Supplementary Material on "Exploiting infinite variance through dummy variables in non-stationary autoregressions"

Giuseppe Cavaliere
Università di Bologna

Iliyan Georgiev
Universidade Nova de Lisboa

S.1 Results on weighted empirical processes

In this supplement we need to work with a slightly more general version of Proposition A.1.

As in the Appendix of the paper, let $\mathcal{F}_{Tt}$ ($T \in \mathbb{N}, 1 \leq t \leq T$) be the σ-algebra generated by $\{y_1, ..., y_t\}$ (y_0, and $\Delta \mathbf{y}_0$ of section 6, are assumed fixed constants). Consider an array $\{\eta_t, \xi_{Tt}, \psi_{Tt}, \zeta_{Tt}, \tau_{Tt}, \gamma_{Tt}\}_{t=1}^{T}$ of r.v.'s such that $\eta_t \in \{1, \varepsilon_t, \varepsilon_t^+, \varepsilon_t^-\}$ (the same choice for all t, T), and $(\xi_{Tt}, \psi_{Tt}, \zeta_{Tt}, \tau_{Tt}, \gamma_{Tt})$ is $\mathcal{F}_{t-1}$-measurable and a.s. finite for all t, T. Define $m(\theta, z) := E(\eta_1 \mathbb{I}_{\{|\varepsilon_1 - z| \leq \theta\}})$,

$$U_T(\theta) := T^{-1/2} \sum \gamma_{Tt} \left[\eta_t \mathbb{I}_{\{|\varepsilon_t - \xi_{Tt} - \psi_{Tt}| \leq \theta + \zeta_{Tt} + \tau_{Tt}\}} - m(\theta + \zeta_{Tt} + \tau_{Tt}, \xi_{Tt} + \psi_{Tt})\right],$$
$$U_T^*(\theta) := T^{-1/2} \sum \gamma_{Tt} \left[\eta_t \mathbb{I}_{\{|\varepsilon_t| \leq \theta\}} - m(\theta, 0)\right].$$

Proposition A.3 (generalized A.1) *In addition to Assumptions $\mathcal{E}$, and $\mathcal{Y}'$ or $\mathcal{Y}(k)$, let the following hold:*

$$T^{-1} \sum \gamma_{Tt}^2 \xrightarrow{w} \gamma < \infty \text{ a.s.},$$
$$T^{-1/2} \sum (|\gamma_{Tt}| + \gamma_{Tt}^2)(|\psi_{Tt}| + |\tau_{Tt}| + \psi_{Tt}^2 + \tau_{Tt}^2) = O_P(1),$$
$$\max_{1 \leq t \leq T} \{|\xi_{Tt}| + |\zeta_{Tt}| + T^{-1/2} |\gamma_{Tt}| (1 + |\psi_{Tt}| + |\tau_{Tt}|)\} = o_P(1).$$

Then the processes $\{U_T\}$ and $\{U_T^*\}$ are tight in the uniform metric of $C[0, A]$ for all $A > 0$, and

$$\sup_{\theta \in [0, A]} |U_T(\theta) - U_T^*(\theta)| = o_P(1).$$

Clearly, Proposition A.1 obtains by choosing $\zeta_{Tt} = \tau_{Tt} = 0$, all t, T.

Proof of Proposition A.3. It suffices to consider $\eta_t \in \{1, \varepsilon_t^+, \varepsilon_t^-\} \geq 0$, since for $\eta_t = \varepsilon_t$ the processes U_T and U_T^* can be decomposed additively using the identity $\varepsilon_t = \varepsilon_t^+ - \varepsilon_t^-$.

We propose a proof which mimics that of Theorem 1.1 in Koul and Ossiander (1994) [hereafter KO94] and uses their notation, whenever notation is not defined here. Specifically, we need the following adjustments. Instead of $n, i, U_n(x)$, $x \in \mathbb{R}$, ξ_{ni}, A_{na}, B_{nb}, $C_{n\eta}$ and d_b, we consider respectively $T, t, U_T(\theta)$, $\theta \in [0, A]$, $(\xi_{Tt}, \psi_{Tt}, \zeta_{Tt}, \tau_{Tt})$, $A_{Ta} := \{\max_{1 \leq t \leq T} \Gamma_{Tt} \leq T^{1/2} a\}$ with $\Gamma_{Tt} := 2|\gamma_{Tt}|(1 + A + |\xi_{Tt}| + |\zeta_{Tt}| + |\psi_{Tt}| + |\tau_{Tt}|)$, $B_{Tb} := \{\max_{1 \leq t \leq T}(|\xi_{Tt}| + |\zeta_{Tt}|) \leq b\}$, $C_{T\eta} := \{T^{-1} \sum \gamma_{Tt}^2 \leq \eta/2\}$ and

$$\rho_b(\theta_1, \theta_2) := \sup_{0 \leq |z|, w \leq b} \{|V_0(\theta_1 + w, z) - V_0(\theta_2 + w, z)|^{1/2}, |V_1(\theta_1 + w, z) - V_1(\theta_2 + w, z)|^{1/2}\}$$

for $\theta_1, \theta_2 \in [0, A]$, where $V_i(\theta, z) := E(\eta_1^{2i} \mathbb{I}_{\{|\varepsilon_t - z| \leq \theta\}})$ $(i = 0, 1;\ 0^0 := 1)$. For repeated use below, notice that for all $\theta \in [0, A]$, $i = 0, 1$ and outcomes in B_{Tb}, it holds that

$$\begin{aligned} & |V_i(\theta + \zeta_{Tt} \pm |\psi_{Tt}| \pm |\tau_{Tt}|, \xi_{Tt}) - V_i(\theta + \zeta_{Tt}, \xi_{Tt})| \\ \leq\ & (|\psi_{Tt}| + |\tau_{Tt}|)\{\|f\|_\infty + (\theta + |\xi_{Tt}| + |\zeta_{Tt}| + |\psi_{Tt}| + |\tau_{Tt}|)\|id \times f\|_\infty\} \\ \leq\ & \chi_{Tt} q \end{aligned} \tag{S.1}$$

with $\chi_{Tt} := |\psi_{Tt}| + |\tau_{Tt}| + \psi_{Tt}^2 + \tau_{Tt}^2$ and $q := \|f\|_\infty + (A + b + 2)\|id \times f\|_\infty < \infty$, and similarly,

$$|V_i(\theta + \epsilon\{|\zeta_{Tt}| + |\xi_{Tt}| + |\psi_{Tt}| + |\tau_{Tt}|\}, 0) - V_i(\theta + \epsilon\{|\zeta_{Tt}| + |\xi_{Tt}|\}, 0)| \leq \chi_{Tt} q,\ \epsilon \in \{-1, 1\},$$

where $T^{-1/2} \sum (|\gamma_{Tt}| + \gamma_{Tt}^2)\chi_{Tt} = O_P(1)$ under the hypotheses of the proposition. Introduce further the events $G_{TM} = \{\max_{1 \leq t \leq T}(|\psi_{Tt}| + |\tau_{Tt}|) \leq M_T\}$ for real M_T to be determined later. The only role of M_T will be to ensure integrability of expressions involving ψ_{Tt}, τ_{Tt} for every fixed T.

First, as in Lemma 2.3 and Corollary 2.1 of KO94, we obtain the pointwise convergence $|U_T(\theta) - U_T^*(\theta)| = o_P(1)$ for every $\theta \in [0, A]$. For $\varepsilon, a, b > 0$ with $V_1(\theta + b, 0) - V_1(\theta - b, 0) \leq a$ (feasible as $V_1(\cdot, 0)$ is continuous), the quadratic variation of $U_T(\theta) - U_T^*(\theta)$ on $A_{Ta} \cap B_{Tb} \cap G_{TM}$ (where, in particular, integrability is automatic), is

$$\begin{aligned} & T^{-1} \sum \gamma_{Tt}^2 Var\left[\eta_t(\mathbb{I}_{\{|\varepsilon_t - \xi_{Tt} - \psi_{Tt}| \leq \theta + \zeta_{Tt} + \tau_{Tt}\}} - \mathbb{I}_{\{|\varepsilon_t| \leq \theta\}}) | \mathcal{F}_{T,t-1}\right] \\ \leq\ & T^{-1} \sum \gamma_{Tt}^2 E\left[\eta_t^2(\mathbb{I}_{\{|\varepsilon_t - \xi_{Tt} - \psi_{Tt}| \leq \theta + \zeta_{Tt} + \tau_{Tt}\}} - \mathbb{I}_{\{|\varepsilon_t| \leq \theta\}})^2 | \mathcal{F}_{T,t-1}\right] \\ \leq\ & T^{-1} \sum \gamma_{Tt}^2 E\left[\eta_t^2 \mathbb{I}_{\{\theta - |\zeta_{Tt}| - |\xi_{Tt}| - |\psi_{Tt}| - |\tau_{Tt}| \leq |\varepsilon_t| \leq \theta + |\zeta_{Tt}| + |\xi_{Tt}| + |\psi_{Tt}| + |\tau_{Tt}|\}} | \mathcal{F}_{T,t-1}\right] \\ \leq\ & T^{-1} \sum \gamma_{Tt}^2 [V_1(\theta + |\zeta_{Tt}| + |\xi_{Tt}| + |\psi_{Tt}| + |\tau_{Tt}|, 0) - V_1(\theta + |\zeta_{Tt}| + |\xi_{Tt}|, 0)] \\ & + T^{-1} \sum \gamma_{Tt}^2 [V_1(\theta - |\zeta_{Tt}| - |\xi_{Tt}|, 0) - V_1(\theta - |\zeta_{Tt}| - |\xi_{Tt}| - |\psi_{Tt}| - |\tau_{Tt}|, 0)] \\ & + T^{-1} \sum \gamma_{Tt}^2 [V_1(\theta + |\zeta_{Tt}| + |\xi_{Tt}|, 0) - V_1(\theta - |\zeta_{Tt}| - |\xi_{Tt}|, 0)] \\ \leq\ & 2T^{-1} \sum \gamma_{Tt}^2 \chi_{Tt} + T^{-1} \sum \gamma_{Tt}^2 [V_1(\theta + b, 0) - V_1(\theta - b, 0)] \\ \leq\ & 2qT^{-1} \sum \gamma_{Tt}^2 \chi_{Tt} + aT^{-1} \sum \gamma_{Tt}^2. \end{aligned}$$

