Exploiting Infinite Variance through Dummy Variables in an AR model

Giuseppe Cavaliere
Università di Bologna

Iliyan Georgiev
Universidade Nova de Lisboa

October 14, 2008

Abstract

In this paper we consider estimation and unit root testing in AR(1) models with infinite variance innovations. Specifically, we study the asymptotic properties of estimators obtained by dummying out "large" innovations, i.e. exceeding a given threshold. These estimators reflect the common practice of dealing with large residuals by including impulse dummies in the estimated regression. Iterative versions of the dummy-variable estimator are also discussed. We provide conditions on the threshold and on the preliminary parameter estimator which ensure that (i) the dummy-based estimator is consistent at higher rates than the basic OLS estimator, and (ii) an asymptotically normal test statistic for the unit root hypothesis can be derived. Our results are related to those obtained for a class of M-estimators in Knight (1989) where, however, their existence and computability is not discussed. Also, due to its non-smoothness, the dummy-variable estimator does not fall within the class of estimators considered in Knight (1989).

1 The problem

A traditional role of impulse dummy variables in time-series econometrics - to correct for a small number of observations that are not well described by a maintained model - has been complemented in recent research by situations where dummy variables are used in number proportional to the sample size and solely as a means to construct estimators or test statistics with certain properties (see Hendry and Santos (2006), Santos et al. (2008)). Such features are also present in this paper: we dummy out the large time-series innovations drawn from a stable distribution although they perfectly fit the maintained stable model and although they are numerous. The justification for doing so are the desirable properties of the resulting estimator, at least in the setup we study.

Knight (1989,1991) proved that in autoregressive (AR) time-series models with a unit root and infinite-variance α-stable innovations a class of M-estimators has faster consistency rate than the OLS estimator and induces asymptotically Gaussian inference under the unit root hypothesis. Knight assumed that the M-estimators in question exist and are sufficiently close to the true parameter value but discussed no conditions or situations where these assumptions are actually satisfied.

In the present paper, on the other hand, we focus on a particular M-estimator which is not covered by Knight's setup but has the advantage of straightforward computability using dummy variables. We analyze it as defined by a particular iterative computational procedure rather than as a usual M-estimator in order to make sure that our results are practically relevant. We then compare the obtained asymptotics with those of Knight.

The focus is on the AR(1) process

$$\Delta y_t = \phi y_{t-1} + \varepsilon_t, \tag{1}$$

where ε_t are infinite-variance i.i.d. and belong to the domain of attraction of an α-stable distribution. The object of study is the dummy-variable estimator $\tilde{\phi}$ of ϕ defined by

$$\tilde{\phi}(\hat{\phi}, \hat{\theta}) := \frac{\sum y_{t-1} \Delta y_t \mathbf{1}_{\{|\hat{\varepsilon}_t| \le \hat{\theta}\}}}{\sum y_{t-1}^2 \mathbf{1}_{\{|\hat{\varepsilon}_t| \le \hat{\theta}\}}}.$$

Here $\hat{\phi}$ is a preliminary estimator of ϕ, $\hat{\varepsilon}_t = \Delta y_t - \hat{\phi} y_{t-1}$ $(t = 1, ..., T)$ are residuals based on this preliminary estimator and $\hat{\theta}$ is an estimator of some scale parameter θ^* of the distribution of ε_1. The benchmark in terms of asymptotic properties is $\tilde{\phi}\,(0, \theta^*)$. Under the hypothesis that $\phi = 0$ and under the assumptions we shall make in the next section it holds that

$$T^{1/2} a_T \tilde{\phi}\,(0, \theta^*) \xrightarrow{w} \frac{\{V\,(\theta^*)\}^{1/2}}{p_{\theta_*}\,(0)} \frac{N\,(0,1)}{(\int S^2)^{1/2}} \quad \text{and} \quad \frac{1}{\hat{\zeta}} \left(\sum y_{t-1}^2 \right)^{1/2} \tilde{\phi}\,(0, \theta^*) \xrightarrow{w} N\,(0,1)\,, \tag{2}$$

where a_T is a normalization sequence which depends on the distribution of ε_1 and diverges faster than $T^{1/2}$, $V\,(\theta^*) := \mathrm{E}(\varepsilon_1^2 \mathbf{1}_{\{|\varepsilon_1| \le \theta^*\}})$ denotes a truncated second moment of ε_1, $p_{\theta_*}\,(0) := P\,\{|\varepsilon_1| \le \theta^*\}$ is assumed positive, S is an α-stable process on $[0,1]$ independent of the Gaussian random variable in the limits and $\hat{\zeta}$ is a consistent estimator of $\{V\,(\theta^*)\}^{1/2}/p_{\theta_*}(0)$. The rate of convergence and the asymptotic distribution are the same as for the M-estimators studied by Knight. This conclusion is not surprising since $\tilde{\phi}\,(0, \theta^*)$ solves the equation $\sum_{t=1}^T y_{t-1} \psi\,(\Delta y_t - \phi y_{t-1}) = 0$ with $\psi = \mathbf{1}_{[-\theta,\theta]} id$, which is similar to the equations studied by Knight, although our ψ does not satisfy the smoothness conditions that Knight assumes.

With respect to $\tilde{\phi}(\hat{\phi}, \hat{\theta})$, in contrast with $\tilde{\phi}\,(0, \theta^*)$, there are two issues to tackle: the fact that θ is estimated and the fact that $\{\varepsilon_t\}$ are estimated. The former issue is also discussed by Knight in the context of his setup. The latter issue is more delicate, and in Knight's work is hidden behind the assumption that an estimator with a sufficiency good consistency rate exists. To clarify ideas, we discuss the two issues (estimation of θ and ϕ) separately.

Throughout integrals are over the interval $[0, 1]$ and summations are for t running from 1 to T, unless otherwise specified.

2 Assumptions

The following assumptions are maintained in what follows.

Assumptions.

A. $\{\varepsilon_t\}_{t=1}^{\infty}$ is an i.i.d. sequence of random variables which have $E\varepsilon_1^2 = \infty$ and belong to the domain of attraction of a stable distribution with index $\alpha \in (0, 2)$.

B. ε_1 has density f with respect to Lebesgue measure and f is a continuous even function.

C. $\{y_t\}_{t=1}^{\infty}$ satisfies eq. (1) with $\phi = 0$ and $y_0 = 0$.

These assumptions can be traced back either to the robust statistics literature or to the literature on empirical processes. The latter connection is due to our interest in statistics defined through residuals.

