ *given in (9) is consistent,* $\hat{\lambda} \xrightarrow{p} \lambda$, *and*

$$T^{(\kappa-1)/\kappa}(\hat{\lambda} - \lambda) \xrightarrow{d} D(\sigma^2, W(2)), \tag{14}$$

where $W(2)$ and κ are as in Lemma 1 and the function D is given subsequently in equations (A.1)–(A.3). If additionally Assumption (A.3) holds, then

$$\sqrt{T}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0,V), \quad \textit{where} \tag{15}$$

$$V = \frac{\partial D(\sigma^2,\rho(1),\rho(2))}{\partial(\sigma^2,\rho(1),\rho(2))}\begin{bmatrix} V_{\sigma^2} & O_{1\times 2} \\ O_{2\times 1} & V_\rho(2)\end{bmatrix}\left(\frac{\partial D(\sigma^2,\rho(1),\rho(2))}{\partial(\sigma^2,\rho(1),\rho(2))}\right)^{\top}. \tag{16}$$

An estimator of the covariance matrix V in (16) can be obtained by first estimating V_{σ^2} and $V_\rho(2)$ using heteroskedasticity and autocorrelation consistent (HAC) variance estimators (see Robinson and Velasco, 1997) and then substituting $\hat{\sigma}^2$, $\hat{\rho}(1)$, and $\hat{\rho}(2)$ into $\partial D(\sigma^2,\rho(1),\rho(2))/\partial(\sigma^2,\rho(1),\rho(2))$. One can alternatively use the analytic expressions of $\rho(k)$ to obtain an estimator of V_{σ^2}.

4. EFFICIENCY ISSUES

We here give a brief discussion of how one may improve on the efficiency of the closed-form estimator, in terms of both convergence rate and asymptotic variance. The basic idea is to perform a number of NR iterations using either the Whittle objective function or the Gaussian quasi-likelihood. Given closed-form estimates, one may wish to proceed to the QMLE or the Whittle estimator,[3] using the initial estimates as a starting point in the numerical optimization; this may help reduce numerical problems because our preliminary estimates are consistent. Alternatively, one can perform a number of NR iterations that do not necessitate the use of any numerical optimization procedure. We define the following sequence of NR estimators:

$$\hat{\lambda}^{\mathrm{NR}}_{k+1} = \hat{\lambda}^{\mathrm{NR}}_k - H_T^{-1}(\hat{\lambda}^{\mathrm{NR}}_k)S_T(\hat{\lambda}^{\mathrm{NR}}_k), \qquad k \geq 1,$$

with the initial value being the closed-form estimator, $\hat{\lambda}^{\mathrm{NR}}_1 = \hat{\lambda}$, $S_T(\lambda) = \partial Q_T(\lambda)/\partial\lambda$, $H_T(\lambda) = \partial^2 Q_T(\lambda)/(\partial\lambda\partial\lambda^{\top})$, and $Q_T(\lambda)$ the criterion function.[4] In the case of the QMLE, $Q_T(\lambda)$ is the Gaussian likelihood when assuming that $z_t \sim$ i.i.d. $N(0,1)$, whereas the Whittle estimator has $Q_T(\lambda)$ as the discrete frequency form (cf. Giraitis and Robinson, 2001, pp. 611–612). For the QMLE, the NR estimator takes the form of a generalized least squares (GLS) type estimator: defining $\hat{\sigma}^2_{k,t} = \hat{\omega}^{\mathrm{NR}}_k + \hat{\beta}^{\mathrm{NR}}_k\hat{\sigma}^2_{k,t-1} + \hat{\alpha}^{\mathrm{NR}}_k y^2_{t-1}$, and $X_{k,t-1} = (1, y^2_{t-1}, \hat{\sigma}^2_{k,t-1})'$, one can show that

$$\hat{\lambda}^{\mathrm{NR}}_{k+1} = \left(\sum_{t=1}^{n}\hat{\sigma}^{-4}_{k,t}X_{k,t-1}X'_{k,t-1}\right)^{-1}\left(\sum_{t=1}^{n}\hat{\sigma}^{-4}_{k,t}X_{k,t-1}y_t^2\right). \tag{17}$$