In this proof, denote the mean function by μ instead of m, to avoid confusion with m of the proof of KO94. Then, with $\Lambda_T := A_{Ta} \cap B_{Tb} \cap C_{T,2\eta/3} \cap G_{TM} \cap \{T^{-1} \sum \gamma_{Tt}^2 \chi_T \leq a\eta/3q\}$, by Lemma 2.2 in KO94, the above evaluation of the quadratic variation yields

$$\begin{aligned} & P\left([|U_T(\theta) - U_T^*(\theta)| > \varepsilon] \cap \Lambda_T\right) \\ & \leq P\Big(\Big[\Big|T^{-1/2} \sum \gamma_{Tt} \mathbb{I}_{\{\Gamma_{Tt} \leq T^{1/2} a\}} \mathbb{I}_{\{|\xi_{Tt}| + |\zeta_{Tt}| \leq b\}} \mathbb{I}_{\{|\psi_{Tt}| + |\tau_{Tt}| \leq M_T\}} \\ & \qquad \times \left\{\eta_t(\mathbb{I}_{\{|\varepsilon_t - \xi_{Tt} - \psi_{Tt}| \leq \theta + \zeta_{Tt} + \tau_{Tt}\}} - \mathbb{I}_{\{|\varepsilon_t| \leq \theta\}}) + \mu(\theta, 0) - \mu(\theta + \zeta_{Tt} + \tau_{Tt}, \xi_{Tt} + \psi_{Tt})\right\}\Big| > \varepsilon\Big] \\ & \qquad \cap \left[2qT^{-1} \sum \gamma_{Tt}^2 \chi_{Tt} + aT^{-1} \sum \gamma_{Tt}^2 \leq a\eta\right]\Big) \\ & \leq \exp\left(-\varepsilon^2 / 2a(\varepsilon + \eta)\right). \end{aligned}$$

So,

$$
\begin{aligned}
P\left(\left|U_T(\theta)-U_T^*(\theta)\right|>\varepsilon\right) \leq\ & \exp\left(-\varepsilon^2/2a\left(\varepsilon+\eta\right)\right) \\
& +P\left(\max_{1\leq t\leq T}(|\psi_{Tt}|+|\tau_{Tt}|)>M_T\right)+P\left(T^{-1}\sum\gamma_{Tt}^2>\eta/3\right) \\
& +P\left(A_{Ta}^c\cup B_{Tb}^c\cup\left\{T^{-1}\sum\gamma_{Tt}^2\chi_{Tt}>a\eta/3q\right\}\right).
\end{aligned}
$$

For a given arbitrary $\omega>0$, choose η and M_T such that $\lim_T P\left(T^{-1}\sum\gamma_{Tt}^2>\eta/3\right)<\omega/3$ and $\limsup_T P(\max_{1\leq t\leq T}(|\psi_{Tt}|+|\tau_{Tt}|)>M_T)<\omega/3$, and then a such that $\exp\left(-\varepsilon^2/2a\left(\varepsilon+\eta\right)\right)<\omega/3$; these choices are possible since $T^{-1}\sum\gamma_{Tt}^2\overset{w}{\to}\gamma<\infty$ a.s. and $|\psi_{Tt}|,|\tau_{Tt}|<\infty$ a.s. With these choices it holds that $\limsup_T P\left(\left|U_T(\theta)-U_T^*(\theta)\right|>\varepsilon\right)<\omega$, as the last probability in the preceding display vanishes under the hypothesis of the Proposition.

Second, concerning tightness, we note that an analogue of Proposition 2.1 of KO94 is valid: for any $T\geq 1$, $b\geq 0$, $\eta\geq 1$ and $\delta>0$, with $\delta/\left(1+N\left(\delta,b\right)\right)^{1/2}\geq 4\left(\eta/T\right)^{1/2}$,

$$
\begin{aligned}
& P\Big(\Big[\sup_{\theta\in[0,A]}|U_T(\theta)-U_T(\pi_{\delta b}(\theta))|>(c_1+\eta c_2)\left(\delta+I_{T\eta}(\delta,b)\right)\Big] \\
& \cap\left[\max_{1\leq t\leq T}|\Gamma_{Tt}|\leq T^{1/2}\delta/\left(1+N\left(\delta,b\right)\right)^{1/2}\right]\cap B_{Tb}\cap C_{T\eta}\cap G_{TM}\cap J_T\Big) \\
\leq\ & c_3e^{-\eta},
\end{aligned}
\tag{S.2}
$$

where $J_T:=\cap_{i\in\{0,1\}}\left\{T^{-1/2}\sum|\gamma_{Tt}|^{1+i}\chi_{Tt}\leq\eta_0\eta^{1+i/2}/(2q)\right\}$ and $c_{1,2,3}$ are universal constants. In KO94's proof, ζ_{ni}^{1k} and ζ_{ni}^{2k} (see KO94, p.552) are replaced respectively by

$$
\begin{aligned}
\zeta_{Tt}^{1k} &:= \mathbb{I}_{\{|\Gamma_{Tt}|\leq a_kT^{1/2}\}},\ 0\leq k\leq m, \\
\zeta_{Tt}^{2k} &:= \begin{cases}\mathbb{I}_{\{a_{k+1}T^{1/2}<|\Gamma_{Tt}|\leq a_kT^{1/2}\}}, & 0\leq k<m\\ \zeta_{Ti}^{1m}, & k=m\end{cases},
\end{aligned}
$$

where a_k and m are used according to the definition of KO94, p.552 (not to be confused with the normalization sequence a_T for the innovations, which never appears in this proof). Further, $D_{ni}\left(x\right)$ and $R_{nk}\left(x\right)$ (see KO94, p.553) are replaced respectively by

$$
\begin{aligned}
D_{Tt}\left(\theta\right) &:= \eta_t\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}-\psi_{Tt}|\leq\theta+\zeta_{Tt}+\tau_{Tt}\}}-\mu(\theta+\zeta_{Tt}+\tau_{Tt},\xi_{Tt}+\psi_{Tt}),\ 1\leq t\leq T, \\
R_{Tk}\left(\theta\right) &:= T^{-1/2}\sum|\gamma_{Tt}|\zeta_{Tt}^{2k}(\eta_t\mathbb{I}_{\{\theta_k'+\zeta_{Tt}+\tau_{Tt}<|\varepsilon_t-\xi_{Tt}-\psi_{Tt}|\leq\theta_k+\zeta_{Tt}+\tau_{Tt}\}} \\
&\quad +\mu(\theta_k+\zeta_{Tt}+\tau_{Tt},\xi_{Tt}+\psi_{Tt})-\mu(\theta_k'+\zeta_{Tt}+\tau_{Tt},\xi_{Tt}+\psi_{Tt})]),\ 0\leq k\leq m,
\end{aligned}
$$

and the evaluation of $|D_{ni}\left(x_k\right)-D_{ni}\left(x\right)|$ between (2.7) and (2.8) on p.553 of KO94, by

$$
\begin{aligned}
|D_{Tt}\left(\theta_k\right)-D_{Tt}\left(\theta\right)| \leq\ & \eta_t\mathbb{I}_{\{\theta_k'+\zeta_{Tt}+\tau_{Tt}<|\varepsilon_t-\xi_{Tt}-\psi_{Tt}|\leq\theta_k+\zeta_{Tt}+\tau_{Tt}\}} \\
& +\mu(\theta_k+\zeta_{Tt}+\tau_{Tt},\xi_{Tt}+\psi_{Tt})-\mu(\theta_k'+\zeta_{Tt}+\tau_{Tt},\xi_{Tt}+\psi_{Tt})
\end{aligned}
$$

for $\theta_k'<\theta\leq\theta_k$ and $\eta_t\geq 0$. The quadratic variation of $S_{Tk}\left(\theta\right):=T^{-1/2}\sum\gamma_{Tt}\zeta_{Tt}^{1k}[D_{Tt}(\theta_{k-1})-$

$D_{Tt}(\theta_k)]$ on the set $B_{Tb} \cap C_{T\eta} \cap G_{TM} \cap J_T$ (cf. p.554) is given by

$$\begin{aligned}
Q_{Tk}(\theta) &= T^{-1} \sum \gamma_{Tt}^2 \zeta_{Tt}^{1k} Var(\eta_t \mathbb{I}_{\{\theta_{k-1}+\zeta_{Tt}+\tau_{Tt}<|\varepsilon_t-\xi_{Tt}-\psi_{Tt}|\leq\theta_k+\zeta_{Tt}+\tau_{Tt}\}} | \mathcal{F}_{T,t-1}) \\
&\leq T^{-1} \sum \gamma_{Tt}^2 \zeta_{Tt}^{1k} [V_1(\theta_k + \zeta_{Tt} + |\tau_{Tt}| + |\psi_{Tt}|, \xi_{Tt}) - V_1(\theta_{k-1} + \zeta_{Tt} + |\tau_{Tt}| - |\psi_{Tt}|, \xi_{Tt})] \\
&= T^{-1} \sum \gamma_{Tt}^2 \zeta_{Tt}^{1k} [V_1(\theta_k + \zeta_{Tt} + |\psi_{Tt}| + |\tau_{Tt}|, \xi_{Tt}) - V_1(\theta_k + \zeta_{Tt}, \xi_{Tt})] \\
&\quad -T^{-1} \sum \gamma_{Tt}^2 \zeta_{Tt}^{1k} [V_1(\theta_{k-1} + \zeta_{Tt} - |\psi_{Tt}| - |\tau_{Tt}|, \xi_{Tt}) - V_1(\theta_{k-1} + \zeta_{Tt}, \xi_{Tt})] \\
&\quad +T^{-1} \sum \gamma_{Tt}^2 \zeta_{Tt}^{1k} [V_1(\theta_k + \zeta_{Tt}, \xi_{Tt}) - V_1(\theta_{k-1} + \zeta_{Tt}, \xi_{Tt})] \\
&\leq 2qT^{-1} \sum \gamma_{Tt}^2 \chi_{Tt} + \rho^2(\theta_k, \theta_{k-1}) T^{-1} \sum \gamma_{Tt}^2 \\
&\leq 2qT^{-1} \sum \gamma_{Tt}^2 \chi_{Tt} + 5\delta_k^2 T^{-1} \sum \gamma_{Tt}^2 \\
&\leq 2T^{-1/2} \eta_0 \eta^{3/2} + (5/2)\delta_k^2 \eta \\
&\leq 2\delta_k^2 \eta + (5/2)\delta_k^2 \eta < 5\delta_k^2 \eta,
\end{aligned}$$

the last inequality by the definition of m and δ_k (see KO94, p.552). Thus, (2.9) of KO94 remains valid upon adding $\cap G_{TM} \cap J_T$ inside the evaluated probability. $R_n(x)$ (see p.554 of KO94) is replaced by

$$R_T(\theta) := 2T^{-1/2} \sum_t \sum_{0\leq k\leq m} |\gamma_{Tt}| \zeta_{Tt}^{2k} \{\mu(\theta_k + \zeta_{Tt} + \tau_{Tt}, \xi_{Tt} + \psi_{Tt}) - \mu(\theta_k' + \zeta_{Tt} + \tau_{Tt}, \xi_{Tt} + \psi_{Tt})\}$$

with

$$\begin{aligned}
|R_T(\theta)| &\leq 2T^{-1/2} \sum_t \sum_{0\leq k\leq m} |\gamma_{Tt}| \zeta_{Tt}^{2k} \\
&\quad \times |V_0(\theta_k + \zeta_{Tt} + |\psi_{Tt}| + |\tau_{Tt}|, \xi_{Tt}) - V_0(\theta_k' + \zeta_{Tt} - |\psi_{Tt}| - |\tau_{Tt}|, \xi_{Tt})|^{1/2} \\
&\quad \times |V_1(\theta_k + \zeta_{Tt} + |\psi_{Tt}| + |\tau_{Tt}|, \xi_{Tt}) - V_1(\theta_k' + \zeta_{Tt} - |\psi_{Tt}| - |\tau_{Tt}|, \xi_{Tt})|^{1/2} \\
&\leq 2T^{-1/2} \sum_t \sum_{0\leq k\leq m} |\gamma_{Tt}| \zeta_{Tt}^{2k} [2q\chi_{Tt} + \rho^2(\theta_k, \theta) + \rho^2(\theta_k', \theta)] \\
&\leq 2T^{-1/2} \sum_t \sum_{0\leq k\leq m} |\gamma_{Tt}| \zeta_{Tt}^{2k} [2q\chi_{Tt} + 2\delta_k^2],
\end{aligned}$$

the first inequality by Cauchy-Schwartz. So, in place of (2.10) of KO94, for outcomes in