Assumption A and the symmetry part of Assumption B imply the existence of a process S in $D[0,1]$ and a normalizing sequence $a_T = T^{1/\alpha} l(T)$, with $l(\cdot)$ standing for a slowly varying function, such that $a_T^{-1} \sum_{t=1}^{\lfloor T \cdot \rfloor} \varepsilon_t \xrightarrow{w} S$ in $D[0,\infty)$ as $T \to \infty$.

In view of Assumption B it holds that $E\left(\varepsilon_1 \mathbf{1}_{\{|\varepsilon_1| \leq \theta\}}\right) = 0$ for all θ. With $\psi = \mathbf{1}_{[-1,1]} id$ this can be written as $E\{\psi(\theta\varepsilon_1)\} = 0$, which is a part of Knight's assumptions in his analysis of scale-parameter estimation (with different choices of ψ, however).

The i.i.d. assumption is common in the literature on empirical processes (see, e.g., Lee and Wei, 1999, and Engler and Nielsen, 2007). In this literature the smoothness assumptions on f are often more restrictive than just continuity, but in our case the simplicity of the model makes stronger assumptions unnecessary.

Throughout, together with $V(\theta^*) := \mathrm{E}(\varepsilon_1^2 \mathbf{1}_{\{|\varepsilon_1| \leq \theta^*\}})$ we use also the following notation related to the distribution of $\{\varepsilon_t\}$:

$$p_\theta(x) := \mathrm{E}\left(\mathbf{1}_{\{|\varepsilon_1 - x| \leq \theta\}}\right), \qquad m_\theta(x) := \mathrm{E}\left(\varepsilon_1 \mathbf{1}_{\{|\varepsilon_1 - x| \leq \theta\}}\right)$$

and F for the cumulative distribution function.

3 Preliminary estimation of a scale parameter

Here we study the behaviour of

$$\tilde{\phi}(0, \hat{\theta}) = \frac{\sum y_{t-1} \varepsilon_t \mathbf{1}_{\{|\varepsilon_t| \leq \hat{\theta}\}}}{\sum y_{t-1}^2 \mathbf{1}_{\{|\varepsilon_t| \leq \hat{\theta}\}}} \tag{3}$$

which obtains for $\hat{\phi} = 0$, i.e., for the true value, and where $\hat{\theta}$ is an estimator of a scale parameter (e.g., two times the median of $|\varepsilon_1|$). So the true innovations are observable and the goal is to compare the asymptotics for $\tilde{\phi}$, under appropriate conditions on $\hat{\theta}$, with the benchmark case (2).

This is achieved in the next proposition, for the sake of whose formulation we need to state a fact regarding the numerator of $\tilde{\phi}(0, \theta)$ as a function in θ. For $\alpha \in (0, 4/3]$ it holds that, as $T \to \infty$,

$$\Phi_{T,1}(\cdot) := \left(\sum y_{t-1}^2\right)^{-1/2} \sum y_{t-1} \varepsilon_t \mathbf{1}_{\{|\varepsilon_t| \leq (\cdot)\}} \xrightarrow{w} B(V(\cdot))$$

on $D[0,\infty)$, where B is a standard Brownian motion independent of S (see Lemma 1(c) in the Appendix).

Proposition 1 *Let Assumptions A-C hold and let a be a constant such that* $P(\{|\varepsilon_1| \leq a\}) > 0$. *Then:*

a. For $\alpha \in (4/3, 2)$, if $T^{3/2-2/\alpha}(\hat{\theta}-\theta^*) = O_P(1)$ and $\theta^* \geq a$, it holds that as $T \to \infty$

$$T^{1/2} a_T \tilde{\phi}(0, \hat{\theta}) \xrightarrow{w} \frac{\{V(\theta^*)\}^{1/2}}{p_{\theta_*}(0)} \frac{N(0,1)}{(\int S^2)^{1/2}} \quad \text{and} \quad \frac{1}{\hat{\zeta}} \left(\sum y_{t-1}^2\right)^{1/2} \tilde{\phi}(0, \hat{\theta}) \xrightarrow{w} N(0,1),$$

where $\hat{\zeta}$ is any consistent estimator of $\{V(\theta^*)\}^{1/2}/p_{\theta_*}(0)$ and the Gaussian variable in the limit is independent of S.

b. For $\alpha \in (0, 4/3]$, if $(\Phi_{T,1}(\cdot), a_T^{-1} y_{\lfloor T \cdot \rfloor}, \hat{\theta}) \xrightarrow{w} (B(V(\cdot)), S(\cdot), \Theta)$ in $D^2[0,\infty) \times \mathbb{R}$, with Θ possibly random and such that $P(\Theta \geq a) = 1$, it holds that as $T \to \infty$

$$T^{1/2} a_T \tilde{\phi}(0, \hat{\theta}) \xrightarrow{w} \frac{1}{p_{\theta_*}(0)} \frac{B(V(\Theta))}{(\int S^2)^{1/2}}.$$

If additionally B is independent of (S, Θ), then

$$T^{1/2} a_T \tilde{\phi}(0, \hat{\theta}) \xrightarrow{w} \frac{\{V(\Theta)\}^{1/2}}{p_{\Theta}(0)} \frac{N(0,1)}{(\int S^2)^{1/2}} \quad \text{and} \quad \frac{1}{\hat{\zeta}} \left(\sum y_{t-1}^2\right)^{1/2} \tilde{\phi}(0, \hat{\theta}) \xrightarrow{w} N(0,1),$$

where $\hat{\zeta}$ is any consistent estimator of $\{V(\Theta)\}^{1/2}/p_{\Theta}(0)$ and the Gaussian variable in the limit is independent of S.

Some remarks are due.

Remark. The discussion is restricted to $\theta^* \geq a$ ($\Theta \geq a$ a.s.) in order to ensure that $\tilde{\phi}(0, \hat{\theta})$ is well-defined with probability approaching one. If the density f is not identically zero in any neighbourhood of 0, any positive a works and the requirements of the proposition reduce to $\theta^* > 0$ ($\Theta > 0$ a.s.). This is the case if $\{\varepsilon_t\}$ are α-stable, instead of just being in an α-stable domain of attraction.

Remark. For $\alpha \in (0, 4/3]$, if $\hat{\theta} \xrightarrow{P} \theta^* = const \geq a$ the requirements of part (b) are trivially satisfied with $\Theta = \theta^*$, and the same asymptotics obtain for $T^{1/2} a_T \tilde{\phi}(0, \hat{\theta})$ as in part (a). However, note that in part (b) there is no requirement on the convergence rate of $\hat{\theta}$, whereas in part (a) this requirement is weaker the further away is α from two.

