Statistical Papers (2021) 62:117–135
https://doi.org/10.1007/s00362-019-01086-y

REGULAR ARTICLE

Asymptotic normality of the MLE in the level-effect ARCH model

Christian M. Dahl[1] **· Emma M. Iglesias**[2]

Received: 10 October 2017 / Revised: 19 December 2018 / Published online: 19 January 2019

Abstract

We establish consistency and asymptotic normality of the maximum likelihood estimator in the level-effect ARCH model of Chan et al. (J Financ 47(3):1209–1227, 1992). Furthermore, it is shown by simulations that the asymptotic properties also apply in finite samples.

Keywords Level-ARCH · Asymptotic normality · Asymptotic theory · Consistency · Stationarity · Maximum likelihood estimation

Mathematics Subject Classification 62F12 · 62M10 · 62P20

JEL Classification C12 · C13 · C22

1 Introduction

After Engle (1982) initiated the literature on autoregressive conditional heteroskedasticity (ARCH) and the model proved itself to be very useful in empirical applications,

We wish to thank the Co-Editor and two referees for very helpful comments. E. M. Iglesias is very grateful for the financial support from the Spanish Ministry of Science and Innovation, Project ECO2015-63845-P. Webpage: www.gcd.udc.es/emma.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00362-019-01086-y) contains supplementary material, which is available to authorized users.

✉ Emma M. Iglesias
emma.iglesias@udc.es

Christian M. Dahl
cmd@sam.sdu.dk

1 Department of Business and Economics, COHERE, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark

2 Department of Economics, Facultade de Economía y Empresa, Universidade da Coruña, Campus de A Coruña, 15071 A Coruña, Spain

an immense amount of research has been directed towards extending Engle's original ideas empirically as well as theoretically (see for example Dias-Curto et al. (2009) as an example of the usefulness of GARCH models in practice). Chan, Karolyi, Longstaff and Sanders (1992, CKLS from now onwards) proposed to introduce the lagged level of the spot interest rate in the conditional variance equation, generating the so called level-effect ARCH model. This model has subsequently been successfully used and extended by Brenner et al. (1996), Andersen and Lund (1997), Ball and Torous (1999) among others. Recently, Maheu and Yang (2016) have estimated with Bayesian methods and using financial data both CKLS model and also the CKLS model with ARCH disturbances; while Bu et al. (2017) have estimated the CKLS in two different regimes also using financial data. However, despite of its empirical success, the asymptotic behavior of the quasi-maximum likelihood (QML) estimator associated with the level-effect ARCH model has, to the best of our knowledge, not been formally established yet. Most papers on conditional heteroskedastic time series, see, e.g., Berkes and Horváth (2004), Straumann and Mikosch (2006), Hamadeh and Zakoïan (2011) and Francq et al. (2018), do not allow for the introduction of the level of a series, such as the interest rate, in the conditional variance equation. The double autoregressive model of Ling (2004) is an exception. Triffi (2006) illustrates the convergence results for the Constant Elasticity of Variance (CEV)-ARCH model of Fornari and Mele (2006), but from the best of our knowledge, the asymptotic normality and consistency of the QML/ML estimator for the level-ARCH model is still unknown. In the following sections we will present the model and provide a simple proof of asymptotic normality and consistency of the ML estimator within the traditional level-effect ARCH(1) setting. The simulation section confirms our theoretical results and finally, we conclude.

2 The level-effect ARCH model

Consider the discrete-time approximation of the CKLS model

$$y_t^* = \Delta y_t - (a + by_{t-1}) = \sigma_t \left| y_{t-1} \right|^{\gamma} z_t, \tag{1}$$

$$\sigma_t^2 = w + \alpha \left(\frac{y_{t-1}^*}{\left| y_{t-2} \right|^{\gamma}} \right)^2, \tag{2}$$

for $t = 1, \ldots, T$ that we denote the level-effect ARCH.[1] Let us denote the parameter vector of interest by $\theta = (\gamma, w, \alpha)'$ and let the true parameter values be given by $\theta_0 = (\gamma_0, w_0, \alpha_0)'$. Further, in many applications y_t^* is chosen as a transformation of current and lagged values of y_t, such as $y_t^* = y_t^* (y_t, y_{t-1}, y_{t-2}, \ldots; \delta)$ where δ is a vector of parameters. One such specification could be $y_t^* = \Delta y_t - (a + by_{t-1})$, where $\delta = (a, b)'$, such as the one considered in Andersen and Lund (1997, p. 354)

[1] Alternative representations for a level-effect ARCH model may be considered such as

$$y_t = \sigma_t z_t$$
$$\sigma_t^2 = w + \alpha y_{t-1}^2 + \gamma y_{t-1}$$

although they are outside the scope of this paper.

and in Broze et al (1995, Eq. 2). However, also note that model (1), (2) is not exactly that of Andersen and Lund (1997), as they use $\Delta y_t = (a + by_{t-1}) + \sigma_t y_{t-1}^{\gamma} z_t$ instead of $\Delta y_t = (a + by_{t-1}) + \sigma_t |y_{t-1}|^{\gamma} z_t$. The absolute value is needed, as otherwise the discrete-time model does not ensure that y_t is non-negative. In practice, δ can be pre-estimated in a first stage. This pre-estimation approach is very common in empirical research, particularly, when modelling spot interest rates, see, for example, Ball and Torous (1999, p. 2349). In order to avoid additional complexity of the proofs, we assume throughout this paper that δ is known, see also Remark 2. [2] It should be noted that (1)–(2) is a generalization of Frydman (1994), who consider a discrete-time process, but where $\alpha_0 = 0$. Broze, Scaillet and Zakoïan (1995, Eq. 2 in p. 202) have also analyzed the regular level effect model but without the ARCH component, but where they introduce also $|y_{t-1}|^{\gamma} z_t$ in (1). Moreover, for the case where γ_0 is known, (1), (2) is the standard linear model for which a complete characterization of the estimation theory has been developed by Jensen and Rahbek (2004a, b) and in Kristensen and Rahbek (2005, 2008). The "quasi"-log likelihood function (conditional on past values of y_t) associated with (1), (2) is given as

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta) = -\frac{1}{2}\sum_{t=1}^{T} \ln\left[y_{t-1}^{2\gamma}\left(w + \alpha\left(\frac{y_{t-1}^*}{|y_{t-2}|^{\gamma}}\right)^2\right)\right] - \frac{1}{2}\sum_{t=1}^{T} \frac{\left(y_t^*\right)^2}{y_{t-1}^{2\gamma}\left(w + \alpha\left(\frac{y_{t-1}^*}{|y_{t-2}|^{\gamma}}\right)^2\right)}. \tag{3}$$

We proceed under the following set of maintained assumptions:

Assumption A **A1** $z_t \sim N.i.i.d.\,(0, 1)$,
A2 $\infty > w_0 > 0$, $\infty > \alpha_0 > 0$,
A3 $E\left[\left(\frac{y_{t-1}^*}{|y_{t-2}|^{i}}\right)^{\varphi} |\ln(|y_{t-1}|)|^3\right] < \infty$, $E\left[\left(\frac{y_{t-1}^*}{|y_{t-2}|^{i}}\right)^{\varphi} (\ln(|y_{t-1}|))^2 |\ln(|y_{t-2}|)|\right] < \infty$, $E\left[\left(\frac{y_{t-1}^*}{|y_{t-2}|^{i}}\right)^{\varphi} (\ln(|y_{t-2}|))^2 |\ln(|y_{t-1}|)|\right] < \infty$, for both $\varphi = 0$ and for some $\varphi > 0$ for $i = 0, 1$.

Assumption B **B1** $E\ln\left(\alpha_0 z_t^2\right) < 0$,
B2 $1 \geq \gamma_0 \geq 0$, $|b+1| < 1$, $E\ln\left|b + 1 + \alpha_0 z_t^2\right| < 0$.

