ECORE
INTERNATIONAL ASSOCIATION FOR RESEARCH AND TEACHING
Economics, Finance, Operations Research, Econometrics and Statistics

++ research ++ teaching ++

ECORE DISCUSSION PAPER

2013/57

# An Almost Closed Form Estimator for the EGARCH model

Christian HAFNER
Oliver LINTON

CORE
Université catholique de Louvain

ECARES
Université Libre de Bruxelles

CORE DISCUSSION PAPER
2013/22

# An almost closed form estimator for the EGARCH model

Christian M. HAFNER [1] and Oliver LINTON[2]

April 2013

**Abstract**

The EGARCH is a popular model for discrete time volatility since it allows for asymmetric effects and naturally ensures positivity even when including exogenous variables. Estimation and inference is usually done via maximum likelihood. Although some progress has been made recently, a complete distribution theory of MLE for EGARCH models is still missing. Furthermore, the estimation procedure itself may be highly sensitive to starting values, the choice of numerical optimation algorithm, etc. We present an alter- native estimator that is available in a simple closed form and which could be used, for example, as starting values for MLE. The estimator of the dynamic parameter is inde- pendent of the innovation distribution. For the other parameters we assume that the innovation distribution belongs to the class of Generalized Error Distributions (GED), profiling out its parameter in the estimation procedure. We discuss the properties of the proposed estimator and illustrate its performance in a simulation study.

**Keywords**: autocorrelations, generalized error distribution, method of moments estimator, Newton-Raphson

**JEL Classification**: C12, C13, C14

[1] Université catholique de Louvain, CORE and Institut de statistique, biostatistique et sciences actuarielles, B-1348 Louvain-la-Neuve, Belgium. E-mail: Christian.hafner@uclouvain.be. This author is also member of ECORE, the association of ECORE and ECARES.
[2] Faculty of Economics, Cambridge University, U.K. E-mail: ob120@cam.ac.uk

We thank Piotr Fryzlewicz, Dimitra Kyriakopoulou, Enno Mammen, Paolo Zaffaroni and Jean-Michel Zakoian for helpful comments. The second author thanks to the ERC for financial support.

# 1 Introduction

The exponential GARCH (EGARCH) model introduced by Nelson (1991) remains one of the most popular GARCH type models for modelling the volatility of financial time series. Its advantages over the classical ARCH model of Engle (1982) and GARCH model of Bollerslev (1986) are manyfold. For example, the EGARCH model allows for asymmetric effects of positive and negative innovations. Furthermore, the conditional variance is positive by construction, which allows one to include exogenous variables in the volatility equation. And finally, stochastic properties of the model such as stationarity are naturally comparable to linear models of the conditional mean, while for classical GARCH models this is not the case. The popular software package Eviews offers EGARCH as one of its main volatility models.

Despite these methodological advantages, there are some technical issues with EGARCH due to the inherent difficulty of deriving a concise theory for estimation and inference. In particular, the maximum likelihood estimator proposed by Nelson (1991) is particularly difficult to analyze, due to the invertibility issue. Some recent progress has been made by Straumann and Mikosch (2006) but only for a special case and even then the regularity conditions are hard to interpret. Wintenberger (2012) proves consistency using continuous invertibility, and provides sufficient conditions for this to hold, which however seem to be restrictive. Similarly, Demos and Kyriakopoulou (2013) give sufficient conditions for asymptotic normality, which also restrict the admissible parameter space.

As an alternative to MLE, Zaffaroni (2009) proposes a Whittle estimator and shows consistency and asymptotic normality. Both MLE and Whittle require multiparameter numerical optimization, and the convergence of optimization algorithms often depends on a judicious choice of starting values.

Instead of using estimators that require numerical optimization, one may want to consider estimators that are available in closed form, for example based on some moment conditions of the model. Such closed form estimators are likely to be less efficient, but have the advantage of being immediately available and as such could be used, for example, as starting value for estimators that do require numerical optimization. As they are $\sqrt{n}$-consistent, they can also be used as such in very large samples, considering that estimators involving numerical optimization often require substantial computational effort to achieve convergence in those situations, which are not rare in financial applications.

In the classical GARCH(1,1) model, Kristensen and Linton (2006) have introduced a closed form estimator. The classical GARCH model is, in many respects, simpler than the EGARCH model. For example, unconditional moments such as the unconditional variance do not depend on the innovation distribution in the GARCH model, whereas they do in the EGARCH model.

In this paper, we propose an estimator of the EGARCH model which is in closed form for a given innovation distribution in the class of generalized error distributions (GED). For the parameter that describes the persistence of shocks to volatility, we propose an estimator that is independent of the innovation distribution and also of form of the news impact function used in the specification of the volatility process. If the particular GED distribution is unknown, we provide two ways of estimating the parameter of this distribution, using a profiled likelihood estimator and a moment estimator, respectively. As the estimator of the model parameters, apart from the persistence parameter, depends on the parameter of the innovation distribution, we say that our estimator is in almost closed form. We derive the asymptotic properties of our moment-based estimator and illustrate its small sample performance via a simulation study. We then

# 2 The EGARCH model

Consider the following exponential GARCH (EGARCH) model for the observed zero mean process $y_t$

$$y_t = e^{h_t/2}\xi_t \tag{1}$$

$$h_t = \omega + g(\xi_{t-1}) + \beta h_{t-1}, \tag{2}$$

where the following conditions are satisfied

Assumption 1. *$\xi_t$ is i.i.d. with density $f$, where $E(\xi_t)=0$ and $\text{var}(\xi_t)=1$, while $g(\cdot)$ is a measurable function such that $\text{E}[g(\xi_t)]=0$ and $E[|g(\xi_t)|^2]<\infty$. The parameter $|\beta|<1$.*

Under these conditions the linear process $h_t$ is strongly and weakly stationary, as well as geometrically ergodic (mixing), Nelson (1991).

This defines a semiparametric model with regard to the error density $f$ and the news impact curve $g$. We will later specialize according to very popular choices for $f$ and $g$, but without further assumptions we record below the second order properties of the series $z_t = \log y_t^2$. In the sequel, we need the following moments of $\xi_t$: $C_1(f) = E[\log \xi_t^2]$, $C_2(f) = \text{var}(\log \xi_t^2)$, $C_3(f;g) = E[\log(\xi_t^2)g(\xi_t)]$, $C_4(f) = \text{var}(|\xi_t|)$, $C_5(f) = E|\xi_t|$, and $C_6(f) = \text{cov}(\log(\xi_t^2), |\xi_t|)$.

Following Lütkepohl (1993, p.233) it is easy to show that $z_t$ has an ARMA(1,1) representation which implies the following result.