Remark. An example of a possible non-random Θ in part (b) obtains for $\hat{\theta} = a_T^{-2} \sum \varepsilon_t^2$. It holds that $\hat{\theta} \xrightarrow{w} [S]_1$, the quadratic variation of S at 1, and $(S, [S]_1)$ is independent of B since S is independent of B. As $[S]_1$ is a.s. positive, if f is not identically zero in any neighbourhood of 0, the convergences in part (b) hold with $\Theta = [S]_1$.

Remark. Two examples of consistent estimators $\hat{\zeta}$ are

$$\hat{\zeta}_1 = T^{1/2} \frac{(\sum \varepsilon_t^2 \mathbf{1}_{\{|\varepsilon_t| \leq \hat{\theta}\}})^{1/2}}{\sum \mathbf{1}_{\{|\varepsilon_t| \leq \hat{\theta}\}}}, \quad \hat{\zeta}_2 = T^{-1/2} \frac{(\sum y_t^2)(\sum \varepsilon_t^2 \mathbf{1}_{\{|\varepsilon_t| \leq \hat{\theta}\}})^{1/2}}{\sum y_t^2 \mathbf{1}_{\{|\varepsilon_t| \leq \hat{\theta}\}}}.$$

Both can be used in parts (a) and (b).

4 Preliminary estimation of ϕ

Given a preliminary estimator $\hat{\phi}$, consider the procedure of (i) calculating residuals, (ii) introducing dummies for those of them which exceed a fixed threshold θ and (iii) reestimating ϕ. The procedure could be iterated once or several times. Here we study the properties of the iteration and, in particular, discuss under what conditions does it conduct to the same asymptotics as in the previous section: $T^{1/2}a_T$-consistency rate and Gaussian asymptotic inference under $\phi = 0$.

By introducing

$$\tilde{\phi}_\theta(\cdot) := \tilde{\phi}(\cdot, \theta) = \frac{\sum y_{t-1}\varepsilon_t \mathbf{1}_{\{|\varepsilon_t - (\cdot)y_{t-1}| \le \theta\}}}{\sum y_{t-1}^2 \mathbf{1}_{\{|\varepsilon_t - (\cdot)y_{t-1}| \le \theta\}}},$$

the iteration can be written as $\hat{\phi}^{(0)} = \hat{\phi}$, $\hat{\phi}^{(i)} := \tilde{\phi}_\theta(\hat{\phi}^{(i-1)})$, $i \in \mathbb{N}$. The next proposition establishes two uniform approximations of $\tilde{\phi}_\theta$. For the statement of the proposition we use the notation $h_\theta := 2\theta f(\theta) \{F(\theta) - F(-\theta)\}^{-1}$ for $\theta > 0$.

Proposition 2 *Let b_T be a positive real sequences. For every fixed $A > 0$ and θ such that $P\{|\varepsilon_1| \le \theta\} > 0$ it holds that:*

a. If $a_T = o(b_T)$, then

$$\tilde{\phi}_\theta(u) = L_{T,\theta} + u\{h_\theta + o_P(1)\} + \begin{cases} o_P(T^{-1/2}a_T^{-1}), & \text{if } T^{1/2}a_T = O(b_T) \\ o_P(\{(T^{-1/2}a_T^{-1})(b_T^{-1})\}^{1/2}), & \text{if } b_T = o(T^{1/2}a_T) \end{cases}$$

uniformly for $|u| < b_T^{-1}A$, where as $T \to \infty$

$$T^{1/2}a_T L_{T,\theta} \stackrel{w}{\to} \frac{\{V(\theta)\}^{1/2}}{p_\theta(0)} \frac{N(0,1)}{(\int S^2)^{1/2}},$$

with the Gaussian variable in the limit independent of S.

b. If $b_T = a_T$, then

$$\tilde{\phi}_\theta(u) = Q_{T,\theta}(u) + o_P(T^{-1/4}a_T^{-1})$$

uniformly for $|u| \le a_T^{-1}A$, where as $T \to \infty$

$$a_T Q_{T,\theta}(a_T^{-1}(\cdot)) \stackrel{w}{\to} \frac{\int S m_\theta((\cdot)S)}{\int S^2 p_\theta((\cdot)S)}$$

in $D[-A, A]$. Furthermore, if f is strictly monotone on $(-\infty, 0)$ and $(0, \infty)$, then there exist random variables $H_{T,\theta} \in [0, 1)$ such that $\sup_{|u| \le a_T^{-1}A} |\mathbf{1}_{\{u \neq 0\}} u^{-1} Q_{T,\theta}(u)| \le H_{T,\theta}$ and $H_{T,\theta}$ converge weakly as $T \to \infty$ to a random variable $H_\theta < 1$ a.s.

The approximation in part (a) is formally similar to the one obtained by Johansen and Nielsen (2008) for $b_T = T^{1/2}$ and $\{\varepsilon_t\}$ with finite fourth moment. It allows us to discuss the sequence of iterates $\{\hat{\phi}^{(i)}\}$ as the solution of a first order stochastic linear difference equation with a random but well-behaved autoregressive coefficient. It is useful in the case where the consistency rate of the preliminary estimator $\hat{\phi}_0$ is better than a_T. Being non-linear, the approximation in part (b) is less tractable. We will apply it to the case where the preliminary consistency rate is a_T, which is the bridge towards less satisfactory preliminary estimators, and in view of the difficulties posed by the non-linearity will only address the special case of unimodal f.

4.1 The case of a good preliminary estimator

In his discussion of M-estimators Knight (1989,1991) obtains consistency at the rate of $T^{1/2}a_T$ and Gaussian asymptotic inference only under the assumption that the M-estimator, say $\check{\phi}$, is consistent at a rate better than a_T, i.e., $a_T\check{\phi} = o_P(1)$. In our analysis an analogous situation occurs in the case where the preliminary estimator $\hat{\phi}^{(0)}$ is consistent at a rate b_T such that $a_T/b_T \to 0$ as $T \to \infty$. Proposition 2(a) suggests that in this case the behaviour of the sequence of iterates depends crucially on the quantity h_θ. If $h_\theta < 1$, iterations improve upon the preliminary estimator, leading to the concentration of probability mass around $L_{T,\theta}/(1-h_\theta)$ which upon normalization by $T^{1/2}a_T$ is approximately Gaussian for large T. In contrast, for $h_\theta > 1$ the iteration procedure is unstable and deteriorates the properties of the preliminary estimator. A similar phenomenon occurs if $\check{\phi}$ is consistent exactly at the rate a_T, though in this case it is not possible to determine the outcome of the iteration by knowing only a one-dimensional quantity like h_θ. Before we make these observations more precise, we comment on h_θ.

