Econometric Theory, **33**, 2017, 1013–1038.
doi:10.1017/S0266466616000256

AN ALMOST CLOSED FORM ESTIMATOR FOR THE EGARCH MODEL

Christian M. Hafner
Université catholique de Louvain

Oliver Linton
University of Cambridge

The exponential GARCH (EGARCH) model introduced by Nelson (1991) 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 are 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 optimization algorithm, etc. We present an alternative 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 independent 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 and an empirical example.

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.

We thank Eric Renault, Piotr Fryzlewicz, Dimitra Kyriakopoulou, Enno Mammen, Paolo Zaffaroni, Jean-Michel Zakoian, and three referees for helpful comments. Financial support of the *Académie universitaire Louvain* is gratefully acknowledged. Thanks to the ERC for financial support. Address correspondence to Christian M. Hafner, CORE and Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium; e-mail: christian.hafner@uclouvain.be.

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 (2013) proves consistency using continuous invertibility, and provides sufficient conditions for this to hold, which however seem to be restrictive. Similarly, Kyriakopoulou (2015) gives sufficient conditions for asymptotic normality, which also restrict the admissible parameter space. As an alternative to the MLE, Zaffaroni (2009) proposes a Whittle estimator and shows consistency and asymptotic normality of the identified parameters under weak conditions.

Unfortunately, the calculation of both these estimators requires the use of numerical multiparameter optimization procedures since closed form expressions are not available. Therefore, the resulting estimator depends on the implementation, with different optimization techniques leading to potentially different estimators. This has been demonstrated in Brooks, Burke, and Persand (2001) and McCullough and Renfro (1999) where different commercially available software packages were used to estimate EGARCH models by QML. Both studies reported markedly different outputs across the various packages, reflecting the different initialization and algorithmic strategies employed.

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. This line of work in GARCH models was started by Engle and Mezrich (1996) who introduced the so-called "target variance" approach, whereby the unconditional variance was estimated by the sample variance and as a consequence one less parameter had to be estimated from the resulting likelihood. In the classical GARCH(1,1) model, Kristensen and Linton (2006) have introduced a closed form estimator, which was used for example by Andrews and Guggenberger (2009) in their simulations on computational grounds. 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 parameters of the EGARCH model that is in closed form for a given innovation distribution. 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 the form of the news impact function used in the specification of the volatility process. We consider the special case where the innovation is in the class of generalized error distributions (GED) and we provide a moment-based estimator of this parameter. We derive the asymptotic properties of our moment-based estimator and illustrate its small sample performance via a simulation study. Finally, we apply the alternative estimators to a long series of S&P 500 daily returns.

2. THE SEMIPARAMETRIC EGARCH MODEL

Consider the following exponential GARCH (EGARCH(1,1)) model for the observed zero mean process y_t

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

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

where the following conditions are satisfied:

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

Under these conditions, the AR(1) process h_t is strongly and weakly stationary, as well as ergodic (Nelson, 1991, Theorem 2.1). If f has support $\mathbb{R}$, meaning it is positive everywhere, then Carrasco and Chen (2002, Corollary 8) have shown that h_t and y_t are strictly stationary and geometrically β-mixing. It follows that "instantaneous" functions of the series i.e., measurable functions of $(y_t, y_{t-1}, \ldots, y_{t-m})$ for finite m are also stationary and geometrically mixing, see Davidson (1994, Theorem 14.1).

Assumption 1 and equations (1) and (2) define a semiparametric model with regard to the error density f and the news impact curve g. We will later specialize according to popular choices for f and g, but without further assumptions we record below the second order properties of the series

$$z_t = \log y_t^2 = h_t + \log\xi_t^2, \tag{3}$$

which is the sum of an AR(1) process and an iid process (and hence an ARMA(1,1) process in the weak sense, see Proposition 1 below), and strongly and weakly stationary and ergodic. Define the mean zero vector of shocks $x_t = (v_t, g_t, \widetilde{u}_t)^{\intercal}$, where $v_t = \log\xi_t^2 - E\log\xi_t^2$, $\widetilde{u}_t = u_t - Eu_t$, where $u_t = \mathrm{sign}(y_t) = \mathrm{sign}(\xi_t)$, and $g_t = g(\xi_t)$. Then, with: $C_0(f) = E(u_t)$, $C_1(f) = E\left[\log\xi_t^2\right]$, $C_2(f) = \sigma_v^2 = \mathrm{var}(v_t)$, $D_1(f;g) = \mathrm{var}(g_t)$, $D_2(f;g) = \mathrm{cov}(v_t, g_t)$, $D_3(f;g) = \mathrm{cov}(u_t, v_t)$, $D_4(f;g) = \mathrm{cov}(u_t, g_t)$, we have

$$Ex_t x_t^{\intercal} = \begin{bmatrix} C_2(f) & D_2(f;g) & D_3(f;g) \\ D_2(f;g) & D_1(f;g) & D_4(f;g) \\ D_3(f;g) & D_4(f;g) & 1 - C_0^2(f) \end{bmatrix}.$$

Then define $\delta = D_2(f;g)/C_2(f)$, and $\sigma_\epsilon^2 = D_1(f;g) - D_2^2(f;g)/C_2(f)$. Then let $\phi = \delta - \beta$, and:

$$d = \frac{\phi\sigma_v^2}{\sigma_\epsilon^2 + \sigma_v^2(1+\phi^2)} \in [-1/2, 1/2]$$

$$\pi = \frac{1-\sqrt{1-4d^2}}{2d} \in [-1, 1]$$

$$\sigma_e^2 = \frac{\sigma_\epsilon^2 + \sigma_v^2(1+\phi^2)}{(1+\pi^2)} \geq 0.$$

PROPOSITION 1. *Under Assumption 1, for $\omega^* = \omega + (1-\beta)C_1(f)$ and π given above, we have* ("*to second order*")

$$z_t = \omega^* + \beta z_{t-1} + e_t + \pi e_{t-1}, \tag{4}$$

where e_t is iid with mean zero and finite variance σ_e^2. Furthermore, with $\omega^{**} = \omega^*/(1-\beta)$ and $\vartheta_0 = 1$, $\vartheta_j = (\pi+\beta)\beta^{j-1}$, $j \geq 1$,

$$z_t = \omega^{**} + \sum_{j=0}^{\infty} \vartheta_j e_{t-j}.$$

Finally, $\gamma_u(k) = \text{cov}(u_t, u_{t-k}) = 0$, and:

$$\mu = E[z_t] = C_1(f) + \frac{\omega}{1-\beta} \tag{5}$$

$$\gamma_z(0) = \text{var}(z_t) = \frac{D_1(f;g)}{1-\beta^2} + C_2(f) \tag{6}$$

$$\gamma_z(k) = \text{cov}[z_t, z_{t-k}] = \beta^{k-1}\left(\frac{\beta D_1(f;g)}{1-\beta^2} + D_2(f;g)\right), \quad k \geq 1 \tag{7}$$

$$\gamma_{zu}(k) = \text{cov}(z_t, u_{t-k}) = \begin{cases} 0 & \text{if } k < 0 \\ D_3(f;g) & \text{if } k = 0 \\ \beta^{k-1} D_4(f;g) & \text{if } k > 0. \end{cases} \tag{8}$$

The autocorrelations are defined through $\rho_z(k) = \gamma_z(k)/\gamma_z(0)$. Depending on the relative magnitudes of $D_1(f;g)$ and $D_2(f;g)$, the autocovariances and autocorrelations of z can in principle be positive or negative, although we would expect them to be positive for financial data. In the pure MA case ($\beta = 0$), $\rho_z(k) \in (-1/2, 1/2)$, but in the general case there is no such restriction on the autocorrelations. For $\beta < 1$, the autocovariances and autocorrelations decay geometrically. The cross autocorrelations depend on the sign of $D_4(f;g)$.

It follows from (7) that

$$\beta = \frac{\gamma_z(k+1)}{\gamma_z(k)} = \frac{\rho_z(k+1)}{\rho_z(k)} \tag{9}$$

for all $k \geq 1$, which identifies the parameter β regardless of f, g.

It follows from (4) to (8) that we can identify at most four parameters from the first and second order properties of z_t alone (and we need $\pi \neq 0$ and $\pi \neq -\beta$ in order to do this). The information in $\gamma_{zu}(k)$ permits the identification of one further parameter at most. The functions $g(.)$ and $f(.)$ are not separately identifiable from this information without further structure that reduces the number of free parameters.

The second order structure of z 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 the parameters of Nelson's model (which is a special case of (1) and (2)) and one must consider the bivariate process (z_t, u_t) as we have done. We turn to Nelson's special case next.

3. NELSON'S NEWS IMPACT CURVE

We now specialize to the model considered by Nelson (1991). He proposed a specific parametric news impact function g,

$$g(\xi_t) = \theta\xi_t + \alpha(|\xi_t| - \mathrm{E}|\xi_t|), \tag{10}$$

where θ and α are unknown parameters. In total there are four parameters in h_t : $\omega, \beta, \alpha,$ and θ. Proposition 1 suggests we may identify all these parameters from the second order properties of (z_t, u_t) given knowledge of f. First, we have

$$\omega(f) = (1-\beta)(\mu - C_1(f)), \tag{11}$$

which identifies ω given the density f. By Jensen's inequality, $C_1(f) \le 0$, so that $(1-\beta)\mu \le \omega$. Let $C_3(f) = \mathrm{var}(|\xi_t|)$, $C_4(f) = E|\xi_t|$, and $C_5(f) = \mathrm{cov}(\log\left(\xi_t^2\right), |\xi_t|)$. We have $C_3(f) = 1 - C_4^2(f)$. Provided the error distribution is symmetric about zero:

$$D_1(f;g) = \theta^2 + \alpha^2 C_3(f)$$

$$D_2(f;g) = \alpha C_5(f) \quad ; \quad D_3(f;g) = 0 \quad ; \quad D_4(f;g) = \theta C_4(f).$$

It is then straightforward to show that $\gamma_{zu}(1) = \theta C_4(f)$. Hence, write

$$\theta(f) = \frac{\gamma_{zu}(1)}{C_4(f)}. \tag{12}$$

We may also use the ratios $\theta(f) = \gamma_{zu}(k)/C_4(f)\beta^{k-1}$ for $k = 2, 3, \ldots$.

