COMMENT ON 'AUTOREGRESSIVE CONDITIONAL DURATION: A NEW MODEL FOR IRREGULARLY SPACED TRANSACTION DATA'

GIUSEPPE CAVALIERE[a], THOMAS MIKOSCH[b], ANDERS RAHBEK[c] AND FREDERIK VILANDT[c]

First draft: May 25, 2023; Second version: March 3, 2024; This version: September 30, 2024

Abstract

Engle and Russell (1998, Econometrica, 66:1127–1162) apply results from the GARCH literature to prove consistency and asymptotic normality of the (exponential) (Q)MLE for the autoregressive conditional duration (ACD) model, under the assumption of strict stationarity and ergodicity of the durations. Using novel arguments based on renewal process theory, we show in this note that their results hold under the stronger requirement that durations have finite expectation. We also show that this is not the case in general, and provide a counterexample where the MLE is asymptotically mixed normal and converges at a rate significantly slower than usual, even if the durations are strictly stationary and ergodic. The main difference between ACD and GARCH asymptotics is that the former needs to account for the fact that the number of durations in a given time span is random. As a by-product, we provide a new lemma which can be applied to analyze asymptotic properties of extremum estimators when the number of observations is random.

KEYWORDS: autoregressive conditional duration (ACD); quasi maximum likelihood.

1 INTRODUCTION

IN THE SEMINAL PAPER by Engle and Russell (1998, ER henceforth), autoregressive conditional duration (ACD) models were introduced for modeling inter-arrival times, or durations, between financial transactions. Given some observation period $[0, T]$, say, with $n(T)$ observed event times $0 < t_1 < t_2 < \cdots < t_{n(T)} \leq T$, the durations x_i are given by $x_i = t_i - t_{i-1}$ and modeled as

$$x_i = \psi_i(\theta)\,\varepsilon_i, \quad i = 1, \ldots, n(T), \tag{1.1}$$

$$\psi_i(\theta) = \omega + \alpha x_{i-1} + \beta\psi_{i-1}(\theta) \tag{1.2}$$

where the innovations $\{\varepsilon_i\}$ are strictly positive, i.i.d. and unit mean, $\mathbb{E}[\varepsilon_i] = 1$.

[a]Department of Economics, University of Bologna, Italy and Department of Economics, University of Exeter, UK.
[b]Department of Mathematical Sciences, University of Copenhagen, Denmark.
[c]Department of Economics, University of Copenhagen, Denmark.
We thank Guido Imbens for his encouragement and valuable suggestions. We also thank Matias Cattaneo, David Harris and seminar participants at Princeton U, TSE, UPF, U Auckland, U Cambridge, U Essex, U Exeter, U Melbourne, U Surrey, U Sydney and U York for comments. We gratefully acknowledge support from the Independent Research Fund Denmark (DFF Grant 10.46540/3099-00076B). Cavaliere and Rahbek also acknowledge support from the Italian Ministry of University and Research (PRIN 2020 Grant 2020B2AKFW). Correspondence to: Giuseppe Cavaliere, Department of Economics, University of Bologna, email giuseppe.cavaliere@unibo.it.

The quasi maximum likelihood estimator (QMLE) of $\theta = (\omega, \alpha, \beta)' \in \Theta \subset [0, \infty)^3$ is defined as $\hat{\theta}_T = \arg\max_{\theta \in \Theta} \mathcal{L}_{n(T)}(\theta)$, with $\mathcal{L}_{n(T)}(\theta)$ the exponential likelihood

$$\mathcal{L}_{n(T)}(\theta) = \sum_{i=1}^{n(T)} \ell_i(\theta), \qquad \ell_i(\theta) = -\Big(\log \psi_i(\theta) + \frac{x_i}{\psi_i(\theta)}\Big), \qquad T \geq 0, \tag{1.3}$$

with initial values x_0, $\psi_0(\theta)$. The true parameter value is denoted by $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$.

ER note that the likelihood function in (1.3) is identical to the likelihood function of the GARCH(1,1) model with Gaussian innovations. Hence, for their main result (p.1135), ER refer to Lee and Hansen (1994, LH henceforth) to conclude that under the condition of strict stationarity and ergodicity of the durations x_i, i.e., $\mathbb{E}[\log(\alpha_0 \varepsilon_i + \beta_0)] < 0$, $\hat{\theta}_T$ is consistent and asymptotically normal at the usual $\sqrt{T}$ rate.

As we argue in this note, the machinery in LH cannot be applied to the ACD setup unless additional arguments are used and further assumptions imposed. In particular, with θ_0 such that the strict stationarity and ergodicity condition holds, i.e., $\theta_0 \in \{\theta \in (0, \infty)^3 : \mathbb{E}[\log(\alpha \varepsilon_i + \beta)] < 0\}$, we argue that, in contrast to the GARCH case, the behavior of the QML estimator $\hat{\theta}_T$ depends on whether (i) $\alpha_0 + \beta_0 < 1$, (ii) $\alpha_0 + \beta_0 > 1$ or (iii) $\alpha_0 + \beta_0 = 1$. Specifically, results regarding rates of convergence, asymptotics of the QMLE, convergence of the score and sample information all depend on which of the three cases above holds. Key is that modifications of arguments are needed due to randomness of the number of durations $n(T)$.

To preview why, consider the score and information, evaluated at the true value $\theta = \theta_0$,

$$\mathcal{S}_{n(T)} = \frac{\partial \mathcal{L}_{n(T)}(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} = \sum_{i=1}^{n(T)} \xi_i\,, \qquad \xi_i = \frac{\partial \ell_i(\theta)}{\partial \theta}\Big|_{\theta=\theta_0}, \tag{1.4}$$

$$\mathcal{I}_{n(T)} = -\frac{\partial^2 \mathcal{L}_{n(T)}(\theta)}{\partial \theta \partial \theta'}\Big|_{\theta=\theta_0} = \sum_{i=1}^{n(T)} \zeta_i\,, \quad \zeta_i = -\frac{\partial^2 \ell_i(\theta)}{\partial \theta \partial \theta'}\Big|_{\theta=\theta_0}. \tag{1.5}$$

To establish asymptotic normality of $\hat{\theta}_T$, standard theory usually requires that these satisfy a central limit theorem (CLT) and a law of large numbers (LLN), respectively. The ACD setting, however, is not standard as the number of observations $n(T)$ is *random,* and not independent of the sequences $\{\xi_i\}$ and $\{\zeta_i\}$.

Note in this respect that the CLT and the LLN for deterministic number n of observations, that is (with N denoting the Gaussian distribution)

$$n^{-1/2}\mathcal{S}_n = n^{-1/2} \sum_{i=1}^{n} \xi_i \to_{\mathrm{d}} \mathrm{N}(0, \Omega_S)\,, \tag{1.6}$$

$$n^{-1}\mathcal{I}_n = n^{-1} \sum_{i=1}^{n} \zeta_i \to_{\mathrm{a.s.}} \Omega_I, \tag{1.7}$$

are not sufficient for their random $n(T)$-analogues in (1.4)-(1.5) to hold. That is, it does not follow from (1.6) that $n(T)^{-1/2} \sum_{i=1}^{n(T)} \xi_i$ is asymptotically Gaussian even if $n(T) \to_{\mathrm{a.s.}} \infty$; see, e.g., Chapter 1.3 in Gut (2009). Likewise, it does not follow that $g(T)^{-1} \sum_{i=1}^{n(T)} \xi_i$ is asymptotically Gaussian for some increasing deterministic sequence $g(T)$. Hence, the arguments based on LH, which are based on $n(T) = n$ deterministic, do not apply.