Assumptions A1, A2 and B1 are very common in the traditional ARCH literature (see e.g. Jensen and Rahbek (2004a, b), p. 1205). In A1 we need to impose Gaussianity as we make use in our proofs of the results in Broze et al. (1995). Also, A1 implies that $E\left(\left(1 - z_t^2\right)^2\right) = \zeta = 2$. Note also, A3 -an assumption very specific for the level-effect ARCH model- explains the necessity to introduce $|y_{t-1}|^{\gamma}$ in (1) as in Broze et al

[2] If δ is estimated, this will affect the asymptotic properties of the estimators of the volatility parameters, but we leave this for further future research.

(1995, Eq. 2).[3] In B2, we require $1 \geq \gamma_0 \geq 0$-see Broze et al (1995, Proposition 3) for more details-. The condition for stationarity of σ_t^2, $\left(y_t^*/y_{t-1}^{\gamma}\right)$, y_t^* and y_t is given by the following two Lemmas:

Lemma 1 *Let Assumption A hold. A necessary and sufficient condition for strict stationarity of σ_t^2 and $\left(y_t^*/\left|y_{t-1}\right|^{\gamma_0}\right)$ as generated by (1), (2) is given by*

$$E \ln\left(\alpha_0 z_t^2\right) < 0.$$

Proof of Lemma 1 Lemma 1 is not a new result in the literature since if $\left(y_t^*/\left|y_{t-1}\right|^{\gamma_0}\right)$ is known, the level-effect ARCH model reduces to the ARCH(1) and the condition for strict stationarity is well known (see e.g. Jensen and Rahbek 2004a). □

Lemma 2 *Let Assumption A and B1 hold. A necessary and sufficient condition for ergodicity and second order stationarity of y_t^* and y_t as generated by (1), (2) is given by*

$$1 > \gamma_0 \geq 0, \left|b+1\right| < 1, E \ln\left|b+1+\alpha_0 z_t^2\right| < 0.$$

When $\gamma_0 = 1$, the previous condition is a sufficient condition for second order stationarity and ergodicity.

Proof of Lemma 2 Given in the proof of Proposition 3 of Broze et al (1995, Appendix D) but replacing $\sigma_{0,h}$ in Broze et al. (1995) by the required strict stationarity condition of σ_t^2 that was proved in Lemma 1, and where the Gaussianity of z_t is used in the definition of the transition density.

Next, the main result of the paper regarding the limiting distribution of the ML estimator in the level-effects ARCH model can be established. □

Theorem 1 *Define $u_{1t}\left(\theta_0\right) = \left(\ln\left|y_{t-1}\right| - w_0\left(\frac{1}{w_0} - \frac{1}{\sigma_t^2\left(\theta_0\right)}\right)\ln\left|y_{t-2}\right|\right)$, $u_{2t}\left(\theta_0\right) = \left(\frac{1}{\sigma_t^2\left(\theta_0\right)}\right)$ and $u_{3t}\left(\theta_0\right) = \left(\frac{w_0}{\alpha_0}\right)\left(\frac{1}{w_0} - \frac{1}{\sigma_t^2\left(\theta_0\right)}\right)$, let Assumptions A and B hold, θ_0 is assumed to be an interior point of the parameter space and assume that $\delta = (a, b)'$ is known. Consider the log likelihood function given by (3). Then, there exists a fixed open neighborhood $U = U\left(\theta_0\right)$ of θ_0 such that with probability tending to one as $T \longrightarrow \infty$, $L_T\left(\theta\right)$ has a unique maximum point $\widehat{\theta}$ in U. In addition, the ML estimator $\widehat{\theta}$ is consistent and asymptotically normal*

$$\sqrt{T}\left[\widehat{\theta} - \theta_0\right]' \xrightarrow{d} N\left(0, \left(\zeta^2/4\right)\Lambda^{-1}\right),$$

[3] See for example Brown et al. (1996) and Bouezmarni and Rombouts (2010) as examples showing the relevance of studying positive time series. See also Broze et al (1995, p. 202) for a discussion of models that preclude negative values under some parameter restrictions for continuous-time processes.

where

$$\Lambda = \zeta \begin{bmatrix} \overline{m}_{11} & \frac{1}{2}\overline{m}_{12} & \frac{1}{2}\overline{m}_{13} \\ \frac{1}{2}\overline{m}_{12} & \frac{1}{4}\overline{m}_{22} & \frac{1}{4}\overline{m}_{23} \\ \frac{1}{2}\overline{m}_{13} & \frac{1}{4}\overline{m}_{23} & \frac{1}{4}\overline{m}_{33} \end{bmatrix} > 0,$$

$\zeta = 2$ *and* $\overline{m}_{ij} = E\left(u_{it}(\theta_0)\, u_{jt}(\theta_0)\right)$ *for* $i = 1, 2, 3$ *and* $j = 1, 2, 3$.[4]

Proof of Theorem 1 The proof of Theorem 1 is given in the Appendix.

Importantly, Theorem 1 applies to the MLE of the stationary level-effect ARCH(1) process. However, if γ_0 is known under Assumption A but $E \ln\left(\alpha_0 z_t^2\right) > 0$ then the asymptotics for the ML estimator of α can still be established. To see this, simply notice that the model (1), (2) in this case can be rewritten as

$$\widetilde{y}_t = \sigma_t z_t$$
$$\sigma_t^2 = \omega + \alpha \widetilde{y}_{t-1}^2$$

for $\widetilde{y}_t \equiv \frac{y_t^*}{|y_{t-1}|^{\gamma_0}}$. This representation of the process $\widetilde{y}_t$ is exactly identical to the model given by equation (1) in Jensen and Rahbek (2004a). Consequently, when $E \ln\left(\alpha_0 z_{t-1}^2\right) > 0$ (Assumption B fails) then $\widetilde{y}_t^2 \xrightarrow{a.s.} \infty$ from Lemma 1 as $\sigma_t^2 \xrightarrow{a.s.} \infty$ (see also Klüppelberg et al. (2004)) and the asymptotic results follows directly from Jensen and Rahbek (2004a, Lemmas 1–5).[5] The case of $E \ln\left(\alpha_0 z_{t-1}^2\right) = 0$ implies, as shown in Klüppelberg et al. (2004) for an ARCH(1) (under suitable conditions), that $\sigma_t^2 \xrightarrow{p} \infty$ and therefore different arguments are required in this case (see Pedersen and Rahbek (2016)). Three remarks should be added: □

Remark 1 It is well known, that the stationary level-effect ARCH model, can be estimated by non-parametric techniques, since the variance function is smooth and only depends on y_{t-1}. In the nonstationary case, Han and Zhang (2009) consider ARCH models by applying the results of Wang and Phillips (2009a, b), although they do not allow for a level-effect.

Remark 2 In model (1), (2), we are assuming that $\delta = (a, b)'$ is known. In practice, δ can be pre-estimated in a first stage. If δ and θ are estimated jointly in mean and variance equation (contrary to Ball and Torous (1999)), then our proof will need to be extended to account for the joint estimation. At this stage we do not know if this would require stronger assumptions than those in Assumptions A and B.

---

4 Note that the data generating process (DGP) is assumed to be ergodic. It might be possible to relax this assumption about ergodicity and simply assume that the DGP is initiated in some fixed value and that the DGP has an ergodic solution (see e.g. Kristensen and Rahbek (2005) and Jensen and Rahbek (2007)) and we leave that for further research.

5 We let $\xrightarrow{a.s.}$ denote convergence "almost surely" as $T \to \infty$. Also note that in this case $\frac{\partial}{\partial\alpha} L_T(\theta) = -\sum_{t=1}^T \frac{1}{2}\left(1 - \frac{\widetilde{y}_t^2}{\sigma_t^2}\right)\frac{\widetilde{y}_{t-1}^2}{\sigma_t^2}$; $\frac{\partial^2}{\partial\alpha^2} L_T(\theta_0) = \frac{1}{2}\sum_{t=1}^T \left(1 - 2\frac{\widetilde{y}_t^2}{\sigma_t^2}\right)\frac{\widetilde{y}_{t-1}^4}{\sigma_t^4}$; $\frac{\partial^3}{\partial\alpha^3} L_T(\theta_0) = -\sum_{t=1}^T \left(1 - 3\frac{\widetilde{y}_t^2}{\sigma_t^2}\right)\frac{\widetilde{y}_{t-1}^6}{\sigma_t^6}$ as shown in Results 1, 2 and 3 in the Technical Appendix and they do correspond to Eqs. (4), (5) and (6) of Jensen and Rahbek (2004a).