In general, the NR estimator will satisfy

$$\|\hat{\lambda}^{\mathrm{NR}}_{k+1} - \hat{\lambda}^{\mathrm{OP}}\| = O_P(\|\hat{\lambda}^{\mathrm{NR}}_1 - \hat{\lambda}^{\mathrm{OP}}\|^{2^k}), \tag{18}$$

where $\hat{\lambda}^{OP} = \arg\max Q_T(\lambda)$ is the actual M estimator (cf. Robinson, 1988, Thm. 2). Under regularity conditions, the M estimator will satisfy

$$\sqrt{T}(\hat{\lambda}^{\mathrm{OP}} - \lambda) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1}), \tag{19}$$

where $H = E[H_T(\lambda_0)]$ and $\Sigma = E[S_T(\lambda_0)S_T(\lambda_0)^\top]$ (see, e.g., Lee and Hansen, 1994, Thm. 2; Giraitis and Robinson, 2001, Thm. 1). Combining (18) and (19), we obtain the following result.

THEOREM 3. *Assume that (19) holds together with (A.1)–(A.2). Then, with κ as in Lemma 1, for $k \geq 2 + [\log(\kappa/2(\kappa - 1))/\log(2)]$,*

$$\sqrt{T}(\hat{\lambda}_k^{\mathrm{NR}} - \lambda) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1}).$$

If additionally (A.3) is satisfied, the preceding result holds for $k \geq 1$.

5. A SIMULATION STUDY

We now examine the quality of our estimator in finite sample through a Monte Carlo study. In particular, we are interested in its behavior when the fourth moment does not exist. We simulate the GARCH process given in (1)–(2) with $z_t \sim$ i.i.d. $N(0,1)$ for four different sets of parameter values. For each choice of parameter values, we simulated 5,000 data sets with T = 200, 400, 800, and 1,000 observations. For each data set, we obtained the QMLE using the Matlab GARCH Toolbox,[5] our ARMA estimator, and the GLS estimator in (17). To estimate ϕ, we used (10) with $w_j = \frac{1}{3}$, $j = 1,2,3$, and $w_j = 0$, $j > 3$, and Winsorized at ϵ and $1 - \epsilon$ with $\epsilon = 0.001$. The GLS estimator reported here is based on two iterations. For the two first sets of parameter values, $\lambda = (0.2, 0.15, 0.25)^\top$ and $\lambda = (0.2, 0.25, 0.35)^\top$, both the fourth and eighth moments exist; for the third, $\lambda = (0.2, 0.35, 0.45)^\top$, only the fourth moment is well defined; and for the fourth, $\lambda = (0.2, 0.10, 0.89)^\top$, neither the fourth nor the eighth moment exists, but the second does.

The results in terms of bias, variance, and mean squared error (MSE) are reported in Tables 1–4 for the four different sets of parameters. For these results, we have discarded a small number of data sets (eight in total) where the Matlab GARCH Toolbox stalled and did not terminate the optimization procedure. We have also set the ARMA estimator equal to zero whenever it returned negative estimates.

For the first set of parameters, the ARMA and GLS estimators are serious competitors to the QMLE. For small GARCH effects, the closed-form estimators seem to be a good alternative to the QMLE in terms of MSE. For the remaining three sets of parameters, the QMLE has significantly better performance compared to the two other estimators as expected. As α and/or β increases the variances of the ARMA and GLS estimators in general increase; this is consis-