Furthermore, we have $\gamma_z(1) - \beta\gamma_z(0) = D_2(f;g) - \beta C_2(f) = \alpha C_5(f) - \beta C_2(f)$, from which we obtain

$$\alpha(f) = \frac{\gamma_z(1) - \beta(\gamma_z(0) - C_2(f))}{C_5(f)}. \tag{13}$$

Therefore, given knowledge of the symmetric density f, we can identify the true parameters α, θ of this news impact curve as well as the parameters ω, β. In the absence of this information, we have a mapping from $f \mapsto \omega(f), \alpha(f), \theta(f)$ that defines "pseudo true" parameter values. Specifically,

$$\omega(f) = \omega_o + (C_1(f_o) - C_1(f))(1-\beta_o) \quad ; \quad \theta(f) = \theta_o\frac{C_4(f_o)}{C_4(f)}$$

$$\alpha(f) = \alpha_o\frac{C_5(f_o)}{C_5(f)} + \beta_o\frac{C_2(f_o) - C_2(f)}{C_5(f)},$$

where subscript o denotes true value. If $f = f_o$, then: $\omega(f) = \omega_o$, $\alpha(f) = \alpha_o$, and $\theta(f) = \theta_o$.

In Section 4 we discuss estimation of $\eta(f) = (\omega(f), \beta, \theta(f), \alpha(f))$ based on the expressions (9), (11), (12), and (13).

4. THE CLOSED FORM ESTIMATOR

Suppose we have a sample $y_1, \ldots, y_n$ from (1) and (2). Define the sample mean and autocovariance function (for $k = 0, 1, 2, \ldots$) :

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} z_t$$

$$\hat{\gamma}_z(k) = \frac{1}{n}\sum_{t=k+1}^{n} \left(z_t - \hat{\mu}\right)\left(z_{t-k} - \hat{\mu}\right)$$

$$\hat{\gamma}_{zu}(k) = \frac{1}{n}\sum_{t=k+1}^{n} \left(z_t - \hat{\mu}\right) u_{t-k},$$

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

Motivated by Proposition 1 and (9), we propose the following moment estimators for β

$$\hat{\beta} = \sum_{k=1}^{p} \frac{\hat{\gamma}_z(k+1)}{\hat{\gamma}_z(k)} w_k = \sum_{k=1}^{p} \frac{\widehat{\rho}_z(k+1)}{\widehat{\rho}_z(k)} w_k, \tag{14}$$

where $p \geq 1$ and w_k are some known weights such that $\sum_{k=1}^{p} w_k = 1$ (for example, $w_k = \lambda^{k-1}/\sum_{j=1}^{p} \lambda^{j-1}$ for some $\lambda \in (0,1)$). This is similar to Kristensen and Linton (2006). Note that $\hat{\beta}$ is independent of the specification of $g(\cdot)$ and of the innovation distribution f. In practice, the median of the ratios or a trimmed mean may provide superior performance, especially when p is large (so that some correlations may be small), and this is discussed further in the numerical section.

The remaining parameters of (2) and (10) depend on the error distribution f. Using (5)–(12), we obtain

$$\hat{\omega}(f) = \left(\hat{\mu} - C_1(f)\right)\left(1 - \hat{\beta}\right)$$

$$\hat{\theta}(f) = \frac{1}{C_4(f)} \sum_{k=1}^{p} \frac{\hat{\gamma}_{zu}(k)}{\widehat{\beta}^{k-1}} w'_k$$

$$\widehat{\alpha}(f) = \frac{1}{C_5(f)} \left\{ \sum_{k=1}^{p} \frac{\widehat{\gamma}_z(k)}{\widehat{\beta}^{k-1}} w''_k - \widehat{\beta}\left(\widehat{\gamma}_z(0) - C_2(f)\right) \right\},$$

for $p \geq 1$ and w'_k, w''_k some known weights such that $\sum_{k=1}^{p} w'_k = \sum_{k=1}^{p} w''_k = 1$. If f were known, e.g., when f is standard Gaussian, this would be a complete estimation procedure. If f is not known, then we are estimating the pseudo-true parameters $\omega(f), \theta(f)$, and $\alpha(f)$.

5. ESTIMATION OF PARAMETERS OF THE GED ERROR DENSITY

Nelson (1991) also assumed that ξ_t followed a standardized generalized error distribution, $\xi_t \sim GED(\nu)$, with mean zero, variance one, and with density function given by

$$f_\nu(\xi) = \frac{\nu \exp\left\{-(1/2)|\xi/\lambda(\nu)|^\nu\right\}}{\lambda(\nu)2^{1+1/\nu}\Gamma(1/\nu)}, \quad \nu > 0 \tag{15}$$

where $\lambda(\nu) = \left\{2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)\right\}^{1/2}$ and Γ is the gamma function. The GED includes the normal as a special case ($\nu = 2$), but allows for fat tails ($\nu < 2$) while maintaining finiteness of all moments of ξ_t. 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 the only member of this family that is strictly log-concave is $\nu = 2$, see Example 2.14 of Saumard and Wellner (2014). It is infinitely differentiable in ν except at the point $\xi = 0$.

With an abuse of notation we now denote $C_1(\nu) = E\left[\log \xi_t^2\right]$, $C_2(\nu) = \operatorname{var}\left(\log \xi_t^2\right)$, $D_2(\nu; g) = \operatorname{cov}\left(\log\left(\xi_t^2\right), g(\xi_t)\right)$, $C_3(\nu) = \operatorname{var}(|\xi_t|) = 1 - C_4^2(\nu)$, $C_4(\nu) = E|\xi_t|$, and $C_5(\nu) = \operatorname{cov}\left(\log\left(\xi_t^2\right), |\xi_t|\right)$. These quantities can be computed numerically for any ν and have “almost closed form” expressions, such as:

$$C_1(\nu) = \frac{2}{\nu}\psi(1/\nu) + \ln\Gamma(1/\nu) - \ln\Gamma(3/\nu), \tag{16}$$

$$C_4(\nu) = \lambda(\nu)2^{1/\nu}\Gamma(2/\nu)/\Gamma(1/\nu), \tag{17}$$

$$C_2(\nu) = (2/\nu)^2\Psi(1/\nu) \quad ; \quad C_5(\nu) = (2/\nu)C_4(\nu)(\psi(2/\nu) - \psi(1/\nu)), \tag{18}$$

where ψ, Ψ are, respectively, the digamma and trigamma functions (the first and second derivative of $\log\Gamma$). Equation (16) is shown in the appendix, (17) is given by (A1.8) of Nelson (1991), and (18) by Zaffaroni (2009, eq. 19). These are smooth functions of ν.[1]

We next consider how to estimate ν from the data in Nelson's model (10) and (15). So far we have used two pieces of information: $\gamma_{zu}(1)$ and $\gamma_z(1) - \beta\gamma_z(0)$ to identify the parameters θ and α given the value of ν. We propose to identify ν (and hence the parameters θ and α (and ω)) from the remaining information contained in $\gamma_z(0)$. We define our approach for the estimation of ν given our closed form estimators of $\eta(\nu) = (\omega(\nu), \beta, \theta(\nu), \alpha(\nu))$.

5.1. The Moment Estimator

We propose to estimate ν from the remaining information contained in $\gamma_z(0)$ using a purely moment-based estimator, see the review of Renault (2009) for other applications of moment-based estimators to volatility models. Let

$$\begin{aligned} M(\nu) &= \left(1-\beta_o^2\right)\left\{\gamma_z(0) - C_2(\nu)\right\} - \theta^2(\nu) - \alpha^2(\nu)C_3(\nu), \\ &= \theta_o^2 - \theta^2(\nu) + \alpha_o^2 C_3(\nu_o) - \alpha^2(\nu)C_3(\nu) + \left(1-\beta_o^2\right)\left\{C_2(\nu_o) - C_2(\nu)\right\}, \end{aligned}$$

where subscript o denotes true value, $\alpha(\nu) = \left[\gamma_z(1) - \beta_o(\gamma_z(0) - C_2(\nu))\right]/C_5(\nu)$, and $\theta(\nu) = \gamma_{zu}(1)/C_4(\nu)$. We have $M(\nu_o) = 0$. We will assume that $M(\nu) \neq 0$ at least in a neighborhood of ν_o. In Figure 1 we show $M(\nu)$ for three different values of $\nu_o : \nu_o = 1.5$, $\nu_o = 2$, and $\nu_o = 2.5$, and the parameter values $\beta_o = 0.97$, $\alpha_o = 0.5, \theta_o = -0.1$. In each case there is a unique crossing of zero on the range $[1,3]$. Note that the slope of M at ν_o, $M'(\nu_o)$, decreases from the case $\nu_o = 1.5$ to the case $\nu_o = 2.5$, although even in that case there is a well defined unique solution on this range of parameter values.