This note makes the following contributions.

First, we provide a new lemma which can be applied to analyze asymptotic properties of extremum estimators when the number of observations $n(T)$ is random. The arguments in its proof use renewal theory and are thus different from LH/ER.

Second, we apply this result and show in Section 2 that under the *additional condition* $\alpha_0+\beta_0 < 1$ which implies $\mathbb{E}[x_i] < \infty$, $n(T)$ and T are proportional in the sense that $n(T)/T \rightarrow_{\text{a.s.}} c > 0$. The latter result can be used to prove that normalizing the score $\mathcal{S}_{n(T)} = \sum_{i=1}^{n(T)} \xi_i$ by either $\sqrt{T}$, $\sqrt{n(T)}$, or the sample information, $\mathcal{I}_{n(T)}^{1/2}$, leads to asymptotic normality, establishing asymptotic normality of $\sqrt{T}(\hat{\theta}_T - \theta_0)$, $\sqrt{n(T)}(\hat{\theta}_T - \theta_0)$, and $\mathcal{I}_{n(T)}^{1/2}(\hat{\theta}_T - \theta_0)$; see Theorem 2.1.

Third, to illustrate that these results do not hold in general we present in Section 3 a counterexample, with x_i stationary and ergodic, but where $\alpha_0 + \beta_0 > 1$, and hence $\mathbb{E}[x_i] = \infty$. With exponential innovations ε_i we show that $\hat{\theta}_T$ converges at the rate $T^{\kappa_0/2}$ for some $\kappa_0 \in (0,1)$, which is significantly slower than the usual $\sqrt{T}$ rate. Moreover, $\hat{\theta}_T$ has a mixed normal (MN) limiting distribution. Hence, the limiting distribution of the MLE $\hat{\theta}_T$ differs from the classical, LH/ER form $\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow_{\text{d}} \text{N}$. Importantly, the MN limit theory implies that different normalizations lead to distinct asymptotic distributions.

Finally, we note that the arguments in the counterexample are specific to the MLE, and hence there is no guarantee that they can be generalized to the QMLE for the case of $\alpha_0+\beta_0 > 1$, and neither to the 'unit root' case of $\alpha_0 + \beta_0 = 1$.

2 Main result

In this section we show that the asymptotic normality of the QMLE can be obtained by imposing the additional condition $\alpha_0 + \beta_0 < 1$ which implies $\mu = \mathbb{E}[x_i] < \infty$. The key insight is that if $\mu < \infty$, the random number of durations $n(T)$ over the observation period $[0, T]$ satisfies

$$n(T)/T \rightarrow_{\text{a.s.}} 1/\mu \text{ as } T \rightarrow \infty. \tag{2.1}$$

This in turn (as $n(T) \rightarrow_{\text{a.s.}} \infty$) is sufficient for the deterministic n LLN in (1.7) to imply that its random $n(T)$-analogue holds. To establish the random $n(T)$-CLT for $\mathcal{S}_{n(T)}$ in (1.4), the deterministic n-CLT in (1.6) is replaced by its stronger functional version

$$n^{-1/2}\mathcal{S}_n(\cdot) = n^{-1/2}\sum_{i=1}^{\lfloor n\cdot\rfloor} \xi_i \rightarrow_{\text{w}} \Omega_S^{1/2} B(\cdot)$$

where B is a standard multivariate Brownian motion and convergence is on the space of càdlàg functions on $[0, \infty)$ equipped with the J_1-topology.

To derive the asymptotic distribution of the QMLE presented in Theorem 2.1, we make use of the following general lemma which extends the results in LH to allow for a random number of observations.

Lemma 2.1 *Let $Q_n(\varphi) \in \mathbb{R}$ be a random function of the parameter $\varphi \in \Phi \subseteq \mathbb{R}^k$, indexed by $n \in \mathbb{N}$. Assume that $Q_n(\cdot)$ is three times continuously differentiable, and that for φ_0 in the interior of Φ, as $n \rightarrow \infty$:*

(C.1) $n^{-1/2}\partial Q_{\lfloor n\cdot\rfloor}(\varphi_0)/\partial\varphi \rightarrow_w \Omega_S^{1/2}B(\cdot),\ \ \Omega_S > 0,$

(C.2) $-n^{-1}\partial^2 Q_n(\varphi_0)/\partial\varphi\partial\varphi' \rightarrow_{a.s.} \Omega_I > 0,$

(C.3) $\max_{h,i,j=1,\ldots,k}\sup_{\varphi\in\mathcal{N}(\varphi_0)}\left|n^{-1}\frac{\partial^3 Q_n(\varphi)}{\partial\varphi_h\partial\varphi_i\partial\varphi_j}\right| \leq \tau_n \rightarrow_{a.s.} \tau$

where $B(\cdot)$ is a k-dimensional Brownian motion, $\mathcal{N}(\varphi_0)$ is a closed neighborhood of φ_0, and $0 < \tau < \infty$. Moreover, with $\{n(t)\}_{t\geq 0}$ a counting process defined on the same probability space as $Q_n(\cdot)$, assume that for some constant $c \in (0,\infty)$:

(C.4) $n(T)/T \rightarrow_{a.s.} c$ *as* $T \rightarrow \infty$.

With $Q_{n(T)}(\varphi) = Q_n(\varphi)|_{n=n(T)}$, there exists an open neighborhood $U(\varphi_0) \subseteq \mathcal{N}(\varphi_0)$, such that, as $T \rightarrow \infty$:

- (i) *With probability tending to one there exists a unique maximum point $\hat{\varphi}_T$ of $Q_{n(T)}(\varphi)$ in $U(\varphi_0)$, $Q_{n(T)}(\varphi)$ is concave on $U(\varphi_0)$ and $\partial Q_{n(T)}(\hat{\varphi}_T)/\partial\varphi = 0$;*
- (ii) $\hat{\varphi}_T \rightarrow_p \varphi_0$;
- (iii) $\sqrt{T}(\hat{\varphi}_T - \varphi_0) \rightarrow_d \mathrm{N}(0,\Sigma)$, $\Sigma = c^{-1}\Omega_I^{-1}\Omega_S\Omega_I^{-1}$.

All proofs of the results in this note are provided in Section 4. Note that Assumption (C.4) can be replaced by $n(T) \rightarrow_{\text{a.s.}} \infty$ and $n(T)/T \rightarrow_{\text{p}} c, 0 < c < \infty$.

Our main result is as follows.