$B_{Tb} \cap C_{T\eta} \cap G_{TM} \cap J_T$ and η_k defined on p.552 of KO94, we find that

$$
\begin{aligned}
|R_T(\theta)| &\leq 2T^{-1}\sum_t \sum_{0\leq k<m} (\gamma_{Tt}^2/a_{k+1})[2q\chi_{Tt} + 2\delta_k^2] + 2T^{-1/2}\sum_t |\gamma_{Tt}|[2q\chi_{Tt} + 2\delta_m^2] \\
&\leq 2T^{-1}\sum_t \sum_{0\leq k<m} \gamma_{Tt}^2[2q\chi_{Tt}/a_{k+1} + 2^3\eta_{k+1}] + 2T^{-1/2}\sum_t |\gamma_{Tt}|[2q\chi_{Tt} + 2\delta_m^2] \\
&\leq 4qT^{-1}\sum_t \gamma_{Tt}^2\chi_{Tt}\left(\sum_{0\leq k<m} 1/a_{k+1}\right) + 2^4T^{-1}\sum_t \gamma_{Tt}^2\left(\sum_{0\leq k<m}\eta_{k+1}\right) \\
&\quad +4qT^{-1/2}\sum_t |\gamma_{Tt}|\chi_{Tt} + 4\left(\sum_t |\gamma_{Tt}|^2\right)^{1/2}\delta_m^2 \\
&\leq 2\eta\left(\sum_{0\leq k<m}\delta_m^2/a_{k+1}\right) + 2^3\eta\left(\sum_{0\leq k<m}\eta_{k+1}\right) + 2\eta_0\eta + 4T^{1/2}(\eta/2)^{1/2}\delta_m^2 \\
&\leq 10\eta\left(\sum_{0\leq k<m}\eta_{k+1}\right) + (2+2^{7/2})\eta_0\eta \leq 16\eta\left(\sum_{0\leq k\leq m}\eta_k\right),
\end{aligned}
$$

which is the same upper bound as in KO94. Finally, the quadratic variation of the conditionally centered $R_{Tk}(\theta)$ is bounded (on $B_{Tb} \cap C_{T\eta} \cap G_{TM} \cap J_T$) by

$$
\begin{aligned}
& T^{-1}\sum \gamma_{Tt}^2\zeta_{Tt}^{2k} Var(\eta_t \mathbb{I}_{\{\theta_k'+\zeta_{Tt}+\tau_{Tt}<|\varepsilon_t-\xi_{Tt}-\psi_{Tt}|\leq\theta_k+\zeta_{Tt}+\tau_{Tt}\}}|\mathcal{F}_{T,t-1}) \\
\leq\; & T^{-1}\sum \gamma_{Tt}^2\zeta_{Tt}^{2k}[V_1(\theta_k+\zeta_{Tt}+|\psi_{Tt}|+|\tau_{Tt}|,\xi_{Tt}) - V_1(\theta_k'+\zeta_{Tt}-|\psi_{Tt}|-|\tau_{Tt}|,\xi_{Tt})] \\
\leq\; & 2qT^{-1}\sum \gamma_{Tt}^2\chi_{Tt} + 2\delta_k^2 T^{-1}\sum\gamma_{Tt}^2 \\
\leq\; & T^{-1/2}\eta_0\eta^{3/2} + \delta_k^2\eta \leq 2\delta_k^2\eta,
\end{aligned}
$$

so (2.11) of KO94 remains valid upon the insertion of absolute values and $\cap G_{TM} \cap J_T$ inside the evaluated probability. By combining the above results, (S.2) follows.

Next, Assumption $\mathcal{E}$(ii) and the compactness of $[0,A]$ imply that Assumption (A5) of KO94 is satisfied automatically; cf. KO94, p.544, with $(\rho_b,[0,A])$ in place of $(d_b,\mathbb{R})$. Then, as in Corollary 2.4 of KO94, for every $\eta\geq 1$, $b\geq 0$ and $\delta>0$,

$$
\begin{aligned}
& \limsup_T P\Big(\sup_{\theta\in[0,A]}|U_T(\theta)-U_T(\pi_{\delta b}(\theta))| > (c_1+\eta c_2)\,(\delta + \int_0^\delta (1+N(u,b))^{1/2}du)\Big) \\
\leq\; & P(\gamma>\eta/2) + \limsup_T P(G_{TM}^c) + \limsup_T P(J_T^c) + c_3 e^{-\eta}
\end{aligned}
$$

As $\gamma, |\psi_{Tt}|, |\tau_{Tt}| < \infty$ a.s. and $T^{-1/2}\sum|\gamma_{Tt}|^{1+i}\chi_{Tt} = O_P(1)$, η and M_T can be chosen sufficiently large to make the above expression as small as desired. The same evaluation holds for $U_T^*(\theta)$ as in Corollary 2.5 of KO94, so the proof of the convergence $\sup_{\theta\in[0,A]}|U_T(\theta) - U_T^*(\theta)| = o_P(1)$ can be completed as the proof of (1.9) in KO94 (pp. 556-557). ■

Proof of Lemma A.2(b,c).To prove part (b), define $\iota_t(u) := \mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)|\leq\theta\}} - p_\theta(\xi_{Tt}(u))$ and, for $M>0$, also

$$r_5^M(u) := \sum y_{t-1}^2 \mathbb{I}_{\{|a_T^{-1}y_{t-1}|\leq M\}}\iota_t(u).$$

Then, for any $M, \eta_T > 0$,

$$P\left(\sup_{|u|\le A} |r_5(b_T u)| > \eta_T\right) \le P\left(\sup_{|u|\le A} |r_5^M(b_T u)| > \eta_T\right) + P\left(\max_{t\le T} |a_T^{-1} y_t| > M\right).$$

As $\max_{t\le T} |a_T^{-1} y_t| \xrightarrow{w} \max_{[0,1]} |S_c^\kappa| < \infty$ a.s., M can be chosen such that $P(\max_{t\le T} |a_T^{-1} y_t| > M)$ be as small as desired. So the sought relations for r_5 will follow once we show that they hold for r_5^M. To this aim we check, first, that for $c_T := T^{3/4}(a_T^5 b_T)^{1/2}$ and for every fixed u, $T^{1/4} c_T^{-1}\{r_5^M(b_T u) - r_5^M(0)\} = O_P(1)$, and hence, $c_T^{-1}\{r_5^M(b_T u) - r_5^M(0)\} = o_P(1)$. Second, we check that $c_T^{-1}\{r_5^M(b_T(\cdot)) - r_5^M(0)\}$ is tight. The two facts jointly imply that $c_T^{-1}\sup_{|u|\le A}\{r_5^M(b_T(\cdot)) - r_5^M(0)\} = o_P(1)$. The proof is completed by noting that $T^{-1/2} a_T^{-1} r_5^M(0) = O_P(1)$ since it equals $T^{-1/2} a_T^{-1} \sum y_{t-1}^2 \mathbb{I}_{\{|a_T^{-1} y_{t-1}| \le M\}} \left[\mathbb{I}_{\{|\varepsilon_t|\le\theta\}} - p_\theta(0)\right] + o_P(1)$ (by Proposition A.1), where the normalized summation converges weakly to an a.s. finite random variable (see Lemma 1 of Knight (1989)). The details follow.

Let $\mathcal{F}_t := \sigma(y_0, ..., y_t)$, and for $u_1, u_2 \in \mathbb{R}$, let $\Delta p_\theta(u_2, u_1) := p_\theta(\xi_{Tt}(u_1)) - p_\theta(\xi_{Tt}(u_2))$ and $\Delta F_\theta(u_2, u_1) := F(\theta + \xi_{Tt}(u_2)) - F(\theta + \xi_{Tt}(u_1))$. Conditionally on $\mathcal{F}_{t-1}$, the r.v. $\iota_t(u_2) - \iota_t(u_1)$ a.s. takes one value among $\Delta p_\theta(u_2, u_1)$ and $\Delta p_\theta(u_2, u_1) \pm 1$, so we find (considering for concreteness $u_2 y_{t-1} \ge u_1 y_{t-1}$) that

$$\begin{aligned} \mathrm{E}\left[\{\iota_t(u_2) - \iota_t(u_1)\}^2 | \mathcal{F}_{t-1}\right] &\le \{\Delta p_\theta(u_2,u_1)\}^2 + \{\Delta p_\theta(u_2,u_1) + 1\}^2 \Delta F_\theta(u_2,u_1) \\ &\quad + \{\Delta p_\theta(u_2,u_1) - 1\}^2 \Delta F_{-\theta}(u_2,u_1) \\ &\le \Delta p_\theta(u_2,u_1) + 2\{\Delta F_\theta(u_2,u_1) + \Delta F_{-\theta}(u_2,u_1)\}, \end{aligned}$$

the last inequality since $|\Delta p_\theta + \omega| \le 1$ ($\omega \in \{-1, 0, 1\}$). By the mean-value theorem (MVT),

$$\mathrm{E}\left[\{\iota_t(u_2) - \iota_t(u_1)\}^2 | \mathcal{F}_{t-1}\right] \le 6\|f\|_\infty (u_2 - u_1) y_{t-1} = 6\|f\|_\infty |u_2 - u_1||y_{t-1}|. \tag{S.3}$$

Now the fact that $T^{1/4} c_T^{-1}\{r_5^M(b_T u) - r_5^M(0)\} = O_P(1)$ follows from Chebyshev's inequality: as $\mathbb{I}_{\{|a_T^{-1} y_{t-1}|\le M\}}\{\iota_t(b_T u) - \iota_t(0)\}$ is an $\mathcal{F}_t$ martingale difference, it holds that $\mathrm{E}\{r_5^M(b_T u) - r_5^M(0)\} = 0$ and

$$\begin{aligned} \mathrm{E}\{r_5^M(b_T u) - r_5^M(0)\}^2 &= \mathrm{E}\sum \left(y_{t-1}^4 \mathbb{I}_{\{|a_T^{-1} y_{t-1}|\le M\}} \mathrm{E}\left[\{\iota_t(b_T u) - \iota_t(0)\}^2 | \mathcal{F}_{t-1}\right]\right) \\ &\le T a_T^5 b_T M^5 6\|f\|_\infty |u| = T^{-1/2} c_T^2 M^5 6 \|f\|_\infty |u|, \end{aligned}$$

the inequality from (S.3).