We now turn to estimation. Define for each $\nu \in V$

$$M_n(\nu) = \left(1 - \widehat{\beta}^2\right)\left\{\widehat{\gamma}_z(0) - C_2(\nu)\right\} - \widehat{\theta}(\nu)^2 - \widehat{\alpha}(\nu)^2 C_3(\nu). \tag{19}$$

We define $\widehat{\nu}$ as any zero of $M_n(\nu)$ over V or (if there is no zero) more generally

$$\widehat{\nu} = \arg\min_{\nu \in V} |M_n(\nu)|,$$

which always exists. This can also be computed by a univariate grid search. For the purpose of defining a numerical grid-search algorithm, we shall restrict attention to $\nu \in V$, where $V = \left[\underline{\nu}, \overline{\nu}\right] \subset (0,\infty)$ is a compact set with arbitrary lower and upper bounds. It is convenient, but not necessary under our assumptions, to set $\underline{\nu} = 1$, as this ensures a finite unconditional variance of y_t by Theorem A1.2 of Nelson (1991). For the numerical results of Section 7 we shall use $V = [1,3]$.

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

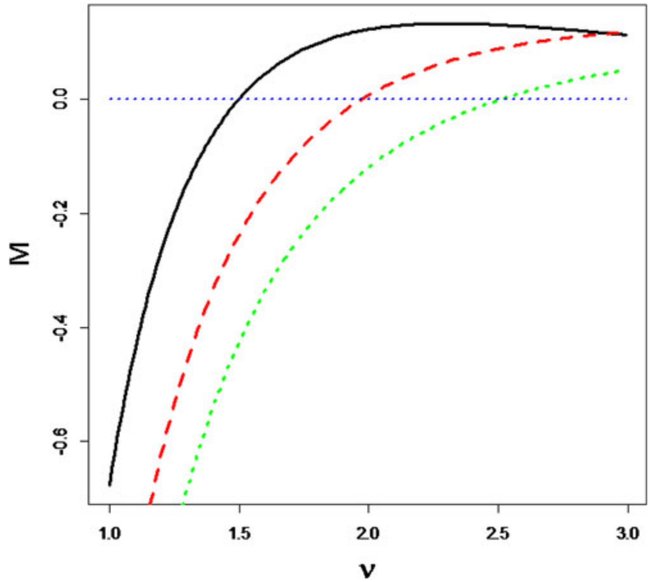

FIGURE 1. The solid line corresponds to $\nu_o = 1.5$, the dashed to $\nu_o = 2$, and the dotted to $\nu_o = 2.5$.

based on estimated parameters. The theoretical properties are much easier to handle: The quantity $M_n(\nu)$ is a smooth function of the autocovariances of z and the cross autocovariances with u_t as well as a nonlinear function of ν. Consequently, one can obtain consistency and asymptotic normality of all the parameter estimates by standard methods. On the other hand, these estimators are not fully efficient. We suppose that our estimators may be used in their own right where computational and theoretical reasons impel, but they also may be used as starting values to compute the MLE, as in Kristensen and Linton (2006) proposed for GARCH models. In the working paper version of this paper, we outlined a specific Newton–Raphson algorithm for this purpose. There are some theoretical results that support the idea that a one step of this algorithm can achieve very close approximation to the MLE, see Robinson (1988, Theorem 2).

6. ASYMPTOTIC PROPERTIES

We present some properties of the estimators defined in Sections 4 and 5. We first give the assumptions we shall use, we then consider the closed form/profiled estimators of Section 4, and then consider in particular the estimator of the full parameter vector defined in Section 5.1.

6.1. Assumptions

To recap, we have defined a semiparametric EGARCH model with two submodels: one where g is parametrically specified, and the second where both g and f are parametrically specified. We strengthen the moment conditions in Assumption 1 for the purposes of conducting inference. The three different models are defined through the following assumptions.

Assumption 1i. ξ_t is i.i.d. with bounded density f, which has support $\mathbb{R}$, where $E(\xi_t) = 0$ and $\mathrm{var}(\xi_t) = 1$ and $E\left[|\xi_t|^4\right] < \infty$, while $g(\cdot)$ is a measurable function such that $\mathrm{E}\left[g(\xi_t)\right] = 0$ and $0 < E\left[|g(\xi_t)|^4\right] < \infty$. The parameter β satisfies $\beta \neq 0$ and $|\beta| < 1$.

Assumption 2i. In addition to Assumption 1i, the unknown density f is symmetric about zero and $g(\xi_t) = \theta\xi_t + \alpha(|\xi_t| - \mathrm{E}|\xi_t|)$, where $C_5(f) \neq 0$.

Assumption 3i. In addition to Assumption 2i, the density f is GED with unknown parameter $\nu \in V$, where V is a compact subset of $(0, \infty)$.

The moment conditions on the innovations are quite mild and there is no restriction on the implied moments for y_t so far, unlike in Kristensen and Linton (2006). The boundedness of the density of ξ_t implies that all moments of $\log \xi_t^2$ exist.[2] This assumption could be weakened to allow the density to grow at the origin at rate less than unity.

The estimator $\hat{\beta}$ is a smooth function of the first $p+1$ autocorrelations of z_t 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

 https://doi.org/10.1017/S0266466616000256

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 to be consistent. 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 ξ_t). This is true without the GED assumption, although the probability limit would obviously depend on the underlying distribution f.

6.2. Asymptotics of the Closed Form/Profiled Estimator

Define the vector $\chi_t = \left[a_t^\intercal, b_t^\intercal, c_t^\intercal\right]^\intercal$ with components

$$a_t = \begin{bmatrix} z_t - \mu \\ (z_t-\mu)^2 \end{bmatrix};\ b_t = \begin{bmatrix} (z_t-\mu)(z_{t-1}-\mu) \\ \vdots \\ (z_t-\mu)(z_{t-p-1}-\mu) \end{bmatrix};\quad c_t = \begin{bmatrix} (z_t-\mu)u_{t-1} \\ \vdots \\ (z_t-\mu)u_{t-p} \end{bmatrix},$$

and let $a_t = (a_{1t}, a_{2t})^\intercal$, $b_t = (b_{1t}, \ldots, b_{pt})^\intercal$, and $c_t = (c_{1t}, \ldots, c_{pt})^\intercal$. Then define the $(2p+2)\times(2p+2)$ matrix Ξ as the long run variance of χ_t

$$\begin{aligned}\Xi(f,g) &= \lim_{n\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\chi_t\right) = \sum_{j=-\infty}^{\infty}\Gamma_\chi(j) \\ &= \begin{bmatrix} \Xi_{aa} & \Xi_{ab} & \Xi_{ac} \\ \Xi_{ba} & \Xi_{bb} & \Xi_{bc} \\ \Xi_{ca} & \Xi_{cb} & \Xi_{cc} \end{bmatrix} = \begin{bmatrix} \Xi_{LL} & \Xi_{LN} \\ \Xi_{NL} & \Xi_{NN} \end{bmatrix}, \end{aligned} \tag{20}$$

where $\Gamma_\chi(j)$ denotes the autocovariance matrix of the vector χ_t. Here, Ξ_{LL} is the submatrix of Ξ corresponding to the vector $\left[a_t^\intercal, b_t^\intercal\right]^\intercal$. The limit in (20) exists by virtue of the moment and mixing properties of the process χ_t. One could in principle derive an explicit formula for Ξ in terms of the parameters ω, β, and the first four moments of the vector of shocks $x_t = \left(\upsilon_t, g_t, \widetilde{u}_t\right)^\intercal$ by extending the calculations in Proposition 1 to consider fourth order properties of the process $(z_t, u_t)^\intercal$. This would be similar to the calculations performed in Francq, Horváth, and Zakoïan (2011, Theorem 2.1) for the variance targeting GARCH estimator, except more complicated due to the more complicated news impact curve in this case. For example, the long run variance of $a_{t1} = z_t - \mu$ is

$$\begin{aligned}\gamma_z(0) + 2\sum_{j=1}^{\infty}\gamma_z(j) &= \frac{D_1(f;g)}{1-\beta^2} + C_2(f) + 2\sum_{j=1}^{\infty}\beta^{j-1}\left(\frac{\beta D_1(f;g)}{1-\beta^2} + D_2(f;g)\right) \\ &= \frac{D_1(f;g)}{1-\beta^2} + C_2(f) + \frac{2}{1-\beta}\left(\frac{\beta D_1(f;g)}{1-\beta^2} + D_2(f;g)\right).\end{aligned}$$

For the other quantities, similar expressions could be obtained except that they involve additional constants depending on f, g. One key difference from

Francq et al. (2011), however, is that in the EGARCH model, we do not have consistent estimates of the parameters $\omega_o, \theta_o, \alpha_o$ unless we know f. Even for the parameter β, where knowledge of f, g is not necessary for consistent estimation, it is required for evaluation of the asymptotic variance because the relevant part of $\Xi(f,g)$ even in that case depends on f, g.