Remark. Under Assumption B there exists a $\theta' \in (0,\theta]$ such that $h_\theta = f(\theta)/f(\theta')$. If f is unimodal, then $\theta' \in (0,\theta)$ and $f(\theta) > f(\theta')$, so $h_\theta < 1$ for every $\theta > 0$. This will be the case if, for instance, $\{\varepsilon_t\}$ are α-stable, since symmetric α-stable densities are known to be unimodal. Even if f is plurimodal, for large θ it will necessarily hold that $h_\theta < 1$, because $h_\theta \to 0$ as $\theta \to \infty$. So we tend to think of the case $h_\theta < 1$ as the rule rather than the exception.

Remark. Nevertheless, distributions satisfying Assumptions A-B and having $h_\theta > 1$ for some $\theta > 0$ do exist. An example, to be used in a Monte Carlo simulation later, could be constructed as follows. Let f be the density of a symmetric α-stable distribution. Define

$$\tilde{f}(x) = \frac{3}{3+2f(0)}\left\{f(x+1)\,\mathbf{1}_{(-\infty,-1)}(x) + f(0)\,x^2\mathbf{1}_{[-1,1]}(x) + f(x-1)\,\mathbf{1}_{(1,\infty)}(x)\right\}. \quad (4)$$

Then $\tilde{f}$ is a density in the domain of attraction of f and $\tilde{h}_\theta = 3$ for $\theta \in (0,1]$.

We turn to the case where $h_\theta < 1$ and consider two kinds of limits: first, an iterated limit where $T \to \infty$ after i has gone to infinity, and second, a path-wise limit, where the path of growth of i is given as a function of T. The latter approach provides results for a wider class of preliminary estimators.

Proposition 3 *Let $h_\theta < 1$ and b_T be a positive real sequence such that $b_T\hat{\phi}^{(0)} = O_P(1)$ and $a_T = o(b_T)$. Let also $\hat{\phi}^{(i)} := \tilde{\phi}_\theta(\hat{\phi}^{(i-1)})$, $i \in \mathbb{N}$. Under Assumptions A-C:*

a. If $T^{1/2}a_T = O(b_T)$, then $\limsup_{i\to\infty}|\hat{\phi}^{(i)} - (1-h_\theta)^{-1}L_{T,\theta}| = o_P(T^{-1/2}a_T^{-1})$, so for every desired numerical precision $\varepsilon > 0$ there exists an $I \in \mathbb{N}$ such that, as $T \to \infty$,

$$P\left(T^{1/2}a_T \sup_{i>I}\left|\hat{\phi}^{(i)} - (1-h_\theta)^{-1}L_{T,\theta}\right| < \varepsilon\right) \to 1,$$

where $L_{T,\theta}$ is as in Proposition 2(a).

b. If $\psi(T)$ is a natural-valued function such that $(h_\theta+\omega)^{\{\psi(T)\}^{1/2}} = o(T^{-1/2}a_T^{-1})$ for some $\omega > 0$ (for example, $\psi(T) = \lfloor T^\nu \rfloor$, $\nu > 0$), then $\hat{\phi}^{(\psi(T))} = (1-h_\theta)^{-1}L_{T,\theta} + o_P(T^{-1/2}a_T^{-1})$ and, *as $T \to \infty$,*

$$T^{1/2} a_T \hat{\phi}^{(\psi(T))} \overset{w}{\to} \frac{\{V(\theta)\}^{1/2}}{(1-h_\theta) p_\theta(0)} \frac{N(0,1)}{(\int S^2)^{1/2}} \quad \textit{and} \quad \frac{1}{\hat{\zeta}} \left(\sum y_{t-1}^2 \right)^{1/2} \hat{\phi}^{(\psi(T))} \overset{w}{\to} N(0,1), \tag{5}$$

where $\hat{\zeta}$ is a consistent estimator of $\{V(\theta)\}^{1/2}(1-h_\theta)^{-1} p_\theta(0)^{-1}$ and the Gaussian variable is independent of S.

Remark. The conclusion in part (a) is established for preliminary estimators which are at least $T^{1/2}a_T$-consistent. The iteration initialized at the true value $\hat{\phi}^{(0)} = 0$ is an example. In part (b) any initial consistency rate better than a_T is covered. For instance, the T-rate of consistency of the OLS estimator (see Chan and Tran, 1989) makes it an admissible $\hat{\phi}^{(0)}$ in part (b) for any $\alpha \in (1,2)$. A result similar to Proposition 3(b) holds also for $b_T = a_T$, provided that the graph of f is bell-shaped. The case of Cauchy errors and $\hat{\phi}^{(0)}$ chosen equal to the OLS estimator is covered by this result, which is formulated next.

Proposition 4 *Let $a_T \hat{\phi}^{(0)} = O_P(1)$ and the density f be strictly monotone on $(-\infty, 0)$ and $(0, \infty)$. Let also $\psi(T) = \lfloor T^\nu \rfloor$ for some $\nu > 0$. Under Assumptions A-C, the convergences (5) hold.*

Remark. Propositions 3 and 4 justify approximate inference based on the iterated estimator and the quantiles of the Gaussian distribution. However, in Proposition 3(a) the number of iterations that guarantees a desired proximity to the approximately Gaussian variable $T^{1/2}a_T(1-h_\theta)^{-1}L_{T,\theta}$ can be chosen independently of T, whereas in Propositions 3(b) and 4 it is a function of T. This difference is of little relevance from practical point of view.

Remark. Typical numerical criteria would declare $\hat{\phi}^{(i)}$ to convergence if, given some $\varepsilon > 0$, they find that $T^{1/2}a_T|\phi^{(i)} - \phi^{(i+1)}| < \varepsilon$ or $|\phi^{(i)} - \phi^{(i+1)}|/|\phi^{(i)}| < \varepsilon$. Under the hypotheses of Propositions 3 and 4 the iteration will be declared to converge with probability approaching one, with respect to both criteria. Thus, in Proposition 3(a) there exists a $J \in \mathbb{N}$ such that, as $T \to \infty$,

$$P\left(\sup_{i,j>J} |T^{1/2}a_T(\phi^{(i)} - \phi^{(j)})| < \varepsilon \right) \to 1,$$

whereas in Propositions 3(b) and 4, as can be seen from the proof,

$$P\left(\sup_{i,j>\psi(T)} |T^{1/2}a_T(\phi^{(i)} - \phi^{(j)})| < \varepsilon \right) \to 1,$$

and similarly for $|\phi^{(i)} - \phi^{(j)}|/|\phi^{(i)}|$ instead of $|T^{1/2}a_T(\phi^{(i)} - \phi^{(j)})|$. Again, the difference is in the dependence, or not, of the number of iterations on T.