We turn to tightness and wish to apply a criterion in $D[-A, A]$. This is not directly possible, given that the sample paths of r_5^M are not càdlàg due to the terms $\mathbb{I}_{\{|\varepsilon_t - (\cdot) y_{t-1}|\le\theta\}}$, which are not càdlàg. If we substitute them by

$$\begin{aligned} \mathbb{I}_{\{|\varepsilon_t - (\cdot) y_{t-1}| \lhd \theta\}} &:= \mathbb{I}_{\{-\theta < \varepsilon_t - (\cdot) y_{t-1} \le \theta\}} \mathbb{I}_{\{y_{t-1} > 0\}} \\ &\quad + \mathbb{I}_{\{-\theta \le \varepsilon_t - (\cdot) y_{t-1} < \theta\}} \mathbb{I}_{\{y_{t-1} < 0\}} + \mathbb{I}_{\{|\varepsilon_t| \le \theta\}} \mathbb{I}_{\{y_{t-1} = 0\}}, \end{aligned}$$

a càdlàg modified process, say $\tilde{r}_5^M$, is obtained. The set of points at which the sample paths of r_5^M and $\tilde{r}_5^M$ differ is $\{(\theta - \varepsilon_t)/y_{t-1} : y_{t-1} > 0; t = 1, ..., T\} \cup \{-(\theta + \varepsilon_t)/y_{t-1} : y_{t-1} < 0; t =$

$1,...,T\}$. Since the distribution of ε_t is absolutely continuous, a.s. at each of these points only one indicator is affected, so

$$\sup_{|u|\le A}|r_5^M(b_T(\cdot)) - \tilde{r}_5^M(b_T(\cdot))| \le \max_{t\le T} y_t^2 = O_P(a_T^2) = o_P(c_T).$$

It is therefore enough to establish the tightness of $c_T^{-1}\{\tilde{r}_5^M(b_T(\cdot)) - \tilde{r}_5^M(0)\}$ in $D\,[-A, A]$. Since we argue in terms of expectations, which are unaffected by the change from r_5^M to $\tilde{r}_5^M$, we continue writing in terms of r_5^M.

For a fixed M and $u_2 > u_m > u_1 \ge 0$,

$$G_{u_2,u_1,u_m} := \mathrm{E}\left(\{r_5^M(b_T u_2) - r_5^M(b_T u_m)\}^2\{r_5^M(b_T u_m) - r_5^M(b_T u_1)\}^2\right) = \sum_{t_1,t_2,t_3,t_4} G_{t_1,t_2,t_3,t_4}$$

with $G_{t_1,t_2,t_3,t_4} := \mathrm{E}\prod_{i=1}^4[y_{t_i-1}^2\mathbb{I}_{Mt_i}\Delta_{t_i}^{(i)}]$, $\mathbb{I}_{Mt} := \mathbb{I}_{\{|a_T^{-1}y_{t-1}|\le M\}}$, $\Delta_t^{(1)} := \Delta_t^{(2)} := \iota_t(b_T u_2) - \iota_t(b_T u_m)$ and $\Delta_t^{(3)} := \Delta_t^{(4)} := \iota_t(b_T u_m) - \iota_t(b_T u_1)$. If only one t_i $(i = 1,...,4)$ equals $\max_{i=1,...,4} t_i$, then $G_{t_1,t_2,t_3,t_4} = 0$ by the independence of $\{\varepsilon_t\}$ and since $\mathrm{E}\Delta_t^{(i)} = 0$ $(i = 1,...,4)$. There remain at most T^3 non-zero G_{t_1,t_2,t_3,t_4}, which can be evaluated as follows, depending on how many subscripts equal $\max_{i=1,...,4} t_i$. Say, first, that $t_k = t_s = \max_{i=1,...,4} t_i \ne t_l, t_n$ and $\{k, s, l, n\} = \{1, 2, 3, 4\}$. Then, with $\beta_1 = \beta_2 = 6\,\|f\|_\infty |u_2 - u_m|$ and $\beta_3 = \beta_4 = 6\,\|f\|_\infty |u_m - u_1|$, using the Cauchy-Schwartz inequality and (S.3) we find that

$$\begin{aligned}
G_{t_1,t_2,t_3,t_4} &\le a_T^8 M^8 \mathrm{E}\left[\mathbb{I}_{Mt_l}\mathbb{I}_{Mt_n}|\Delta_{t_l}^{(l)}||\Delta_{t_n}^{(n)}|\mathbb{I}_{Mt_k}\mathbb{I}_{Mt_s}\mathrm{E}(|\Delta_{t_k}^{(k)}||\Delta_{t_s}^{(s)}||\mathcal{F}_{t_s-1})\right] \\
&\le a_T^8 M^8 \mathrm{E}\left[\mathbb{I}_{Mt_l}\mathbb{I}_{Mt_n}|\Delta_{t_l}^{(l)}||\Delta_{t_n}^{(n)}|\{\mathrm{E}(|\Delta_{t_k}^{(k)}|^2|\mathcal{F}_{t_k-1})\}^{1/2}\{\mathrm{E}(|\Delta_{t_s}^{(s)}|^2|\mathcal{F}_{t_s-1})\}^{1/2}\right] \\
&\le a_T^9 b_T M^9 \beta_k^{1/2}\beta_s^{1/2}\mathrm{E}\left[\mathbb{I}_{Mt_l}\mathbb{I}_{Mt_n}|\Delta_{t_l}^{(l)}||\Delta_{t_n}^{(n)}|\right] \\
&\le a_T^9 b_T M^9 \beta_k^{1/2}\beta_s^{1/2}[\mathrm{E}\{\mathbb{I}_{Mt_l}\mathrm{E}(|\Delta_{t_l}^{(l)}|^2|\mathcal{F}_{t_l-1})\}]^{1/2}[\mathrm{E}\{\mathbb{I}_{Mt_s}\mathrm{E}(|\Delta_{t_n}^{(n)}|^2|\mathcal{F}_{t_n-1})\}]^{1/2} \\
&\le a_T^{10} b_T^2 M^{10}\beta_k^{1/2}\beta_s^{1/2}\beta_l^{1/2}\beta_n^{1/2} \\
&= a_T^{10} b_T^2 M^{10} 36\,\|f\|_\infty^2 |u_2 - u_m||u_2 - u_m| \\
&\le a_T^{10} b_T^2 M^{10} 9\,\|f\|_\infty^2 (u_2 - u_1)^2.
\end{aligned}$$

If, second, $t_s = t_k = t_n = \max_{i=1,...,4} t_i$, we need to evaluate $\mathrm{E}(|\Delta_{t_s}^{(s)}||\Delta_{t_k}^{(k)}||\Delta_{t_n}^{(n)}|^i|\mathcal{F}_{t_s-1})$, $i = 1, 2$. Say for concreteness that $s = 1$, $k = 2$, $n = 3$ and $u_2 y_{t-1} > u_m y_{t-1} > u_1 y_{t-1}$ (the other possibilities can be considered analogously). Then $|\Delta_{t_1}^{(1)}|^2|\Delta_{t_3}^{(3)}|^i$ takes the values $|\Delta p_\theta(u_m, u_2)+1|^2|\Delta p_\theta(u_1, u_m)\pm 1|^i$ and $|\Delta p_\theta(u_m, u_2)-1|^2|\Delta p_\theta(u_1, u_m)-1|^i$ with zero probability, whereas it takes the value $|\Delta p_\theta(u_m, u_2)-1|^2|\Delta p_\theta(u_1, u_m)+1|^i$ with positive probability iff $b_T u_2 y_{t-1} - \theta > b_T u_m y_{t-1} + \theta > b_T u_m y_{t-1} - \theta > b_T u_1 y_{t-1} + \theta$, in which case this positive probability is $F(\theta + \xi_{Tt}(u_m)) - F(-\theta + \xi_{Tt}(u_m)) < F(\xi_{Tt}(u_2)) - F(\xi_{Tt}(u_1))$. Enumerating the other values taken by $|\Delta_{t_1}^{(1)}|^2|\Delta_{t_3}^{(3)}|^i$ with positive probability and recalling that $|\Delta p_\theta| \le 1$, we obtain that

$$\begin{aligned}
\mathrm{E}(|\Delta_{t_1}^{(1)}|^2|\Delta_{t_3}^{(3)}|^i|\mathcal{F}_{t_1-1}) &< \mathbb{I}_{\{b_T|u_2-u_1||y_{t_1-1}|>2\theta\}}16\{F(\xi_{Tt}(u_2)) - F(\xi_{Tt}(u_1))\} \\
&\quad + |\Delta p_\theta(u_m, u_2)|^2|\Delta p_\theta(u_1, u_m)|^i \\
&\quad + 4|\Delta p_\theta(u_m, u_2)|^2\{\Delta F_\theta(u_m, u_1) + \Delta F_{-\theta}(u_m, u_1)\} \\
&\quad + 4|\Delta p_\theta(u_1, u_m)|^i\{\Delta F_\theta(u_2, u_m) + \Delta F_{-\theta}(u_2, u_m)\}.
\end{aligned}$$

Using the MVT and $|\Delta p_\theta| \le 1$,

$$\begin{aligned} \mathrm{E}(|\Delta_{t_1}^{(1)}|^2|\Delta_{t_3}^{(3)}|^i|\mathcal{F}_{t_1-1}) &< \mathbb{I}_{\{b_T|u_2-u_1||y_{t_1-1}|>2\theta\}} b_T 16\,\|f\|_\infty |u_2-u_1||y_{t_1-1}| \\ &\quad + b_T^2 33\,\|f\|_\infty^2 |u_2-u_m||u_m-u_1||y_{t_1-1}|^2. \end{aligned}$$