Remark 3 The generalization of the asymptotic results when going from ARCH(1) to GARCH(1,1) can most likely be provided in a similar fashion as the extension from Jensen and Rahbek (2004a) to Jensen and Rahbek (2004b), with the added complexity in the proofs.

3 Simulations

In this section we evaluate and discuss, based on simulations, how well the asymptotic results of $\widehat{\theta}$ given by Theorem 1 can approximate the finite sample properties. We also report simulation results to check the consequences when some of the assumptions are violated. We set $\delta = (a, b)' = (0, -1)'$ in all simulations and we do not estimate it.

In Panel A of Table 1, results are reported on point estimates, their associated biases and the root mean squared errors (RMSE's) when Assumption B1 holds, i.e., the volatility process is stationary. In Panel A, three alternative data generating processes are considered: The first is characterized by having standard Gaussian distributed innovations whereas the two remaining processes have fatter tails as the innovations are standard t-distributed with 5 degrees of freedom. The results show that under all three data generating processes, the MLE is relatively accurate with small biases and small RMSE's even at small sample sizes, i.e., $T = 1000$. As expected from Theorem 1, the biases and the RMSE's are decreasing for all the estimators as sample sizes increase. Note that in Theorem 1, we show the asymptotic theory for the MLE, however in this simulation section we also show what happens under alternative distribution functions for the innovations.

In Panel B of Table 1, two data generating processes, both in violation with Assumption B1, are considered. As noted in Jensen and Rahbek (2004a, b) ω is not identified in this case hence is fixed at its true value in the population. The results of Panel B, Table 1, clearly illustrates that the consistency properties of the MLE of α still holds when σ_t^2 and $\left(y_t^*/y_{t-1}^{\gamma_0}\right)$ are nonstationary. The main theoretical results of this paper are silent about the asymptotic properties of γ when B1 is violated. This is due to the complexity of the expressions of the first, second and third order derivatives of the loglikelihood with respect to γ, and we can only analyze the case when B1 holds. However the simulations seem to indicate strongly that consistency of the MLE of γ is maintained also when Assumption B1 does not hold.

In Table 2, consistency of the estimated standard errors of the MLE in finite samples is illustrated. The term SE $(\widehat{\gamma})$ denotes the "feasible" standard error of $\widehat{\gamma}$ computed according to the expression for the asymptotic variance-covariance function derived in Theorem 1, but where all the population parameters are replaced by sample estimates (sample analogy estimation). ASE $(\widehat{\gamma})$ denotes the asymptotic standard error of $\widehat{\gamma}$. It is defined similar to SE $(\widehat{\gamma})$ but computed based on the true populations parameters.

We see that for all the models presented in Table 2, there is a very close correspondence between SE $(\widehat{\gamma})$ and ASE $(\widehat{\gamma})$. These results are very encouraging, implying that the asymptotic variance covariance matrix provides a good approximation of the parameter estimation uncertainty also in finite sample. It is also noticeable, that in the cases of relative fat tailed t-distributed innovations (when using the misspecified

Table 1 Simulation results on the QMLE, i.e., $\widehat{\theta} = (\widehat{\gamma}, \widehat{\omega}, \widehat{\alpha})$

	$\widehat{\gamma}$	$\widehat{\omega}$	$\widehat{\alpha}$	bias ($\widehat{\gamma}$)	bias ($\widehat{\omega}$)	bias ($\widehat{\alpha}$)	RMSE ($\widehat{\gamma}$)	RMSE ($\widehat{\omega}$)	RMSE ($\widehat{\alpha}$)
Panel A: Assumption B1 holds									
$(\gamma, \omega, \alpha) = (0.6, 0.05, 2)$, $z \sim N(0, 1)$									
T = 1000	0.599	0.0563	1.98	−0.000869	0.00631	−0.02004	0.0265	0.01632	0.1212
T = 2000	0.599	0.0538	1.99	−0.000629	0.00376	−0.00977	0.0211	0.01322	0.0929
T = 4000	0.598	0.0536	1.99	−0.001680	0.00365	−0.01471	0.0160	0.00987	0.0670
$(\gamma, \omega, \alpha) = (0.6, 0.05, 2)$ $z \sim t(df = 5)/\sqrt{5/3}$									
T = 1000	0.601	0.0543	1.96	0.00126	0.00434	−0.0429	0.0759	0.0782	0.333
T = 2000	0.593	0.0554	1.95	−0.00673	0.00543	−0.0484	0.0532	0.0432	0.216
T = 4000	0.594	0.0521	1.98	−0.00597	0.00207	−0.0210	0.0364	0.0251	0.157
$(\gamma, \omega, \alpha) = (0.6, 0.05, 3.7)$, $z \sim t(df = 5)/\sqrt{5/3}$									
T = 1000	0.612	0.0569	3.61	0.01156	0.00692	−0.09021	0.0364	0.0536	0.348
T = 2000	0.604	0.0573	3.65	0.00419	0.00726	−0.04952	0.0265	0.0197	0.245
T = 4000	0.598	0.0511	3.69	−0.00212	0.00112	−0.00779	0.0204	0.0148	0.174
Panel B: Assumption B1 fails									
$(\gamma, \omega, \alpha) = (0.6, 0.05, 3.7)$, $z \sim N(0, 1)$									
T = 1000	0.598	0.050	3.69	−0.001895	·	−0.0136	0.01610	·	0.166
T = 2000	0.599	0.050	3.68	−0.000740	·	−0.0165	0.01192	·	0.119
T = 4000	0.599	0.050	3.67	−0.000735	·	−0.0314	0.01003	·	0.101
$(\gamma, \omega, \alpha) = (0.6, 0.05, 5.0)$, $z \sim t(df = 5)/\sqrt{5/3}$									
T = 1000	0.602	0.050	4.93	0.00237	·	−0.0704	0.0234	·	0.390
T = 2000	0.601	0.050	4.94	0.00148	·	−0.0575	0.0201	·	0.357
T = 4000	0.602	0.050	4.94	0.00163	·	−0.0555	0.0188	·	0.313

The reported results are based on $M = 1000$ replications. The terms bias ($\widehat{\gamma}$) and RMSE ($\widehat{\gamma}$) are computed as $\frac{1}{M}\sum_{m=1}^{M}(\widehat{\gamma}_m - \gamma_0)$ and $\left(\frac{1}{M}\sum_{m=1}^{M}(\widehat{\gamma}_m - \gamma_0)^2\right)^{1/2}$ respectively. The terms bias($\widehat{\omega}$), bias($\widehat{\alpha}$) and RMSE ($\widehat{\omega}$), RMSE ($\widehat{\alpha}$) are defined similarly

Table 2 Simulation results on the standard errors of the QMLE

	SE ($\widehat{\gamma}$)	SE ($\widehat{\omega}$)	SE ($\widehat{\alpha}$)	ASE ($\widehat{\gamma}$)	ASE ($\widehat{\omega}$)	ASE ($\widehat{\alpha}$)
Panel A: Assumption B1 holds						
$(\gamma, \omega, \alpha) = (0.6, 0.05, 2.0)$, $z \sim N(0, 1)$						
T = 1000	0.0323	0.01867	0.1320	0.0343	0.01762	0.1366
T = 2000	0.0231	0.01267	0.0935	0.0242	0.01241	0.0964
T = 4000	0.0165	0.00904	0.0666	0.0170	0.00874	0.0677
$(\gamma, \omega, \alpha) = (0.6, 0.05, 2.0)$, $z \sim t(df = 5)/\sqrt{5/3}$						
T = 1000	0.0399	0.02275	0.1737	0.0446	0.02295	0.1838
T = 2000	0.0287	0.01570	0.1232	0.0317	0.01631	0.1306
T = 4000	0.0211	0.01161	0.0896	0.0224	0.01155	0.0924
$(\gamma, \omega, \alpha) = (0.6, 0.05, 3.7)$, $z \sim t(df = 5)/\sqrt{5/3}$						
T = 1000	0.0218	0.01811	0.1966	0.0222	0.01624	0.202
T = 2000	0.0151	0.01243	0.1382	0.0157	0.01147	0.142
T = 4000	0.0110	0.00809	0.0988	0.0111	0.00808	0.100
Panel B: Assumption B1 fails						
$(\gamma, \omega, \alpha) = (0.6, 0.05, 3.7)$, $z \sim N(0, 1)$						
T = 1000	0.01632	·	0.1657	0.01618	·	0.1659
T = 2000	0.01154	·	0.1176	0.01152	·	0.1173
T = 4000	0.00848	·	0.0852	0.00841	·	0.0836
$(\gamma, \omega, \alpha) = (0.6, 0.05, 5.0)$, $z \sim t(df = 5)/\sqrt{5/3}$						
T = 1000	0.0166	·	0.234	0.0164	·	0.237
T = 2000	0.0119	·	0.170	0.0116	·	0.168
T = 4000	0.0108	·	0.146	0.0110	·	0.167