Define the weight vectors:

$$\pi_0(w) = \left(-\frac{w_1\rho_z(2)}{\rho_z(1)^2}, \frac{w_2}{\rho_z(1)} - \frac{w_3\rho_z(3)}{\rho_z(2)^2}, \dots, \frac{w_p}{\rho_z(p)} - \frac{w_{p+1}\rho_z(p)}{\rho_z(p-1)^2}, -\frac{w_{p+1}}{\rho_z(p)}\right)$$

$$\pi_1(w,\beta) = \left(w_1, \frac{w_2}{\beta}, \dots, \frac{w_p}{\beta^{p-1}}, 0\right)^{\top},$$

$$\pi_2(w,\beta) = -\left(0, \frac{w_2}{\beta^2}, \dots, \frac{(p-1)w_p}{\beta^p}\right)^{\top}.$$

Then define the $4 \times (2p+2)$ matrix

$$\mathcal{A}(f) = \begin{bmatrix} 1-\beta & 0 & -(\mu - C_1(f))\pi_0(w)^{\top} & 0 \\ 0 & 0 & \pi_0(w)^{\top} & 0 \\ 0 & 0 & \frac{i^{\top}\pi_2(w',\beta)}{C_4(f)}\pi_0^{\top} & \frac{1}{C_4(f)}\pi_1\left(w',\beta\right)^{\top} \\ 0 & -\frac{\beta}{C_5(f)} & \frac{1}{C_5(f)}\pi_1(w,\beta)^{\top} - \left(\frac{\pi_2(w'',\beta)^{\top}\gamma_z}{C_5(f)} + \frac{\gamma_z(0)-C_2(f)}{C_5(f)}\right)\pi_0^{\top} & 0 \end{bmatrix}, \quad \textbf{(21)}$$

where $\gamma_z = (\gamma_z(1), \dots, \gamma_z(p+1))^{\top}$ and $i = (1,1,\dots,1)^{\top}$. Define $\widehat{\eta}(f) = \left(\widehat{\omega}(f), \widehat{\beta}, \widehat{\theta}(f), \widehat{\alpha}(f)\right)^{\top}$ for the given f, and let $\eta(f) = (\omega(f), \beta_o, \theta(f), \alpha(f))^{\top}$ be defined as the probability limit of $\widehat{\eta}(f)$ under the given density f. Note that we do not assume that $f = f_o$ the true error density.

THEOREM 1. *Suppose that Assumption 1i holds. Then,*

$$\sqrt{n}\left[\widehat{\beta} - \beta_o\right] \Longrightarrow N\left(0, \pi_0(w)^{\top}\,\Xi_{bb}(f,g)\pi_0(w)\right). \quad \textbf{(22)}$$

Suppose further that Assumption 2i holds. Then

$$\sqrt{n}\left[\widehat{\eta}(f) - \eta(f)\right] \Longrightarrow N\left(0, \mathcal{A}(f)\,\Xi(f,g)\mathcal{A}(f)^{\top}\right). \quad \textbf{(23)}$$

For the construction of consistent confidence intervals, it suffices to apply a suitable nonparametric estimator of the long run variance of the estimable series χ_t or to use the subsampling method applied directly to the data (Politis and Romano, 1994). If one assumes knowledge of f, then there is a simpler option—to use the parametric bootstrap and simulate from this error density, see below for more discussion in the concrete GED case.

If Assumption 3i holds, then:

$$\omega(\nu) = \omega_o + (C_1(\nu_o) - C_1(\nu))(1-\beta_o) \quad ; \quad \theta(\nu) = \theta_o\frac{C_4(\nu_o)}{C_4(\nu)}$$

$$\alpha(\nu) = \alpha_o\frac{C_5(\nu_o)}{C_5(\nu)} + \beta_o\frac{C_2(\nu_o) - C_2(\nu)}{C_5(\nu)},$$

where subscript o denotes true value. If ν_0 were known, Theorem 1 could be used to provide standard errors and conduct inference about $\eta(\nu_o)$. Standard errors can be constructed by the methods discussed above. On the other hand, if $\nu \in V$ is unknown, then one could use (23) to conduct conservative inference. Specifically, suppose that $\mathcal{C}_n(\nu)$ is a consistent confidence region for some subvector of $\eta(\nu)$, say $\omega(\nu)$, i.e., for each ν

$$\Pr\left[\omega(\nu) \in \mathcal{C}_n(\nu)\right] \to 1-\alpha.$$

Then $\cup_{\nu \in V}\mathcal{C}_n(\nu)$ is a conservative confidence region for $\omega(\nu_o)$, i.e.,

$$\lim_{n\to\infty} \Pr\left[\omega(\nu_o) \in \cup_{\nu \in V}\mathcal{C}_n(\nu)\right] \geq 1-\alpha.$$

6.3. Asymptotics of the Full Parameter Estimator

We now turn to the properties of the estimator of all parameters proposed in Sections 3 and 4.2 under the full Nelson specification. Let $\widehat{\phi}_m = \left(\widehat{\omega}\left(\widehat{\nu}_m\right), \widehat{\beta}, \widehat{\theta}\left(\widehat{\nu}_m\right), \widehat{\alpha}\left(\widehat{\nu}_m\right), \widehat{\nu}_m\right)^\top$ and $\phi_o = (\omega_o, \beta_o, \theta_o, \alpha_o, \nu_o)^\top$. Define:

$$\Omega_{\phi\phi}(\phi_o) = \begin{bmatrix} \Omega_{\eta\eta} & \Omega_{\eta\nu} \\ \Omega_{\eta\nu}^\top & \Omega_{\nu\nu} \end{bmatrix} \quad ; \quad \Omega_{\nu\nu}(\phi_o) = \frac{\mathcal{B}(\nu_o)^\top \Xi(\phi_o) \mathcal{B}(\nu_o)}{[M'(\nu_o)]^2}$$

$$\Omega_{\eta\eta}(\phi_o) = \left[\mathcal{A}(\nu_o) + \eta'(\nu_o)\left[M'(\nu_o)\right]^{-1}\mathcal{B}(\nu_o)^\top\right] \Xi(\phi_o) \left[\mathcal{A}(\nu_o) + \eta'(\nu_o)\left[M'(\nu_o)\right]^{-1}\mathcal{B}(\nu_o)^\top\right]^\top$$

$$\Omega_{\eta\nu}(\phi_o) = \frac{1}{M'(\nu_o)}\left[\mathcal{A}(\nu_o) + \eta'(\nu_o)\left[M'(\nu_o)\right]^{-1}\mathcal{B}(\nu_o)^\top\right] \Xi(\phi_o)\mathcal{B}(\nu_o)$$

$$\mathcal{B}(\nu_o) = \begin{bmatrix} 0 \\ (1-\beta_o^2) + 2\alpha_o \frac{\beta_o C_3(\nu_o)}{C_5(\nu_o)} \\ \begin{bmatrix} -2\alpha_o C_3(\nu_o) \frac{1}{C_5(\nu_o)} \pi_1(w, \beta_o)^\top \\ -\left(2\beta_o(\gamma_z(0) - C_2(\nu_o)) + 2\theta_o \frac{i^\top \pi_2(w', \beta_o)}{C_4(\nu_o)} + \frac{\pi_2(w'', \beta_o)^\top \gamma_z}{C_5(\nu_o)} + \frac{\gamma_z(0) - C_2(\nu_o)}{C_5(\nu_o)}\right)\pi_0(w)^\top \\ -\frac{2\theta_o}{C_4(\nu_o)} \pi_1(w', \beta_o)^\top \end{bmatrix} \end{bmatrix}.$$

The quantities $\eta'(\nu_o)$ and $M'(\nu_o)$ can be calculated by numerical differentiation, but we give explicit formulae in the appendix.

THEOREM 2. *Suppose that Assumption 3i holds and suppose that $M'(\nu_o) \neq 0$. Then*

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

For inference we may consider the parametric bootstrap (Andrews, 1997; Bai, 2003). The conditional distribution of y_t given the past is parametric, i.e., $y_t|\mathcal{F}_{t-1} \sim F(\cdot|\mathcal{F}_{t-1}, \phi) = F_\nu\left(e^{-h_t(\eta)/2}y_t\right)$, where F_ν is the c.d.f. of the GED distribution with parameter ν. Therefore, we may proceed as follows. For given $\widehat{\phi}_m$ we generate a sample recursively by sampling repeatedly

$$y_t^* \sim F\left(\cdot\,|\mathcal{F}_{t-1}^*, \widehat{\phi}_m\right),$$

where $\mathcal{F}_{t-1}^{*} = \{y_{t-1}^{*}, \ldots, y_{1}^{*}\}$. We then compute the estimator $\widehat{\phi}_{m}^{*}$ from the sample $\{y_{T}^{*}, \ldots, y_{1}^{*}\}$, and repeat to obtain B bootstrap samples, see Kheifets (2015) for recent discussion and theoretical analysis. Mineo and Ruggieri (2005) provide **R** code for simulating from the GED that is publicly available.

7. SOME EXTENSIONS

7.1. A Pivotal Test of the Leverage Effect

We next show that one can carry out a test of leverage in the model (10) without knowledge of ν or even f (so long as it is symmetric about zero), i.e., we are operating under Assumption 2i. Our estimator $\hat{\theta}(f)$, for any f, can be used to test for a leverage effect within the Nelson model (10). The reason is that in constructing the t-ratio the constant term $C_4(f)$ is cancelled out. That is, we may compute

$$t = \frac{\hat{\theta}(f)}{\mathrm{se}\left(\hat{\theta}(f)\right)} = \frac{\frac{1}{n}\sum_{t=1}^{n}\zeta_t}{\sqrt{n \mathrm{lrvar}(\zeta_t)}},$$

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

$$\mathrm{var}(z_t u_{t-1}) = E z_t^2 - E^2 z_t u_{t-1} = \frac{\theta^2 + \alpha^2 C_3(f)}{1-\beta^2} + C_2(f) + \left(C_1(f) + \frac{\omega}{1-\beta}\right)^2 - \theta^2 C_4^2(f).$$

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

$$\widehat{t} = \frac{\sqrt{n}\overline{\zeta}}{\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(\zeta_t - \overline{\zeta}\right)^2}}, \quad \overline{\zeta} = \frac{1}{n}\sum_{t=1}^{n}\zeta_t. \tag{24}$$

Then under the null hypothesis of no leverage, $\theta = 0$, and under Assumption 2i, the statistic $\widehat{t}$ is asymptotically standard normal.