Remark. One could pose the question of choosing θ optimally as a minimizer of $\{V(\theta)\}^{1/2}(1-h_\theta)^{-1}p_\theta(0)^{-1}$, if a minimum is achieved. For example, for the Cauchy distribution the optimal θ is approximately 2.03.

The next result shows how an inappropriate choice of θ can deteriorate the properties of a good preliminary estimator, even if it is the true value of 0.

Proposition 5 *Let* $h_\theta > 1$ *and* b_T *be a positive real sequence such that* $b_T\hat{\phi}^{(0)} = O_P(1)$ *and* $a_T = o(b_T)$. *Let also* $\hat{\phi}^{(i)} := \tilde{\phi}_\theta(\hat{\phi}^{(i-1)})$, $i \in \mathbb{N}$. *Under Assumptions A-C:*

a. If $T^{1/2}a_T = o(b_T)$, *then for every* $A > 0$

$$P\left(|T^{1/2}a_T\phi^{(i)}| < A \textit{ for all } i \in \mathbb{N}\right) \to 0 \quad \textit{as } T \to \infty,$$

i.e., the sequence $\{T^{1/2}a_T\hat{\phi}^{(i)}\}_{i=0}^{\infty}$ *is unbounded.*

b. If either (i) $b_T = o(T^{1/2}a_T)$ *and* $b_T\hat{\phi}^{(0)}$ *is bounded away from zero in probability or (ii),* $b_T = T^{1/2}a_T$ *and* $T^{1/2}a_T\{\hat{\phi}^{(0)} - (1-h_\theta)^{-1}L_{T,\theta}\}$ *is bounded away from zero in probability, then for every* $A > 0$

$$P\left(|b_T\phi^{(i)}| < A \textit{ for all } i \in \mathbb{N}\right) \to 0 \quad \textit{as } T \to \infty.$$

5 Some simulations

In Section 4 we concluded that the asymptotics for the iterated dummy-variable estimator (with fixed θ) have two main determinants: the consistency rate of the preliminary estimator $\hat{\phi}^{(0)}$ and the threshold θ beyond which observations are dummied out, the influence of θ being summarized by the quantity $h_\theta = 2\theta f(\theta)\{F(\theta) - F(-\theta)\}^{-1}$. Here we illustrate the importance of these two determinants numerically.

For each of the data generation processes we use, samples of size $T \in \{100\lfloor e^\tau \rfloor : \tau = 0, ..., 4\} = \{100, 271, 738, 2008, 5459\}$ are considered, and for each sample size 10 000 replications are performed. The exponential determination of the sample size is related to our interest in evaluating the convergence rate, expressed as a power of T, of the iterated estimator. The iteration is repeated until $|\phi^{(i)} - \phi^{(i+1)}|/|\phi^{(i)}| < 10^{-6}$ is achieved.

The importance of the preliminary estimator. Data are generated according to eq. (1) with $\phi = 0$ and two different error distributions, both symmetric α-stable, with α equal to 1/2 and 3/2. The targeted convergence rates of $T^{1/2}a_T = T^{1/2+1/\alpha}$ are respectively $T^{5/2}$ and $T^{7/6}$ ($l(\cdot)$ can be taken constant for both error distributions). For both error distributions it holds that $h_\theta < 1$ for all $\theta > 0$, so by Proposition 3 these rates should be achieved for $\hat{\phi}^{(0)} = 0$ independently of the choice of θ. On the other hand, if $\hat{\phi}^{(0)}$ is equal to the OLS estimator, the conditions of Proposition 3 are met for $\alpha = 3/2$ but not for $\alpha = 1/2$ and we expect the rate of $T^{1/2+1/\alpha}$ to be achieved only for $\alpha = 3/2$. For both error distributions we use $\theta = 3$ in the iteration.

In Figure 1 we plot the simulated mean absolute error (MAE) of the iterated dummy-variables estimator against the sample size, with a logarithmic scale on both axes. Results for $\alpha = 3/2$ are in the right panel. It is seen that the MAE is essentially the same no matter whether the iteration is started at 0 or at the OLS estimator. Actually, from the proof of Proposition 3 it follows that in this case the iterated estimators corresponding to the two different preliminary estimators are asymptotically equivalent, which explains why the two lines in the right panel almost coincide. Their estimated slopes are -1.18 (zero initial value) and -1.19 (OLS initial value), which is quite close to $-7/6 \approx -1.17$.

In contrast, for $\alpha = 1/2$ the outcome of the iteration exhibits clear sensitivity to the initial value. The line corresponding to the "good" initial value of 0 has estimated slope of -2.52 in agreement with the theoretical prediction, whereas the line corresponding to the OLS initial value has an estimated slope of -1.21 (-1 is in the 95% confidence interval, -2.5 is not) and the two iterated estimators are clearly not asymptotically equivalent, as the MAE for the OLS initial value is between 2 500 and 420 000 times larger than that for $\hat{\phi}^{(0)}$.

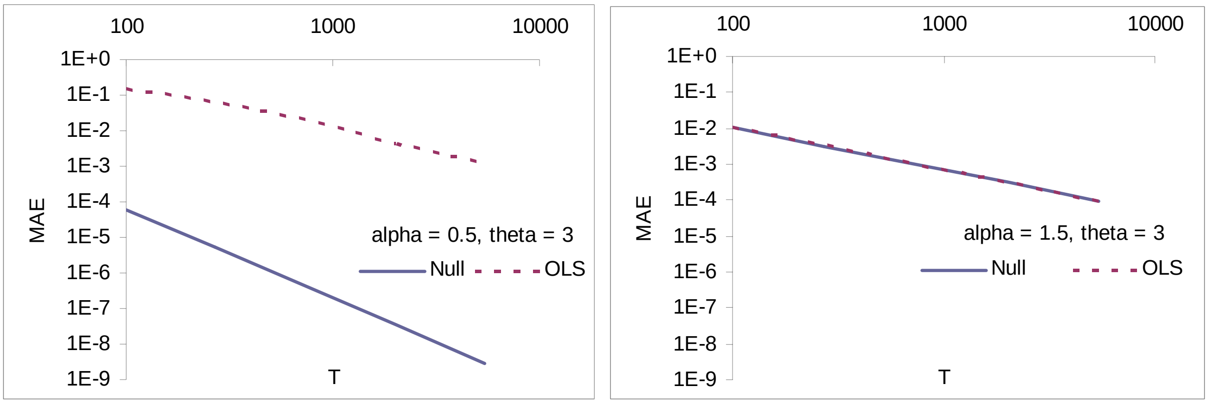

Figure 1. Sensitivity of the iterated dummy-variables estimator to the consistency rate of the preliminary estimator.