THEOREM 2.1 *For the ACD model (1.1)-(1.2) with true parameter θ_0, assume:*

- (i) *$\{\varepsilon_i\}$ is an i.i.d. sequence of random variables with support $(0,\infty)$, pdf $f_\varepsilon(\cdot)$ bounded away from zero on compact subsets of $(0,\infty)$, $\mathbb{E}[\varepsilon_i] = 1$ and $\mathbb{E}[\varepsilon_i^2] < \infty$,*
- (ii) *θ_0 is an interior point satisfying $\alpha_0 + \beta_0 < 1$.*

As $T \rightarrow \infty$, the maximizer $\hat{\theta}_T$ of $\mathcal{L}_{n(T)}(\theta)$ in (1.3) is consistent and asymptotically normal:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow_d \mathrm{N}(0,\Sigma)$$

where $\Sigma = \mu\Omega_I^{-1}\Omega_S\Omega_I^{-1}$. Here $\mu = \mathbb{E}[x_i] = \omega_0/(1-\alpha_0-\beta_0) < \infty$, and $\Omega_I = \mathbb{E}[\zeta_i]$, $\Omega_S = \mathbb{E}[\xi_i\xi_i']$ are given by (1.7) and (1.6), respectively.

REMARK 2.1 *Theorem 2.1 shows that, if the strict stationarity condition $\mathbb{E}[\log(\alpha_0\varepsilon_i + \beta_0)] < 0$ is strengthened with the additional restriction $\alpha_0 + \beta_0 < 1$, then $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is asymptotically normal as $T \rightarrow \infty$. In particular, $n(T)/T \rightarrow_{a.s.} 1/\mu > 0$. In this case, using $\sqrt{n(T)}$ as normalization instead, Theorem 2.1 and (2.1) jointly imply that $\sqrt{n(T)}(\hat{\theta}_T - \theta_0) \rightarrow_d \mathrm{N}(0,(1/\mu)\Sigma)$. Hence, up to a scaling factor, $\sqrt{T}(\hat{\theta}_T - \theta_0)$ and $\sqrt{n(T)}(\hat{\theta}_T - \theta_0)$ have the same asymptotic distribution. Likewise, when normalizing by the sample information, we find $\mathcal{I}_{n(T)}^{1/2}(\hat{\theta}_T - \theta_0) \rightarrow_d \mathrm{N}(0,I)$ for the MLE as then $\Omega_S = \Omega_I$.*

REMARK 2.2 *The proof of Theorem 2.1 relies on the new Lemma 2.1, which may also be used to establish asymptotic theory for the more general ACD(m,q) models mentioned in ER (p.1133) which allow $m \geq 1$ lags of x_i and $q \geq 0$ lags of ψ_i in (1.2), including the simple stylized ACD$(1,0)$ model considered in Cavaliere, Mikosch, Rahbek and Vilandt (2024).*

REMARK 2.3 *Asymptotic normality of the estimator is not guaranteed to hold when Assumption (ii) does not hold, see Section 3 for a counterexample.*

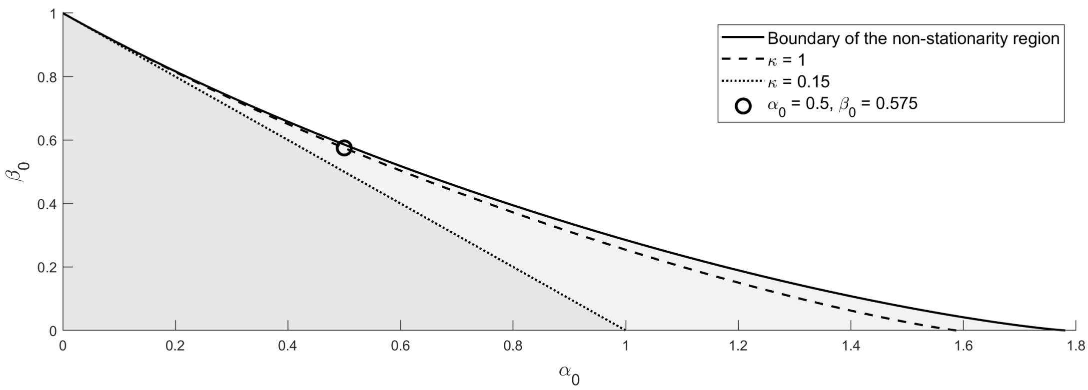

Figure 1: $\kappa_0 = 1$, $\kappa_0 = 0.15$ and the stationary region (grey).

3 Non-standard asymptotics

Here we present a counterexample which shows that if $\alpha_0 + \beta_0 > 1$, implying $\mathbb{E}[x_i] = \infty$, the asymptotic distribution of $\hat{\theta}_T$ is not normal, even under strict stationarity and ergodicity. Specifically, different normalizations (e.g., using a deterministic function of T, or a random normalization such as the sample information) may lead to different asymptotics.

Consider the ACD model given by (1.1)-(1.2), under the assumption that the ε_i's are exponentially distributed with $\mathbb{E}[\varepsilon_i] = 1$. We have the following result.

Theorem 3.1 *Consider the exponential ACD model with true parameter θ_0 being an interior point satisfying the strict stationarity condition $\mathbb{E}[\log(\alpha_0\varepsilon + \beta_0)] < 0$ and $\alpha_0 + \beta_0 > 1$. As $T \to \infty$, the maximizer $\hat{\theta}_T$ of $\mathcal{L}_{n(T)}(\theta)$ in (1.3) is consistent and asymptotically mixed normal:*

$$T^{\kappa_0/2}(\hat{\theta}_T - \theta_0) \to_d \mathrm{MN}(0, V) \tag{3.1}$$

where $V = (1/\lambda_{\kappa_0})\Omega_I^{-1}$. Here, $\kappa_0 = \kappa(\alpha_0, \beta_0) \in (0,1)$ is the unique solution of the equation $\mathbb{E}[(\alpha_0\varepsilon + \beta_0)^\kappa] = 1$, $\Omega_I = \mathbb{E}[\zeta_i]$ is defined in (1.7) and the random variable $\lambda_{\kappa_0} > 0$ is given in Lemma 4.1.

The proof of Theorem 3.1 makes use of the key result that, when $\alpha_0 + \beta_0 > 1$,

$$n(T)/T^{\kappa_0} \to_{\mathrm{d}} \lambda_{\kappa_0} \tag{3.2}$$

rather than in probability to a positive constant; see Lemma 4.1. This is non-standard and, importantly, implies the need for a different approach to show convergence results for sums of the form $\sum_{i=1}^{n(T)} \xi_i$; see in particular Section 4.

In line with Remark 2.1 the following corollary for $\hat{\theta}_T$ normalized by the sample information $\mathcal{I}_{n(T)}$ or by $n(T)$ holds.