From here and since $|\Delta_{t_4}^{(4)}|^{2-i} \le 4$,

$$\begin{aligned} G_{t_1,t_2,t_3,t_4} &\le a_T^8 M^8 \mathrm{E}\left[\mathbb{I}_{Mt_4}|\Delta_{t_4}^{(4)}|^{2-i}\mathbb{I}_{Mt_1}\mathrm{E}(|\Delta_{t_1}^{(1)}|^2|\Delta_{t_3}^{(3)}|^i|\mathcal{F}_{t_1-1})\right] \\ &\le a_T^9 b_T 4M^9\left[16\,\|f\|_\infty |u_2-u_1|\mathrm{E}\left(\mathbb{I}_{Mt_1}\mathbb{I}_{\{b_T|u_2-u_1||y_{t_1-1}|>2\theta\}}\right) + a_T b_T 9M\,\|f\|_\infty^2 |u_2-u_1|^2\right]. \end{aligned}$$

By Markov's inequality,

$$\mathrm{E}\left(\mathbb{I}_{Mt_1}\mathbb{I}_{\{b_T|u_2-u_1|y_{t_1-1}>2\theta\}}\right) \le a_T b_T(2\theta)^{-1}|u_2-u_1|\mathrm{E}\left[\mathbb{I}_{Mt_1}a_T^{-1}|y_{t_1-1}|\right] \le a_T b_T (2\theta)^{-1}M|u_2-u_1|,$$

so

$$G_{t_1,t_2,t_3,t_4} \le a_T^{10} b_T^2 4M^{10}\left[16\,\|f\|_\infty (2\theta)^{-1} + 9\,\|f\|_\infty^2\right](u_2-u_1)^2.$$

As this evaluation was found to hold also when precisely two among t_i $(i = 1, ..., 4)$ equal $\max_{i=1,...,4} t_i$, it holds for all non-zero G_{t_1,t_2,t_3,t_4}, so

$$G_{u_2,u_1,u_m} \le (T^3 a_T^{10} b_T^2) 4M^{10}\left[16\,\|f\|_\infty (2\theta)^{-1} + 9\,\|f\|_\infty^2\right](u_2-u_1)^2. \tag{S.4}$$

Since $c_T^{-1}\{r_5^M(b_T u) - r_5^M(0)\} = o_P(1)$ for fixed u, from (S.4) and Theorem 15.6 of Billingsley (1968) it follows that $c_T^{-1}\sup_{|u|\le A}|r_5^M(b_T u) - r_5^M(0)| = o_P(1)$. In view of the previous argument about $r_5^M(0)$ this proves part (b).

For part (c), we first derive an inequality analogous to (S.4), with $\upsilon_t(u) := \varepsilon_t \mathbb{I}_{\{|\varepsilon_t - \xi_{Tt}(u)|\le\theta\}} - m_\theta(\xi_{Tt}(u))$ instead of $\iota_t(u)$. Now

$$\begin{aligned} \mathrm{E}\left[\{\upsilon_t(u_2) - \upsilon_t(u_1)\}^2|\mathcal{F}_{t-1}\right] &\le \{m_\theta(\xi_{Tt}(u_2)) - m_\theta(\xi_{Tt}(u_1))\}^2 \\ &\quad + 3\{\theta + (|u_1|\vee|u_2| + d_T^{-1}|c|)|y_{t-1}|\}^2\left\{|\Delta F_\theta(u_2,u_1)| + |\Delta F_{-\theta}(u_2,u_1)|\right\}, \end{aligned}$$

where the first term after the inequality sign corresponds to integration over values of ε_t such that $\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u_1)|\le\theta\}} = \mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u_2)|\le\theta\}}$, and the second term – over the remaining values of ε_t. As $|m_\theta'(x)| \le 2\,\|f\|_\infty(\theta + |x|)$, from the MVT

$$\begin{aligned} &\mathrm{E}\left[\{\upsilon_t(u_2) - \upsilon_t(u_1)\}^2|\mathcal{F}_{t-1}\right] \\ \le\; &\{\theta + (|u_1|\vee|u_2| + d_T^{-1}|c|)|y_{t-1}|\}^2\{4\,\|f\|_\infty^2 |u_2-u_1|y_{t-1}^2 + 6\,\|f\|_\infty |y_{t-1}|\}|u_2-u_1|. \end{aligned}$$

Recalling that $a_T b_T = o(1)$ and $a_T = o(d_T)$, we conclude that for $|u_1|, |u_2| \le A$ and large T,

$$\mathrm{E}\left[\mathbb{I}_{\{|a_T^{-1}y_{t-1}|\le M\}}\{\upsilon_t(b_T u_2) - \upsilon_t(b_T u_1)\}^2|\mathcal{F}_{t-1}\right] \le C a_T b_T |u_2-u_1|, \tag{S.5}$$

with $C = (\theta+1)^2(1 + 6\,\|f\|_\infty M)$. Introducing

$$r_2^M(u) = \sum y_{t-1}\mathbb{I}_{\{|a_T^{-1}y_{t-1}|\le M\}}\upsilon_t(u),$$

using (S.5) and the independence of $\{\varepsilon_t\}$, it follows by an argument like for the process r_5^M that for some $L > 0$

$$\mathrm{E}\left(\{r_2^M(b_T u_2) - r_2^M(b_T u_m)\}^2\{r_2^M(b_T u_m) - r_2^M(b_T u_1)\}^2\right) \leq (T^3 a_T^6 b_T^2) L (u_2 - u_1)^2.$$

By Theorem 15.6 of Billingsley (1968), $T^{-3/4}(a_T^3 b_T)^{-1/2}\{r_2^M(b_T(\cdot)) - r_2^M(0)\}$ is tight in $D[-A, A]$ for every fixed M (more precisely, the process can be modified like r_5^M earlier so that a tight càdlàg sequence is obtained). Since $T^{-3/4}(a_T^3 b_T)^{-1/2}\{r_2^M(b_T u) - r_2^M(0)\} = o_P(1)$ for every fixed u (as $\mathrm{E}\{r_2^M(b_T u) - r_2^M(0)\} = 0$ and $\mathrm{E}\{r_2^M(b_T u) - r_2^M(0)\}^2 \leq T a_T^3 b_T M^2 C|u|$ using (S.5)), by tightness the convergence is uniform on $[-A, A]$, as asserted in part (c). ■

Proof of Proposition A.2. The assertions follow from Proposition A.1 by adapting the compactness and monotonicity arguments of Koul (2002) for his Theorem 7.2.1. We only discuss the necessary modifications.

Note first that under Assumptions $\mathcal{E}$ and $\mathcal{Y}(k)$, y_t has the decomposition $y_t = Q\sum_{i=1}^t \varepsilon_i + \sum_{i=0}^{t-1} q_i \varepsilon_{t-i} + O_P(T^{1/2})$ uniformly in $t = 1, ..., T$, where $Q = 1 - \sum_{i=1}^k \partial_i \neq 0$ and $\{q_i\}_{i=0}^\infty$ decrease exponentially, whereas $\Delta y_t = \Delta y_t^0 + c\iota_t$, where $\Delta y_t^0 = \sum_{i=0}^{t-1} \tilde{q}_i \varepsilon_{t-i}$ with exponentially decreasing $\{\tilde{q}_i\}$ and $\sum |\iota_t| = O_P(T^{1/2})$. The decomposition of y_t can be used to show that $T^{-1}\sum |a_T^{-1} y_{t-1}| \xrightarrow{w} |Q| \int |S| \in (0, \infty)$ a.s. Furthermore, we shall use the evaluations

$$T^{-1} a_T^{-1} \sum |\gamma_{Tt}||y_{t-1}| = O_P(1) \text{ and } T^{-1/2} b_T^{-1} \sum |\gamma_{Tt}|\|\Delta \mathbf{y}_{t-1}\| = o_P(1), \tag{S.6}$$

the former one since the left-hand side is bounded by $(\max_{1\leq t\leq T} |\gamma_{Tt}|)(T^{-1}\sum |a_T^{-1} y_{t-1}|) = O_P(1)$, and the latter one, since

$$T^{-1/2} b_T^{-1} \sum |\gamma_{Tt}|\|\Delta \mathbf{y}_{t-1}\| \leq \max_{1\leq t\leq T} |\gamma_{Tt}| \left(T^{-1/2} b_T^{-1} \sum \|\Delta \mathbf{y}_{t-1}\|\right) = b_T^{-1} O_P(T^{1/2} \vee T^{-1/2} a_T). \tag{S.7}$$

Here we have evaluated $\sum \|\Delta \mathbf{y}_{t-1}\| \leq \sum \|\Delta \mathbf{y}_{t-1}^0\| + O_P(T^{1/2})$ as $O_P(T)$ for distributions with $\mathrm{E}|\varepsilon_1| < \infty$ by Markov's inequality, and as $O_P(a_T)$ for $\mathrm{E}|\varepsilon_1| = \infty$, since then

$$\begin{aligned} \sum \|\Delta \mathbf{y}_{t-1}^0\| &\leq k \sum_{t=1}^T \sum_{i=0}^{t-1} |\tilde{q}_i|\,|\varepsilon_{t-i}| \\ &= k \sum_{t=1}^T \sum_{i=0}^{t-1} |\tilde{q}_i|\,|\varepsilon_{t-i}|\,\mathbb{I}_{\{|\varepsilon_{t-i}|\leq a_T\}} + k \max_{1\leq t\leq T} |\varepsilon_t| \sum_{t=1}^T \sum_{i=0}^{t-1} |\tilde{q}_i|\,\mathbb{I}_{\{|\varepsilon_{t-i}|> a_T\}}. \end{aligned}$$

with $\mathrm{E}(\sum_{t=1}^T \sum_{i=0}^{t-1} |\tilde{q}_i|\,|\varepsilon_{t-i}|\,\mathbb{I}_{\{|\varepsilon_{t-i}|\leq a_T\}}) \leq T\mathrm{E}(|\varepsilon_1|\mathbb{I}_{\{|\varepsilon_1|\leq a_T\}}) \sum_{i=0}^\infty |\tilde{q}_i| = O(a_T)$ by Karamata's theorem, $\max_{1\leq t\leq T} |\varepsilon_t| = O_P(a_T)$ and $\mathrm{E}(\sum_{t=1}^T \sum_{i=0}^{t-1} |\tilde{q}_i|\,\mathbb{I}_{\{|\varepsilon_{t-i}|> a_T\}}) \leq TP(|\varepsilon_1| > a_T) \sum_{i=0}^\infty |\tilde{q}_i| = O(1)$.