TABLE 1. MSE of the QMLE and ARMA estimator for $\lambda = (0.2, 0.15, 0.25)$

	QMLE, $T =$				ARMA estimator, $T =$				GLS estimator, $T =$			
	200	400	800	1,000	200	400	800	1,000	200	400	800	1,000
$\omega = 0.20\ (\times 10^{-3})$	5.28	5.71	5.75	5.78	22.93	18.20	13.46	12.24	0.22	1.79	4.09	4.47
	14.94	13.42	12.21	11.93	6.63	8.19	9.53	9.70	41.10	29.19	18.39	17.66
	20.22	19.13	17.96	17.70	29.56	26.39	22.99	21.94	41.32	30.98	22.48	22.12
$\alpha = 0.15\ (\times 10^{-3})$	3.58	3.54	3.41	3.43	6.55	6.18	5.45	5.28	4.24	3.89	3.61	3.62
	10.65	8.21	6.94	6.64	5.53	5.57	5.72	5.66	10.16	8.21	6.94	6.60
	14.23	11.75	10.35	10.07	12.08	11.75	11.17	10.94	14.40	12.10	10.55	10.21
$\beta = 0.25\ (\times 10^{-3})$	13.17	10.91	11.03	10.87	26.98	13.44	2.80	1.20	23.23	17.82	13.36	12.47
	53.40	47.99	37.23	35.02	159.20	140.20	112.55	102.24	36.47	37.76	35.20	34.44
	66.56	58.90	48.26	45.89	186.18	153.63	115.34	103.44	59.70	55.58	48.56	46.91

Notes: The three elements in each cell are, from top to bottom, squared bias, variance, and MSE. The autocorrelation coefficient $\rho(1) = 0.157$, whereas the kurtosis $\kappa_4 = 0.352$.

TABLE 2. MSE of the QMLE and ARMA estimator for $\lambda = (0.2, 0.25, 0.35)$

	QMLE, T =				ARMA estimator, T =				GLS estimator, T =			
	200	400	800	1,000	200	400	800	1,000	200	400	800	1,000
$\omega = 0.20$ ($\times 10^{-3}$)	0.47	0.18	0.05	0.03	2.58	0.40	0.02	0.00	8.27	1.64	0.11	0.05
	9.83	5.71	3.13	2.56	18.71	14.39	9.15	7.56	82.38	64.20	32.03	22.80
	10.29	5.90	3.18	2.59	21.29	14.79	9.17	7.57	90.65	65.84	32.14	22.85
$\alpha = 0.25$ ($\times 10^{-3}$)	0.00	0.00	0.00	0.00	9.28	4.67	1.86	1.24	0.25	0.10	0.04	0.03
	12.77	6.36	3.26	2.61	10.58	10.22	7.49	6.34	13.60	7.48	3.93	3.16
	12.78	6.36	3.26	2.61	19.86	14.89	9.34	7.57	13.86	7.58	3.97	3.19
$\beta = 0.35$ ($\times 10^{-3}$)	2.62	1.00	0.27	0.16	67.08	24.23	6.31	3.27	9.36	2.30	0.12	0.02
	51.53	32.24	18.16	14.87	90.51	84.25	58.06	48.39	59.68	44.97	29.17	24.53
	54.15	33.23	18.42	15.03	157.58	108.48	64.37	51.65	69.04	47.27	29.29	24.55

Notes: The three elements in each cell are, from top to bottom, squared bias, variance, and MSE. The autocorrelation coefficient $\rho(1) = 0.281$, whereas the kurtosis $\kappa_4 = 0.932$.