7.2. A Unit Root Test

The development so far has assumed that $|\beta| < 1$, we now consider what happens in the unit root case, i.e., when $\beta = +1$, but otherwise Assumption 1i holds. In this case, the correlogram of the level series is not informative. Consider the differenced series

$$\begin{aligned}\Delta z_t &= \omega + (\beta - 1)h_{t-1} + g_{t-1} + \log \xi_t^2 - \log \xi_{t-1}^2 \\ &= \omega + (\beta - 1)z_{t-1} + g_{t-1} + \log \xi_t^2 - \beta \log \xi_{t-1}^2.\end{aligned}$$

Under the unit root hypothesis, $\Delta z_t = \omega + g_{t-1} + v_t - v_{t-1}$ is a stationary first order moving average process

PROPOSITION 2. *Suppose that Assumption 1i holds except that* $\beta = 1$. *Then, the first two moments of* Δz_t *are given by*

$$\mu_{\Delta z} = E[\Delta z_t] = \omega,$$
$$\gamma_{\Delta z}(0) = \mathrm{var}(\Delta z_t) = D_1(f,g) + 2C_2(f) - D_2(f,g),$$
$$\gamma_{\Delta z}(k) = \mathrm{cov}[\Delta z_t, \Delta z_{t-k}] = \begin{cases} D_2(f,g) - C_2(f) & \text{if } k = 1 \\ 0 & \text{if } k > 1. \end{cases}$$

Therefore we can test the unit root hypothesis by examining $\gamma_{\Delta z}(k)$, $k = 2, \ldots$. Specifically, we let

$$\tau = n \sum_{k=1}^{p} \widehat{\rho}^2_{\Delta z}(2k), \tag{25}$$

where $\rho_{\Delta z}(k) = \gamma_{\Delta z}(k)/\gamma_{\Delta z}(0)$ and $\widehat{\rho}_{\Delta z}(k)$ is the sample quantity (for an MA(1) process the autocorrelations $\widehat{\rho}_{\Delta z}(2k)$, $k = 1, 2, \ldots$ are asymptotically independent). Then under the null hypothesis (that $\beta = 1$), $\tau \Longrightarrow \chi^2(p)$.

8. NUMERICAL RESULTS

8.1. 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 β. The first, $\hat{\beta}_e$, is a simple mean of ratios, i.e., it is the estimator in (14) with $w_j = 1/p$; the second, $\hat{\beta}_w$, is a mean of ratios with linearly declining weights, i.e., it is the estimator in (14) with $w_j = 2(1 - j/(p+1))/p$; the third is the median: $\hat{\beta}_{rob} = \mathrm{med}\{\hat{\gamma}(j+1)/\hat{\gamma}(j)\}$, and the fourth is the OLS without intercept regression estimator given by (26)

$$\widehat{\beta}_R = (a^{\intercal} a)^{-1} a^{\intercal} e, \tag{26}$$

where $e = \left(\hat{\gamma}_z(2) w_1^{1/2}, \ldots, \hat{\gamma}_z(p+1) w_p^{1/2}\right)^{\intercal}$, $a = \left(\hat{\gamma}_z(1) w_1^{1/2}, \ldots, \hat{\gamma}_z(p) w_p^{1/2}\right)^{\intercal}$. All four estimators depend on the number of terms p included, which we increase in steps of ten from 10 to 50. For higher values of p, the estimates of the ACF of z_t become too noisy and all estimates of β suffer from high variability. Results are reported in Table 1. It is remarkable that both the weighted and unweighted means underperform for higher p, due to high variability of 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.

We next consider the performance of the closed form estimator of the remaining model parameters. For the estimator of β we use a fixed order $p = 10$ and equal weights, while for the estimator of α we use $q = 1$. Experiments with higher q

Table 1. Simulation results: estimation of β using simple mean $(\hat{\beta}_e)$, 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, $\beta = 0.9, \omega = -0.3, \theta = -0.1$, and $\alpha = 0.5$. s.d. is the standard deviation, and the number of replications is 1000

p	$\hat{\beta}_e$	s.d.	$\hat{\beta}_w$	s.d.	$\hat{\beta}_{rob}$	s.d.	$\hat{\beta}_{ols}$	s.d.
				$n = 1,000$				
10	0.957	0.542	0.950	0.324	0.907	0.099	0.868	0.060
20	0.795	3.647	0.892	2.038	0.875	0.131	0.845	0.070
30	0.643	5.838	0.762	2.946	0.831	0.161	0.824	0.080
40	0.518	6.418	0.694	3.857	0.805	0.173	0.808	0.087
50	0.516	5.379	0.624	4.322	0.788	0.177	0.798	0.092
				$n = 10,000$				
10	0.905	0.015	0.904	0.012	0.900	0.024	0.897	0.013
20	0.918	0.216	0.912	0.042	0.900	0.024	0.894	0.012
30	0.899	2.030	0.928	0.552	0.892	0.035	0.892	0.013
40	0.747	4.314	0.880	1.451	0.874	0.050	0.889	0.013
50	0.729	3.894	0.838	2.251	0.856	0.061	0.887	0.013

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 ν we use the estimator proposed in Section 5.1, found by a grid search on the interval $[1,3]$. We compare with two other estimators, the full EGARCH MLE that maximizes $L(\phi) = \sum_{t=1}^n \ell_t(\phi)$ with respect to $\phi = (\omega, \beta, \theta, \alpha, \nu)^\top \in \mathbb{R}^5$, where:

$$\ell_t(\phi) = \ln\left(\nu/\lambda 2^{1+1/\nu}\Gamma(1/\nu)\right) - \frac{1}{2}\left\{|\xi_t(\phi)/\lambda|^\nu + h_t(\phi)\right\},$$

$$h_t(\phi) = \omega + \theta\xi_{t-1}(\phi) + \alpha(|\xi_{t-1}(\phi)| - C_4(\nu)) + \beta h_{t-1}(\phi),$$

where $\xi_t(\phi) = y_t \exp(-h_t(\phi)/2)$ and $h_1(\phi) = \omega/(1-\beta)$. We also compare with a profiled Likelihood method that uses our closed form estimators $\widehat{\eta}(\nu) = \left(\widehat{\omega}(\nu), \widehat{\beta}, \widehat{\theta}(\nu), \widehat{\alpha}(\nu)\right)$ to define a profiled likelihood $\widehat{L}(\nu) = \sum_{t=1}^n \widehat{\ell}_t(\nu)$, where:

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

where $\hat{\xi}_t(\nu) = \exp\left\{-\hat{h}_t(\nu)/2\right\} y_t$, and:

$$\hat{h}_t(\nu) = \hat{\omega}(\nu) + \hat{\theta}(\nu)\hat{\xi}_{t-1}(\nu) + \hat{\alpha}(\nu)\left(|\hat{\xi}_{t-1}(\nu)| - C_4(\nu)\right) + \hat{\beta}\hat{h}_{t-1}(\nu) \quad \textbf{(27)}$$

$$\hat{h}_1(\nu) = \frac{\hat{\omega}(\nu)}{1-\hat{\beta}} = (\hat{\mu} - C_1(\nu)). \quad \textbf{(28)}$$

We maximize $\widehat{L}(\nu)$ w.r.t. ν using a univariate grid search over the compact set V. Let $\widehat{\nu}_L$ denote the maximizer of $\widehat{L}(\nu)$. We argue that this estimator could be considered as similar in spirit to the "target variance" estimators of Engle and Mezrich (1996), because all parameters except ν are profiled out based on the first and second moment structure of z_t and the remaining parameter is found by maximizing the profiled likelihood.[3] Unfortunately, it is difficult to prove rigorously the consistency even of this estimator of ν. The issue is due to the nonlinear recursive equation (27), which are very difficult to analyze away from the true parameter value, which is exactly the same problem as for the original EGARCH MLE. Although we do not provide analytical results regarding this estimator it may be of value in terms of its computational benefits relative to the full MLE, so we include it in our numerical comparison.

In Table 2, we report the performance of the estimators. The results corroborate the theoretical finding that the estimators are consistent. Estimation of ω, the scale parameter, and θ, the sign effect, seems rather unaffected by the estimation of ν. That is, bias and variance of $\hat{\omega}$ and $\hat{\theta}$ are almost identical under the likelihood and

TABLE 2. Simulation results for estimated EGARCH(1,1) processes. True parameters are $\beta = 0.9$, $\omega = -0.3$, $\theta = -0.1$, and $\alpha = 0.5$. s.d. is the standard deviation, and the number of replications is 1000

	Gaussian ($\nu = 2$)				GED with $\nu = 1.5$			
	$n = 1,000$		$n = 10,000$		$n = 1,000$		$n = 10,000$	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
β	0.933	0.115	0.904	0.016	0.931	0.140	0.904	0.015
Profiled moment estimator								
ω	−0.196	0.333	−0.285	0.047	−0.214	0.405	−0.286	0.047
θ	0.105	0.189	−0.098	0.060	−0.091	0.201	−0.098	0.069
α	0.275	0.275	0.481	0.047	0.159	0.333	0.481	0.048
ν	2.355	0.535	2.024	0.182	1.879	0.495	1.518	0.098
Profiled likelihood estimator								
ω	−0.211	0.412	−0.284	0.048	−0.218	0.423	−0.299	0.050
θ	−0.096	0.197	−0.099	0.061	−0.096	0.218	−0.099	0.071
α	0.560	0.234	0.501	0.042	0.542	0.223	0.504	0.038
ν	1.796	0.344	1.964	0.123	1.398	0.232	1.485	0.078
Full MLE								
β	0.891	0.049	0.900	0.006	0.892	0.045	0.899	0.007
ω	−0.323	0.133	−0.297	0.019	−0.323	0.126	−0.302	0.022
θ	−0.099	0.031	−0.101	0.009	−0.099	0.037	−0.100	0.012
α	0.501	0.079	0.497	0.016	0.495	0.080	0.498	0.020
ν	2.005	0.160	2.003	0.043	1.510	0.105	1.500	0.033

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, α, and the GED parameter ν. The moment estimator of α 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 ν, 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.