The reported results are based on $M = 1000$ replications. The term SE ($\widehat{\gamma}$) denotes the "feasible" standard error of $\widehat{\gamma}$ computed according to the expression for the asymptotic variance-covariance function derived in Theorem 1, but where all the population parameters are replaced by sample estimates (sample analogy estimation). ASE ($\widehat{\gamma}$) denotes the asymptotic standard error of $\widehat{\gamma}$. It is defined similar to SE ($\widehat{\gamma}$) but computed based on the true populations parameters. SE ($\widehat{\omega}$), SE ($\widehat{\alpha}$), ASE ($\widehat{\omega}$) and ASE ($\widehat{\alpha}$) are defined similarly

Gaussian quasi-likelihood) the standard errors of the MLE increase as expected, relative to the cases where innovations are normally distributed. But the observed increase in parameter estimation uncertainty seems to be only of limited magnitude. Similar results hold for SE ($\widehat{\omega}$), SE ($\widehat{\alpha}$), ASE ($\widehat{\omega}$) and ASE ($\widehat{\alpha}$).[6]

4 Conclusion

In this paper we establish consistency and asymptotic normality of the ML estimator in the level-effect ARCH model. We also show in simulations that the asymptotic theory provides a good approximation in finite samples.

[6] In a Supplementary Appendix that is available upon request from any of the authors, we provide additional simulation results, where $\delta = (a, b)'$ is estimated jointly with the variance parameters. According to those simulation results, when δ is also estimated and assumptions A and B hold, we conjecture that the ML estimator seem to follow also an asymptotically normal distribution.

Appendix

The analytical expressions for the first, second and third order derivatives of the quasi log likelihood function are given in a Supplementary Appendix available upon request from any of the authors. We provide now three important propositions that we need in order to prove Theorem 1. The proof technique for the MLE utilizes the classic Cramér type conditions for consistency and asymptotic normality (central limit theorem for the score, convergence of the Hessian and uniformly bounded third-order derivatives); see e.g. Lehmann (1999).

Proposition 1 *Let $u_{jt}(\theta_0)$ be defined as in Theorem 1. Under Assumptions A and B, the joint distribution of the score functions evaluated at $\theta = \theta_0$ are asymptotically Gaussian,*

$$\frac{1}{\sqrt{T}} \frac{\partial}{\partial \theta} L_T(\theta_0) \xrightarrow{d} N(0, \Lambda),$$

where

$$\Lambda = \zeta \begin{bmatrix} \overline{m}_{11} & \frac{1}{2}\overline{m}_{12} & \frac{1}{2}\overline{m}_{13} \\ \frac{1}{2}\overline{m}_{12} & \frac{1}{4}\overline{m}_{22} & \frac{1}{4}\overline{m}_{23} \\ \frac{1}{2}\overline{m}_{13} & \frac{1}{4}\overline{m}_{23} & \frac{1}{4}\overline{m}_{33} \end{bmatrix} > 0,$$

and $\overline{m}_{ij} = E\left(u_{it}(\theta_0)\, u_{jt}(\theta_0)\right)$ for $i = 1, 2, 3$ and $j = 1, 2, 3$.

Proof of Proposition 1 For the proof of Proposition 1, we need first the following 2 Lemmas. □

Lemma A *Let Assumptions A and B hold and define $u_{1t}(\theta_0) = \left(\ln|y_{t-1}| - w_0\left(\frac{1}{w_0} - \frac{1}{\sigma_t^2(\theta_0)}\right)\ln|y_{t-2}|\right)$,*
$u_{2t}(\theta_0) = \left(\frac{1}{\sigma_t^2(\theta_0)}\right)$ and $u_{3t}(\theta_0) = \left(\frac{w_0}{\alpha_0}\right)\left(\frac{1}{w_0} - \frac{1}{\sigma_t^2(\theta_0)}\right)$. Then $u_{it}(\theta_0)$ is a stationary and ergodic sequence. In addition $\frac{1}{T}\sum_{t=1}^{T} u_{it}(\theta_0) \xrightarrow{p} E(u_{it}(\theta_0)) \equiv \overline{u}_i$ and $\frac{1}{T}\sum_{t=1}^{T} u_{it}^2(\theta_0) \xrightarrow{p} E\left(u_{it}^2(\theta_0)\right) \equiv \overline{m}_{ii}$ for $i = 1, 2, 3$.

Proof of Lemma A Define $I_t = \{y_t, z_t, y_{t-1}, z_{t-1}, y_{t-2}, z_{t-2}, \ldots\}$. Note first that

$$|u_{1t}(\theta_0)| \leq |\ln|y_{t-1}|| + w_0 |\ln|y_{t-2}|| \left(\frac{1}{w_0} + \frac{1}{\sigma_t^2(\theta_0)}\right) \leq |\ln|y_{t-1}|| + 2|\ln|y_{t-2}||,$$

hence

$$E|u_{1t}(\theta_0)| \leq 3E(\ln|y_t|) < \infty,$$

where we have used assumptions A and B and where the last inequality follows from A3 where the first two moments of $\ln|y_t|$ are assumed to be bounded. Hence we can write

$$u_{1t}(\theta_0) \equiv g_1\left(y_{t-1}, y_{t-2}, \sigma_t^2(\theta_0)\right),$$

where g_1 is a I_t-measurable function and where all arguments y_{t-1}, y_{t-2} and $\sigma_t^2(\theta_0)$ are stationary and ergodic as a consequence of Lemmas 1 and 2. This implies that $u_{1t}(\theta_0)$ is stationary and ergodic by Theorem 3.35 in White (1984). Consequently $\frac{1}{T}\sum_{t=1}^{T} u_{1t}(\theta_0) \xrightarrow{p} E(u_{1t}(\theta_0))$ follows by the Ergodic Theorem. Similarly, it follows straightforwardly that $E|u_{2t}(\theta_0)| \le \left(\frac{1}{w_0}\right)$ and $E|u_{3t}(\theta_0)| \le \left(\frac{2}{\alpha_0}\right)$. We can write $u_{2t}(\theta_0) \equiv g_2\left(\sigma_t^2(\theta_0)\right)$ and $u_{3t}(\theta_0) \equiv g_3\left(\sigma_t^2(\theta_0)\right)$ and as above conclude that $(u_{2t}(\theta_0), u_{3t}(\theta_0))$ is stationary and ergodic, and hence $\frac{1}{T}\sum_{t=1}^{T} u_{it}(\theta_0) \xrightarrow{p} E(u_{it}(\theta_0))$ for $i = 2, 3$. Second, notice that

$$\begin{aligned}
\left|u_{1t}^2(\theta_0)\right| &= |\ln^2|y_{t-1}| - 2w_0\left(\frac{1}{w_0} - \frac{1}{\sigma_t^2(\theta_0)}\right)\ln|y_{t-2}|\ln|y_{t-1}| \\
&\quad + (\ln|y_{t-2}|)^2 w_0^2\left(\frac{1}{w_0} - \frac{1}{\sigma_t^2(\theta_0)}\right)^2| \\
&\le \ln^2|y_{t-1}| + \frac{4w_0^2}{\sigma_t^4(\theta_0)}\ln^2|y_{t-2}| + 4\frac{w_0}{\sigma_t^2(\theta_0)}|\ln|y_{t-1}|\ln|y_{t-2}|| \\
&\le \ln^2|y_{t-1}| + 4\ln^2|y_{t-2}| + 4|\ln|y_{t-1}|\ln|y_{t-2}||,
\end{aligned}$$

such that

$$E\left|u_{1t}^2(\theta_0)\right| \le 5E\left((\ln|y_t|)^2\right) + 4E|\ln|y_t|\ln|y_{t-1}|| < \infty.$$