Proposition 1. *The first two moments of $z_t$ are given by*

$$\mu = E[z_t] = C_1(f) + \frac{\omega}{1-\beta} \tag{3}$$

$$\sigma^2 = \text{var}(z_t) = \frac{\text{var}(g(\xi_t))}{1-\beta^2} + C_2(f) \tag{4}$$

$$\gamma(k) = \text{cov}[z_t, z_{t-k}] = \beta^{k-1}\left(\frac{\beta\text{var}(g(\xi_t))}{1-\beta^2} + C_3(f;g)\right), \quad k \geq 1. \tag{5}$$

It follows that

$$\beta = \frac{\gamma(k+1)}{\gamma(k)} = \frac{\rho(k+1)}{\rho(k)}$$

for all $k \geq 1$, which identifies the parameter $\beta$. Here, $\rho(.)$ denotes the autocorrelation function. The remaining quantities $\omega, g$, and $f$ are not separately identifiable from this information without further structure.

This second order structure was used in Zaffaroni (2009) to form the Whittle likelihood to estimate a subset of the parameters of Nelson's model. As he remarked, it contains insufficient information to identify all parameters of Nelson's model, which is a special case of (1) and (2) that we turn to next.

## 2.1 Nelson's Model

We now specialize to the model considered by Nelson (1991). He assumed that $\xi_t$ followed a standardized generalized error distribution, $\xi_t \sim GED(\nu)$, with mean zero, variance one, and with density function given by

$$f(\xi) = \frac{\nu \exp\{-(1/2)|\xi/\lambda|^{\nu}\}}{\lambda 2^{1+1/\nu}\Gamma(1/\nu)}, \tag{6}$$

where $\lambda = \left\{2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)\right\}^{1/2}$. The GED includes the normal as a special case ($\nu = 2$), but allows for fat tails ($\nu < 2$) while maintaining finiteness of unconditional moments such as the unconditional variance. This density is also called the EPD (Exponential power distribution) and the Subbotin distribution, Subbotin (1923). For $\nu > 1$, it is a log concave density, although not strictly log-concave, see Wellner (2012). We shall restrict attention to $\nu \in V$, where $V \subset (1, \infty)$ is a compact set. We now denote $C_1(\nu) = E[\log \xi_t^2]$, $C_2(\nu) = \text{var}(\log \xi_t^2)$, $C_3(\nu; g) = E[\log(\xi_t^2)g(\xi_t)]$, $C_4(\nu) = \text{var}(|\xi_t|)$, $C_5(\nu) = E|\xi_t|$, and $C_6(\nu) = \text{cov}(\log(\xi_t^2), |\xi_t|)$. These quantities can be computed numerically for any $\nu$ and in some cases have "almost closed form" expressions, such as $C_5(\nu) = 2^{1/\nu}\Gamma(2/\nu)/\Gamma(1/\nu)$, see (A1.8) of Nelson (1991). They are smooth functions of $\nu$ over $V$ for some choice of $V$. One can see that $\partial C_j/\partial\nu \neq 0$.

Nelson (1991) also proposed a specific parametric news impact function $g$

$$g(\xi_t) = \theta\xi_t + \alpha(|\xi_t| - \mathrm{E}|\xi_t|), \tag{7}$$

where $\theta$ and $\alpha$ are unknown parameters. For this specification we have $\text{var}(g(\xi_t)) = \theta^2 + \alpha^2 C_4(\nu)$.[1] It is then straightforward to show that

$$E[z_t \text{sgn}(y_{t-1})] = \theta C_5(\nu). \tag{8}$$

[1] Actually, the GED specification is not needed for this result, only the symmetry of $f$.

Furthermore, under the GED distribution (actually under any symmetric distribution)

$$C_3(\nu; g) = \alpha C_6(\nu). \tag{9}$$

We introduce one further expression. Nelson (1991, Theorem A1.2) shows that provided $\nu > 1$,

$$\begin{aligned} C_7(\nu; \omega, \beta, \gamma, \theta) &= E \exp(h_t/2) \qquad (10) \\ &= \exp\left(\frac{\omega}{2(1-\beta)}\right) \prod_{j=1}^{\infty} \exp(-\beta^{j-1}\gamma\Gamma(2/\nu)\lambda(\nu)2^{(1-v)/\nu}/\Gamma(1/\nu)) \\ & \sum_{k=0}^{\infty}(2^{(1-v)/\nu}\lambda(\nu)\beta^{j-1})^k \left[(\gamma+\theta)^k + (\gamma-\theta)^k\right] \frac{\Gamma((k+1)/\nu)}{2\Gamma(1/\nu)\Gamma(k+1)} \\ &< \infty. \end{aligned}$$

This is obviously a very complicated nonlinear expression and it depends on all the parameters. It is also a smooth function over the range of parameter values for which the moment exists.

We next discuss estimation of $\phi = (\nu, \omega, \alpha, \beta, \gamma, \theta)$ based on the expressions (3), (4), (5), (8), (9), and (10).

# 3 The closed form estimator

Suppose we have a sample $y_1, \ldots, y_n$ and the parameters are such that the process is stationary and ergodic. Define the sample mean and autocovariance function (for $k = 0, 1, 2, \ldots$)

$$\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{t=1}^{n} z_t \\ \hat{\gamma}(k) &= \frac{1}{n-k}\sum_{t=k+1}^{n}(z_t - \hat{\mu})(z_{t-k} - \hat{\mu}), \end{aligned}$$

and define the autocorrelation function $\widehat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$.

Equation (5) implies that $\gamma(j+1) = \beta\gamma(j)$ and $\rho(j+1) = \beta\rho(j)$, $j = 1, 2, \ldots$. Then, motivated by Proposition 1, we propose the following moment estimators for $\beta$

$$\hat{\beta} = \sum_{j=1}^{p} \frac{\hat{\gamma}(j+1)}{\hat{\gamma}(j)} w_j = \sum_{j=1}^{p} \frac{\widehat{\rho}(j+1)}{\widehat{\rho}(j)} w_j, \tag{11}$$

where $p \geq 1$ and $w_j$ some known weights such that $\sum_{j=1}^{p} w_j = 1$. This is just like Kristensen and Linton (2006). Note that $\hat{\beta}$ is independent of the specification of $g(\cdot)$ and of the innovation distribution. In practice, the median of the ratios or a trimmed mean may provide superior performance. Another method is based on regression. Let $y = (\hat{\gamma}(2), \ldots, \hat{\gamma}(p+1))^\intercal$, $z = (\hat{\gamma}(1), \ldots, \hat{\gamma}(p))^\intercal$, and $x = (i_p, \widehat{\gamma})$, where $i_p$ is the $p-$vector of ones. Then let $\widehat{\beta} = (z^\intercal z)^{-1} z^\intercal y$ or even the slope coefficient from the intercept regression $(x^\intercal x)^{-1} x^\intercal y$.

The remaining parameters depend on the error distribution parameter $\nu$ and we shall therefore profile this. Using (3)-(8), we obtain

$$\begin{aligned} \hat{\omega}(\nu) &= (\hat{\mu} - C_1(\nu))(1 - \hat{\beta}) \\ \hat{\theta}(\nu) &= \frac{1}{C_5(\nu)} \frac{1}{n} \sum_{t=1}^{n} z_t \mathrm{sgn}(y_{t-1}) \\ \widehat{\alpha}(\nu) &= \frac{1}{C_6(\nu)} \left\{ \sum_{j=1}^{q} \frac{\widehat{\gamma}(j)}{\widehat{\beta}^{j-1}} \omega_j - \widehat{\beta} \left( \frac{1}{n} \sum_{t=1}^{n} (z_t - \hat{\mu})^2 - C_2(\nu) \right) \right\}. \end{aligned}$$

for $q \geq 1$ and $\omega_j$ some known weights such that $\sum_{j=1}^{q} \omega_j = 1$. If $\nu$ were known, e.g., when $\nu = 2$, this would be a complete estimation procedure.