8.2. Application

We investigated the performance of our different estimators of β on a large dataset, the demeaned daily (close to close) return on the S&P500 index from 1950 to 2012, a total of 15,757 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
β	0.9866	0.00135
ω	−0.2542	0.01729
θ	−0.0685	0.00367
α	0.1353	0.00650
ν	1.3726	0.01248

The process is quite persistent, which agrees with much earlier work. We carried out the unit root test of Section 6.2 and found the following test statistics with associated p-values: $\tau_5 = 6.91$ (0.227), $\tau_{10} = 14.37$ (0.157), and $\tau_{25} = 47.15$ (0.0047). The evidence against the unit root hypothesis only comes out once long lags are considered.

The estimated 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 against the null of no leverage effect. However, the nonparametric test based on (24) 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 β 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 2, we show the value of the estimated β against the number of lags

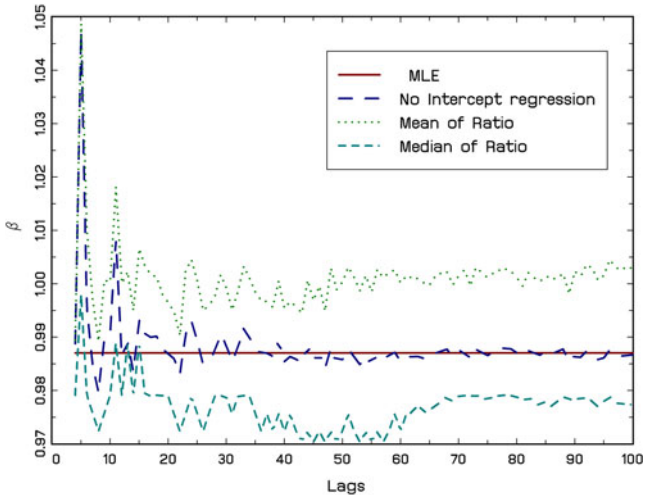

FIGURE 2. S&P 500: Three estimators of β as a function of the number of lags. The solid reference line is the maximum likelihood estimator.

p used for $p = 4, \ldots, 100$. 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

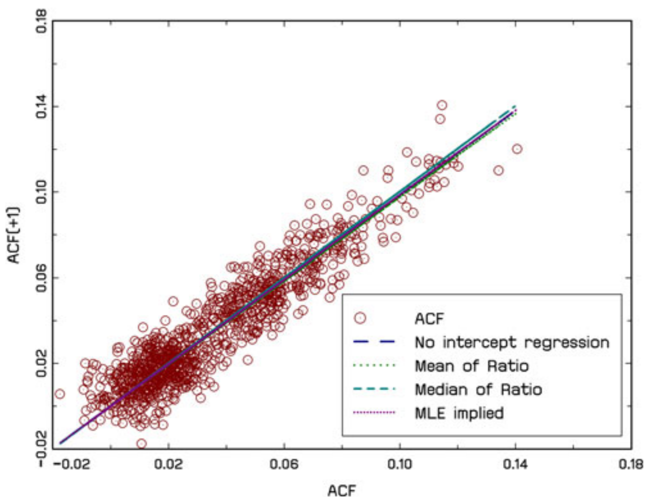

FIGURE 3. S&P 500: Scatter plot of empirical autocorrelations of log squared returns of order $k+1$ (vertical axis) and k (horizontal axis), for $k = 1, \ldots, 1001$, and regression lines corresponding to the alternative estimators.

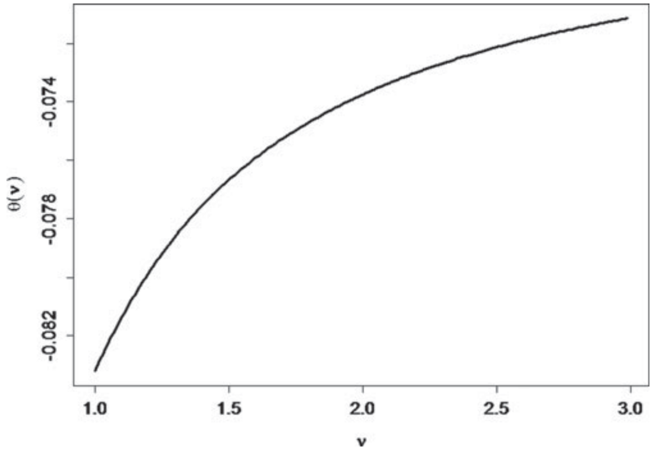

FIGURE 4. S&P 500: Profiled estimator of θ as a function of ν.

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.

We next show a scatter plot of the empirical autocorrelations along with the fitted regression line, see Figure 3. We show the first 1,001 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 β 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.

Based on the least squares estimator for β, the remaining parameters are estimated using the order $p = q = 100$ as follows: $\hat{\omega} = -0.1437$, $\hat{\theta} = -0.0760$, and $\hat{\alpha} = 0.1955$ for a given tail thickness estimator $\hat{\nu} = 1.58$ which minimizes $|M_n(\nu)|$. These estimates are quite close to the MLE. We computed the profiled estimators as functions of ν: ω and θ are relatively stable w.r.t. ν, while α is strongly varying, changing sign at about $\nu \approx 1.8$. We show in Figure 4 below the profiled estimator of θ.

9. CONCLUSIONS

We have shown that a simple closed form estimator of the EGARCH model is consistent, asymptotically normally distributed, and has reasonable finite sample properties. We recommend this estimator in large samples, or as starting values for estimators requiring numerical optimization. In applications involving a rolling window scheme and/or Monte Carlo experiments where one may have to compute parameter estimates many times, this method yields a substantial computational saving over the MLE.

NOTES

1. We have for any $x > 0$ that

$$\Psi(x) = \sum_{i=0}^{\infty} \frac{1}{(x+i)^2},$$

and this function is analytic over the positive real line.

2. This is because $E\left[\log \xi_t^2\right]^{2r} < \infty$ holds for uniformly distributed random variables ξ. To see this, consider the moment generating function (MGF) for $\log \xi^2$: $E\left[\exp\left\{t \cdot \left[\log \xi^2\right]\right\}\right] = E\left[\xi^{2t}\right]$, which is finite for all $t > -1/2$. Therefore the MGF of $\log \xi^2$ exists and all moments exist.

3. To be precise, it matches the mean of z_t, the correlation of z_t and u_{t-k}, and the autocovariance of z_t (but not the variance), whereas the original variance targeting just profiles out one parameter and links it to the unconditional variance. So it is not the same, but the idea is similar, and it simplifies the calculation of the likelihood.

REFERENCES

Andrews, D.W.K. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097–1128.

Andrews, D.W.K. & P. Guggenberger (2009) Hybrid and size-corrected subsampling methods. *Econometrica* 77, 721–762.

Bai, J. (2003) Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics* 85, 531–554.

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

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

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

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

Doukhan, P. (1994) *Mixing: Properties and Examples*. Springer-Verlag.

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

Engle, R.F. & J. Mezrich (1996) GARCH for groups. *Risk* 9, 36–40.

Francq, C., L. Horváth, & J.-M. Zakoïan (2011) Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* 9, 619–656.

Gradshteyn, I.S. & I.M. Ryzhik (2007) *Tables of Integrals, Series, and Products*. 7th ed. Academic Press.

Kheifets, I (2015) Specification tests for nonlinear dynamic models. *Econometrics Journal* 18, 67–94.

Kristensen, D. & O. Linton (2006) A closed form estimator for the GARCH(1,1) model. *Econometric Theory* 22, 323–337.

Kyriakopoulou, D. (2015) Asymptotic Normality of the QML Estimator of the EGARCH(1,1) Model. SSRN Working paper. Available at SSRN: http://ssrn.com/abstract=2236055 or http://dx.doi.org/10.2139/ssrn.2236055.

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.

Mineo, A.M. & M. Ruggieri (2005) A software tool for the exponential power distribution: The normalp package. *The Journal of Statistical Software* 12, 1–24.

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

Politis, D.N. & J.P. Romano (1994) Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* 22, 2031–2050.

Renault, E. (2009) Moment-based estimation of stochastic volatility models. In T. Andersen, R. Davis, J.-P. Kreiss, & T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 269–311, Springer Verlag.

Robinson, P.M. (1988) The stochastic difference between econometric statistics. *Econometrica* 56, 531–548.

Saumard, A. & J.A. Wellner (2014) Log-concavity and strong log-concavity: A review. *Statistics Surveys* 8, 45–114.

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

Subbotin, M.T. (1923) On the law of frequency of errors. *Matematicheskii Sbornik* 31, 296–301.

Wintenberger, O. (2013) Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* 40, 846–867.