On the right hand side of the first inequality we have used Lemmas 1 and 2 and the second inequality follows from A3 (existence of second order moments). In addition, $E\left|u_{2t}^2(\theta_0)\right| \le \left(\frac{1}{w_0^2}\right)$ and $E\left|u_{3t}^2(\theta_0)\right| \le \left(\frac{4}{\alpha_0^2}\right)$. We can therefore conclude, by Theorem 3.35 in White (1984), that since $u_{it}(\theta_0)$ is stationary and ergodic then so is $u_{it}^2(\theta_0)$ for $i = 1, 2, 3$. Furthermore as $E|u_{it}^2(\theta_0)|$ is bounded then $\frac{1}{T}\sum_{t=1}^{T} u_{1t}^2(\theta_0) \xrightarrow{p} E\left(u_{1t}^2(\theta_0)\right)$ for $i = 1, 2, 3$ follows from the ergodicity theorem. This completes the proof of Lemma A. □

Lemma B *Under Assumptions A and B, the marginal distributions of the score functions given by Eqs. (9)–(11) evaluated at $\theta = \theta_0$ are asymptotically Gaussian,*

$$\frac{1}{\sqrt{T}}\frac{\partial}{\partial\gamma}L_T(\theta_0) = \frac{-1}{\sqrt{T}}\sum_{t=1}^{T}\left(1 - z_t^2\right)u_{1t}(\theta_0) \xrightarrow{d} N(0, \zeta\overline{m}_{11}), \tag{4}$$

$$\frac{1}{\sqrt{T}}\frac{\partial}{\partial w}L_T(\theta_0) = \frac{-1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{2}\left(1 - z_t^2\right)u_{2t}(\theta_0) \xrightarrow{d} N(0, \zeta\overline{m}_{22}), \tag{5}$$

$$\frac{1}{\sqrt{T}}\frac{\partial}{\partial\alpha}L_T(\theta_0)=\frac{-1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{2}\left(1-z_t^2\right)u_{3t}(\theta_0)\xrightarrow{d}N\left(0,\zeta\overline{m}_{33}\right), \tag{6}$$

where $\overline{m}_{ii}$, $i=1,2,3$ *and* ζ *are defined by Lemma A and A3 respectively.*

Proof of Lemma B We will prove (4) in detail. The results in (5) and (6) hold by identical arguments. Define again $I_t=\{y_t,z_t,y_{t-1},z_{t-1},y_{t-2},z_{t-2},...\}$ and recall from Result 1 that

$$s_{1t}(\theta_0)=-\left(1-z_t^2\right)u_{1t}(\theta_0).$$

Consequently

$$\begin{aligned}E\left(s_{1t}|I_{t-1}\right)&=-E\left(\left(1-z_t^2\right)u_{1t}(\theta_0)\,|I_{t-1}\right)=-E\left(\left(1-z_t^2\right)\right)u_{1t}(\theta_0)\\&=0.\end{aligned} \tag{7}$$

Since $\{s_{1t},I_t\}$ is an adapted stochastic sequence the result in (7) implies that $\{s_{1t},I_t\}$ is a martingale difference sequence according to Definition 3.75 in White (1984). Further, notice that

$$V_{1T}^2(\theta_0)=\sum_{t=1}^{T}E\left(s_{1t}^2(\theta_0)\,|I_{t-1}\right)=\sum_{t=1}^{T}E\left(\left(1-z_t^2\right)^2\right)u_{1t}^2(\theta_0)=\zeta\sum_{t=1}^{T}u_{1t}^2(\theta_0).$$

Hence,

$$E(V_{1T}^2(\theta_0))=\zeta\sum_{t=1}^{T}E\left(u_{1t}^2(\theta_0)\right)=T\zeta\overline{m}_{11}.$$

Furthermore, according to Lemma A we have that

$$\frac{1}{T}\sum_{t=1}^{T}u_{1t}^2(\theta_0)\xrightarrow{p}\overline{m}_{11},$$

implying that

$$\frac{1}{T}V_{1T}^2(\theta_0)\xrightarrow{p}\zeta\overline{m}_{11}.$$

From this we see that

$$\left(V_{1T}^2(\theta_0)\right)\left(E(V_{1T}^2(\theta_0))\right)^{-1}\xrightarrow{p}1. \tag{8}$$

Importantly, the result given by equation (8) corresponds to Condition (1), p. 60 in Brown (1971).[7]

Finally, we need to prove that the Lindeberg type condition, which is Condition (2) in Brown (1971). In particular, we need to show that

$$\left(E(V_{1T}^2(\theta_0))\right)^{-1}\sum_{t=1}^{T}E\left(s_{1t}^2(\theta_0)\,1\left\{|s_{1t}(\theta_0)|>\epsilon\sqrt{E(V_{1T}^2(\theta_0))}\right\}\right)\xrightarrow{p}0,$$

for all $\epsilon > 0$. By inserting the expression for s_{1t}^2 and $E(V_{1T}^2(\theta_0))$ we get

$$\lim_{T\to\infty}\frac{1}{T\zeta\overline{m}_{11}}\sum_{t=1}^{T}E\left(s_{1t}^2(\theta_0)\,1\left\{|s_{1t}(\theta_0)|>\epsilon\sqrt{T\zeta\overline{m}_{11}}\right\}\right)=\lim_{T\to\infty}\frac{1}{\zeta\overline{m}_1}$$
$$E\left(\left(\left(1-z_t^2\right)^2u_{1t}^2(\theta_0)\right)1\left\{\left|\left(1-z_t^2\right)^2u_{1t}^2(\theta_0)\right|>\sqrt{T\zeta\overline{m}_{11}}\right\}\right)\to 0,$$

for all $\zeta\overline{m}_{11}$ because, from Lemma A and A1, $u_{1t}^2(\theta_0)$ and z_t^2 have finite moments and are stationary and ergodic. Consequently, the Lindeberg condition holds.

According to Theorem 2, p. 60, in Brown (1971) we can therefore conclude that

$$\frac{1}{\sqrt{T\zeta\overline{m}_{11}}}\sum_{t=1}^{T}s_{1t}(\theta_0)\xrightarrow{d}N(0,1),$$

which completes the proof. □

Along the same lines

$$\frac{1}{T}\sum_{t=1}^{T}E\left(s_{2t}^2\mid I_{t-1}\right)=\frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\frac{1}{\left(w_0+\alpha_0\left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma}\right)^2\right)}\xrightarrow{p}\frac{\zeta}{4w_0^2}>0,$$

$$\frac{1}{T}\sum_{t=1}^{T}E\left(s_{3t}^2\mid I_{t-1}\right)=\frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\frac{\left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma}\right)^2}{\left(w_0+\alpha_0\left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma}\right)^2\right)}\xrightarrow{p}\frac{\zeta}{4\alpha_0^2}>0.$$

and

$$\frac{1}{T}\sum_{t=1}^{T}E\left(s_{2t}^2 1\left\{|s_{2t}|>\sqrt{T}\delta\right\}\right)$$
$$\le E\left(\left(\frac{(1-z_t^2)^2}{4w_0^2}\right)1\left\{\left|\frac{(1-z_t^2)}{2w_0}\right|>\sqrt{T}\delta\right\}\right)\to 0,$$

[7] Note that since $s_{1t}(\theta_0)$ is a martingale difference sequence, we may use Billingsley (1961)'s Central Limit Theorem (CLT) instead of Brown (1971)'s CLT.