## 4 Estimation of $\nu$

We next consider how to estimate $\nu$ from the data as in Nelson (1991). There are two approaches. First, the profiled likelihood method. We can obtain recursively

$$\hat{\xi}_t(\nu) = \exp\{-\hat{h}_t(\nu)/2\} y_t \tag{12}$$

where

$$\begin{aligned} \hat{h}_t(\nu) &= \hat{\omega}(\nu) + \hat{\theta}(\nu)\hat{\xi}_{t-1}(\nu) + \hat{\alpha}(\nu)(|\hat{\xi}_{t-1}(\nu)| - C_5(\nu)) + \hat{\beta}\hat{h}_{t-1}(\nu) && (13) \\ \hat{h}_1(\nu) &= \frac{\hat{\omega}(\nu)}{1 - \hat{\beta}} = (\hat{\mu} - C_1(\nu)). && (14) \end{aligned}$$

The profiled log likelihood is then given by

$$\widehat{L}(\nu) = \sum_{t=1}^{n} \widehat{\ell}_t(\nu)$$

$$\widehat{\ell}_t(\nu) = \ln(\nu/\lambda 2^{1+1/\nu} \Gamma(1/\nu)) - \frac{1}{2} \left\{ |\hat{\xi}_t(\nu)/\lambda|^{\nu} + \hat{h}_t(\nu) \right\},$$

which is maximized w.r.t. $\nu$ using a grid search over the compact set $V$. Let $\widehat{\nu}_L$ denote the maximizer. Unfortunately, it is difficult to prove rigorously the consistency even of this estimator of $\nu$. The issue is due to the nonlinear recursive equations (11)-(13), which are very difficult to analyze away from the true value, exactly the same problem as for the original EGARCH MLE.

Instead, we consider an alternative method based on the method of moments using the expression (10). Specifically, find $\widehat{\nu}_m \in V$ to solve the nonlinear equation

$$m_n = \frac{1}{T}\sum_{t=1}^{n} |y_t| = C_8(\nu; \widehat{\omega}(\nu), \widehat{\beta}, \widehat{\gamma}(\nu), \widehat{\theta}(\nu)), \tag{15}$$

where $C_8(\nu;\omega,\beta,\gamma,\theta) = C_5(v)C_7(\nu;\omega,\beta,\gamma,\theta)$ and $C_7(\nu;\omega,\beta,\gamma,\theta)$ is defined in (10). This can also be done by a univariate grid search. In practice, one may prefer to take logarithms of both sides to obtain expressions only involving summations, which may then be suitably truncated to achieve whatever degree of accuracy is required. The advantage of this method is that it is purely based on sample moments of observables and so there is no need to define a recursive dynamic equation based on estimated parameters. The theoretical properties are much easier to handle. Specifically, one can obtain consistency and asymptotic normality of all the parameter estimates.[2]

# 5 Asymptotic Properties

We first discuss some properties of the estimators defined in section 3. Under Assumption 1, $z_t = \log y_t^2$ is a linear process that satisfies the conditions of Theorem 7.2.2. of Brockwell and Davies (2006) and so the sample autocorrelations are asymptotically normal at rate $\sqrt{n}$. The estimator $\hat{\beta}$ is a smooth function of the first $p+1$ autocorrelations and so (provided $\beta \in (0,1)$) the asymptotic distribution of $\hat{\beta}$ follows without the additional structure provided by Nelson's model. To repeat, the estimator $\hat{\beta}$ is robust to both $f$ and $g$, in the sense that it is consistent for a large class of these functions, unlike the GED MLE proposed by Nelson (1991), which certainly requires correct specification of the news impact curve and may also require at least symmetry of the true density for the GED based MLE to be consistent.

Stengthening the moment conditions of Assumption 1 to $E[|\xi_t|^4] < \infty$ and $E[|g(\xi_t)|^4] < \infty$ we may apply Theorem 7.2.1 of Brockwell and Davies (2006) to obtain the root-n consistency and asymptotic normality of the sample autocovariances used in section 3.[3] This in turn implies the root-n asymptotic normality of $\widehat{\omega}(\nu)$ and $\widehat{\alpha}(\nu)$ around some limiting value (the argument for $\widehat{\theta}(\nu)$ is similar as it only depends additionally on the sign of $\xi_t$). This is true without the GED assumption, although the probability limit would obviously depend on the underlying distribution $f$. Define $\widehat{\eta}(\nu) = (\widehat{\omega}(\nu), \widehat{\beta}, \widehat{\theta}(\nu), \widehat{\alpha}(\nu))^\top$ for each $v \in V$, and let $\eta(\nu) = (\omega(\nu), \beta_o, \theta(\nu), \alpha(\nu))^\top$ be defined as the probability limit of $\widehat{\eta}(\nu)$. Then it is straightforward to

[2]Nelson (1991, Theorem A1.2) shows that provided $\nu > 1$, all moments of $y$ exist and so the asymptotic normality of $n^{-1/2}\sum_{t=1}^{n}|y_t| - E|y_t|$ follows under the stationarity and geometric ergodicity of $y$.

[3]The moment conditions on the innovation are quite mild and there is no restriction on the implied moments for $y_t$ so far, unlike in Kristensen and Linton (2006).

show (see the appendix) that for some positive definite covariance matrix $\Sigma(\nu)$

$$\delta_n(\nu) = \sqrt{n}\,[\widehat{\eta}(\nu) - \eta(\nu)] \Longrightarrow N(0, \Sigma(\nu)). \tag{16}$$

Furthermore, the process $\delta_n(\nu)$ is tight in $\nu$ (it is continuously differentiable in $\nu$).

We now assume that Nelson's model holds in its entirety. In this case $\xi_t$ and $g(\xi_t)$ have all moments existing. Furthermore,

$$\begin{aligned}\omega(\nu) &= \omega_o + (C_1(\nu_o) - C_1(\nu))(1-\beta_o) \quad ; \quad \theta(\nu) = \theta_o \frac{C_5(\nu_o)}{C_5(\nu)} \\ \alpha(\nu) &= \alpha_o \frac{C_6(\nu_o)}{C_6(\nu)} + \frac{1}{C_6(\nu)}\beta_o\{C_2(\nu_o) - C_2(\nu)\},\end{aligned}$$

where subscript o denotes true value. In this case, if $\nu_0$ is known, the result can be used to provide standard errors and conduct inference about the remaining parameters. We next show that one can carry out a test of leverage in the model (7) without knowledge of $\nu$ or even $f$ (so long as it is symmetric about zero).