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

APPENDIX

Proof of Proposition 1. Recall that $z_t = \omega + \beta z_{t-1} + g(\xi_{t-1}) + \log \xi_t^2 - \beta \log \xi_{t-1}^2$. Let $g_t = g(\xi_t)$ and $v_t = \log \xi_t^2 - E \log \xi_t^2$. Then we can write $z_t = \omega^* + \beta z_{t-1} + g_{t-1} + v_t - \beta v_{t-1}$. Consider the moving average part $m_t = g_{t-1} + v_t - \beta v_{t-1}$. We can write $g_t = \delta v_t + \epsilon_t$, where $\epsilon_t = g_t - \delta v_t$ is iid and uncorrelated with v_t, where: $\delta = \gamma_{vg}/\sigma_v^2$ and $\mathrm{var}(\epsilon_t) = \sigma_g^2 - \gamma_{vg}^2/\sigma_v^2 = \sigma_g^2 - \delta^2\sigma_v^2$. Then $m_t = \epsilon_{t-1} + v_t + (\delta - \beta)v_{t-1} = \epsilon_{t-1} + v_t + \phi v_{t-1}$, where $\mathrm{cov}(\epsilon_t, v_t) = 0$. We have $\mathrm{var}(m_t) = \sigma_\epsilon^2 + \sigma_v^2\left(1+\phi^2\right)$ and

$$\mathrm{cov}(m_t, m_{t-k}) = \begin{cases} \phi\sigma_v^2 & \text{if } k = 1 \\ 0 & \text{if } k > 1. \end{cases}$$

Therefore

$$\rho_m(1) = \mathrm{corr}(m_t, m_{t-1}) = \frac{\phi\sigma_v^2}{\sigma_\epsilon^2 + \sigma_v^2\left(1+\phi^2\right)}.$$

We compare with an MA(1) process $m_t = e_t + \pi e_{t-1}$, which has $\mathrm{var}(m_t) = \sigma_e^2\left(1+\pi^2\right)$ and $\mathrm{cov}(m_t, m_{t-k}) = \pi\sigma_e^2$. Equating parameters $\sigma_\epsilon^2 + \sigma_v^2\left(1+\phi^2\right) = \sigma_e^2\left(1+\pi^2\right)$ and $\pi\sigma_e^2 = \phi\sigma_v^2$, we obtain

$$\frac{\pi}{1+\pi^2} = \frac{\phi\sigma_v^2}{\sigma_\epsilon^2 + \sigma_v^2\left(1+\phi^2\right)} = d \in [-1/2, 1/2],$$

from which (4) is shown. ■

This representation shows that the second order properties of z_t identify three parameters: β, π, and σ_e^2, the mean identifies ω^*. Note that the process z_t is a (vector) linear process with

$$z_t = \omega^{**} + \sum_{j=0}^{\infty} a_j^{\top} \begin{bmatrix} g_t \\ v_t \end{bmatrix}$$

for some sequence of vectors a_j.

Proof of Theorem 1. Note that

$$\widehat{\beta} = Q\big(\widehat{\rho}_z(1), \ldots, \widehat{\rho}_z(p+1)\big),$$

where Q is differentiable at the point $(\rho_z(1), \ldots, \rho_z(p+1))$ assuming that $\min_{1 \le k \le p+1} |\rho_z(k)| > 0$ (which is implied by $\beta \ne 0$). We have by the CLT for stationary mixing process (Doukhan, 1994, p. 46)) that $\sqrt{n}\big(\widehat{\rho}_z(1) - \rho_z(1), \ldots, \widehat{\rho}_z(p+1) - \rho_z(p+1)\big)$ is jointly asymptotically normal. We can then apply the delta method to obtain the limiting distribution for $\sqrt{n}(\widehat{\beta} - \beta)$. Since we have to give the same argument for the other parameter estimates, and in that case we will use the autocovariance function, which will require stronger moment conditions, we will conclude here. Note that the moment conditions for the autocorrelation function based estimators are weaker than for those based on the autocovariance. However, the estimator $\widehat{\alpha}(f)$ depends also on the sample variance of z_t, and so the asymptotics for the full vector requires a theory for the sample autocovariance function to which we now turn.

Define the vector of sample mean and autocovariances $U_n = n^{-1} \sum_{t=1}^{n} \chi_t$ and the corresponding population quantity $U = E\chi_t$. From the stationarity and ergodicity, we can apply Theorem 13.12 of Davidson (1994) to obtain $U_n \to U$ in probability. Furthermore, under Assumption 1i, for the symmetric positive definite matrix Ξ, (using the CLT for geometrically mixing processes (Doukhan, 1994, p. 46), we have

$$\sqrt{n}\big[U_n - U\big] \Longrightarrow N(0, \Xi). \qquad \textbf{(A.1)}$$

We can write $\widehat{\eta}(f) = F(U_n; f)$ for some function F of the vector U_n and the error density f, and $\eta(f) = F(U; f)$. For example,

$$\omega(f) = (\mu - C_1(f)) \left(\sum_{k=1}^{p} \frac{\gamma_z(k+1)}{\gamma_z(k)} w_k \right).$$

The function F is twice continuously differentiable in first argument at the point U. By the continuous mapping theorem, $F(U_n; f) = \widehat{\eta}(f) \longrightarrow \eta(f) = F(U; f)$ in probability. Furthermore, by (A.1) and the delta method

$$\sqrt{n}\big(\widehat{\eta}(f) - \eta(f)\big) = \frac{\partial F}{\partial U^{\top}}(U; f)\sqrt{n}(U_n - U) + o_p(1), \qquad \textbf{(A.2)}$$

where the error term in (A.2) can be denoted $R_n(f)$. We give some more detail on this expansion. We have

$$\widehat{\omega}(f) - \omega(f) = (1-\beta)\frac{1}{n}\sum_{t=1}^{n}(a_{1t} - Ea_{1t}) - (\mu - C_1(f))\left(\widehat{\beta} - \beta\right) + o_p\left(n^{-1/2}\right)$$

$$\widehat{\beta} - \beta = \pi_0^{\top}(w)\frac{1}{n}\sum_{t=1}^{n}(b_t - Eb_t) + o_p\left(n^{-1/2}\right),$$

$$\widehat{\theta}(f)-\theta(f)=\frac{1}{C_4(f)}\frac{1}{n}\sum_{t=1}^{n}\pi_1\left(w',\beta\right)^{\mathsf{T}}(c_t-Ec_t)$$
$$+\frac{1}{C_4(f)}i^{\mathsf{T}}\pi_2\left(w',\beta\right)\left(\widehat{\beta}-\beta\right)+o_p\left(n^{-1/2}\right),$$

$$\widehat{\alpha}(f)-\alpha(f)=\frac{1}{C_5(f)}\frac{1}{n}\sum_{t=1}^{n}\pi_1\left(w'',\beta\right)^{\mathsf{T}}(b_t-Eb_t)-\frac{\pi_2\left(w'',\beta\right)^{\mathsf{T}}\gamma_z}{C_5(f)}\left(\widehat{\beta}-\beta\right) \tag{A.3}$$
$$-\frac{\gamma_z(0)-C_2(f)}{C_5(f)}\left(\widehat{\beta}-\beta\right)-\frac{\beta}{C_5(f)}\frac{1}{n}\sum_{t=1}^{n}(a_{2t}-Ea_{2t})+o_p\left(n^{-1/2}\right).$$

We can collect this as

$$\sqrt{n}\left(\widehat{\eta}(f)-\eta(f)\right)=\mathcal{A}(f)\frac{1}{n}\sum_{t=1}^{n}(\chi_t-E\chi_t)+o_p(1), \tag{A.4}$$

and the limiting distribution follows as above. ■

Proof of Theorem 2. We first establish the properties of $\widehat{\nu}_m$. We can write $M_n(\nu)=H(U_n;\nu)$ with H a twice continuously differentiable function in both its arguments. It follows that

$$\sup_{\nu\in V}|M_n(\nu)-M(\nu)|=o_p(1). \tag{A.5}$$

By the identifiable uniqueness assumption, if $\left|\widehat{\nu}_m-\nu_o\right|>\delta$, then $\left|M\left(\widehat{\nu}_m\right)\right|\geq\epsilon(\delta)$. Consequently

$$\Pr\left(\left|\widehat{\nu}_m-\nu_o\right|>\delta\right)\leq\Pr\left(\left|M\left(\widehat{\nu}_m\right)\right|\geq\epsilon(\delta)\right),$$

and it is sufficient to prove that for any $\epsilon(\delta)>0$, the latter probability goes to zero. But

$$\begin{aligned}\left|M\left(\widehat{\nu}_m\right)\right| &\leq \left|M\left(\widehat{\nu}_m\right)-M_n\left(\widehat{\nu}_m\right)\right|+\left|M_n\left(\widehat{\nu}_m\right)\right|, \text{(triangle inequality)}\\ &\leq \sup_{\nu\in V}|M(\nu)-M_n(\nu)|+\left|M_n\left(\widehat{\nu}_m\right)\right|, \text{(set inclusion)}\\ &\leq o_p(1)+\left|M_n\left(\widehat{\nu}_m\right)\right|, \text{(from the ULLN)}\\ &\leq o_p(1)+\left|M_n\left(\nu_o\right)\right|, \text{(from the definition of the estimator and set inclusion)}\\ &\leq o_p(1), \text{(from the ULLN)}.\end{aligned}$$

It follows that $\widehat{\nu}_m$ is (weakly) consistent.