$$\frac{1}{T}\sum_{t=1}^{T} E\left(s_{3t}^2 1\left\{|s_{3t}| > \sqrt{T}\delta\right\}\right)$$
$$\leq E\left(\left(\frac{\left(1-z_t^2\right)^2}{4\alpha_0^2}\right) 1\left\{\left|\frac{\left(1-z_t^2\right)}{2\alpha_0}\right| > \sqrt{T}\delta\right\}\right) \to 0,$$

for some $\delta > 0$ and as T tends to ∞ . □

Proof of Proposition 1 In order to fully characterize the asymptotic distribution we need to determine the off-diagonal elements of the variance covariance matrix of the score vectors given by Λ. In particular, because $u_{1t}(\theta_0)$, $u_{2t}(\theta_0)$ and $u_{3t}(\theta_0)$ are all stationary and ergodic with finite first moments (from Lemma A) it follows straightforwardly that

$$\frac{1}{T}\sum_{t=1}^{T} s_{1t}(\theta_0)\, s_{2t}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\left(1-z_t^2\right)^2 u_{1t}(\theta_0)\, u_{2t}(\theta_0) \xrightarrow{p} \frac{1}{2}\zeta\overline{m}_{12},$$
$$\frac{1}{T}\sum_{t=1}^{T} s_{1t}(\theta_0)\, s_{3t}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\left(1-z_t^2\right)^2 u_{1t}(\theta_0)\, u_{3t}(\theta_0) \xrightarrow{p} \frac{1}{2}\zeta\overline{m}_{13},$$
$$\frac{1}{T}\sum_{t=1}^{T} s_{2t}(\theta_0)\, s_{3t}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{4}\left(1-z_t^2\right)^2 u_{2t}(\theta_0)\, u_{3t}(\theta_0) \xrightarrow{p} \frac{1}{4}\zeta\overline{m}_{23}.$$

Since all the elements in the score vector are asymptotically normal (see Lemma B), the result follows directly from application of the Cramer-Wold device, see for example Proposition 5.1 in White (1984), which completes the proof. □

Proposition 2 *Let $u_{jt}(\theta_0)$ be defined as in Theorem 1. Under Assumptions A and B, the observed information evaluated at $\theta = \theta_0$ converges in probability, i.e.,*

$$-\frac{1}{T}\frac{\partial^2}{\partial\theta\partial\theta'}L_T(\theta_0) \xrightarrow{p} \Omega,$$

where

$$\Omega = \begin{bmatrix} 2\overline{m}_{11} & \overline{m}_{12} & \overline{m}_{13} \\ \overline{m}_{12} & \frac{1}{2}\overline{m}_{22} & \frac{1}{2}\overline{m}_{23} \\ \overline{m}_{13} & \frac{1}{2}\overline{m}_{23} & \frac{1}{2}\overline{m}_{33} \end{bmatrix} > 0,$$

and $\overline{m}_{ij} = E\left(u_{it}(\theta_0)\, u_{jt}(\theta_0)\right)$ for $i = 1, 2, 3$ and $j = 1, 2, 3$.

Proof of Proposition 2 Recall from Result 2 (see Supplementary Appendix) that

$$-\frac{1}{T}\frac{\partial^2}{\partial\gamma^2}L_T(\theta_0) = 2\frac{1}{T}\sum_{t=1}^{T} z_t^2 u_{1t}^2(\theta_0) - \frac{2}{T}\sum_{t=1}^{T}\left(1-z_t^2\right)$$

$$(\ln|y_{t-2}|)^2 w_0^2 \left(\frac{1}{w_0} - \frac{1}{\sigma_t^2(\theta_0)}\right)^2 \\ +\frac{2}{T}\sum_{t=1}^{T}\left(1-z_t^2\right) w_0 \left(\frac{1}{w_0} - \frac{1}{\sigma_t^2(\theta_0)}\right)(\ln|y_{t-2}|)^2 .$$

Since z_t^2 and $u_{1t}^2(\theta_0)$ are independent, the first term on the right hand side converges to $2\overline{m}_{11}$ by Lemma A. Furthermore, since $(\ln|y_{t-2}|)^2 w_0^2 \left(\frac{1}{w_0} - \frac{1}{\sigma_t^2(\theta_0)}\right)^2$ and $w_0\left(\frac{1}{w_0} - \frac{1}{\sigma_t^2(\theta_0)}\right)(\ln|y_{t-2}|)^2$ have bounded moments, they are ergodic and stationary and since $E\left(1-z_t^2\right)=0$, it follows from the ergodic theorem that the last term on the right hand side converges in probability to zero. Therefore, the result follows. Using identical arguments we find

$$-\frac{1}{T}\frac{\partial^2}{\partial w^2}L_T(\theta_0) = -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(1-2z_t^2\right)u_{2t}^2(\theta_0) \xrightarrow{p} \frac{1}{2}\overline{m}_{22}, -\frac{1}{T}\frac{\partial^2}{\partial\alpha^2}L_T(\theta_0) \\ = -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(1-2z_t^2\right)u_{3t}^2(\theta_0) \xrightarrow{p} \frac{1}{2}\overline{m}_{33}, -\frac{1}{T}\frac{\partial^2}{\partial\gamma\partial w}L_T(\theta_0) \\ = \frac{1}{T}\sum_{t=1}^{T} z_t^2 u_{1t}(\theta_0)\,u_{2t}(\theta_0) \\ +\frac{1}{T}\sum_{t=1}^{T}\left(1-z_t^2\right)\left((\ln|y_{t-2}|)\left(\frac{w_0}{\sigma_t^2(\theta_0)}\right)\left(\frac{1}{w_0}-\frac{1}{\sigma_t^2(\theta_0)}\right)\right) \xrightarrow{p} \overline{m}_{12}, \\ -\frac{1}{T}\frac{\partial^2}{\partial\gamma\partial\alpha}L_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} z_t^2 u_{1t}(\theta_0)\,u_{3t}(\theta_0) \\ -\frac{1}{T}\sum_{t=1}^{T}\left(1-z_t^2\right) w_0 (\ln|y_{t-2}|)\left(\frac{w_0}{\alpha_0}\right)\left(\frac{1}{w_0}-\frac{1}{\sigma_t^2(\theta_0)}\right)^2 \\ \xrightarrow{p} \overline{m}_{13}, -\frac{1}{T}\frac{\partial^2 L_T(\theta_0)}{\partial w\partial\alpha} = -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(1-2z_t^2\right)u_{2t}(\theta_0)\,u_{3t}(\theta_0) \xrightarrow{p} \frac{1}{2}\overline{m}_{23}.$$

We proceed now to show that Λ is positive definite. Λ will be positive definite if for any non-zero column vector z with entries a, b and c, we show that $z^T\Lambda z > 0$. In our case

$$z^T\Lambda z = \zeta \begin{pmatrix} aE\left(u_{1t}^2\right) + \frac{b}{2}E(u_{1t}u_{2t}) + \frac{c}{2}E(u_{1t}u_{3t}) \\ \frac{a}{2}E(u_{1t}u_{2t}) + \frac{b}{4}E\left(u_{2t}^2\right) + \frac{c}{4}E(u_{2t}u_{3t}) \\ \frac{a}{2}E(u_{1t}u_{3t}) + \frac{b}{4}E(u_{2t}u_{3t}) + \frac{c}{4}E\left(u_{3t}^2\right) \end{pmatrix}^T \begin{pmatrix} a \\ b \\ c \end{pmatrix} \\ = \zeta\left[a^2E\left(u_{1t}^2\right) + \frac{ab}{2}E(u_{1t}u_{2t}) + \frac{ac}{2}E(u_{1t}u_{3t}) + \frac{ab}{2}E(u_{1t}u_{2t}) \right.\\ \left. + \frac{b^2}{4}E\left(u_{2t}^2\right) + \frac{bc}{4}E(u_{2t}u_{3t})\right.$$

$$+\frac{ac}{2}E\left(u_{1t}u_{3t}\right)+\frac{bc}{4}E\left(u_{2t}u_{3t}\right)+\frac{c^2}{4}E\left(u_{3t}^2\right)\Bigg]$$
$$=\zeta\Bigg[a^2E\left(u_{1t}^2\right)+\frac{b^2}{4}E\left(u_{2t}^2\right)+\frac{c^2}{4}E\left(u_{3t}^2\right)$$
$$+abE\left(u_{1t}u_{2t}\right)+acE\left(u_{1t}u_{3t}\right)+\frac{bc}{2}E\left(u_{2t}u_{3t}\right)\Bigg]$$