Our estimator $\hat{\theta}(\nu)$, for any $\nu$, can be used to test for a leverage effect within the Nelson model. The reason is that in constructing the t-ratio the constant term $C_5(\nu)$ is cancelled out. That is, we may compute

$$t = \frac{\hat{\theta}(\nu)}{\mathrm{se}(\hat{\theta}(\nu))} = \frac{\frac{1}{n}\sum_{t=1}^{n} u_t}{\sqrt{n \mathrm{lrvar}(u_t)}},$$

where $u_t = z_t \mathrm{sgn}(y_{t-1})$ and $\mathrm{lrvar}(x_t)$ denotes the long run variance of a series $x_t$. In fact, the series $z_t\mathrm{sgn}(y_{t-1}) = z_t\mathrm{sgn}(\xi_{t-1})$ is serially uncorrelated, so that $\mathrm{lrvar}(u_t) = \mathrm{var}(u_t)$. By direct calculation we have

$$\mathrm{var}(z_t\mathrm{sgn}(y_{t-1})) = Ez_t^2 - E^2 z_t \mathrm{sgn}(y_{t-1}) = \frac{\theta^2 + \alpha^2 C_4(\nu)}{1-\beta^2} + C_2(\nu) + \left(C_1(\nu) + \frac{\omega}{1-\beta}\right)^2 - \theta^2 C_5^2(\nu).$$

This can be estimated by the plug in method or from the sample variance of $u_t$ itself, which in practice is easier. Let

$$\widehat{t} = \frac{\sqrt{n}\overline{u}}{\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(u_t - \overline{u})^2}}, \quad \overline{u} = \frac{1}{n}\sum_{t=1}^{n} u_t. \tag{17}$$

Then under the null hypothesis of no leverage, $\widehat{t}$ is asymptotically standard normal. This test is robust to the value of $v$ and indeed is valid for any symmetric error distribution.

Finally, we turn to the properties of the estimator of all parameters proposed in section 4. Let $\phi_o = (\omega_o, \beta_o, \theta_o, \alpha_o, \nu_o)^\top$ and let $\widehat{\phi}_m = (\widehat{\omega}(\widehat{\nu}_m), \widehat{\beta}, \widehat{\theta}(\widehat{\nu}_m), \widehat{\alpha}(\widehat{\nu}_m), \widehat{\nu}_m)^\top$.

Theorem 1. *Suppose that Assumption 1 holds, that Nelson's model (6) and (7) holds, and that $\beta > 0$. Then there exists a positive definite matrix $\Omega_{\phi\phi}$ such that*

$$\sqrt{n}(\widehat{\phi}_m - \phi_o) \Longrightarrow N(0, \Omega_{\phi\phi}).$$

In fact one can calculate the variance matrix $\Omega_{\phi\phi}$ exactly in terms of the parameters $\phi_o$ (along the lines of Francq, Horvath, and Zakoian (2011)).

# 6 Approximating the MLE

The closed form estimators we have proposed are not fully efficient. We give a brief discussion on how one may improve on the efficiency of the closed form estimator. The basic idea is to perform a number of Newton-Raphson (NR) iterations using the full EGARCH likelihood function. One may wish to proceed to the MLE, using the initial estimates as a starting point in the numerical optimization; this may help reduce numerical problems since our preliminary estimates are consistent, i.e., are likely to be close to the true values. Alternatively, one can perform a number of NR-iterations which do not necessitate the use of any numerical optimization procedure. We define the following sequence of NR-estimators,

$$\widehat{\phi}_{k+1}^{\text{NR}} = \widehat{\phi}_{k}^{\text{NR}} - H_n^{-1}(\widehat{\phi}_{k}^{\text{NR}})S_n(\widehat{\phi}_{k}^{\text{NR}}), \quad k \geq 1,$$

with initial value being the closed form estimator, $\widehat{\phi}_{1}^{\text{NR}} = \widehat{\phi}_m$, while $S_n(\phi) = \partial L(\phi)/\partial\phi$, $H_n(\phi) = \partial^2 L(\phi)/\partial\phi\partial\phi^\top$, and $L(\phi)$ is the EGARCH likelihood function:

$$L(\phi) = \sum_{t=1}^{n} \ell_t(\phi)$$

$$\begin{aligned}
\ell_t(\phi) &= \ln(\nu/\lambda 2^{1+1/\nu}\Gamma(1/\nu)) - \frac{1}{2}\left\{|\hat{\xi}_t(\phi)/\lambda(\nu)|^\nu + \hat{h}_t(\phi)\right\} \\
\widehat{\xi}_t(\phi) &= \exp\{-\hat{h}_t(\phi)/2\}y_t \\
\hat{h}_t(\phi) &= \omega + \theta\hat{\xi}_{t-1}(\phi) + \alpha(|\hat{\xi}_{t-1}(\phi)| - C_5(\nu)) + \beta\hat{h}_{t-1}(\phi) \\
\hat{h}_1(\phi) &= \frac{\omega}{1-\beta}.
\end{aligned}$$

Note that $L(\phi)$ is twice continuously differentiable in $\phi$ for a compact subset of $\mathbb{R}^5$.

In general, the NR-estimator will satisfy

$$||\widehat{\phi}_{k+1}^{\text{NR}} - \widehat{\phi}^{\text{OP}}|| = O_P(||\widehat{\phi}_{1}^{\text{NR}} - \widehat{\phi}^{\text{OP}}||^{2^k}), \tag{18}$$

where $\widehat{\phi}^{OP} = \arg\max Q_n(\phi)$ is the actual $M$-estimator, c.f. Robinson (1988, Theorem 2). We expect that a two step estimator should be fully efficient, i.e.,

$$\sqrt{n}(\widehat{\phi}_{2}^{\text{NR}} - \phi_o) \Longrightarrow N\left(0, \left(E\left[\frac{\partial\ell_t(\phi)}{\partial\phi}\frac{\partial\ell_t(\phi)}{\partial\phi^\intercal}\right]_{\phi=\phi_o}\right)^{-1}\right)$$

under some regularity conditions.

# 7 Numerical Results

## 7.1 A simulation study

We explore the finite sample properties of the proposed estimators through a Monte Carlo simulation study. We generate EGARCH processes with Gaussian (i.e., $\nu = 2$) and GED ($\nu = 1.5$) innovations, and the following parameters: $\omega = -0.3$, $\alpha = 0.5$, $\beta = 0.9$, and $\theta = -0.1$, which represent typical values for financial time series. We consider various sample sizes $n$ and use 1000 replications.

We first analyse the properties of alternative estimators of $\beta$: a simple mean of ratios, a weighted mean with linearly declining weights, the median, and the OLS without intercept regression estimator. All four estimators depend on the number of terms $p$ included, which we increase in steps of five from 5 to 50. For higher values of $p$, the estimates of the ACF of $z_t$ become too noisy and all estimates of $\beta$ suffer from high variability. Results are reported in Table 1.

It is remarkable that both the weighted and unweighted means underperform for higher $p$, simply because the variance becomes too high due to some outliers in estimated ACF. Median and OLS estimates are robust to these, and mean square errors are reasonably small. The median has a smaller bias for small values of $p$, while the OLS estimate has generally a smaller standard deviation. A globally good choice of $p$ appears to be 10 or 15 for this type of persistence.