Corollary 3.1 *Under the assumptions of Theorem 3.1, $\sqrt{n(T)}(\hat{\theta}_T - \theta_0) \to_d \mathrm{N}(0, \Omega_I^{-1})$ and $\mathcal{I}_{n(T)}^{1/2}(\hat{\theta}_T - \theta_0) \to_d \mathrm{N}(0, I)$ as $T \to \infty$.*

Remark 3.1 *Consider a drifting sequence of true parameters of the form $\theta_T = \theta_0 + sT^{-\kappa_0/2}$ ($s \in \mathbb{R}^3$). Using the same arguments as in the proof of Theorem 3.1, it holds that $T^{\kappa_0/2}(\hat{\theta}_T -$*

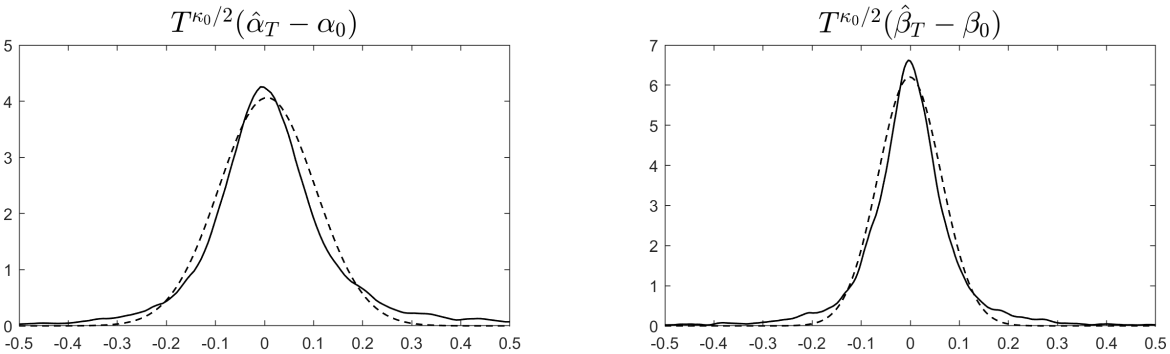

Figure 2: Estimated densities (solid lines) against the Gaussian pdf (dashed lines)

$\theta_0) \to_d -s + \mathrm{MN}(0, V)$, which is mixed normal with (deterministic) non-centrality parameter $-s$. In contrast, for $\hat{\theta}_T$ normalized by the sample information $\mathcal{I}_{n(T)}$ as in Corollary 3.1 we find $\mathcal{I}_{n(T)}^{1/2}(\hat{\theta}_T - \theta_0) \to_d -\lambda_{\kappa_0}^{1/2}\Omega_I^{1/2} s + \mathrm{N}(0, I)$, where the non-centrality parameter $\lambda_{\kappa_0}^{1/2}\Omega_I^{1/2} s$ is now random. The latter result implies that when $\alpha_0 + \beta_0 > 0$ the asymptotic local power of t-ratios is random in the limit, which contrasts with the case $\alpha_0 + \beta_0 < 1$ in Theorem 2.1.

To shed some light on the mixed normality in Theorem 3.1, note that when $\alpha_0 + \beta_0 > 1$ the limiting distribution of $\hat{\theta}_T$ does not have exponential tails; in particular, it has infinite variance. To see this, write the right hand side of (3.1) as $\lambda_{\kappa_0}^{-1/2}\Omega_I^{-1/2} Z$ with $Z \in \mathbb{R}^3$ standard normal, independent of λ_{κ_0}. Then, since $\lambda_{\kappa_0}^{-1/\kappa_0}$ is a κ_0-stable random variable with $\kappa_0 \in (0, 1)$ (see Lemma 4.1), it follows by an application of Breiman's Lemma (see, e.g., Lemma 3.1.11 in Mikosch and Wintenberger, 2024) that for x large, $\mathbb{P}(\lambda_{\kappa_0}^{-1/2}\Omega_I^{-1/2} Z > x) \sim c\mathbb{P}(\lambda_{\kappa_0}^{-1/\kappa_0} > x^{2/\kappa_0}) \sim cx^{-2}$ (component-wise) with c generic positive constants.

To further emphasize the different asymptotic behavior of $\hat{\theta}_T$ when $\alpha_0 + \beta_0 > 1$, consider here a small Monte Carlo study where $10,000$ i.i.d. realizations of $\hat{\theta}_T = (\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T)'$ are generated for large T. In particular, we consider the kernel density estimates for $T^{\kappa_0/2}(\hat{\alpha}_T - \alpha_0)$ and $T^{\kappa_0/2}(\hat{\beta}_T - \beta_0)$ against the normal density function which matches the (sample) median and interquartile range across Monte Carlo realizations. Specifically, we set $(\alpha_0, \beta_0) = (0.500, 0.575)$, corresponding to approximately $\kappa_0 = 0.15$. This particular value is shown in Figure 1, where we also show the values of (α_0, β_0) corresponding to $\kappa_0 = 1$ $(\alpha_0 + \beta_0 = 1)$, $\kappa_0 = 0.15$, as well as those satisfying $\mathbb{E}[\log(\alpha_0 \varepsilon + \beta_0)] = 0$ (boundary of the non-stationarity region). The sample size T is calibrated such that the median number of durations in $[0, T]$ is about $900,000$.

Figure 2 shows that, as predicted by Theorem 3.1, the large sample distributions of both $T^{\kappa_0/2}(\hat{\alpha}_T - \alpha_0)$ and $T^{\kappa_0/2}(\hat{\beta}_T - \beta_0)$ display fatter tails than the Gaussian pdf.

Remark 3.2 *It is important to note that for the case of $\alpha_0 + \beta_0 = 1$, which is not ruled out by the ER conditions, the limiting behavior of $\hat{\theta}_T$ is unknown (for both types of normalizations), even when ε_i is exponential (MLE). The key challenge in this case is that the large sample behavior of $n(T)$ has not been established at present; see, e.g., Mikosch and Resnick (2006). Also, we note that the results in Theorem 3.1 and its corollary hold only for the MLE. Further research is needed to understand the QMLE case.*

4 Proofs and additional results

4.1 Proof of Lemma 2.1

We first consider the asymptotic behavior as $T \to \infty$ for the score, the second order derivative and the third order derivatives of $Q_{n(T)}(\varphi)$. Next, we use these results to establish *(i)–(iii).*
Score: It holds that with $\partial Q_{n(T)} = \partial Q_{n(T)}(\varphi)/\partial\varphi\big|_{\varphi=\varphi_0}$,

$$n(T)^{-1/2}\partial Q_{n(T)} \to_{\mathrm{d}} \mathrm{N}(0, \Omega_S). \tag{4.1}$$

To see this, let $\partial Q_{[Tc]} = \partial Q_{[Tc]}(\varphi)/\partial\varphi\big|_{\varphi=\varphi_0}$, and decompose $n(T)^{-1/2}\partial Q_{n(T)}$ as

$$n(T)^{-1/2}\partial Q_{n(T)} = a_T^{-1/2}([Tc]^{-1/2}\partial Q_{[Tc]}) + a_T^{-1/2} A_T,$$

where $A_T = [Tc]^{-1/2}(\partial Q_{n(T)} - \partial Q_{[Tc]})$ and $a_T = \frac{n(T)}{[Tc]}$. By conditions (C.1) and (C.4), $[Tc]^{-1/2}\partial Q_{[Tc]} \to_{\mathrm{d}} \mathrm{N}(0, \Omega_S)$ and $a_T \to_{\text{a.s.}} 1$. Next, note that for any $M, \delta > 0$,

$$\mathbb{P}(\|A_T\| > M) = \mathbb{P}(\|A_T\| > M,\ |n(T)/T - c| > \delta) + \mathbb{P}(\|A_T\| > M,\ |n(T)/T - c| \le \delta).$$