where we have written $u_{it}(\theta_0)=u_{it}$ for simplicity reasons. Since ζ, by Assumption A1, is always positive and larger than zero, and from Lemma A we have that

$$u_{3t}=\left(\frac{w_0}{\alpha_0}\right)\left(\frac{1}{w_0}-u_{2t}\right),$$

then, we need to show if the following term is strictly positive

$$a^2E\left(u_{1t}^2\right)+\frac{b^2}{4}E\left(u_{2t}^2\right)+\left(\frac{cw_0}{2\alpha_0}\right)^2E\left(\left(\frac{1}{w_0}-u_{2t}\right)^2\right)$$
$$+abE\left(u_{1t}u_{2t}\right)+\frac{acw_0}{\alpha_0}E\left(\frac{u_{1t}}{w_0}-u_{1t}u_{2t}\right)$$
$$+\frac{bcw_0}{2\alpha_0}E\left(\frac{u_{2t}}{w_0}-u_{2t}^2\right)=a^2E\left(u_{1t}^2\right)+\frac{1}{4}\left(b-\frac{cw_0}{\alpha_0}\right)^2$$
$$E\left(u_{2t}^2\right)+\left(\frac{c}{2\alpha_0}\right)^2+\frac{ac}{\alpha_0}E\left(u_{1t}\right)$$
$$+\frac{c}{2\alpha_0}\left(b-\frac{cw_0}{\alpha_0}\right)E\left(u_{2t}\right)+a\left(b-\frac{cw_0}{\alpha_0}\right)E\left(u_{1t}u_{2t}\right)$$
$$=a^2E\left(u_{1t}^2\right)+\frac{1}{4}\left(b-\frac{cw_0}{\alpha_0}\right)^2E\left(u_{2t}^2\right)+a\left(b-\frac{cw_0}{\alpha_0}\right)E\left(u_{1t}u_{2t}\right)$$
$$+\frac{c}{\alpha_0}\left[aE\left(u_{1t}\right)+\frac{1}{2}\left(b-\frac{cw_0}{\alpha_0}\right)E\left(u_{2t}\right)\right]+\left(\frac{c}{2\alpha_0}\right)^2$$
$$=E\Bigg[a^2u_{1t}^2+\frac{1}{4}\left(b-\frac{cw_0}{\alpha_0}\right)^2u_{2t}^2+a\left(b-\frac{cw_0}{\alpha_0}\right)u_{1t}u_{2t}$$
$$+\frac{c}{\alpha_0}\left[au_{1t}+\frac{1}{2}\left(b-\frac{cw_0}{\alpha_0}\right)u_{2t}\right]+\left(\frac{c}{2\alpha_0}\right)^2\Bigg]$$
$$=E\Bigg[\left(au_{1t}+\frac{1}{2}\left(b-\frac{cw_0}{\alpha_0}\right)u_{2t}\right)^2$$

$$+\frac{c}{\alpha_0}\left[au_{1t}+\frac{1}{2}\left(b-\frac{cw_0}{\alpha_0}\right)u_{2t}\right]+\left(\frac{c}{2\alpha_0}\right)^2\Bigg]$$

$$=E\left(au_{1t}+\frac{1}{2}\left(b-\frac{cw_0}{2\alpha_0}\right)u_{2t}+\frac{c}{2\alpha_0}\right)^2>0.$$

Finally notice that since $\Omega=2\Lambda\zeta^{-1}=\Lambda$, then $\Omega>0$. This completes the proof of Proposition 2. □

Proposition 3 *Define the lower and upper values for each parameter in θ_0 as $\gamma_L<\gamma_0<\gamma_U$, $w_L<w_0<w_U$, and $\alpha_L<\alpha_0<\alpha_U$, respectively and the neighborhood $N(\theta_0)$ around θ_0 as*

$$N(\theta_0)=\{\theta\backslash\gamma_L\leq\gamma\leq\gamma_U, w_L\leq w\leq w_U,\ and\ \alpha_L\leq\alpha\leq\alpha_U\}.$$

Under Assumptions A and B, there exists a neighborhood $N(\theta_0)$ for which for $i,j,k=1,2,3$

$$\sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}L_T(\theta)\right|\leq\frac{1}{T}\sum_{t=1}^{T}w_{ijkt},$$

where w_{ijkt} is stationary. Furthermore $\frac{1}{T}\sum_{t=1}^{T}w_{ijkt}\xrightarrow{a.s.}E(w_{ijkt})<\infty$ for $\forall ijk$.

Proof of Proposition 3 Let us start from the components of $\left|\frac{1}{T}\frac{\partial^3}{\partial\gamma^3}L_T(\theta)\right|$ defined in Result 3 (see Supplementary Appendix). Part I (which is also defined in Result 3) can be written as

$$\left|\frac{4}{T}\sum_{t=1}^{T}\frac{\alpha\left(\frac{y^*_{t-1}}{|y_{t-2}|^\gamma}\right)^2(\ln(|y_{t-2}|))^3}{\left(w+\alpha\left(\frac{y^*_{t-1}}{|y_{t-2}|^\gamma}\right)^2\right)}\left(1-\frac{(y^*_t)^2}{y^{2\gamma}_{t-1}\left(w+\alpha\left(\frac{y^*_{t-1}}{|y_{t-2}|^\gamma}\right)^2\right)}\right)\right|$$

$$\leq\left|\frac{4}{T}\sum_{t=1}^{T}\frac{\left(\left(w+\alpha\left(\frac{y^*_{t-1}}{|y_{t-2}|^\gamma}\right)^2\right)-w\right)|\ln(|y_{t-2}|)|^3}{\left(w+\alpha\left(\frac{y^*_{t-1}}{|y_{t-2}|^\gamma}\right)^2\right)}\right.$$
$$\left.\times\left(\frac{(y^*_t)^2}{y^{2\gamma}_{t-1}\left(w+\alpha\left(\frac{y^*_{t-1}}{|y_{t-2}|^\gamma}\right)^2\right)}+1\right)\right|$$

$$\leq \left| \frac{4}{T} \sum_{t=1}^{T} \left(\frac{\left(y_t^*\right)^2}{y_{t-1}^{2\gamma} \left(w + \alpha \left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma} \right)^2 \right)} + 1 \right) |\ln (|y_{t-1}|)|^3 \right|$$

$$\leq \left| \frac{4}{T} \sum_{t=1}^{T} \left(\left(\frac{y_{t-1}^{2\gamma_0} \left(w_0 + \alpha_0 \left(\frac{y_{t-1}^*}{|y_{t-2}|^{\gamma_0}} \right)^2 \right)}{y_{t-1}^{2\gamma} \left(w + \alpha \left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma} \right)^2 \right)} \right) z_t^2 + 1 \right) |\ln (|y_{t-1}|)|^3 \right|$$

$$\leq \left| \frac{4}{T} \sum_{t=1}^{T} \left(\left(\frac{w_0}{w} y_{t-1}^{2(\gamma_0 - \gamma)} + \frac{\alpha_0}{\alpha} \left(\frac{y_{t-1}^*}{|y_{t-2}|} \right)^{2(\gamma_0 - \gamma)} \right) z_t^2 + 1 \right) |\ln (|y_{t-1}|)|^3 \right|$$

$$\leq \left| \frac{4}{T} \sum_{t=1}^{T} \left(\left\{ \frac{w_U}{w_L} \Lambda_{t-1} + \frac{\alpha_U}{\alpha_L} \Lambda_{t-2} \right\} z_t^2 + 1 \right) |\ln (|y_{t-1}|)|^3 \right|$$

$$\leq \left| \frac{4}{T} \sum_{t=1}^{T} \left(\left\{ \frac{w_U}{w_L} \Lambda_{t-1} + \frac{\alpha_U}{\alpha_L} \Lambda_{t-2} \right\} z_t^2 + 1 \right) |\ln (|y_{t-1}|)|^3 \right|$$