In Table 2, we report the performance of the closed form estimator of the remaining model parameters. For the estimator of $\beta$ we use a fixed order $p = 10$ and equal weights, while for the estimator of $\alpha$ we use $q = 1$. Experiments with higher $q$ did not improve the results with $q = 1$ in terms of mean squared error, so we only report the latter results. For the estimation of $\nu$ we use both estimators proposed in Section 4, the profiled likelihood and the profiled moment estimator. Both estimators are found by a grid search on the interval $[1, 3]$. Recall that the estimator of $\beta$ is independent of the estimator of $\nu$, while estimation of the remaining parameters, i.e. $\omega, \theta$ and $\alpha$ depends on the method to estimate $\nu$.

The results corroborate the theoretical finding that both estimators are consistent. Estimation of $\omega$, the scale parameter, and $\theta$, the sign effect, seems rather unaffected by the estimation of $\nu$. That is, bias and variance of $\hat{\omega}$ and $\hat{\theta}$ are almost identical under the likelihood and the moment estimation. The precision of both $\hat{\omega}$ and $\hat{\theta}$ is only slightly higher under Gaussian compared to GED($\nu = 1.5$) innovations.

Quite different results are obtained for the estimation of the size effect, $\alpha$, and the GED parameter $\nu$. The moment estimator of $\alpha$ has a negative bias and higher variance than the likelihood estimator. Moreover, the moment estimator deteriorates under fat tails of the innovation distribution in the sense that the bias aggravates and the variance increases. For the likelihood estimator, on the other hand, exploiting the information of the innovation distribution turns out to be beneficial: both bias and variance decrease under GED compared to Gaussian innovations.

For the estimation of $\nu$, the moment estimator has a positive, the likelihood estimator a negative bias, while the latter has a smaller variance than the former. Again, the likelihood estimator is more precise under GED innovations than under Gaussianity, and now this also holds for the moment estimator.

### 7.2 Application

We investigated the performance of our different estimators of $\beta$ on a large dataset, the demeaned daily (close to close) return on the S&P500 index from 1950-2012, a total of 15757 observations. This data is quite heavy tailed, with a tail thickness parameter around three, which implies that the second moments of returns may exist but the fourth ones do not. Eviews computed the following parameter estimates using the default numerical optimization algorithm. It took 26 iterations to achieve convergence and the results are shown below.

| | estimate | standard error |
|---|---|---|
| $\beta$ | 0.9866 | 0.00135 |
| $\omega$ | -0.2542 | 0.01729 |
| $\theta$ | -0.0685 | 0.003675 |
| $\alpha$ | 0.1353 | 0.00650 |
| $\nu$ | 1.3726 | 0.012485 |

The process is quite persistent, which agrees with much earlier work. The tail thickness parameter is lower than that in Nelson. The parametric test of the leverage effect yields a t-statistic of nearly -19, indicating strong evidence. However, the nonparametric test based on (16) yields a smaller t-statistic of -4.666, which is still significant but less so. This test however, is robust to the choice of error distribution so long as it is symmetric.

We then investigate three estimators of $\beta$ with regard to the choice of $p$ : the mean of the ratio, the median of the ratio, and the no intercept regression estimator. In Figure 1, we show the value of the estimated $\beta$ against the number of lags $p$ used for $p = 4, \dots, 100$. The straight mean of the ratio estimator is generally above one in value. The median estimator is generally below the MLE, while the no intercept regression estimator is much closer to the MLE value than the others. We also looked at using $p$ up to a thousand, and more or less the same outcome is observed, except as may be expected, the mean of the ratio estimator becomes very volatile due to the appearance of occasional small and negative values at the long lags; this affects the median and the regression estimators much less.

| $p$ | $\hat{\beta}$ | s.e. | $\hat{\beta}_w$ | s.e. | $\hat{\beta}_{rob}$ | s.e. | $\hat{\beta}_{ols}$ | s.e. |
|---|---|---|---|---|---|---|---|---|
| | | | | $n = 1000$ | | | | |
| 5 | 0.936 | 0.089 | 0.933 | 0.079 | 0.913 | 0.096 | 0.884 | 0.073 |
| 10 | 0.957 | 0.542 | 0.950 | 0.324 | 0.907 | 0.099 | 0.868 | 0.060 |
| 15 | 0.863 | 3.605 | 0.943 | 1.133 | 0.894 | 0.115 | 0.856 | 0.065 |
| 20 | 0.795 | 3.647 | 0.892 | 2.038 | 0.875 | 0.131 | 0.845 | 0.070 |
| 25 | 0.600 | 3.673 | 0.798 | 2.466 | 0.851 | 0.149 | 0.834 | 0.076 |
| 30 | 0.643 | 5.838 | 0.762 | 2.946 | 0.831 | 0.161 | 0.824 | 0.080 |
| 35 | 0.690 | 6.183 | 0.738 | 3.455 | 0.819 | 0.165 | 0.816 | 0.085 |
| 40 | 0.518 | 6.418 | 0.694 | 3.857 | 0.805 | 0.173 | 0.808 | 0.087 |
| 45 | 0.466 | 5.891 | 0.649 | 4.152 | 0.794 | 0.176 | 0.803 | 0.090 |
| 50 | 0.516 | 5.379 | 0.624 | 4.322 | 0.788 | 0.177 | 0.798 | 0.092 |
| | | | | $n = 5000$ | | | | |
| 5 | 0.905 | 0.028 | 0.905 | 0.027 | 0.902 | 0.037 | 0.896 | 0.027 |
| 10 | 0.911 | 0.023 | 0.908 | 0.017 | 0.900 | 0.035 | 0.894 | 0.019 |
| 15 | 0.734 | 6.087 | 0.870 | 1.432 | 0.901 | 0.032 | 0.891 | 0.018 |
| 20 | 0.888 | 5.773 | 0.848 | 2.948 | 0.898 | 0.038 | 0.889 | 0.019 |
| 25 | 0.878 | 4.876 | 0.861 | 3.524 | 0.893 | 0.044 | 0.886 | 0.019 |
| 30 | 0.990 | 5.094 | 0.895 | 3.780 | 0.882 | 0.055 | 0.883 | 0.019 |
| 35 | 0.848 | 5.278 | 0.893 | 3.906 | 0.870 | 0.063 | 0.881 | 0.020 |
| 40 | 1.109 | 8.086 | 0.943 | 4.354 | 0.857 | 0.072 | 0.878 | 0.021 |
| 45 | 1.096 | 7.963 | 0.978 | 4.801 | 0.846 | 0.077 | 0.876 | 0.021 |
| 50 | 1.030 | 7.759 | 0.987 | 5.169 | 0.833 | 0.087 | 0.874 | 0.022 |
| | | | | $n = 10000$ | | | | |
| 5 | 0.902 | 0.019 | 0.902 | 0.019 | 0.900 | 0.025 | 0.898 | 0.019 |
| 10 | 0.905 | 0.015 | 0.904 | 0.012 | 0.900 | 0.024 | 0.897 | 0.013 |
| 15 | 0.911 | 0.017 | 0.907 | 0.010 | 0.900 | 0.022 | 0.896 | 0.012 |
| 20 | 0.918 | 0.216 | 0.912 | 0.042 | 0.900 | 0.024 | 0.894 | 0.012 |
| 25 | 0.929 | 0.584 | 0.920 | 0.179 | 0.898 | 0.028 | 0.893 | 0.012 |
| 30 | 0.899 | 2.030 | 0.928 | 0.552 | 0.892 | 0.035 | 0.892 | 0.013 |
| 35 | 0.843 | 2.527 | 0.920 | 0.918 | 0.883 | 0.043 | 0.890 | 0.013 |
| 40 | 0.747 | 4.314 | 0.880 | 1.451 | 0.874 | 0.050 | 0.889 | 0.013 |
| 45 | 0.776 | 4.083 | 0.860 | 1.931 | 0.864 | 0.055 | 0.888 | 0.013 |
| 50 | 0.729 | 3.894 | 0.838 | 2.251 | 0.856 | 0.061 | 0.887 | 0.013 |