Here, $\mathbb{P}(\|A_T\| > M, |n(T)/T - c| > \delta) \le \mathbb{P}(|n(T)/T - c| > \delta) \to 0$ by (C.4). Next,

$$\begin{aligned}\mathbb{P}(\|A_T\| > M,\ |n(T)/T - c| \le \delta) &= \mathbb{P}([Tc]^{-1/2}\left\|\partial Q_{n(T)} - \partial Q_{[Tc]}\right\| > M,\ |n(T)/T - c| \le \delta)\\ &\le \mathbb{P}([Tc]^{-1/2} \max_{c-\delta\le u\le c+\delta} \left\|\partial Q_{[Tu]} - \partial Q_{[Tc]}\right\| > M)\\ &\le 2\,\mathbb{P}(\max_{u\le\delta} ||[Tc]^{-1/2}\partial Q_{[Tu]}|| > M) \to 2\,\mathbb{P}(\max_{u\le\delta} ||\Omega_S^{1/2}B(u)|| > c^{1/2}M),\end{aligned}$$

as $T \to \infty$. As δ can be arbitrarily small, it follows that $A_T = o_p(1)$.
Second order derivative: Since (C.4) implies $n(T) \to_{\text{a.s.}} \infty$, then by Gut (2009, Theorem 2.1) it holds that (C.2) implies

$$-n(T)^{-1}\partial^2 Q_{n(T)}(\varphi_0)/\partial\varphi\partial\varphi' \to_{\text{a.s.}} \Omega_I, \quad \Omega_I > 0 \tag{4.2}$$

Third order derivatives: By (C.3),

$$\max_{h,i,j=1,\ldots,k} \sup_{\varphi\in\mathcal{N}(\varphi_0)} \left| n(T)^{-1}\frac{\partial^3 Q_{n(T)}(\varphi)}{\partial\varphi_h\partial\varphi_i\partial\varphi_j}\right| \le \tau_{n(T)} \tag{4.3}$$

and hence, since $\tau_n \to_{\text{a.s.}} \tau$, by (C.4) and again using Gut (2009, Theorem 2.1), $\tau_{n(T)} \to_{\text{a.s.}} \tau$.

Establishing (i)–(iii): These hold by using (4.1)–(4.3) together with the arguments in the proof of Lemma 1 in Jensen and Rahbek (2004), replacing T there by $n(T)$, and setting $\ell_T(\varphi) = -n(T)^{-1}Q_{n(T)}(\varphi)$. Specifically, (4.1) replaces condition (A.1) in Jensen and Rahbek (2004), (4.2) replaces their condition (A.2), and (4.3) replaces their condition (A.3). □

4.2 Proof of Theorem 2.1

We verify conditions (C.1)–(C.4) in Lemma 2.1 for $Q_{n(T)} = \mathcal{L}_{n(T)}$, with $\mathcal{L}_{n(T)}$ the log-likelihood in (1.3), and $x_i = \psi_i\varepsilon_i$ in (1.1)-(1.2), with the corresponding counting process $n(t) = \max\{k : \sum_{i=1}^k x_i \le t\}$, $t \in [0,\infty)$. It is well-known that (i) and (ii) imply that (x_i, ψ_i) is strictly stationary and ergodic (and β-mixing with geometrically decaying rate); see, e.g., Theorem 4.1.9 and Corollary 4.2.8 in Buraczewski, Damek and Mikosch (2016) (henceforth, BDM) and Meitz and Saikkonen (2008). In particular, condition (C.1) holds by standard arguments (see, e.g., LH, proof of Lemma 9), and the strong LLN (see, e.g., Theorem 1 in Jensen and Rahbek, 2007) applies, implying (C.2). For (C.3), let ω_L, ω_U, α_L, α_U, β_L and β_U be strictly positive finite constants such that $\omega_L < \omega_0 < \omega_U$, $\alpha_L < \alpha_0 < \alpha_U$ and $\beta_L < \beta_0 < \beta_U < 1$, and define the closed neighborhood,

$$\mathcal{N}(\theta_0) = \{\theta : \omega_L \leq \omega \leq \omega_U,\ \alpha_L \leq \alpha \leq \alpha_U,\ \beta_L \leq \beta \leq \beta_U\}. \tag{4.4}$$

Then, (C.3) follows as

$$\max_{h,i,j=1,2,3} \sup_{\theta \in \mathcal{N}(\theta_0)} \left| n^{-1} \frac{\partial^3 \mathcal{L}_n(\theta)}{\partial\theta_h \partial\theta_i \partial\theta_j} \right| \leq \tau_n = n^{-1} \sum_{i=1}^{n} u_i \to_{\text{a.s.}} \mathbb{E}[u_i] < \infty, \tag{4.5}$$

with u_i strictly stationary and ergodic, by arguments as in Jensen and Rahbek (2004, proof of Lemma 10).

In order to establish (C.4), note that since $\sum_{i=1}^{n(T)} x_i \leq T < \sum_{i=1}^{n(T)} x_i + x_{n(T)+1}$ we have $0 \leq T/n(T) - \sum_{i=1}^{n(T)} x_i/n(T) < x_{n(T)+1}/n(T)$, where the last term tends to zero a.s. (as $T \to \infty$, and hence, $n(T) \to_{\text{a.s.}} \infty$). Hence, up to a negligible term, $T/n(T)$ equals $\sum_{i=1}^{n(T)} x_i/n(T)$, which by Gut (2009, Theorem 2.1) and the strong LLN converges a.s. to $\mu = \mathbb{E}[x_i]$, as desired.□

4.3 Proof of Theorem 3.1

If $\alpha_0 + \beta_0 > 1$, then the information is random in the limit. The main challenge is to establish that for the score and information

$$(T^{-\kappa_0/2}\mathcal{S}_{n(T)}, T^{-\kappa_0}\mathcal{I}_{n(T)}) \to_{\text{d}} (\lambda_{\kappa_0}^{1/2}\Omega_I^{1/2}Z, \lambda_{\kappa_0}\Omega_I), \tag{4.6}$$

where $Z \sim \text{N}(0, I)$ is independent of the random variable λ_{κ_0} defined in Lemma 4.1. Consistency and (3.1) then hold by an application of Kristensen and Rahbek (2010, Lemma 12), together with the uniform convergence of the information.

To establish (4.6), we apply Theorem 3.1 in Sweeting (1992) which holds under the regularity conditions D1 and D2 therein. Specifically, condition D1 holds if $T^{-\kappa_0}\mathcal{I}_{n(T)} \to_{\text{d}} W = \lambda_{\kappa_0}\Omega_I > 0$ (a.s.), under a sequence of data generating processes (DGPs) with true parameter value $\theta_T = \theta_0 + sT^{-\kappa_0/2}$, $s \in \mathbb{R}^3$. Let $\mathcal{B}(x) = \{\theta : |\theta - \theta_0| \leq x\}$ with $|\theta| = \max_{i=1,2,3} |\theta_i|$. Condition D2 holds if, for any $\delta > 0$, $\sup_{\theta \in \mathcal{B}(\delta T^{-\kappa_0/2})} T^{-\kappa_0} \parallel \mathcal{I}_{n(T)} - \mathcal{I}_{n(T)}(\theta) \parallel \to_{\text{p}} 0$, under the θ_T-sequences of DGPs.