$$\leq \left| \frac{4}{T} \sum_{t=1}^{T} \left(\left\{ \frac{w_U}{w_L} \Lambda_{t-1} + \frac{\alpha_U}{\alpha_L} \Lambda_{t-2} \right\} z_t^2 + 1 \right) |\ln (|y_{t-1}|)|^3 \right|,$$

where we can define the lower bound for all t, $y_L \leq |y_{t-1}|$, $y_L \leq |y_{t-2}|$, $\Lambda_{t-1} = \max \left\{ y_L^{2|\gamma_U - \gamma_L|}, y_{t-1}^{2|\gamma_U - \gamma_L|} \right\}$, $\Lambda_{t-2} = \max \left\{ 1, \left(\frac{y_{t-1}^*}{|y_{t-2}|} \right)^{2|\gamma_U - \gamma_L|} \right\}$ and the result follows by setting $2 |\gamma_U - \gamma_L| = \varphi$, Assumptions A and B and the law of large numbers (see Jensen and Rahbek (2004a), Lemma 5). Part II requires also assumption A3 since

$$\left| \frac{8}{T} \sum_{t=1}^{T} \frac{\alpha^3 \left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma} \right)^6 (\ln (|y_{t-2}|))^3}{\left(w + \alpha \left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma} \right)^2 \right)^3} \left(1 - \frac{3 \left(y_t^*\right)^2}{y_{t-1}^{2\gamma} \left(w + \alpha \left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma} \right)^2 \right)} \right) \right|$$

$$\leq \left| \frac{8}{T} \sum_{t=1}^{T} \left(\frac{3 \left(y_t^*\right)^2}{y_{t-1}^{2\gamma} \left(w + \alpha \left(\frac{y_{t-1}^*}{|y_{t-2}|^\gamma} \right)^2 \right)} - 1 \right) |\ln (|y_{t-2}|)|^3 \right|$$

$$\leq \left| \frac{8}{T} \sum_{t=1}^{T} \left(3 \left\{ \frac{w_U}{w_L} \Lambda_{t-1} + \frac{\alpha_U}{\alpha_L} \Lambda_{t-2} \right\} z_t^2 + 1 \right) |\ln (|y_{t-1}|)|^3 \right|.$$

Parts III, IV, V and VI follow the same argument.

Along the same lines for $\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha^3}L_T(\theta)\right|$

$$\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha^3}L_T(\theta)\right| = \left|\frac{1}{T}\sum_{t=1}^{T}\left(3\frac{\left(y_t^*\right)^2}{y_{t-1}^{2\gamma}\left(w+\alpha\left(\frac{y_{t-1}^*}{|y_{t-2}|^{\gamma}}\right)^2\right)}-1\right)\frac{\left(\frac{y_{t-1}^*}{y_{t-2}^{\gamma}}\right)^6}{\left(w+\alpha\left(\frac{y_{t-1}^*}{|y_{t-2}|^{\gamma}}\right)^2\right)^3}\right|$$

$$\leq \left|\frac{1}{T}\sum_{t=1}^{T}\left(3\frac{\left(y_t^*\right)^2}{y_{t-1}^{2\gamma}\left(w+\alpha\left(\frac{y_{t-1}^*}{|y_{t-2}|^{\gamma}}\right)^2\right)}-1\right)\right|\frac{1}{\alpha_L^3}$$

$$\leq \frac{1}{T}\sum_{t=1}^{T}\left(3\left\{\frac{w_U}{w_L}\Lambda_{t-1}+\frac{\alpha_U}{\alpha_L}\Lambda_{t-2}\right\}z_t^2+1\right)\frac{1}{\alpha_L^3}.$$

The rest of the cases follow directly using the same argument. This completes the proof of Proposition 3. □

Proof of Theorem 1 Given the conditions provided by Propositions 1–3, Theorem 1 follows from Lumsdaine (1996, pp. 593–595, Theorem 3), the ergodic theorem and Lemma 1, p. 1206 in Jensen and Rahbek (2004b). □

References

Andersen TG, Lund J (1997) Estimating continuous time stochastic volatility models of the short term interest rate. J Econom 77:343–377

Ball CA, Torous WN (1999) The stochastic volatility of short-term interest rates: some international evidence. J Financ 54(6):2339–2359

Berkes I, Horváth L (2004) The efficiency of the estimators of the parameters in GARCH processes. Ann Stat 32(2):633–655

Billingsley P (1961) The Lindeberg–Lévy theorem for martingales. Proc Am Math Soc 12:788–792

Bouezmarni T, Rombouts JVK (2010) Nonparametric density estimation for positive time series. Comput Stat Data Anal 54(2):245–261

Brenner R, Harjes R, Kroner K (1996) Another look at models of the short term interest rates. J Financ Quant Anal 31:85–107

Brown BM (1971) Martingale central limit theorems. Ann Math Stat 42:59–66

Brown TC, Feigin PD, Pallant DL (1996) Estimation for a class of positive nonlinear time series models. Stoch Process Appl 63(2):139–152

Broze L, Scaillet O, Zakoïan JM (1995) Testing for continuous time models of the short term interest rate. J Empir Financ 2:199–223

Bu R, Chen J, Hadri K (2017) Specification analysis in regime-switching continuous-time diffusion models for market volatility. Stud Nonlinear Dyn Econom 21:1

Chan KC, Karolyi GA, Longstaff F, Sanders A (1992) An empirical comparison of alternative models of short term interest rates. J Finan 47(3):1209–1227

Dias-Curto J, Castro-Pinto J, Nuno-Tavares G (2009) Modeling stock markets' volatility using GARCH models with normal Student's t and stable Paretian distributions. Stat Pap 50:311–321

Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50:987–1007

Fornari F, Mele A (2006) Approximating volatility diffusions with CEV-ARCH models. J Econ Dyn Control 30(6):931–966

Francq C, Wintenberger O, Zakoïan J-M (2018) Goodness-of-fit tests for Log-GARCH and EGARCH models. TEST 27(1):27–51

Frydman H (1994) Asymptotic inference for the parameters of a discrete-time square-root process. Math Financ 4(2):169–181

Hamadeh T, Zakoïan J-M (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. J Stat Plan Inference 141(1):488–507

Han H, Zhang S (2009) Nonstationary semiparametric ARCH models. Department of Economics, National University of Singapore, Manuscript

Jensen ST, Rahbek A (2004a) Asymptotic normality of the QML estimator of ARCH in the nonstationary case. Econometrica 72(2):641–646

Jensen ST, Rahbek A (2004b) Asymptotic inference for nonstationary GARCH. Econ Theory 20(6):1203–1226

Jensen ST, Rahbek A (2007) On the law of large numbers for (geometrically) Ergodic Markov chains. Econ Theory 23:761–766

Klüppelberg C, Lindner A, Maller R (2004) A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour. J Appl Probab 41:601–622

Kristensen D, Rahbek A (2005) Asymptotics of the QMLE for a class of ARCH(q) models. Econ Theory 21:946–961

Kristensen D, Rahbek A. (2008) Asymptotics of the QMLE for non-linear ARCH models. J Time Series Econ 1(1), Article 2

Lehmann EL (1999) Elements of large sample theory. Springer, New York

Ling S (2004) Estimation and testing stationarity for double-autoregressive models. J R Stat Soc B 66(1):63–78

Lumsdaine RL (1996) Asymptotic properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. Econometrica 64(3):575–596

Maheu JM, Yang Q (2016) An infinite hidden Markov model for short-term interest rates. J Empir Financ 38:202–220

Pedersen RS, Rahbek A (2016) Nonstationary GARCH with t-distributed innovations. Econ Lett 138:19–21

Straumann D, Mikosch T (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. Ann Stat 34(5):2449–2495

Triffi A (2006) Issues of aggregation over time of conditional heteroskedastic volatility models: what type of diffusion do we recover? Stud Nonlinear Dyn Econ 10:4

Wang Q, Phillips PCB (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econ Theory 25:1–29

Wang Q, Phillips PCB (2009b) Structural nonparametric cointegrating regression. Econometrica 77(6):1901–1948

White H (1984) Asymptotic theory for econometricians. Academic Press, New York

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.