The importance of h_θ. Data are again generated according to eq. (1) with $\phi = 0$ but now with Cauchy errors ($\alpha = 1$) and with errors whose density is given by formula (4), where the standard Cauchy density is plugged in for f. The targeted consistency rate is $T^{1+1/\alpha} = T^{3/2}$ ($l(\cdot)$ can be taken constant for both error distributions). Simulations are performed with $\theta \in \{1, 3\}$. We have $h_\theta < 1$ for both values of θ for the Cauchy density and $h_3 < 1$ for the bimodal density. By Proposition 3, if the dummy-variables iteration is initialized at $\hat{\phi}^{(0)} = 0$, the targeted consistency rate should be achieved for these choices of θ and of the error distribution. On the other hand, for the bimodal density it holds that $h_1 > 1$ and, according to Proposition 5, the targeted convergence rates cannot be achieved even if the iteration is started at the true value of 0.

Simulations seem to agree with theory. In the right panel of Figure 2 the outcomes for the Cauchy case are plotted (like earlier, against log scales) and the estimated slopes of the lines are -1.43 and -1.49, respectively for $\theta = 1$ and $\theta = 3$. For the bimodal density (see the left panel) the estimated slope for $\theta = 3$ is -1.48. whereas for $\theta = 1$ the line is noticeably flatter and has an estimated slope of -0.92.

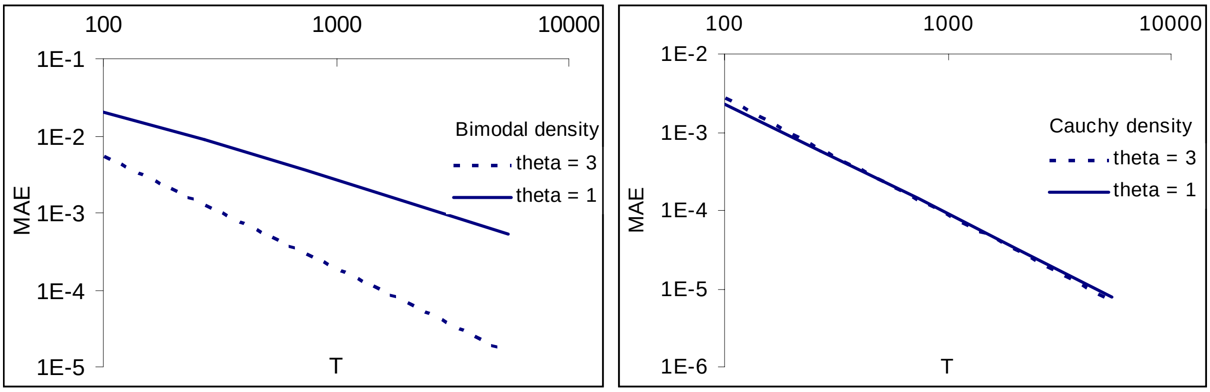

Figure 2. Sensitivity of the iterated dummy-variables estimator to the position of h_θ relative to 1.

6 Appendix

6.1 Preliminary estimation of a scale parameter

Let $\Phi_T(\theta) := p_\theta(0)\tilde{\phi}(0,\theta)$ for $\theta > 0$ and let it be factored into $\Phi_T(\theta) = \Phi_{T,1}(\theta)\,\Phi_{T,2}(\theta)$, where

$$\Phi_{T,1}(\theta) := \frac{\sum y_{t-1}\varepsilon_t \mathbf{1}_{\{|\varepsilon_t|\le\theta\}}}{\left(\sum y_{t-1}^2\right)^{1/2}},\ \Phi_{T,2}(\theta) := \frac{p_\theta(0)\left(\sum y_{t-1}^2\right)^{1/2}}{\sum y_{t-1}^2 \mathbf{1}_{\{|\varepsilon_t|\le\theta\}}}.$$

Decompose further the denominator of $\Phi_{T,2}(\theta)$ into $p_\theta(0)\sum y_{t-1}^2 + \Phi_{T,3}(\theta)$, with $\Phi_{T,3}(\theta) := \sum y_{t-1}^2(\mathbf{1}_{\{|\varepsilon_t|\le\theta\}} - p_\theta(0))$. We establish next some asymptotic properties of $\Phi_{T,1}$ and $\Phi_{T,3}$ as processes in θ. By weak convergence of a functional sequence $X_T(\cdot)$ in $D[-A,A]$ we mean the weak convergence of $X_T(\cdot - A)$ in $D[0,2A]$.

Lemma 1 *Under Assumptions A-C, as $T\to\infty$ it holds that:*

a. The finite-dimensional distributions (fidis) of $\Phi_{T,1}(\cdot)$ converge to those of $B(V(\cdot))$.

b. For $\alpha \in (4/3, 2)$ and $\omega \ge 3/2 - 2/\alpha$, $\Phi_{T,1}(\theta^* + T^{-\omega}(\cdot)) \xrightarrow{w} B(V(\theta^*))$ in $D[-A,A]$ for every $A, \theta^* > 0$.

c. For $\alpha \in (0, 4/3]$, $\Phi_{T,1}(\cdot) \xrightarrow{w} B(V(\cdot))$ in $D[0,\infty)$.

d. $T^{-3/4}a_T^{-2}\sup_{0\le\theta\le A}|\Phi_{T,3}(\theta)| = O_P(1)$ for every $A > 0$.

Proof. For notational ease, we discuss the bivariate fidis of $\Phi_{T,1}$, the generalization being straightforward. For $\theta_1 < \theta_2$, $\varepsilon_t\mathbf{1}_{\{|\varepsilon_t|\le\theta_1\}}$ and $\varepsilon_t\mathbf{1}_{\{\theta_1<|\varepsilon_t|\le\theta_2\}}$ are uncorrelated under the assumption that the distribution of ε_t is symmetric. Thus, using also the absolute continuity of the distribution of $\{\varepsilon_t\}$, it is seen that

$$(B_{T,1}(\cdot), B_{T,2}(\cdot)) := T^{-1/2}\sum_{t=1}^{\lfloor T(\cdot)\rfloor}\varepsilon_t\left(\frac{\mathbf{1}_{\{|\varepsilon_t|\le\theta_1\}}}{\{V(\theta_1)\}^{1/2}}, \frac{\mathbf{1}_{\{\theta_1<|\varepsilon_t|\le\theta_2\}}}{\{V(\theta_2)-V(\theta_1)\}^{1/2}}\right) \xrightarrow{w} (B_1(\cdot), B_2(\cdot)),$$

where B_1 and B_2 are independent standard Brownian motions. Furthermore, B_1 and B_2 are independent of the weak limit S of $a_T^{-1}y_{\lfloor T(\cdot)\rfloor}$ (Resnick and Greenwood, 1979). Therefore,