To verify D1 in Sweeting (1992), note that by Lemma 4.1, $n(T)/T^{\kappa_0} \to_{\text{d}} \lambda_{\kappa_0}$ under θ_T-sequences, and thus D1 follows by Slutsky's Theorem provided $\mathcal{I}_{n(T)}/n(T) \to_{\text{p}} \Omega_I$ under θ_T-sequences, as $T \to \infty$. Consider $\{m_T\}$, with $m_T > 0$ and integer-valued, and $m_T \to \infty$ as $T \to \infty$. As $n(T) \to_{\text{a.s.}} \infty$ the desired result follows from Gut (2009, Theorem 2.1) if $m_T^{-1} \sum_{i=1}^{m_T} \zeta_i \to_{\text{a.s.}} \Omega_I$, under θ_T-sequences as $T \to \infty$. Here $\zeta_i = \zeta_{i,\theta_T}$, where $-\zeta_{i,\theta}$ denotes the i-th term of the second order derivative of the likelihood function evaluated at θ_0, and with the DGP generated by θ. We note

$$\left\| m_T^{-1} \sum_{i=1}^{m_T} \zeta_i - \Omega_I \right\| \leq \left\| m_T^{-1} \sum_{i=1}^{m_T} \zeta_i - \mathbb{E}[\zeta_i] \right\| + \|\mathbb{E}[\zeta_i] - \Omega_I\| = M_{1T} + M_{2T}$$

with $M_{1T} + M_{2T}$ defined implicitly. Since $\theta_T \in \mathcal{B}(\delta)$ for T large, it follows that $M_{1T} \leq \sup_{\theta \in \mathcal{B}(\delta)} \left\| m_T^{-1} \sum_{i=1}^{m_T} \zeta_{i,\theta} - \mathbb{E}[\zeta_{i,\theta}] \right\| \to_{\text{a.s.}} 0$, as $T \to \infty$, by the uniform law of large numbers in, e.g., Straumann (2005, Theorem 2.2.1). Finally, $M_{2T} \to 0$ by dominated convergence as $\theta_T \to \theta_0$ when $T \to \infty$.

To verify D2 in Sweeting (1992), note that by definition,

$$\sup_{\theta\in\mathcal{B}(\delta T^{-\kappa_0/2})} T^{-\kappa_0}\left\|\mathcal{I}_{n(T)}-\mathcal{I}_{n(T)}(\theta)\right\| = n(T)T^{-\kappa_0}\sup_{\theta\in\mathcal{B}(\delta T^{-\kappa_0/2})} n(T)^{-1}\left\|\mathcal{I}_{n(T)}-\mathcal{I}_{n(T)}(\theta)\right\|$$

with $n(T)/T^{\kappa_0}=O_p(1)$ by Lemma 4.1. Next, by the mean value theorem

$$n(T)^{-1}\left\|\mathcal{I}_{n(T)}-\mathcal{I}_{n(T)}(\theta)\right\| \le n(T)^{-1}\sum_{i=1}^{n(T)} v_{T,i}(\theta)\left|\theta-\theta_0\right|,\ v_{T,i}(\theta)=\max_{h,j,k}\left|\frac{\partial^3\ell_i(\bar\theta)}{\partial\theta_h\partial\theta_j\partial\theta_k}\right|,$$

with $\bar\theta$ between θ and θ_0. By arguments as in Jensen and Rahbek (2004, Lemmas 7 and 9), $v_{T,i}(\theta)\le v_i$, with v_i stationary, ergodic and all moments finite. Hence,

$$\sup_{\theta\in\mathcal{B}(\delta T^{-\kappa_0/2})} n(T)^{-1}\left\|\mathcal{I}_{n(T)}-\mathcal{I}_{n(T)}(\theta)\right\| \le \delta T^{-\kappa_0/2}\Big(n(T)^{-1}\sum_{i=1}^{n(T)} v_i\Big) = O_{\text{a.s.}}(T^{-\kappa_0/2}),$$

using Gut (2009, Theorem 2.1) as $n(T)\to_{\text{a.s.}}\infty$ under θ_T-sequences. □

4.4 Proof of Corollary 3.1

By the proof of Theorem 3.1, $(T^{-\kappa_0}\mathcal{I}_{n(T)},\mathcal{I}^{1/2}_{n(T)}(\hat\theta_T-\theta_0))\to_{\text{d}}(\lambda_{\kappa_0}\Omega_I,Z)$, where Z is independent of λ_{κ_0}. Moreover, $n(T)^{1/2}(\hat\theta_T-\theta_0)=\left(n(T)^{-1}\mathcal{I}_{n(T)}\right)^{-1/2}\mathcal{I}^{1/2}_{n(T)}(\hat\theta_T-\theta_0)$ where $n(T)^{-1}\mathcal{I}_{n(T)}\to_{\text{a.s.}}\Omega_I$; this implies the desired result. □

4.5 Additional lemma

Lemma 4.1 *If* $\mathbb{E}\left[\log\left(\alpha_0\varepsilon+\beta_0\right)\right]<0$ *and* $\alpha_0+\beta_0>1$,

$$n(T)/T^{\kappa_0}\xrightarrow{d}\lambda_{\kappa_0} \tag{4.7}$$

where $\lambda_{\kappa_0}^{-1/\kappa_0}$ *is an almost surely positive* κ_0*-stable random variable with parameter* $\kappa_0\in(0,1)$ *the unique positive solution to* $h(\kappa)=E\left[(\alpha_0\varepsilon_i+\beta_0)^\kappa\right]=1$, $\kappa>0$. *The result in (4.7) also holds under* θ_T*-sequences of DGPs.*

Remark 4.1 *We note that a* κ*-stable random variable is given via its characteristics function. It has right tail of the asymptotic order* $x^{-\kappa}$ *as* $x\to\infty$*; see e.g. Samorodnitsky and Taqqu (1994, Chapter 1) for more details.*

Proof of Lemma 4.1: By definition, at the true value θ_0, ψ_i in (1.2) satisfies the stochastic recurrence equation $\psi_i=A_i\psi_{i-1}+B_i$, $i\in\mathbb{Z}$, with the i.i.d. sequence $(A_i,B_i)_{i\in\mathbb{Z}}=(\alpha_0\varepsilon_{i-1}+\beta_0,\omega_0)_{i\in\mathbb{Z}}$. Since the function $h(\kappa)$, $\kappa>0$, is convex with negative right derivative at 0, $h(0)=1$, and A_1 has infinite support we have $h(\kappa)\to\infty$ as $\kappa\to\infty$, and there is a unique value $\kappa_0>0$ such that $h(\kappa_0)=1$. If $\kappa_0>1$ and $\alpha_0+\beta_0>1$ an application of Hölder's inequality leads to a contradiction:

$$\left(\mathbb{E}[(\alpha_0\varepsilon_0+\beta_0)^{\kappa_0}]\right)^{1/\kappa_0}\ge\mathbb{E}[\alpha_0\varepsilon_0+\beta_0]=\alpha_0+\beta_0>1\,.$$