Table 1: Simulation results: estimation of $\beta$ using simple mean ($\hat{\beta}$), weighted mean ($\hat{\beta}_w$), median ($\hat{\beta}_{rob}$), and regression without intercept ($\hat{\beta}_{ols}$). Simulated process: EGARCH(1,1) with GED($\nu = 1.5$) innovations, $\omega = -0.3$, $\alpha = 0.5$, $\beta = 0.9$, and $\theta = -0.1$. The number of replications is 1000.

| | | $n = 1000$ | | $n = 2000$ | | $n = 5000$ | | $n = 10000$ | |
|---|---|---|---|---|---|---|---|---|---|
| | true | mean | s.d. | mean | s.d. | mean | s.d. | mean | s.d. |
| Gaussian | | | | | | | | | |
| $\beta$ | 0.9 | 0.926 | 0.144 | 0.924 | 0.065 | 0.910 | 0.023 | 0.904 | 0.016 |
| Profiled moment estimator | | | | | | | | | |
| $\omega$ | -0.3 | -0.218 | 0.427 | -0.227 | 0.195 | -0.269 | 0.069 | -0.285 | 0.047 |
| $\theta$ | -0.1 | -0.092 | 0.187 | -0.103 | 0.129 | -0.097 | 0.087 | -0.098 | 0.060 |
| $\alpha$ | 0.5 | 0.256 | 0.287 | 0.343 | 0.217 | 0.449 | 0.107 | 0.475 | 0.059 |
| $\nu$ | 2.0 | 2.367 | 0.510 | 2.225 | 0.415 | 2.059 | 0.253 | 2.014 | 0.153 |
| Profiled likelihood estimator | | | | | | | | | |
| $\omega$ | -0.3 | -0.211 | 0.412 | -0.222 | 0.191 | -0.267 | 0.069 | -0.284 | 0.048 |
| $\theta$ | -0.1 | -0.096 | 0.197 | -0.106 | 0.134 | -0.098 | 0.089 | -0.099 | 0.061 |
| $\alpha$ | 0.5 | 0.560 | 0.234 | 0.516 | 0.142 | 0.507 | 0.068 | 0.501 | 0.042 |
| $\nu$ | 2.0 | 1.796 | 0.344 | 1.880 | 0.249 | 1.942 | 0.179 | 1.964 | 0.123 |
| GED with $\nu = 1.5$ | | | | | | | | | |
| $\beta$ | 0.9 | 0.931 | 0.140 | 0.928 | 0.038 | 0.910 | 0.022 | 0.904 | 0.015 |
| Profiled moment estimator | | | | | | | | | |
| $\omega$ | -0.3 | -0.228 | 0.444 | -0.218 | 0.120 | -0.282 | 0.071 | -0.300 | 0.050 |
| $\theta$ | -0.1 | -0.091 | 0.205 | -0.101 | 0.140 | -0.101 | 0.095 | -0.098 | 0.071 |
| $\alpha$ | 0.5 | 0.125 | 0.352 | 0.250 | 0.339 | 0.434 | 0.149 | 0.473 | 0.063 |
| $\nu$ | 1.5 | 1.908 | 0.491 | 1.772 | 0.393 | 1.562 | 0.207 | 1.517 | 0.091 |
| Profiled likelihood estimator | | | | | | | | | |
| $\omega$ | -0.3 | -0.218 | 0.423 | -0.212 | 0.119 | -0.279 | 0.072 | -0.299 | 0.050 |
| $\theta$ | -0.1 | -0.096 | 0.218 | -0.104 | 0.147 | -0.102 | 0.097 | -0.099 | 0.071 |
| $\alpha$ | 0.5 | 0.542 | 0.223 | 0.521 | 0.123 | 0.509 | 0.056 | 0.504 | 0.038 |
| $\nu$ | 1.5 | 1.398 | 0.232 | 1.446 | 0.162 | 1.476 | 0.108 | 1.485 | 0.078 |

Table 2: Simulation results for estimated EGARCH(1,1) processes. Number of replications is 1000.

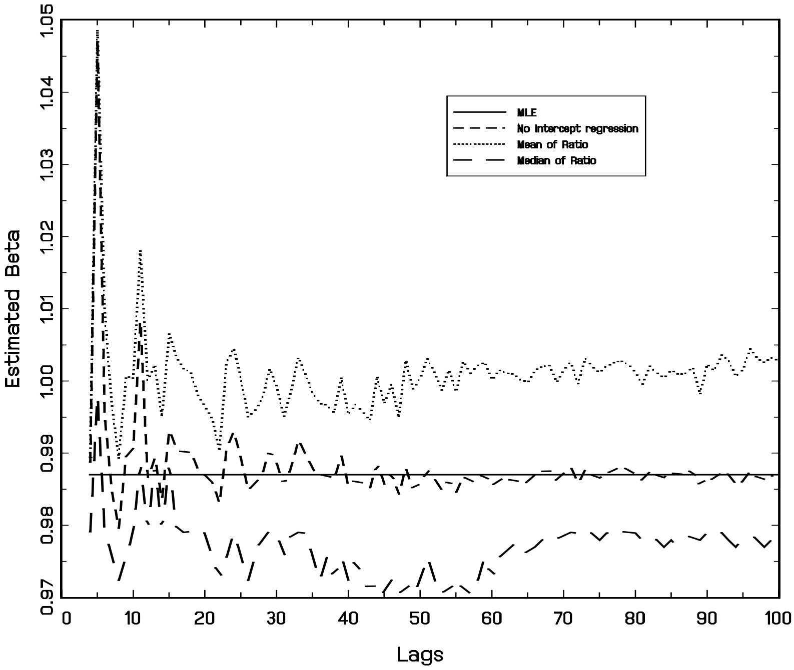


Figure 1: Estimated $\beta$ for the S&P500 returns using the mean of ratios as in (**??**) with flat weights and $p = 4, \ldots, 100$ (horizontal axis), the median of ratios, and the no-intercept regression estimator. The straight line is the MLE.