For notational ease, redefine $\xi_{Tt}(u) := T^{-1/2} a_T^{-1}(u + c) y_{t-1}$ and $\psi_{Tt}(s) := b_T^{-1} s' \Delta \mathbf{y}_{t-1}$, such that the statements to prove become

$$\sup_{(u,s,\theta)\in K} |U_T(u, s, \theta) - U_T(0, 0, \theta)| = o_P(1)$$

with $K := \{(u, s, \theta) : |u| \leq C, \|s\| \leq C, \theta \in [0, A]\}$, and

$$\sup_{(u,s,\theta)\in K} \left|\sum \gamma_{Tt} \left\{\eta_t[\mathbb{I}_{\{|\varepsilon_t - \xi_{Tt}(u) - \psi_{Tt}(s)|\leq\theta\}} - \mathbb{I}_{\{|\varepsilon_t - T^{-1/2} a_T^{-1} c y_{t-1}|\leq\theta\}}] - 2u\theta f(\theta)\, a_T^{-1} T^{-1/2} y_{t-1}\right\}\right| = o_P(T^{1/2}).$$

Let

$$J_T^{\pm,\pm}(u,s,\theta) := T^{-1/2}\sum \gamma_{Tt}^{\pm}\eta_t^{\pm}\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)-\psi_{Tt}(s)|\le\theta\}},$$

where the + and − superscripts of γ_{Tt} and η_t are taken independently. In place of the processes defined by Koul, p.299, we need

$$T^{\pm,\pm}(\theta,u,s,a) := T^{-1/2}\sum \gamma_{Tt}^{\pm}\eta_t^{\pm}\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)-\psi_{Tt}(s)|\le\theta+T^{-1/2}a_T^{-1}a|y_{t-1}|+b_T^{-1}a\|\Delta\mathbf{y}_{t-1}\|\}},$$

$$m^{\pm,\pm}(\theta,u,s,a) := T^{-1/2}\sum \gamma_{Tt}^{\pm}m\left(\theta+T^{-1/2}a_T^{-1}a|y_{t-1}|+b_T^{-1}a\|\Delta\mathbf{y}_{t-1}\|,\xi_{Tt}(u)+\psi_{Tt}(s)\right)$$

and $Z^{\pm,\pm} := T^{\pm,\pm} - m^{\pm,\pm}$, where m is the conditional expectation function matching the choice of η_t^+ or η_t^-. For $|u-v|\le\delta$ and $\|s-r\|\le\delta$ (with $\delta>0$ to be chosen later) the inequalities

$$\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)-\psi_{Tt}(s)|\le\theta-T^{-1/2}a_T^{-1}\delta|y_{t-1}|-b_T^{-1}a\|\Delta\mathbf{y}_{t-1}\|\}} \le \mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(v)-\psi_{Tt}(r)|\le\theta\}}$$
$$\le \mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)-\psi_{Tt}(s)|\le\theta+T^{-1/2}a_T^{-1}\delta|y_{t-1}|+b_T^{-1}a\|\Delta\mathbf{y}_{t-1}\|\}}$$

hold. So

$$\begin{aligned} T^{\pm,\pm}(\theta,u,s,-\delta)-T^{\pm,\pm}(\theta,u,s,0) &\le J_T^{\pm,\pm}(v,r,\theta)-J_T^{\pm,\pm}(u,s,\theta)\\ &\le T^{\pm,\pm}(\theta,u,s,\delta)-T^{\pm,\pm}(\theta,u,s,0)\end{aligned}$$

and,

$$\begin{aligned} |J_T^{\pm,\pm}(v,r,\theta)-J_T^{\pm,\pm}(u,s,\theta)| \le\ & |Z^{\pm,\pm}(\theta,u,s,\delta)-Z^{\pm,\pm}(\theta,0,0,0)| \\ &+|Z^{\pm,\pm}(\theta,u,s,-\delta)-Z^{\pm,\pm}(\theta,0,0,0)|\\ &+2|Z^{\pm,\pm}(\theta,u,s,0)-Z^{\pm,\pm}(\theta,0,0,0)|\\ &+|m^{\pm,\pm}(\theta,u,s,\delta)-m^{\pm,\pm}(\theta,u,s,0)|\\ &+|m^{\pm,\pm}(\theta,u,s,-\delta)-m^{\pm,\pm}(\theta,u,s,0)|.\end{aligned} \tag{S.8}$$

By the MVT, for every (u,s),

$$\sup_{\theta\in[0,A]}|m^{\pm,\pm}(\theta,u,s,\pm\delta)-m^{\pm,\pm}(\theta,u,s,0)| \le 2\delta G T^{-1/2}\sum\gamma_{Tt}^{\pm}(T^{-1/2}a_T^{-1}|y_{t-1}|+b_T^{-1}\|\Delta\mathbf{y}_{t-1}\|),$$

$$\sup_{\theta\in[0,A];|u-v|,\|s-r\|\le\delta}|T^{-1/2}\sum\gamma_{Tt}\left[m(\theta,\xi_{Tt}(u)+\psi_{Tt}(s))-m(\theta,\xi_{Tt}(v)+\psi_{Tt}(r))\right]|$$
$$\le 2\delta G T^{-1/2}\sum|\gamma_{Tt}|(T^{-1/2}a_T^{-1}|y_{t-1}|+b_T^{-1}\|\Delta\mathbf{y}_{t-1}\|),$$

where $G := \sup_{x\in\mathbb{R}}|\max\{x,1\}f(x)|<\infty$ by the continuity of f and by hypothesis, and

$$\begin{aligned} 0 &\le T^{-1}a_T^{-1}\sum\gamma_{Tt}^{\pm}|y_{t-1}| \le T^{-1}a_T^{-1}\sum|\gamma_{Tt}||y_{t-1}| = O_P(1),\\ 0 &\le T^{-1/2}b_T^{-1}\sum\gamma_{Tt}^{\pm}\|\Delta\mathbf{y}_{t-1}\| \le T^{-1/2}b_T^{-1}\sum|\gamma_{Tt}|\|\Delta\mathbf{y}_{t-1}\| = o_P(1)\end{aligned}$$

by (S.6). For arbitrary fixed $\epsilon>0$ this allows us to choose $\delta>0$ so that, for every u and s,

$$\limsup_{T\to\infty} P\left(\sup_{\theta\in[0,A]}|m^{\pm,\pm}(\theta,u,s,\pm\delta)-m^{\pm,\pm}(\theta,u,s,0)|\ge\epsilon\right)<\epsilon, \tag{S.9}$$

$$\limsup_{T\to\infty} P\Big(\sup_{\substack{\theta\in[0,A]\\|u-v|\le\delta\\\|s-r\|\le\delta}}|T^{-1/2}\sum\gamma_{Tt}\left[m(\theta,\xi_{Tt}(u)+\psi_{Tt}(s))-m(\theta,\xi_{Tt}(v)+\psi_{Tt}(r))\right]|\ge\epsilon\Big)<\epsilon \tag{S.10}$$

For this chosen δ and every (u,s), by Proposition A.3,

$$\sup_{\theta\in[0,A]}|Z^{\pm,\pm}(\theta,u,s,\omega\delta)-Z^{\pm,\pm}(\theta,0,0)|=o_P(1),\ \omega\in\{-1,0,1\}. \tag{S.11}$$

In fact, we can check the hypothesis of Proposition A.3 with $\zeta_{Tt}:=T^{-1/2}a_T^{-1}\omega\delta|y_{t-1}|$ and $\tau_{Tt}:=b_T^{-1}\omega\delta\|\Delta\mathbf{y}_{t-1}\|$. It holds every (u,s) that $\max_{1\le t\le T}|\zeta_{Tt}|$ and $\max_{1\le t\le T}|\xi_{Tt}(u)|$ are $O_P\left(T^{-1/2}\right)$, $\max_{1\le t\le T}|\tau_{Tt}|$ and $\max_{1\le t\le T}|\psi_{Tt}(s)|$ are $O_P(b_T^{-1}a_T)=o_P(T^{1/2})$, and

$$\begin{aligned}
&\sum(|\gamma_{Tt}|+\gamma_{Tt}^2)(|\psi_{Tt}(s)|+|\tau_{Tt}|+\psi_{Tt}^2(s)+\tau_{Tt}^2)\\
\le\ & 2(C+\delta+C^2+\delta^2)(1+\max_{1\le t\le T}\gamma_{Tt}^2)\sum(b_T^{-1}\|\Delta\mathbf{y}_{t-1}\|+b_T^{-2}\|\Delta\mathbf{y}_{t-1}\|^2)\\
=\ & O_P(1)\sum(b_T^{-1}\|\Delta\mathbf{y}_{t-1}\|+b_T^{-2}\|\Delta\mathbf{y}_{t-1}\|^2)\\
=\ & O_P(b_T^{-1}T\vee b_T^{-1}a_T)+O_P(b_T^{-2}a_T^2)=o_P(T^{1/2}),
\end{aligned}$$

since $b_T^{-1}T^{1/2}=o\,(1)$ and $b_T^{-1}a_T=o(T^{1/4})$. Here we have used the evaluation of $\sum\|\Delta\mathbf{y}_{t-1}\|$ after (S.7) and the evaluation $\sum\|\Delta\mathbf{y}_{t-1}\|^2\le\sum\|\Delta\mathbf{y}_{t-1}^0\|^2+k(\sum|\iota_t|)^2\le k\sum(\Delta y_{t-1}^0)^2+O_P(T)=O_P(a_T^2)$ since $a_T^{-2}\sum(\Delta y_{t-1}^0)^2\xrightarrow{w}(\sum_{i=0}^\infty\tilde{q}_i^2)[S]_1$.

From (S.8), (S.9) and (S.11) it follows that for every (u,s),

$$\limsup_{T\to\infty}P\left(\sup_{\theta\in[0,A];|u-v|,\|s-r\|\le\delta}|J_T^{\pm,\pm}(v,r,\theta)-J_T^{\pm,\pm}(u,s,\theta)|\ge 6\epsilon\right)<\epsilon,$$

which jointly with (S.10) gives

$$\limsup_{T\to\infty}P\left(\sup_{\theta\in[0,A];|u-v|,\|s-r\|\le\delta}|U_T(v,r,\theta)-U_T(u,s,\theta)|\ge 25\epsilon\right)<\epsilon.$$

As, for every (u,s), $\sup_{\theta\in[0,A]}|U_T(u,s,\theta)-U_T(0,0,\theta)|=o_P(1)$ by Proposition A.1 and the triangle inequality with $U_T^*(\theta)$ as third point, relation (A.9) follows via a compactness argument.

To obtain from (A.9) the second convergence in the proposition, it suffices to show that

$$\lambda_T:=\sup_{(u,s,\theta)\in K}\left|\sum\gamma_{Tt}\left\{[m(\theta,\xi_{Tt}(u)+\psi_{Tt}(s))-m(\theta,\xi_{Tt}(0))]-2u\theta f\,(\theta)\,T^{-1/2}a_T^{-1}y_{t-1}\right\}\right|$$

is $o_P(T^{1/2})$ with m calculated for $\eta_t=\varepsilon_t$. With this choice it holds that $m_2'(\theta,0)=2\theta f\,(\theta)$, where m_2' is the derivative of m w.r.t. its second argument. Thus, using the MVT, it is seen that $\lambda_T\le\lambda_T^{(1)}+\lambda_T^{(2)}$ with

$$\begin{aligned}
\lambda_T^{(1)}&:=T^{-1/2}a_T^{-1}C\sup_{(u,s,\theta)\in K}\left|\sum\gamma_{Tt}y_{t-1}[m_2'(\theta,\tau_{Tt}(u,s))-m_2'(\theta,0)]\right|,\\
\lambda_T^{(2)}&:=b_T^{-1}\sup_{(u,s,\theta)\in K}\left|\sum s'(\gamma_{Tt}\Delta\mathbf{y}_{t-1})m_2'(\theta,\tau_{Tt}(u,s))\right|
\end{aligned}$$

and $\tau_{Tt}(u,s):=T^{-1/2}a_T^{-1}cy_{t-1}+\omega_{Tt}\{T^{-1/2}a_T^{-1}uy_{t-1}+\psi_{Tt}(s)\}$ for appropriate $\omega_{Tt}\in[0,1]$.