Hence $\kappa_0\le1$, but $\kappa_0=1$ corresponds to the case $\mathbb{E}[\alpha_0\varepsilon_0+\beta_0]=\alpha_0+\beta_0=1$ which is excluded as well. This proves that $\kappa_0\in(0,1)$. By Theorem 2.4.4 in BDM, with $\tilde c_{\kappa_0}=\kappa_0^{-1}\mathbb{E}\left[(A_i\psi_{i-1}+B_i)^{\kappa_0}-(A_i\psi_{i-1})^{\kappa_0}\right]/(\mathbb{E}[(A_i)^{\kappa_0}\log(A_i)])$, it holds that $\mathbb{P}(\psi_i>x)\sim\tilde c_{\kappa_0}x^{-\kappa_0}$ as $x\to\infty$. Hence by Lemma B.5.1 in BDM

$$\mathbb{P}(x_i>x)=\mathbb{P}(\varepsilon_i\psi_i>x)\sim c_{\kappa_0}x^{-\kappa_0},\ c_{\kappa_0}=\mathbb{E}[\varepsilon_i^{\kappa_0}]\tilde c_{\kappa_0}, \tag{4.8}$$

as $x \to \infty$. Next, to establish (4.7) under θ_T-sequences, let $x_{i,\theta_0} = \psi_{i,\theta_0}\varepsilon_i$, $\psi_{i,\theta_0} = \omega_0 + \alpha_0 x_{i-1,\theta_0} + \beta_0\psi_{i-1,\theta_0}$. With $Y_m = m^{-1/\kappa_0}\sum_{i=1}^m x_{i,\theta_0}$ and $X_m = m^{-1/\kappa_0}\sum_{i=1}^m (x_i - x_{i,\theta_0})$, we have

$$\mathbb{P}(n(T)/T^{\kappa_0} \le z) = 1 - \mathbb{P}((zT^{\kappa_0})^{-1/\kappa_0}\sum\nolimits_{i=1}^{zT^{\kappa_0}} x_i \le z^{-1/\kappa_0}) = 1 - \mathbb{P}(Y_{(zT^{\kappa_0})} + X_{(zT^{\kappa_0})} \le z^{-1/\kappa_0}).$$

Using the tail asymptotics (4.8) and mixing of $\{x_{i,\theta_0}\}$ with geometric rate, by Theorem 9.2.1 in Mikosch and Wintenberger (2024), $Y_m \to_{\mathrm{d}} \eta_{\kappa_0}$ as $m \to \infty$, where η_{κ_0} is a positive κ_0-stable random variable. Hence (4.7) holds with $\lambda_{\kappa_0} =_{\mathrm{d}} \eta_{\kappa_0}^{-\kappa_0}$, provided $X_{(zT^{\kappa_0})} \to_{\mathrm{p}} 0$ for $T \to \infty$. This again follows by noting that, by definition, $x_i - x_{i,\theta_0} = (\psi_i - \psi_{i,\theta_0})\varepsilon_i$, such that for any $0 < \rho < \kappa_0$,

$$\mathbb{E}|X_{(zT^{\kappa_0})}|^\rho \le T^{\kappa_0-\rho} z^{1-\rho/\kappa_0}\mathbb{E}\left[|\psi_i - \psi_{i,\theta_0}|^\rho\right]\mathbb{E}\left[\varepsilon_i^\rho\right]$$

by stationarity for T large enough. With $M_T = |\omega_T - \omega_0|^\rho + (|\alpha_T - \alpha_0|^\rho\mathbb{E}[\varepsilon_i^\rho] + |\beta_T - \beta_0|^\rho)\mathbb{E}[\psi_i^\rho]$,

$$\mathbb{E}\left[|\psi_i - \psi_{i,\theta_0}|^\rho\right] \le h(\rho)\mathbb{E}\left[|\psi_i - \psi_{i,\theta_0}|^\rho\right] + M_T \le (1 - h(\rho))^{-1} M_T$$

as $h(\rho) < h(\kappa_0) = 1$. Next, $M_T = O(T^{-\rho\kappa_0/2})$ as $\theta_T - \theta_0 = s/T^{\kappa_0/2}$ and

$$\mathbb{E}[\psi_i^\rho] \le \omega_T^\rho \sum_{j=0}^{\infty}(\mathbb{E}[(\alpha_T\varepsilon_i + \beta_T)^\rho])^j = O(1),$$

since $\mathbb{E}[(\alpha_T\varepsilon_i + \beta_T)^\rho] < 1$ for T large as $\mathbb{E}[(\alpha_T\varepsilon_i + \beta_T)^\rho] \to h(\rho)$. We conclude that $\mathbb{E}[|X_{T,(zT^{\kappa_0})}|^\rho] = O(T^{\kappa_0 - \rho(1+\kappa_0/2)})$ which implies $X_{T,(zT^{\kappa})} \to_{\mathrm{p}} 0$ by picking ρ such that $\kappa_0 < \rho(1 + \kappa_0/2)$. □

References

Buraczewski, D., Damek, E. and Mikosch, T. (2016): *Stochastic Models with Power-Law Tails.* Springer, NY.

Engle, R.F. and Russell, J.R. (1998): "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 66:1127–1162.

Cavaliere, G., Mikosch, T., Rahbek, A. and Vilandt, F. (2024): "Tail behavior of ACD models and consequences for likelihood-based estimation," *Journal of Econometrics*, 238:105613.

Gut, A. (2009): *Stopped Random Walks: Limit Theorems and Applications.* Springer, NY.

Jensen, S.T. and Rahbek, A. (2004): "Asymptotic inference for nonstationary GARCH," *Econometric Theory,* 20:1203–1226.

Jensen, S.T. and Rahbek, A. (2007): "On the law of large numbers for (geometrically) ergodic Markov chains," *Econometric Theory,* 23:761–766.

Kristensen, D. and Rahbek, A. (2010): "Likelihood-based inference for cointegration with nonlinear error-correction," *Journal of Econometrics,* 158:78–94.

Lee, S. and Hansen, B. (1994): "Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator," *Econometric Theory,* 10:29–52.

Meitz, M. and Saikkonen, P. (2008): "Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models," *Econometric Theory,* 24:1291–1320.

Mikosch, T. and Resnick, S. (2006): "Activity rates with very heavy tails," *Stochastic Processes and Their Applications*, 116:131–155.

Mikosch, T. and Wintenberger, O. (2024): *Extreme Value Theory for Time Series. Models With Power-Law Tails*. Springer, NY.

Samorodnitsky, G. and Taqqu, M.S. (1994): *Stable Non-Gaussian Random Processes.* Chapman and Hall, NY.

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

Sweeting, T.J. (1992): "Asymptotic ancillarity and conditional inference for stochastic processes," *Annals of Statistics,* 20:580–589.

COMMENT ON 'AUTOREGRESSIVE CONDITIONAL DURATION: A NEW MODEL FOR IRREGULARLY SPACED TRANSACTION DATA'

Additional material (not for publication)

For the benefit of the reviewers, we provide here the details of the proof of Lemma 2.1, results *(i)–(iii)*. As noted, these results are based on a modification of the proof of Lemma 1 in Jensen and Rahbek (2004).