We next show a scatter plot of the empirical autocorrelations along with the fitted regression line, see Figure 2. We show the first 1001 values, where the estimator is determined from the first $p = 100$ of them. The scatter plot shows fairly good agreement with a linear fit. The lines corresponding to the mean or median estimator of $\beta$ are very close to the regression line. We have $\widehat{\beta} = 1.002$, $\widehat{\beta}_{ols} = 0.986$, and $\widehat{\beta}_{rob} = 0.976$, all of which are quite close although the straight average of the ratio violates the stationarity constraint, which could pose problems if it were plugged into a numerical optimization algorithm.

# 8 Conclusions

We have shown that a simple closed form estimator of the EGARCH is consistent, asymptotically normally distributed, and has reasonable finite sample properties. We recommend this

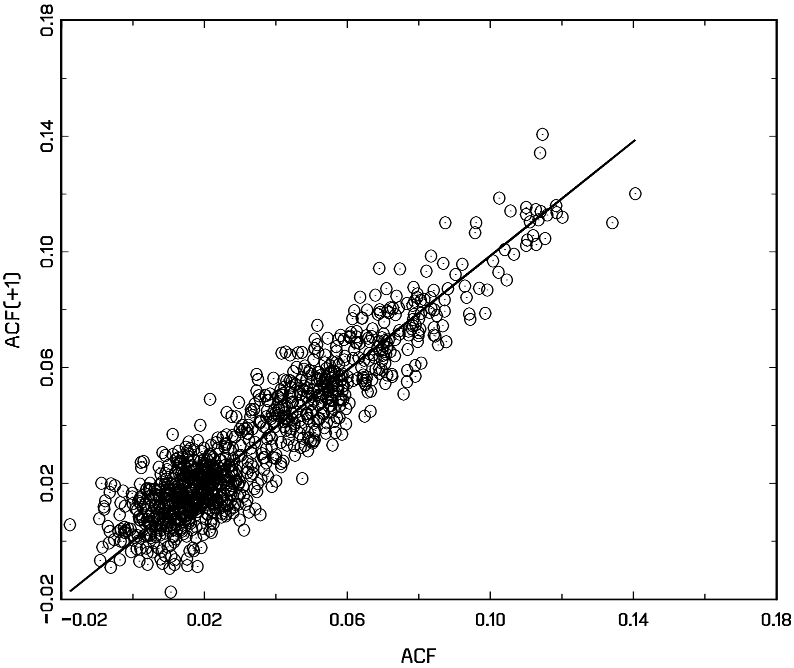


Figure 2: Scatterplot of empirical autocorrelations of $z_t = \log y_t^2$, $\gamma(k+1)$ (vertical axis) vs $\gamma(k)$ (horizontal axis), $k = 1, \ldots, 1000$. The straight line is the no intercept regression line with slope 0.986 obtained using the first 100 observations of $(\gamma(k), \gamma(k+1))$.

estimator in large samples, or as starting values for estimators requiring numerical optimization.

Several directions for future work are possible. For example, one may find out the best way to choose the lag order $p$ and the weighting scheme used in the estimation of $\beta$. The explicit expression of the asymptotic variance covariance matrix can be derived. And finally, empirical applications will show the utility of the estimator.

## Appendix

Proof of Theorem 1. We first note that we can write $\widehat{\eta}(\nu) = h(U_n;\nu)$ for some smooth (meaning, for $x = u, v$, $|\partial h(u,v)/\partial x| \le r(u)$ with $\sup_n Er(U_n) < \infty$) function $h$ of the vector $U_n$ and parameter $\nu$ and $\eta(\nu) = h(EU_n;\nu)$, where

$$U_n = \begin{bmatrix} \frac{1}{n}\sum_{t=1}^n z_t \\ \frac{1}{n}\sum_{t=1}^n z_t^2 \\ \frac{1}{n}\sum_{t=1}^n z_t \mathrm{sgn}(y_{t-1}) \\ \frac{1}{n}\sum_{t=2}^n z_t z_{t-1} \\ \vdots \\ \frac{1}{n}\sum_{t=p+2}^n z_t z_{t-p-1} \end{bmatrix} \equiv \frac{1}{n}\sum_{t=1}^n u_t.$$

From the stationarity and geometric ergodicity, we have $U_n \to EU_n$ with probability one (wp1) and furthermore, for some symmetric positive definite $\Psi$,

$$\sqrt{n}[U_n - EU_n] \Longrightarrow N(0,\Psi).$$

Hence, by the delta method (15) follows.

We now turn to the properties of $\widehat{\nu}_m$. By the smoothness of $h$ with respect to $\nu$, we have wp1

$$\sup_{\nu\in V} |\widehat{\eta}(\nu) - \eta(\nu)| \longrightarrow 0.$$

It follows that (wp1)

$$\sup_{\nu\in V} \left| C_8(\nu;\widehat{\omega}(\nu),\widehat{\beta},\widehat{\gamma}(\nu),\widehat{\theta}(\nu)) - C_8(\nu;\omega(\nu),\beta_o,\gamma(\nu),\theta(\nu)) \right| \longrightarrow 0.$$

Furthermore, $m_n = C_8(\nu_o;\omega_o,\beta_o,\gamma_o,\theta_o) + o(1)$ wp1 provided $Ey_t^2 < \infty$. The function $C_8(\nu;\omega(\nu),\beta_o,\gamma(\nu),\theta(\nu))$ is monotonic in $\nu$ over the range $[1,3]$ (which can be seen numerically). It follows that the equation

$$C_8(\nu;\omega(\nu),\beta_o,\gamma(\nu),\theta(\nu)) = C_8(\nu_o;\omega_o,\beta_o,\gamma_o,\theta_o)$$

has a unique solution at $\nu = \nu_o$. Therefore, $\widehat{\nu}_m$ is consistent.

We have

$$m_n - Em_n + C_8(\nu_o;\omega_o,\beta_o,\gamma_o,\theta_o) = C_8(\widehat{\nu}_m;\widehat{\omega}(\widehat{\nu}_m),\widehat{\beta},\widehat{\gamma}(\widehat{\nu}_m),\widehat{\theta}(\widehat{\nu}_m)).$$

Denote by $\widehat{P}(\nu) = C_8(\nu;\widehat{\omega}(\nu),\widehat{\beta},\widehat{\gamma}(\nu),\widehat{\theta}(\nu))$ the random function of $\nu$ induced by the profile estimation and $P(\nu) = C_8(\nu;\omega(\nu),\beta_o,\gamma(\nu),\theta(\nu))$. By the mean value theorem, for $\widetilde{\nu}$ with