TABLE 3. MSE of the QMLE and ARMA estimator for $\lambda = (0.2, 0.35, 0.45)$

	QMLE, $T =$				ARMA estimator, $T =$				GLS estimator, $T =$			
	200	400	800	1,000	200	400	800	1,000	200	400	800	1,000
$\omega = 0.20\ (\times 10^{-3})$	1.46	0.29	0.06	0.04	1.40	1.33	0.87	0.71	13.22	1.27	0.16	0.33
	13.37	5.56	2.41	1.81	53.69	33.83	21.97	18.98	418.89	742.68	83.66	73.06
	14.83	5.84	2.47	1.85	55.09	35.16	22.84	19.69	432.11	743.95	83.82	73.38
$\alpha = 0.35\ (\times 10^{-3})$	0.00	0.00	0.00	0.00	24.96	14.83	8.90	7.52	0.28	0.01	0.00	0.01
	15.14	7.54	3.72	2.91	16.16	15.17	12.62	12.05	19.40	11.25	6.36	5.22
	15.14	7.54	3.72	2.91	41.12	30.00	21.52	19.56	19.68	11.27	6.36	5.23
$\beta = 0.45\ (\times 10^{-3})$	2.54	0.46	0.09	0.06	24.05	9.89	4.77	3.98	4.58	0.06	0.38	0.50
	30.40	14.44	6.48	5.02	83.22	69.79	52.84	48.23	61.15	39.08	23.74	19.99
	33.53	14.90	6.57	5.08	107.28	79.68	57.62	52.21	65.74	39.14	24.12	20.48

Notes: The three elements in each cell are, from top to bottom, squared bias, variance, and MSE. The autocorrelation coefficient $\rho(1) = 0.464$, whereas the kurtosis $\kappa_4 = 9.391$.

TABLE 4. MSE of the QMLE and ARMA estimator for $\lambda = (0.2, 0.10, 0.89)$

	QMLE, $T =$				ARMA estimator, $T =$				GLS estimator, $T =$			
	200	400	800	1,000	200	400	800	1,000	200	400	800	1,000
$\omega = 0.20\ (\times 10^{-3})$	1×10^3	96.83	10.65	5.80	4×10^3	3×10^3	1×10^3	1×10^3	2×10^4	2×10^3	224.26	99.02
	6×10^3	481.07	49.35	21.18	3×10^4	1×10^4	6×10^3	5×10^3	7×10^4	7×10^3	1×10^3	909.83
	8×10^3	577.90	60.00	26.98	4×10^4	1×10^4	7×10^3	7×10^3	9×10^4	9×10^3	2×10^3	1×10^3
$\alpha = 0.10\ (\times 10^{-3})$	0.01	0.00	0.00	0.00	0.05	0.15	0.27	0.32	0.01	0.07	0.09	0.12
	3.39	1.20	0.54	0.42	3.15	3.17	3.13	3.07	7.81	5.42	3.54	3.01
	3.40	1.20	0.54	0.42	3.19	3.32	3.39	3.38	7.83	5.49	3.63	3.13
$\beta = 0.89\ (\times 10^{-3})$	6.80	0.56	0.06	0.04	10.81	9.21	6.26	5.90	62.22	7.12	0.01	0.16
	29.35	4.42	0.92	0.53	45.34	35.49	22.91	20.16	86.40	38.41	13.55	10.12
	36.15	4.98	0.98	0.56	56.15	44.69	29.17	26.06	148.62	45.53	13.56	10.27

Notes: The three elements in each cell are, from top to bottom, squared bias, variance, and MSE. The autocorrelation coefficient and the kurtosis are not well defined.

tent with the theory, which predicts that the convergence rate of the estimator deteriorates as $\alpha + \beta$ increases. However, contrary to the theory, it seems as if the ARMA estimator of α and β remains consistent even if the fourth moment is not well defined; the performance of the closed-form estimator of ω is very poor, though. The GLS estimator based on two iterations yields a large improvement in the precision relative to the initial ARMA estimator, except for the estimation of ω. The improvement might have been even more pronounced if we had done further iterations.

6. CONCLUDING REMARKS

The procedure easily extends to the case where there is a mean process, say, $y_t = \beta^\top x_t + \sigma_t z_t$ for some covariates x_t, by applying the preceding process to the residuals from some preliminary fitting of the mean. Some aspects of the procedure extend to various multivariate cases and to GARCH(p,q). Specifically, in the multivariate case the relationships (3) and (6) continue to be useful and can be used as in Engle and Sheppard (2001) to reduce the dimensionality of the optimization space.