Note first that by definition $Q_{n(T)}(\varphi)$ is continuous in φ and hence attains its maximum in any compact neighborhood $K(\varphi_0, r) = \{\varphi | \; \|\varphi - \varphi_0\| \leq r\} \subseteq \mathcal{N}(\varphi_0)$ of φ_0. With $v_\varphi = \varphi - \varphi_0$, and φ^* on the line from φ to φ_0, Taylor's formula gives

$$Q_{n(T)}(\varphi) - Q_{n(T)}(\varphi_0) = \partial Q_{n(T)}(\varphi_0) v_\varphi - \left(-\tfrac{1}{2} v_\varphi' \partial^2 Q_{n(T)}(\varphi^*) v_\varphi\right), \tag{B.1}$$

where $\partial Q_{n(T)}(\varphi) = \partial Q_{n(T)}(\varphi)/\partial\varphi$ and $\partial^2 Q_{n(T)}(\varphi) = \partial^2 Q_{n(T)}(\varphi)/\partial\varphi\partial\varphi'$. The second term on the rhs of (B.1), normalized by $n(T)$, can be expressed as

$$v_\varphi' \left[\Omega_I + (-n(T)^{-1}\partial^2 Q_{n(T)}(\varphi_0) - \Omega_I) - n(T)^{-1}(\partial^2 Q_{n(T)}(\varphi^*) - \partial^2 Q_{n(T)}(\varphi_0))\right] v_\varphi. \tag{B.2}$$

Denote by $\rho_{n(T)}$ and ρ, $\rho > 0$, the smallest eigenvalues of $\left[\left(-n(T)^{-1}\partial^2 Q_{n(T)}(\varphi_0)\right) - \Omega_I\right]$ and Ω_I respectively. By (4.2) and the fact that the smallest eigenvalue of a $k \times k$ symmetric matrix M, $\inf_{\{v \in \mathbb{R}^k | \|v\|=1\}} v'Mv$ is continuous in M, $\rho_{n(T)} \to_{\text{a.s.}} 0$. Using (B.1), then (4.1)–(4.3) imply that

$$\begin{aligned}\sup_{\varphi: v_\varphi = r} n(T)^{-1}[Q_{n(T)}(\varphi) - Q_{n(T)}(\varphi_0)] &\leq \|n(T)^{-1}\partial Q_{n(T)}(\varphi_0)\| r - \tfrac{1}{2}\left[\rho + \rho_{n(T)} - \tilde{\tau}_{n(T)} r\right] r^2 \\ &\to_{\text{p}} -\tfrac{1}{2}\left[\rho - \tilde{\tau} r\right] r^2,\end{aligned}$$

where $\tilde{\tau}_{n(T)} = k^{3/2}\tau_{n(T)}$ and $\tilde{\tau} = k^{3/2}\tau$. Therefore, if $r < \rho/\tilde{\tau}$, the probability that $Q_{n(T)}(\varphi)$ attains its maximum on the boundary of $K(\varphi_0, r)$ tends to zero. Next, for $\varphi \in K(\varphi_0, r)$ and $v \in \mathbb{R}^k$, rewriting $v'\partial^2 Q_{n(T)}(\varphi)v$ as in (B.2), $-n(T)^{-1}v'\partial^2 Q_{n(T)}(\varphi)v \geq \|v\|^2(\rho + \rho_T - r\tilde{\tau}_{n(T)})$ which tends in probability to $\|v\|^2(\rho - r\tilde{\tau})$. Hence, if $r < \rho/\tilde{\tau}$ the probability that $Q_{n(T)}(\varphi)$ is strongly concave in the interior of $K(\varphi_0, r)$ tends to 1, and therefore it has at most one stationary point.

This establishes *(i)*: If $r < \rho/\tilde{\tau}$ and $K(\varphi, r) \subseteq \mathcal{N}(\varphi_0)$, there is with probability tending to one exactly one solution $\hat{\varphi}_T$ to the likelihood equation in the interior of $K(\varphi, r)$, $U(\varphi_0)$. It is the unique maximum point of $Q_{n(T)}(\varphi)$ in $U(\varphi_0)$ and, as it is a stationary point, it solves $\partial Q_{n(T)}(\varphi) = 0$.

To establish *(ii),* note that by the same argument, for any δ, $0 < \delta < r$ there is with a probability tending to one a solution to the likelihood equation in $K(\varphi_0, \delta)$. As $\hat{\varphi}_T$ is the unique solution to the likelihood equation in $K(\varphi_0, r)$, it must therefore be in $K(\varphi_0, \delta)$ with a probability tending to 1. Hence we have proved that $\hat{\varphi}_T$ is consistent. That is, for any $0 < \delta < r$, the probability that $\hat{\varphi}_T$, $\|\hat{\varphi}_T - \varphi_0\| \leq \delta$, is a unique solution to $\partial Q_{n(T)}(\varphi) = 0$ in $K(\varphi_0, r)$ tends to one, as desired.

Turning to *(iii)*: By the result in (4.1) and by Taylor's formula to $\partial Q_{n(T)}(\varphi)/\partial\varphi_j$

$$n(T)^{-1/2}\partial Q_{n(T)}(\varphi_0) = (n(T)/T)^{1/2}\left(\Omega_I + A_T(\hat{\varphi}_T)\right) T^{1/2}(\hat{\varphi}_T - \varphi_0), \tag{B.3}$$

for $j = 1, \ldots, k$. Here the elements in the matrix $A_T(\hat{\varphi}_T)$ are of the form $v_1'(-n(T)^{-1}\partial^2 Q_{n(T)}(\varphi_T^*) - \Omega_I)v_2$ with v_1, v_2 unit vectors in $\mathbb{R}^k$ and φ_T^* a point on the line from φ_0 to $\hat{\varphi}_T$ (where φ_T^* depends on the first vector v_1). Next, for any vectors $v_1, v_2 \in \mathbb{R}^k$, and any $\varphi \in \mathcal{B}(\varphi_0)$, using (4.3),

$$n(T)^{-1} \left| v_1' \left(\partial^2 Q_{n(T)}(\varphi) - \partial^2 Q_{n(T)}(\varphi_0) \right) v_2 \right| \leq \|v_1\| \, \|v_2\| \, \|\varphi - \varphi_0\| \, \tilde{\tau}_{n(T)}. \tag{B.4}$$

Using (B.4),

$$|v_1'(-n(T)^{-1}\partial^2 Q_{n(T)}(\varphi_T^*) - \Omega_I)v_2| \leq |v_1'(-n(T)^{-1}\partial^2 Q_{n(T)}(\varphi_0) - \Omega_I)v_2| + \|v_1\|\|v_2\|\|\varphi_T^* - \varphi_0\|\tilde{\tau}_{n(T)}.$$

Since $\varphi_T^* \to_{\mathrm{p}} \varphi_0$ and $\tilde{\tau}_{n(T)} \to_{\mathrm{p}} \tilde{\tau} < \infty$ it follows from (4.2) that the right hand side tends in probability to 0. Hence $A_T(\hat{\varphi}_T) \to_{\mathrm{p}} 0$ and *(iii)* follows by (B.3) using (4.1).