$|\widetilde{\nu} - \nu_o| \leq |\widehat{\nu} - \nu_o|$ we have

$$
\begin{aligned}
& C_8(\widehat{\nu}_m; \widehat{\omega}(\widehat{\nu}_m), \widehat{\beta}, \widehat{\gamma}(\widehat{\nu}_m), \widehat{\theta}(\widehat{\nu}_m)) \\
= \; & C_8(\nu_o; \widehat{\omega}(\nu_o), \widehat{\beta}, \widehat{\gamma}(\nu_o), \widehat{\theta}(\nu_o)) + \left.\frac{\partial \widehat{P}(\nu)}{\partial \nu}\right\}_{\nu=\widetilde{\nu}} (\widehat{\nu}_m - \nu_o) \\
= \; & C_8(\nu_o; \widehat{\omega}(\nu_o), \widehat{\beta}, \widehat{\gamma}(\nu_o), \widehat{\theta}(\nu_o)) + \left.\frac{\partial P(\nu)}{\partial \nu}\right\}_{\nu=\nu_o} (\widehat{\nu}_m - \nu_o) \\
& + \left[ \left.\frac{\partial \widehat{P}(\nu)}{\partial \nu}\right\}_{\nu=\widetilde{\nu}} - \left.\frac{\partial P(\nu)}{\partial \nu}\right\}_{\nu=\nu_o} \right] (\widehat{\nu}_m - \nu_o)
\end{aligned}
$$

By the uniform convergence arguments and the consistency of $\widetilde{\nu}$ we have

$$
\left.\frac{\partial \widehat{P}(\nu)}{\partial \nu}\right\}_{\nu=\widetilde{\nu}} - \left.\frac{\partial P(\nu)}{\partial \nu}\right\}_{\nu=\nu_o} = o_p(1).
$$

Furthermore, by similar arguments

$$
\begin{aligned}
& C_8(\nu_o; \widehat{\omega}(\nu_o), \widehat{\beta}, \widehat{\gamma}(\nu_o), \widehat{\theta}(\nu_o)) \\
\simeq \; & C_8(\nu_o; \omega_o, \beta_o, \gamma_o, \theta_o) + \left.\frac{\partial C_8(\nu_o; \omega, \beta_o, \gamma_o, \theta_o)}{\partial \omega}\right\}_{\omega=\omega_o} [\widehat{\omega}(\nu_o) - \omega_o] \\
& + \left.\frac{\partial C_8(\nu_o; \omega_o, \beta, \gamma_o, \theta_o)}{\partial \beta}\right\}_{\beta=\beta_o} [\widehat{\beta} - \beta_o] \\
& + \left.\frac{\partial C_8(\nu_o; \omega_o, \beta_o, \gamma, \theta_o)}{\partial \gamma}\right\}_{\gamma=\gamma_o} [\widehat{\gamma}(\nu_o) - \gamma_o] \\
& + \left.\frac{\partial C_8(\nu_o; \omega, \beta_o, \gamma_o, \theta)}{\partial \theta}\right\}_{\theta=\theta_o} [\widehat{\theta}(\nu_o) - \theta_o].
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& \sqrt{n}(\widehat{\nu}_m - \nu_o) \\
\simeq \; & \left[ \left.\frac{\partial P(\nu)}{\partial \nu}\right\}_{\nu=\nu_o} \right]^{-1} [m_n - Em_n] \\
& - \left[ \left.\frac{\partial P(\nu)}{\partial \nu}\right\}_{\nu=\nu_o} \right]^{-1} \left[ \left.\frac{\partial C_8(\nu_o; \omega, \beta_o, \gamma_o, \theta_o)}{\partial \omega}\right\}_{\omega=\omega_o} \sqrt{n}[\widehat{\omega}(\nu_o) - \omega_o] \right. \\
& - \left.\frac{\partial C_8(\nu_o; \omega_o, \beta, \gamma_o, \theta_o)}{\partial \beta}\right\}_{\beta=\beta_o} \sqrt{n}[\widehat{\beta} - \beta_o] \\
& - \left.\frac{\partial C_8(\nu_o; \omega_o, \beta_o, \gamma, \theta_o)}{\partial \gamma}\right\}_{\gamma=\gamma_o} \sqrt{n}[\widehat{\gamma}(\nu_o) - \gamma_o] \\
& - \left.\frac{\partial C_8(\nu_o; \omega, \beta_o, \gamma_o, \theta)}{\partial \theta}\right\}_{\theta=\theta_o} \sqrt{n}[\widehat{\theta}(\nu_o) - \theta_o] \\
\equiv \; & A^\intercal \sqrt{n} d(W_n),
\end{aligned}
$$

for some vector $A$ and some smooth function $d$ of the vector $W_n$, where

$$W_n = \begin{bmatrix} \frac{1}{n}\sum_{t=1}^n z_t \\ \frac{1}{n}\sum_{t=1}^n z_t^2 \\ \frac{1}{n}\sum_{t=1}^n z_t \text{sgn}(y_{t-1}) \\ \frac{1}{n}\sum_{t=2}^n z_t z_{t-1} \\ \vdots \\ \frac{1}{n}\sum_{t=p+2}^n z_t z_{t-p-1} \\ \frac{1}{n}\sum_{t=1}^n |y_t| \end{bmatrix} \equiv \frac{1}{n}\sum_{t=1}^n w_t.$$

Under our conditions

$$\sqrt{n}(W_n - EW_n) \Longrightarrow N(0, \Psi),$$

where $\Psi = \text{lrvar}(w_t) < \infty$. Therefore, $\sqrt{n}(\widehat{\nu}_m - \nu_o)$ is asymptotically normal. It follows that $\sqrt{n}(\widehat{\phi}_m - \phi_o)$ by further application of the delta method. □

# References


[1] Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307-327.#

[2] Brockwell, P.J., and R.A. Davis (2006). *Time Series: Theory and Methods.* Springer Verlag, Berlin.

[3] Demos, A. and D. Kyriakopoulou (2013). Asymptotic normality of the QMLE in the EGARCH(1,1) model. *working paper.*

[4] Devroye, L. (1986). Non uniform random number generation

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

[6] Francq, C., Horváth, L. and Zakoïan, J.-M. (2011). Merits and Drawbacks of Variance Targeting in GARCH Models, *Journal of Financial Econometrics*, 9(4): 619-656 first published online July 18, 2011 doi:10.1093/jjfinec/nbr004

[7] Kristensen, D. and O. Linton (2006), A closed form estimator for the GARCH(1,1) model, *Econometric Theory*, 22, 323-337.

[8] Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Springer Verlag, New York.

[9] Nelson, D. (1991), Conditional heteoskedasticity in asset returns: A new approach. *Econometrica* 59, 347-370.

[10] Robinson, P.M. (1988) The Stochastic Difference Between Econometric Statistics. *Econometrica* 56, 531-548.

[11] Straumann, D. and T. Mikosch (2006), Quasi-maximum likelihood estimation in conditionally heteroskedastic time series: A stochastic recurrence equations approach *Annals of Statistics* 34, 2449-2495.

[12] Subbotin, M.T. (1923). “On the Law of Frequency of Errors.” Matematicheskii Sbornik, 31, 296–301.

[13] Wellner, J. (2012). Log-concave distributions: definitions, properties, and consequences. http://www.proba.jussieu.fr/pageperso/picard/Paris-Jan-p1-small.pdf

[14] Wintenberger, O. (2012). QML estimation of the continuously invertible EGARCH(1,1) model. discussion paper.

[15] Zaffaroni P. (2009), Whittle estimation of EGARCH and other exponential volatility models, *Journal of Econometrics*, 151, 190-200.