One may also consider other GARCH models that can be represented as an ARMA-like process. For example, suppose that the conditional variance is given as $\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha y_{t-1}^2 1(y_{t-1}^2 > 0) + \delta y_{t-1}^2 1(y_{t-1}^2 \leq 0)$. Then for $x_t \equiv y_t^2$ we have $x_t = \omega + \phi x_{t-1} + u_t$, where u_t is a martingale difference sequence with respect to $\mathcal{F}_{t-2}$ and $\phi = \beta + \alpha m_2^+ + \delta m_2^-$, where $m_2^+ = E[z_t^2 1(z_t > 0)]$ and $m_2^- = E[z_t^2 1(z_t \leq 0)]$. Furthermore, $\sigma^2 = \omega/(1-\phi)$ so that we can identify ω and ϕ as before from the variance and the second-order covariance. To proceed further one needs to make strong assumptions about the distribution of z_t, for example, symmetry about zero, in which case $\phi = \beta + (\alpha + \delta)/2$.

Our estimator also allows for a simple t-test for no GARCH effect because the asymptotic distribution derived earlier does not require $\alpha, \beta > 0$ in contrast to the QMLE, where the Taylor expansion used requires this.

NOTES

1. If $\beta = 0$, $\rho(1) = \alpha$, and we estimate α by the first-order sample correlation.
2. This corresponds to the invertibility condition of the ARMA(1,1) model.
3. Observe that the Whittle estimator proposed in Giraitis and Robinson (2001) in fact is the quasi-likelihood of the process $\{x_t\}$ assuming that it is Gaussian.
4. We here assume that $H_T(\hat{\lambda}_k^{NR})$ is invertible; Robinson (1988) discusses these issues in detail. Also, the step length is kept constant; it may however be a good idea in practice to use a line search in each iteration.
5. The Matlab package uses a subspace trust region method based on the interior-reflective Newton method of Coleman and Li (1996).

REFERENCES

Baillie, R.T. & H. Chung (2001) Estimation of GARCH models from the autocorrelations of the squares of a process. *Journal of Time Series Analysis* 22, 631–650.

Basrak, B., R.A. Davis, & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95–115.

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

Bollerslev, T. (1988) On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis* 9, 121–131.

Brooks, C., S.P. Burke, & G. Persand (2001) Benchmarks and the accuracy of GARCH model estimation. *International Journal of Forecasting* 17, 45–56.

Coleman, T.F. & Y. Li (1996) An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization* 6, 418–445.

Davis, R.A. & T. Mikosch (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Annals of Statistics* 26, 2049–2080.

Engle, R.F. & K. Sheppard (2001) Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. NBER Working paper w8554.

Fiorentini, G., G. Calzolari, & L. Panattoni (1996) Analytical derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics* 11, 399–417.

Francq, C. & J.-M. Zakoïan (2000) Estimating weak GARCH representations. *Econometric Theory* 16, 692–728.

Galbraith, J.W. & V. Zinde-Walsh (1994) A simple noniterative estimator for moving average models. *Biometrika* 81, 143–155.

Giraitis, L. & P.M. Robinson (2001) Whittle estimation of ARCH models. *Econometric Theory* 17, 608–631.

Gouriéroux, C. & J. Jasiak (2001) *Financial Econometrics: Problems, Models and Methods*. Princeton University Press.

Hannan, E.J. & C.C. Heyde (1972) On limit theorems for quadratic functions of discrete time series. *Annals of Mathematical Statistics* 43, 2058–2066.

Harvey, A.C. (1993) *Time Series Models*, 2nd ed. Harvester Wheatsheaf.

He, C. & T. Teräsvirta (1999) Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15, 824–846.

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

Lee, S.-W. & B. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–53.

Ling, S. & M. McAleer (2003) Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* 19, 280–310.

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

Lumsdaine, R. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.