$$\left(\{V(\theta_1)\}^{1/2}B_{T,1}, \{V(\theta_2)-V(\theta_1)\}^{1/2}B_{T,2}, a_T^{-1}y_{\lfloor T(\cdot)\rfloor}\right) \xrightarrow{w} \left(\sqrt{V(\theta_1)}B_1, \{V(\theta_2)-V(\theta_1)\}^{1/2}B_2, S\right)$$

on $D_3\,[0,\infty)$. From the CMT and Lemma 1 of Knight (1989) it follows that

$$(\Phi_{T,1}(\theta_1), \Phi_{T,1}(\theta_2) - \Phi_{T,1}(\theta_1)) \xrightarrow{w} \left(\int S^2\right)^{-1/2} \left(\{V(\theta_1)\}^{1/2}\int SdB_1, \{V(\theta_2)-V(\theta_1)\}^{1/2}\int SdB_2\right)$$
$$\overset{d}{=} \left(B\left(V\left(\theta_1\right)\right), B\left(V\left(\theta_2\right)\right) - B\left(V\left(\theta_1\right)\right)\right)$$

with B independent of S, the distributional equality by the independence of B_1, B_2 and S. Hence, $(\Phi_{T,1}(\theta_1), \Phi_{T,1}(\theta_2)) \xrightarrow{w} (B\left(V\left(\theta_1\right)\right), B\left(V\left(\theta_2\right)\right))$.

In view of part (a), to prove (b) and (c) we establish below the tightness of $\Phi_{T,1}(\theta^*+T^{-\omega}(\cdot))$ in $D[-A,A]$ (for $\omega \geq 3/2-2/\alpha$ and $\alpha \in (4/3,2)$) and in $D[0,A]$ (for $\omega = 0$ and $\alpha \in (0,4/3]$), for every $A>0$. Then part (b) follows from tightness and the facts that (i) $\Phi_{T,1}(\theta_*) \xrightarrow{w} B(V(\theta^*))$ and (ii) $\Phi_{T,1}(\theta_*+T^{-\omega}\theta)-\Phi_{T,1}(\theta_*) = |V(\theta_*+T^{-\omega}\theta)-V(\theta_*)|^{1/2}\nu_T(\theta)$, where $V(\theta_*+T^{-\omega}\theta)-V(\theta_*) = o(1)$ and $\nu_T(\theta)$ converges in distribution to a stochastic integral, for every $\theta>0$. Part (c) obtains by setting $\theta^*=0$, $\omega=0$.

As $a_T^{-2}T^{-1}\sum y_{t-1}^2 \xrightarrow{w} \int S^2$ by the CMT, it suffices to show that $a_T^{-1}T^{-1/2}\sum y_{t-1}\varepsilon_t\mathbf{1}_{\{|\varepsilon_t|\leq\theta^*+T^{-\omega}(\cdot)\}}$ is tight. For a fixed $A>0$, define

$$y_t^{<} := \sum \varepsilon_t \mathbf{1}_{\{|\varepsilon_t|\leq\theta^*+A\}} \text{ and } y_t^{>} := \sum \varepsilon_t \mathbf{1}_{\{|\varepsilon_t|>\theta^*+A\}}.$$

Using a truncation trick from Knight (1989), for an arbitrary $M>0$ decompose

$$a_T^{-1}T^{-1/2}\sum y_{t-1}\varepsilon_t\mathbf{1}_{\{|\varepsilon_t|\leq\theta^*+T^{-\omega}(\cdot)\}} = \Phi_{T,1}^{<}(\cdot) + \Phi_{T,1}^{><}(\cdot) + \Phi_{T,1}^{>>}(\cdot),$$

with

$$\begin{aligned}
\Phi_{T,1}^{<}(\cdot) &:= a_T^{-1}T^{-1/2}\sum y_{t-1}^{<}\varepsilon_t\mathbf{1}_{\{|\varepsilon_t|\leq\theta^*+T^{-\omega}(\cdot)\}},\\
\Phi_{T,1}^{><}(\cdot) &:= a_T^{-1}T^{-1/2}\sum y_{t-1}^{>}\varepsilon_t\mathbf{1}_{\{|a_T^{-1}y_{t-1}^{>}|\leq M\}}\mathbf{1}_{\{|\varepsilon_t|\leq\theta^*+T^{-\omega}(\cdot)\}},\\
\Phi_{T,1}^{>>}(\cdot) &:= a_T^{-1}T^{-1/2}\sum y_{t-1}^{>}\varepsilon_t\mathbf{1}_{\{|a_T^{-1}y_{t-1}^{>}|>M\}}\mathbf{1}_{\{|\varepsilon_t|\leq\theta^*+T^{-\omega}(\cdot)\}}.
\end{aligned}$$

Since $P(\sup_K|\Phi_{T,1}^{>>}(\cdot)|>0) \leq P\left(\max_{t\leq T}|a_T^{-1}y_{t-1}|>M\right)$ for $K\in\{[-A,A],[0,A]\}$, can be made arbitrarily small by choosing a sufficiently large M (as $\max_{t\leq T}\left|a_T^{-1}y_{t-1}\right|$ converges weakly to an a.s. finite random variable), it is enough to show tightness of $\Phi_{T,1}^{<}$ and $\Phi_{T,1}^{><}(\cdot)$ for a fixed M. We start from the former.

Let $\theta_1<\theta_m<\theta_2\in[-A,A]$ or $[0,A]$. Then

$$\mathrm{E}\left[\left(\Phi_{T,1}^{<}(\theta_2)-\Phi_{T,1}^{<}(\theta_m)\right)^2\left(\Phi_{T,1}^{<}(\theta_m)-\Phi_{T,1}^{<}(\theta_1)\right)^2\right] = a_T^{-4}T^{-2}\sum_{t_1,t_2,t_3,t_4=1}^{T}E_{t_1t_2t_3t_4}^{<},$$

with $E_{t_1t_2t_3t_4}^{<} := \mathrm{E}[\mathbf{1}_{t_1}^{(1)}\mathbf{1}_{t_2}^{(1)}\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}\prod_{i=1}^{4}y_{t_i-1}^{<}\varepsilon_{t_i}]$, $\mathbf{1}_t^{(1)} := \mathbf{1}_{\{\theta^*+T^{-\omega}\theta_1<|\varepsilon_t|\leq\theta^*+T^{-\omega}\theta_m\}}$ and $\mathbf{1}_t^{(2)} := \mathbf{1}_{\{\theta^*+T^{-\omega}\theta_m<|\varepsilon_t|\leq\theta^*+T^{-\omega}\theta_2\}}$.