Let $\varepsilon > 0$ be given. Let

$$A_{TY} := \{T^{-1}a_T^{-1}\sum|\gamma_{Tt}||y_{t-1}| > Y\}, \quad C_{T\Gamma} := \{T^{-1}\sum\gamma_{Tt}^2 > \Gamma\},$$
$$D_{TN} := \{\max_{1\le t\le T}|a_T^{-1}y_{t-1}| > N\}.$$

Choose and fix $Y, \Gamma, N > 0$ such that $\limsup_T P(A_{TY}) \le \varepsilon/3$, $\lim_T P(C_{T\Gamma}) \le \varepsilon/3$ and $\limsup_T P(D_{TN}) \le \varepsilon/3$. Finally, let $b > 0$ be such that $|m_2'(\theta,x) - m_2'(\theta,0)| \le \tilde{\varepsilon} := \varepsilon/(2CY)$ for every $\theta \in [0,A]$ and $|x| \le b$ (b exists by the uniform continuity of f on compacts). Then on the event $\{\max_{1\le t\le T}|T^{-1/2}a_T^{-1}(|c|+C)y_{t-1}| \le b/2\} \cap A_{TY}^c \cap C_{T\Gamma}^c \cap D_{TN}^c$ it holds that

$$\begin{aligned}
\lambda_T^1 &\le T^{-1/2}a_T^{-1}C\sum\left[\tilde{\varepsilon}\mathbb{I}_{\{b_T^{-1}C\|\Delta\mathbf{y}_{t-1}\|\le b/2\}} + 4G\mathbb{I}_{\{b_T^{-1}C\|\Delta\mathbf{y}_{t-1}\|>b/2\}}\right]|\gamma_{Tt}||y_{t-1}| \\
&\le C\left[\tilde{\varepsilon}\left(T^{-1/2}a_T^{-1}\sum|\gamma_{Tt}||y_{t-1}|\right)\right. \\
&\quad \left. + 4G\max_{1\le t\le T}|a_T^{-1}y_{t-1}|\left(T^{-1}\sum\gamma_{Tt}^2\right)^{1/2}\left(\sum\mathbb{I}_{\{b_T^{-1}C\|\Delta\mathbf{y}_{t-1}\|>b/2\}}\right)^{1/2}\right] \\
&\le T^{1/2}\varepsilon/2 + 4CGN\Gamma^{1/2}\left(\sum\mathbb{I}_{\{b_T^{-1}C\|\Delta\mathbf{y}_{t-1}\|>b/2\}}\right)^{1/2},
\end{aligned}$$

since $\sup_{\mathbb{R}^2}|m_2'| \le 2\sup_{\mathbb{R}}|id \times f| \le 2G$, so

$$\begin{aligned}
P\left(T^{-1/2}\lambda_T^{(1)} > \varepsilon\right) &\le P(A_{TY}) + P(C_{T\Gamma}) + P(D_{TN}) \\
&\quad + P\left(\max_{1\le t\le T}|T^{-1/2}a_T^{-1}(|c|+C)y_{t-1}| > b/2\right) \\
&\quad + P\left(\sum\mathbb{I}_{\{b_T^{-1}C\|\Delta\mathbf{y}_{t-1}\|>b/2\}} > \frac{T}{64\Gamma}\left(\frac{\varepsilon}{CGN}\right)^2\right).
\end{aligned} \tag{S.12}$$

The latter two probabilities tend to zero, the second one since $\sum\mathbb{I}_{\{b_T^{-1}C\|\Delta\mathbf{y}_{t-1}\|>b/2\}} = o_P(T)$. Indeed, on the one hand,

$$\begin{aligned}
\sum\mathbb{I}_{\{b_T^{-1}C\|\Delta\mathbf{y}_{t-1}\|>b/2\}} &\le k\sum\mathbb{I}_{\{b_T^{-1}C|\Delta y_t^0|>b/(4k)\}} + k\sum\mathbb{I}_{\{b_T^{-1}C|\iota_t|>b/(4k)\}} \\
&= k\sum\mathbb{I}_{\{b_T^{-1}C|\Delta y_t^0|>b/(4k)\}}
\end{aligned}$$

with probability approaching one, since

$$P\left(\sum\mathbb{I}_{\{b_T^{-1}C|\iota_t|>b/(4k)\}} > 0\right) = P\left(\max_{1\le t\le T}b_T^{-1}C|\iota_t| > b/(4k)\right) \le P\left(\sum|\iota_t| > b_Tb/(4Ck)\right) \to 0$$

for $\sum|\iota_t| = O_P(T^{1/2})$ and $T^{-1/2}b_T \to \infty$, and on the other hand, $\sum\mathbb{I}_{\{b_T^{-1}C|\Delta y_t^0|>b/(4k)\}} = o_P(T)$ by Markov's inequality:

$$\begin{aligned}
\sum P(b_T^{-1}C|\Delta y_t^0| > b/(4k)) &\le TP\left(b_T^{-1}C\sum_{i=0}^{\infty}|\tilde{q}_i||\varepsilon_{t+i}| > b/4\right) \\
&= O(Tl_Tb_T^{-\alpha}) = O(Tl_Ta_T^{-\alpha}(a_T/b_T)^{\alpha}) = O(\tilde{l}_TT^{\alpha/2}) = o(T),
\end{aligned}$$

where l_T and $\tilde{l}_T$ are slowly varying, since the distribution function of $\sum_{i=0}^{\infty}|\tilde{q}_i||\varepsilon_{t+i}|$ is regularly varying with tail index α inherited from $\{\varepsilon_t\}$. Therefore, returning to (S.12), $\limsup_T P(T^{-1/2}\lambda_T^{(1)} > \varepsilon) \le \varepsilon$, and by the arbitrariness of ε, $T^{-1/2}\lambda_T^{(1)} = o_P(1)$.

Regarding $\lambda_T^{(2)}$, it satisfies the inequality

$$\lambda_T^{(2)} \leq b_T^{-1} 2CG \max_{1\leq t\leq T} |\gamma_{Tt}| \sum \|\Delta \mathbf{y}_{t-1}\| = O_P(b_T^{-1}T \vee b_T^{-1}a_T) = o_P(T^{1/2}),$$

see (S.7). Thus, also $T^{-1/2}\lambda_T^{(2)} = o_P(1)$ and $T^{-1/2}\lambda_T = o_P(1)$. This completes the proof of the proposition. ■

S.2 A counter-exemplary result

Proof of Proposition 6. We use (A.4) and the notation introduced in the proof of Theorem 4. Fix an $A > 0$ and let $M_1 := \sup_{|u|\leq A} \left| I_{\{u\neq 0\}} \{(n_T u)^{-1} \phi_2 (n_T u) - h_\theta\} \right| = o_P(1)$ and $M_2 := \sup_{|u|\leq A} |\tilde{\phi}(0,\theta)(\phi_1(n_T u) - 1) + \phi_3(n_T u)| = o_P(\delta_T^{-1})$ by the proof of Proposition 2(a). By the same proposition, the event $\mathcal{A} := \{|n_T^{-1}\hat{\phi}^{(i)}| < A, \forall i \in \mathbb{N}\cup\{0\}\}$ is contained in

$$\mathcal{A}_1 := \{\hat{\phi}^{(i)} = \hat{\phi}^{(i-1)}(h_\theta + R_1^{(i)}) + \tilde{\phi}(0,\theta) + R_2^{(i)}, \forall i \in \mathbb{N}\} \cap \{\sup_{i\in\mathbb{N}} |R_j^{(i)}| \leq M_j,\ j = 1, 2\}.$$

Let $\hat{\Delta}_T := \hat{\phi}^{(0)} - (1-h_\theta)^{-1}\tilde{\phi}(0,\theta)$; then in the decomposition $\hat{\phi}^{(i)} = \hat{\phi}_1^{(i)} + \hat{\phi}_2^{(i)}$ we have

$$\begin{aligned} \hat{\phi}_1^{(i)} &= \hat{\Delta}_T \prod_{j=1}^{i} (h_\theta + R_1^{(i)}) + \tilde{\phi}(0,\theta)\left[(1-h_\theta)^{-1} + \lambda_T^{(i)}\right] \text{ with} \qquad \text{(S.13)} \\ \lambda_T^{(i)} &:= \sum_{j=1}^{i} \Big[\prod_{k=j+1}^{i} (h_\theta + R_1^{(k)}) - h_\theta^{i-j} \Big] + (1-h_\theta)^{-1} \Big[\prod_{j=1}^{i} (h_\theta + R_1^{(i)}) - h_\theta^{i} \Big]. \end{aligned}$$

With $\mathcal{A}_2 := \left\{ |M_1| \leq 2^{-1}(h_\theta - 1) \right\}$, for outcomes in $\mathcal{A}_1 \cap \mathcal{A}_2$ it holds that

$$\left| \prod_{j=1}^{i} (h_\theta + R_1^{(i)}) \right| \geq \left(\frac{h_\theta + 1}{2} \right)^i \quad \text{and} \quad \sum_{j=1}^{i} \prod_{k=j+1}^{i} |h_\theta + R_1^{(k)}| < \frac{(3h_\theta - 1)^i}{2^i (h_\theta - 1)}. \qquad \text{(S.14)}$$

Define

$$I_T := \min\left\{ i \in \mathbb{N} : n_T^{-1}|\hat{\Delta}_T| 2^{-i}(h_\theta+1)^i + n_T^{-1}|\tilde{\phi}(0,\theta)|(1-h_\theta)^{-1} > 3A \right\}.$$