Mayoral, L. (2004) A New Minimum Distance Estimation Procedure of ARFIMA Models. Manuscript, Pompeu Fabra.

Meitz, M. & P. Saikkonen (2004) Ergodicity, Mixing and Existence of Moments of a Class of Markov Models with Applications to GARCH and ACD Models. Stockholm School of Economics, SSE/EFI Working Paper Series in Economics and Finance 573.

Meyn, S.P. & R.L. Tweedie (1993) *Markov Chains and Stochastic Stability*. Communications and Control Engineering Series. Springer-Verlag.

McCullough, B.D. & C.G. Renfro (1999) Benchmarks and software standards: A case study of GARCH procedures. *Journal of Economic and Social Measurement* 25, 59–71.

Mikosch, T. & C. Stărică (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics* 28, 1427–1451.

Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag.
Robinson, P.M. (1988) The stochastic difference between econometric statistics. *Econometrica* 56, 531–548.
Robinson, P.M. & C. Velasco (1997) Autocorrelation-robust inference. In G.S. Maddala & C.R. Rao (eds.), *Robust Inference*, pp. 267–298. North-Holland.
Samorodnitsky, G. & M.S. Taqqu (1994) *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance.* Chapman & Hall.
Tuan, P.-D. (1979) The estimation of parameters for autoregressive-moving average models from sample autocovariances. *Biometrika* 66, 555–560.

APPENDIX: Proofs

Proof of Lemma 1. The consistency part is a simple application of the law of large numbers for stationary and ergodic processes because $E[x_t] < \infty$ and $E[x_t x_{t+k}] < \infty$ under (A.1)–(A.2). The weak convergence result in (12) follows from Mikosch and Stărică (2000, Sec. 5).

The convergence results in (11) and (13) are proved using a standard central limit theorem for Markov chains that are strongly mixing with geometric rate (cf. Meyn and Tweedie, 1993, Thm. 17.0.1). It is easily seen that

$$E[(y_s^2 - \sigma^2)(y_t^2 - \sigma^2)(y_{t+k}^2 - \sigma^2)] = 0,$$

for any $s, t \geq 0$ and $k > 0$. Thus, the asymptotic covariance between $\hat{\sigma}^2$ and $\hat{R}(m)$ is zero. ■

Proof of Theorem 2. We have that $\hat{\lambda} = D(\hat{\sigma}^2, \hat{\rho}(1), \hat{\rho}(2))$, where $D = (D_\alpha, D_\beta, D_\omega)^\top$ is given by

$$D_\alpha(\sigma^2, \rho(1), \rho(2)) = T(\rho(1), F(\rho(1), \rho(2))) + F(\rho(1), \rho(2)), \tag{A.1}$$

$$D_\beta(\sigma^2, \rho(1), \rho(2)) = -T(\rho(1), F(\rho(1), \rho(2))), \tag{A.2}$$

$$D_\omega(\sigma^2, \rho(1), \rho(2)) = \sigma^2(1 - F(\rho(1), \rho(2))), \tag{A.3}$$

with

$$T(\rho, \phi) = \frac{-2\rho\phi + 1 + \phi^2 - \sqrt{4\rho^2\phi^2 + 4\rho\phi - 4\rho\phi^3 + 1 - 2\phi^2 + \phi^4 - 4\rho^2}}{2(\rho - \phi)},$$

and $F(\rho(1), \rho(2)) = \rho(2)/\rho(1)$. It is easily seen that D is a continuously differentiable mapping of σ^2, $\rho(1)$, and $\rho(2)$. The convergence result in (14) now follows from the continuous mapping theorem together with the fact that $\hat{\sigma}^2$ is $\sqrt{T}$-consistent whereas $\hat{\rho}(1)$ and $\hat{\rho}(2)$ converge at a slower rate. The convergence result stated in (15) follows from the standard result for differentiable transformations of asymptotically normally distributed variables. ■