By the triangle inequality,

$$\left|\widehat{\eta}(\widehat{\nu}_m)-\eta(\nu_o)\right|\leq\left|\widehat{\eta}(\widehat{\nu}_m)-\eta(\widehat{\nu}_m)\right|+\left|\eta(\widehat{\nu}_m)-\eta(\nu_o)\right|,$$

where

$$\Pr\left[\left|\widehat{\eta}(\widehat{\nu}_m)-\eta(\widehat{\nu}_m)\right|>\epsilon\right]\leq\Pr\left[\sup_{\nu\in V}\left|\widehat{\eta}(\nu)-\eta(\nu)\right|>\epsilon\right]\to 0,$$

where convergence to zero follows because of the uniformity in ν of the expansion (A.4), which itself follows by the smoothness of the mapping $\nu \mapsto F(U; f_\nu)$. Furthermore, $\eta(\widehat{\nu}_m) - \eta(\nu_o) = o_p(1)$ by the smoothness of the mapping $\nu \mapsto \eta(\nu)$. Therefore, $\widehat{\phi}_m = (\widehat{\nu}_m, \widehat{\eta}(\widehat{\nu}_m))$ is consistent. Then, by the Mean Value Theorem we have

$$0 = M_n(\nu_o) + \frac{\partial}{\partial \nu} M_n(\nu_m^*)(\widehat{\nu}_m - \nu_o),$$

where ν_m^* lies between $\widehat{\nu}_m$ and ν_o. It follows that

$$\sqrt{n}(\widehat{\nu}_m - \nu_o) = [M'(\nu_m^*)]^{-1}\sqrt{n}M_n(\nu_o) = [M'(\nu_o)]^{-1}\sqrt{n}M_n(\nu_o) + o_p(1),$$

where $\sqrt{n}M_n(\nu_o)$ is asymptotically normal applying the same CLT used in Theorem 1. Specifically,

$$\begin{aligned}
M_n(\nu_o) &= (1-\beta_o^2)(\widehat{\gamma}_z(0) - \gamma_z(0)) - 2\beta_o\left(\widehat{\beta} - \beta_o\right)(\gamma_z(0) - C_2(\nu_o)) \\
&\quad - 2\theta(\nu_o)\left(\widehat{\theta}(\nu_o) - \theta(\nu_o)\right) - 2\alpha(\nu_o)(\widehat{\alpha}(\nu_o) - \alpha(\nu_o))C_3(\nu_o) + o_p\left(n^{-1/2}\right) \\
&= \left(1-\beta_o^2\right)\frac{1}{n}\sum_{t=1}^n a_{2t} - 2\beta_o(\gamma_z(0) - C_2(\nu_o)), \pi_0(w)^\intercal \frac{1}{n}\sum_{t=1}^n b_t \\
&\quad - 2\theta(\nu_o)\frac{i^\intercal \pi_2(w', \beta_o)}{C_4(\nu)}\pi_0(w)^\intercal \frac{1}{n}\sum_{t=1}^n b_t \\
&\quad - 2\theta(\nu_o)\frac{1}{C_4(\nu)}\pi_1(w', \beta)^\intercal \frac{1}{n}\sum_{t=1}^n c_t + 2\alpha(\nu_o)C_3(\nu_o)\frac{\beta}{C_5(\nu)}\frac{1}{n}\sum_{t=1}^n a_{2t} \\
&\quad - 2\alpha(\nu_o)C_3(\nu_o)\frac{1}{C_5(\nu)}\pi_1(w, \beta_o)^\intercal \frac{1}{n}\sum_{t=1}^n b_t \\
&\quad - \left(\frac{\pi_2(w'', \beta_o)^\intercal \gamma_z}{C_5(\nu)} + \frac{\gamma_z(0) - C_2(\nu)}{C_5(\nu)}\right)\pi_0(w)^\intercal \frac{1}{n}\sum_{t=1}^n b_t + o_p\left(n^{-1/2}\right) \\
&= \mathcal{B}(\nu_o)^\intercal \frac{1}{n}\sum_{t=1}^n \chi_t + o_p\left(n^{-1/2}\right).
\end{aligned} \tag{A.6}$$

Note that

$$\begin{aligned}
M'(\nu) &= -2\theta(\nu)\theta'(\nu) - 2\alpha(\nu)\alpha'(\nu)C_3(\nu) - \alpha(\nu)^2 C_3'(\nu) - \left(1-\beta_o^2\right)C_2'(\nu) \\
&= 2\frac{C_4'(\nu)}{C_4(\nu)}\theta(\nu)^2 - 2\beta\alpha(\nu)\frac{C_2'(\nu)C_3(\nu)}{C_5(\nu)} + \alpha(\nu)^2\left[\frac{2C_5'(\nu)C_3(\nu) - C_5(\nu)C_3'(\nu)}{C_5(\nu)}\right] \\
&\quad - \left(1-\beta_o^2\right)C_2'(\nu) \\
&= 2\frac{C_4'(\nu_o)}{C_4(\nu_o)}\theta_o^2 - 2\beta_o\alpha_o\frac{C_2'(\nu_o)C_3(\nu_o)}{C_5(\nu_o)} + \alpha_o^2\left[\frac{2C_5'(\nu_o)C_3(\nu_o) - C_5(\nu_o)C_3'(\nu_o)}{C_5(\nu_o)}\right] \\
&\quad - \left(1-\beta_o^2\right)C_2'(\nu_o) \;\; [\text{at } \nu = \nu_o],
\end{aligned}$$

where:

$$\theta'(\nu) = \frac{-C_4'(\nu)}{C_4^2(\nu)} \sum_{k=1}^{p} \frac{\gamma_{zu}(k)}{\beta^{k-1}} w_k' = \frac{-C_4'(\nu)}{C_4(\nu)} \theta(\nu)$$

$$\alpha'(\nu) = \frac{-C_5'(\nu)}{C_5^2(\nu)} \left\{ \sum_{k=1}^{p} \frac{\gamma_z(k)}{\beta^{k-1}} w_k'' - \beta\left(\gamma_z(0) - C_2(\nu)\right) \right\} + \frac{\beta C_2'(\nu)}{C_5(\nu)}$$
$$= \frac{\beta C_2'(\nu) - C_5'(\nu)\alpha(\nu)}{C_5(\nu)}.$$

Then we have

$$\begin{aligned}
\sqrt{n}\left(\widehat{\eta}(\widehat{\nu}_m) - \eta(\nu_o)\right) &= \sqrt{n}\left(\widehat{\eta}(\nu_o) - \eta(\nu_o)\right) + \frac{\partial \widehat{\eta}}{\partial \nu}(\nu_m^*)\sqrt{n}\left(\widehat{\nu}_m - \nu_o\right) \\
&= \sqrt{n}\left(\widehat{\eta}(\nu_o) - \eta(\nu_o)\right) + \frac{\partial \eta}{\partial \nu}(\nu_o)\sqrt{n}\left(\widehat{\nu}_m - \nu_o\right) + o_p(1) \\
&= \sqrt{n}\left(\widehat{\eta}(\nu_o) - \eta(\nu_o)\right) + \frac{\partial \eta}{\partial \nu}(\nu_o)\left[\frac{\partial}{\partial \nu} M(\nu_o)\right]^{-1} \sqrt{n} M_n(\nu_o) + o_p(1) \\
&= \left[\mathcal{A}(\nu) + \frac{\partial \eta}{\partial \nu}(\nu_o)\left[\frac{\partial}{\partial \nu} M(\nu_o)\right]^{-1} \mathcal{B}(\nu_o)^{\intercal}\right] \frac{1}{n}\sum_{t=1}^{n} \chi_t + o_p(1),
\end{aligned}$$

and so the result follows from the joint asymptotic normality of $\left[\sqrt{n}\left(\widehat{\eta}(\nu_o) - \eta(\nu_o)\right), \sqrt{n} M_n(\nu_o)\right]$, which itself follows from the expansions (A.2) and (A.6). Note that $\eta'(\nu) = (\omega'(\nu), 0, \theta'(\nu), \alpha'(\nu))^\top$, where $\theta'(\nu), \alpha'(\nu)$ are given above and $\omega'(\nu) = -C_1'(\nu)(1-\beta)$. ■

Proof of (16). We want to calculate

$$T = E \log \xi^2,$$

where ξ is a unit variance symmetric GED with density function

$$f_\nu(\xi) = \frac{\nu \exp\left\{-(1/2)|\xi/\lambda|^\nu\right\}}{\lambda 2^{1+1/\nu}\Gamma(1/\nu)},$$

where $\lambda = \left\{2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)\right\}^{1/2}$. We have

$$T = \frac{4\nu}{\lambda 2^{1+1/\nu}\Gamma(1/\nu)} \int_0^\infty \log(x)\exp\left(-a x^\nu\right) dx$$

with $a = (1/2)\lambda^{-\nu}$. We change variable from $x \mapsto y = x^\nu$, which yields $dy = \nu x^{\nu-1} dx = \nu y^{(\nu-1)/\nu} dx$ and $\log(x) = \log\left(y^{1/\nu}\right) = (1/\nu)\log y$. Therefore,

$$T = \frac{4}{\lambda(\nu)\nu 2^{1+1/\nu}\Gamma(1/\nu)} \int_0^\infty y^{-(\nu-1)/\nu} \log(y)\exp(-ay)\, dy.$$

Equation 4.352:1 of Gradshteyn and Ryzhik (2007) says that for $\alpha > -1, \beta > 0$,

$$\int_0^\infty x^\alpha \log(x)\exp(-\beta x)\, dx = \frac{1}{\beta^{\alpha+1}}\Gamma(\alpha+1)\left[\psi(\alpha+1) - \ln\beta\right],$$

where $\psi(t) = d\ln\Gamma(t)/dt$. Therefore,

$$T = \frac{2}{\nu}\left[\psi(1/\nu) + \ln 2 + \nu \ln(\lambda)\right] = \frac{2}{\nu}\psi(1/\nu) + \ln\Gamma(1/\nu) - \ln\Gamma(3/\nu)$$

$$\ln\lambda = \frac{1}{2}\left[\frac{-2}{\nu}\ln 2 + \ln\Gamma(1/\nu) - \ln\Gamma(3/\nu)\right].$$

■