Then $I_T = O_P(1)$ because $n_T^{-1}\tilde{\phi}(0,\theta) = O_P(1)$ by Proposition 1(a), $n_T^{-1}\hat{\Delta}_T$ is bounded away from zero in probability under the hypotheses of parts (a) and (b), and $h_\theta > 1$. Further, for outcomes in $\mathcal{A}_1 \cap \mathcal{A}_2$

$$\lambda_T^{(i)} \leq \sum_{j=1}^{i} [(h_\theta + M_1)^{i-j} - h_\theta^{i-j}] + (1-h_\theta)^{-1}[(h_\theta - M_1)^i - h_\theta^i]$$

and a similar evaluation of $\lambda_T^{(i)}$ from below holds, so that

$$\frac{(h_\theta - M_1)^i - 1}{h_\theta - M_1 - 1} - \frac{(h_\theta + M_1)^i - 1}{h_\theta - 1} \leq \lambda_T^{(i)} \leq \frac{(h_\theta + M_1)^i - 1}{h_\theta + M_1 - 1} - \frac{(h_\theta - M_1)^i - 1}{h_\theta - 1}.$$

As $I_T = O_P(1)$ and $M_1 = o_P(1)$, it follows that $\lambda_T^{(I_T)} = o_P(1)$, and since $n_T^{-1}\tilde{\phi}(0,\theta) = O_P(1)$, $P(\mathcal{A}_3) \to 1$ for $\mathcal{A}_3 := \{|n_T^{-1}\tilde{\phi}(0,\theta)\hat{\lambda}_T^{(I_T)}| < A\}$. Recalling eq. (**??**) and the definition of I_T, we can conclude that for outcomes in $\cap_{i=1}^{3}\mathcal{A}_i$, $|n_T^{-1}\hat{\phi}_1^{(I_T)}| > 2A$. Recalling also (**??**), for such outcomes $|n_T^{-1}\hat{\phi}_2^{(I_T)}| \leq n_T^{-1}M_2\{2^{-1}(3h_\theta - 1)\}^{I_T}(h_\theta - 1)^{-1} =: \kappa_T$, and since $P(\mathcal{A}_4) \to 1$ for $\mathcal{A}_4 := \{\kappa_T < A\}$, we finally obtain that, for outcomes in $\cap_{i=1}^{4}\mathcal{A}_i$, $|n_T^{-1}\hat{\phi}^{(I_T)}| \geq |n_T^{-1}\hat{\phi}_1^{(I_T)}| - |n_T^{-1}\hat{\phi}_2^{(I_T)}| > A$. Therefore, $\mathcal{A} \subset \left(\cap_{i=1}^{4}\mathcal{A}_i\right)^c$ (with c denoting complement). Recalling that $\mathcal{A} \subset \mathcal{A}_1$, we find that $\mathcal{A} \subset \left(\cap_{i=2}^{4}\mathcal{A}_i\right)^c$, where $P\left(\cap_{i=2}^{4}\mathcal{A}_i\right) \to 1$. This implies that $P(\mathcal{A}) \to 0$, which jointly with $n_T^{-1}\hat{\phi}^{(0)} = O_P(1)$ (by hypothesis) and

$$P\left(|n_T^{-1}\hat{\phi}^{(i)}| < A, \forall i \in \mathbb{N}\right) \leq P(\mathcal{A}) + P(|n_T^{-1}\hat{\phi}^{(0)}| \geq A)$$

yields $\lim_{A\to\infty}\limsup_{T\to\infty} P(|n_T^{-1}\hat{\phi}^{(i)}| < A, \forall i \in \mathbb{N}) = 0$. As $\limsup_{T\to\infty} P(|n_T^{-1}\hat{\phi}^{(i)}| < A, \forall i \in \mathbb{N})$ is increasing in A, it follows that $\lim_{T\to\infty} P(|n_T^{-1}\hat{\phi}^{(i)}| < A, \forall i \in \mathbb{N}) = 0$ for all $A > 0$. ■

S.3 Further simulations

The focus in the paper was on iterative estimation of the AR parameter ϕ, resulting in efficiency gains and associated local power gains of UR tests. To evaluate whether iterative estimation of the threshold enhances, diminishes or does not affect these gains, we carry out a further simulation exercise.

The same data generation processes as in section 7 are used. Two groups of experiments are run, with threshold equal respectively to the 75th percentile of the set of residual absolute values, and to a self-normalized residual standard deviation as in (5.19). In experiments with a fixed threshold, the threshold $\hat{\theta}$ is computed from the residuals associated with the preliminary estimator of ϕ, and is not updated in the iteration over ϕ. This gives rise to the statistics $\xi_T^{(N)}(0) := \xi_T^{(N)}(0, \hat{\theta})$ and $\xi_T^{(\sqrt{T})}(\phi_{LS}) := \xi_T^{(\sqrt{T})}(\phi_{LS}, \hat{\theta})$ respectively for the zero and the OLS preliminary estimator of ϕ. In experiments with joint iteration over ϕ and the threshold, the threshold is updated from each new set of residuals; see section 5. As in Table 1, three preliminary estimators of ϕ for the joint iteration are considered - 0 and the two estimators obtained by fixed-threshold iteration. Empirical size and power for all the experiments are reported in Table S1.

The following regularities are observed.

First, for experiments with threshold equal to the self-normalized residual standard deviation, iteration of the threshold does not appear to affect size and power.

Second, for the quantile threshold, empirical size tends to be smaller when the threshold is iterated, thus counteracting the frequent slight oversize of fixed-threshold tests, and sometimes transforming it into a slight undersize. Also the empirical rejection frequencies under the local alternative are somewhat smaller. However, size-adjusted power (not reported) is virtually independent of whether the thershold is iterated or not.

An exception occurs for $\alpha = \frac{1}{2}$ (representative for $\alpha < 1$) when estimation is initialized with the OLS estimator. Its convergence rate is too slow compared to the magnitude order of $\{y_{t-1}\}$, resulting in large residuals. In particular, the 75th residual percentile is unbounded in probability. Nevertheless, the Gaussian approximation seems appropriate under the null. Under the local alternative, fixing the threshold at the OLS residual quantile yields rejection frequencies decreasing in T, whereas iterating the threshold gives rise to the same rejection frequencies as under the assumptions of our theory. This interesting outcome does not occur if the usual residual standard deviation (normalized by T) is used as a threshold, although for the OLS residuals it has the same stochastic magnitude order as the quantile. Rather, power decreasing in T is observed both with and without iteration of the threshold. Likewise, this outcome does not occur if a stochastically bounded threshold is used (see Table 1 and the block in Table S1 for the self-normalized standard deviation). Thus, the quantile's robustness is preferable to both the insensitivity of stochastically bounded thresholds and the high sensitivity of the usual standard deviation. Iterations over ϕ improve its magnitude order gradually, which is appropriately captured by the quantile and is enhanced in further iterations. We do not explore this issue further.

Summarizing, for the empirically most relevant case $\alpha > 1$ iterations of the threshold seem to make little practical difference.

A secondary conclusion from the simulation is that among the two thresholds, the self-normalized residual standard deviation and the residual quantile, the quantile appears to have a slight advantage in terms of both size and power.

References

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: Wiley.

Koul, H. & M. Ossiander (1994) Weak convergence of randomly weighted dependent residual empirical processes with applications to autoregression. *Annals of Statistics* 22, 540-562.

Koul, H. (2002) *Weighted Empirical Processes in Dynamic Nonlinear Models.* Springer.

Knight, K. (1989) Limit theory for autoregressive parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics* 17, 261-278.

Table S1. Empirical Size and Power for Fixed and Iterated $\hat{\theta}$

	$\hat{\theta}$ =self-normalized residual s.d.					$\hat{\theta}$ =75th residual percentile				
	Fixed $\hat{\theta}$		Joint iteration started at			Fixed $\hat{\theta}$		Joint iteration started at		
T	$\xi_T^{(N)}(0)$	$\xi_T^{(\sqrt{T})}(\phi_{LS})$	0	$\tilde{\phi}_{\hat{\theta}}^{(N)}(0)$	$\tilde{\phi}_{\hat{\theta}}^{(\sqrt{T})}(LS)$	$\xi_T^{(N)}(0)$	$\xi_T^{(\sqrt{T})}(\phi_{LS})$	0	$\tilde{\phi}_{\hat{\theta}}^{(N)}(0)$	$\tilde{\phi}_{\hat{\theta}}^{(\sqrt{T})}(LS)$
					Empirical size					
	$\alpha = 3/2$									
100	5.3	7.3	5.3	5.4	7.4	5.9	6.3	4.6	5.5	6.1
500	5.5	6.3	5.4	5.5	6.2	5.7	5.5	5.2	5.7	5.3
	$\alpha = 1$ (Cauchy)									
100	3.9	10.5	3.9	4.0	10.2	5.5	5.5	4.2	4.8	5.2
500	4.1	7.3	4.1	4.2	7.2	5.1	5.3	4.5	4.9	5.1
10^4	4.5	5.3	4.5	4.5	5.3					
	$\alpha = 1$ (Bimodal)									
100	*22.7*	*32.0*	*22.8*	*22.9*	*32.2*	5.8	7.4	4.3	5.0	6.0
500	*33.3*	*41.7*	*33.3*	*33.3*	*41.7*	5.3	5.6	4.7	5.0	5.3
	$\alpha = 1/2$									
100	3.1	*42.6*	3.1	3.1	*42.4*	5.8	*6.8*	4.2	4.6	*5.8*
500	3.9	*48.8*	3.9	3.9	*48.8*	5.2	*6.3*	4.3	4.6	*5.9*
10^4							*5.4*			*5.5*
					Empirical rejection frequencies for $\phi = -7/d_T$					
	$\alpha = 3/2$									
100	38.8	46.8	38.2	39.1	46.6	47.3	46.8	42.1	45.3	46.6
500	49.2	51.6	49.3	49.6	51.5	53.3	51.7	51.3	53.0	51.5
10^4	58.8	60.4	58.8	58.9	60.4	61.9	60.8	61.5	62.0	60.7
	$\alpha = 1$ (Cauchy)									
100	52.1	62.6	52.1	52.3	62.7	67.3	67.5	64.6	66.1	66.9
500	64.3	67.8	64.3	64.4	67.8	71.3	70.7	70.2	70.8	71.0
10^4	71.1	70.7	71.1	71.1	70.7	72.8	71.8	72.5	72.7	72.6
	$\alpha = 1$ (Bimodal)									
100	11.7	30.7	*11.6*	*11.7*	*30.9*	43.0	44.4	37.6	39.2	40.6
500	22.9	40.2	*22.8*	*22.8*	*40.2*	44.7	47.0	43.1	43.9	44.2
10^4	35.9	49.0	*35.8*	*35.8*	*49.0*	46.9	48.7	46.6	47.0	47.0
	$\alpha = 1/2$									
100	54.5	*49.8*	54.5	54.5	*49.9*	56.3	*35.8*	53.8	54.5	*54.1*
500	62.8	*52.3*	62.8	62.9	*52.3*	56.7	*24.7*	55.7	56.1	*56.5*
10^4	72.6	*52.6*	72.6	72.6	*52.5*	58.5	*12.3*	58.3	58.4	*56.8*

Notes. Monte Carlo results based on 10,000 replications. Italics match with those in Table 1.