If the largest two among t_i ($i=1,...,4$) are distinct (and, say, $t_4=\max_{i=1,\dots,4}t_i$), from the serial independence of $\{\varepsilon_t\}$ it follows that $E_{t_1t_2t_3t_4}^{<}$ equals the product of $\mathrm{E}[\mathbf{1}_{t_4}^{(2)}\varepsilon_{t_4}]$ and the expectation of the remaining factors. As the distribution of $\mathbf{1}_{t_4}^{(2)}\varepsilon_{t_4}$ is symmetric about zero, it holds that $\mathrm{E}[\mathbf{1}_{t_4}^{(2)}\varepsilon_{t_4}]=0$ and $E^{<}_{t_1t_2t_3t_4}=0$. So there are at most T^3 possibly non-zero $E^{<}_{t_1t_2t_3t_4}$'s, each of them satisfying, in view of the Cauchy-Schwartz inequality,

$$\begin{aligned}|E^{<}_{t_1t_2t_3t_4}| &\le [\mathrm{E}\{(y_T^{<})^2\}]^2\{V(\theta^*+T^{-\omega}\theta_2)-V(\theta^*+T^{-\omega}\theta_m)\}\{V(\theta^*+T^{-\omega}\theta_1)-V(\theta^*+T^{-\omega}\theta_m)\}\\ &\le T^2(\theta^*+A)^2 4^{-1}\{V(\theta^*+T^{-\omega}\theta_2)-V(\theta^*+T^{-\omega}\theta_1)\}^2\\ &\le T^{2-2\omega}C\left(\theta_2-\theta_1\right)^2,\end{aligned}$$

the last line from the Mean Value Theorem, with $C=4(\theta^*+A)^4\left\|f\right\|_\infty^2$, $\left\|f\right\|_\infty:=\sup_{\mathbb{R}}|f|<\infty$. This gives

$$\mathrm{E}\left[\left(\Phi^{<}_{T,1}\left(\theta_2\right)-\Phi^{<}_{T,1}(\theta_m)\right)^2\left(\Phi^{<}_{T,1}\left(\theta_m\right)-\Phi^{<}_{T,1}(\theta_1)\right)^2\right]\le a_T^{-4}T^{3-2\omega}C(\theta_2-\theta_1)^2,$$

and since $a_T^{-4}T^{3-2\omega}=O_P(1)$ for the considered ω and α, by Theorem 15.6 of Billingsley (1968) $\Phi^{<}_{T,1}\left(\cdot\right)$ is tight.

Similarly,

$$E^{>}_{\theta_1,\theta_2,\theta_m}:=\mathrm{E}\left[\left(\Phi^{><}_{T,1}\left(\theta_2\right)-\Phi^{><}_{T,1}(\theta_m)\right)^2\left(\Phi^{><}_{T,1}\left(\theta_m\right)-\Phi^{><}_{T,1}(\theta_1)\right)^2\right]=a_T^{-4}T^{-2}\sum_{t_1,t_2,t_3,t_4=1}^{T}E^{>}_{t_1t_2t_3t_4},$$

where $E^{>}_{t_1t_2t_3t_4}:=\mathrm{E}[\mathbf{1}_{t_1}^{(1)}\mathbf{1}_{t_2}^{(1)}\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}\prod_{i=1}^4\mathbf{1}_{\{|a_T^{-1}y^{>}_{t_i-1}|\le M\}}y^{<}_{t_i-1}\varepsilon_{t_i}]$. As before, $E^{>}_{t_1t_2t_3t_4}=0$ if the two largest t_i are distinct. The same holds if at least one among t_1 and t_2 equals at least one among t_3 and t_4, since then $\mathbf{1}_{t_1}^{(1)}\mathbf{1}_{t_2}^{(1)}\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}=0$.

Further, if $t_1=t_2>t_3,t_4$ and $t_3\ne t_4$, again $E^{>}_{t_1t_2t_3t_4}=0$. Indeed, conditionally on $\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}=0$ the expectation is trivially zero, whereas conditionally on $\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}=1$ it holds, by the definition of $y_t^{>}$ and by the serial independence of $\{\varepsilon_t\}$, that $\prod_{i=1}^4\mathbf{1}_{\{|a_T^{-1}y^{>}_{t_i-1}|\le M\}}y^{<}_{t_i-1}$ is independent of $\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}\varepsilon_{t_3}\varepsilon_{t_4}$, yielding

$$E^{>}_{t_1t_2t_3t_4}=\mathrm{E}\left[\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}\varepsilon_{t_3}\varepsilon_{t_4}\middle|\,\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}=1\right]\mathrm{E}\left[\prod_{i=1}^4\mathbf{1}_{\{|a_T^{-1}y^{>}_{t_i-1}|\le M\}}y^{<}_{t_i-1}\middle|\,\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}=1\right]\mathrm{E}\left[\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}\right],$$

which is zero since $\mathrm{E}[\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}\varepsilon_{t_3}\varepsilon_{t_4}|\mathbf{1}_{t_3}^{(2)}\mathbf{1}_{t_4}^{(2)}=1]=0$.

For the same reason, if $t_3=t_4>t_1,t_2$ and $t_1\ne t_2$, again $E^{>}_{t_1t_2t_3t_4}=0$. It remains that

$$\begin{aligned}E^{>}_{\theta_1,\theta_2,\theta_m} &= a_T^{-4}T^{-2}\sum_{\substack{t_1,t_3=1\\ t_1\ne t_3}}^{T}E^{>}_{t_1t_1t_3t_3}\\ &\le M^4\{V(\theta^*+T^{-\omega}\theta_2)-V(\theta^*+T^{-\omega}\theta_m)\}\{V(\theta^*+T^{-\omega}\theta_1)-V(\theta^*+T^{-\omega}\theta_m)\}\\ &\le T^{-2\omega}C\left(\theta_2-\theta_1\right)^2\end{aligned}$$

with $C=4M^4(\theta^*+A)^2\left\|f\right\|_\infty^2$. As $\omega\ge 0$, also $\Phi^{><}_{T,1}\left(\cdot\right)$ is tight, and so is $\Phi_{T,1}\left(\cdot\right)$.

As $\Phi_{T,3}(0) = 0$ and $T^{-1/2}a_T^{-2}\Phi_{T,3}(\theta^*)$ converges weakly to a stochastic integral, part (d) will follow if we show that $T^{-1/2}a_T^{-2}\Phi_{T,3}(\cdot)$ is tight in $D\,[0, A]$. Introducing

$$\Phi_{T,3}^{\leq}(\theta) := \sum y_{t-1}^2 \mathbf{1}_{\{|a_T^{-1}y_{t-1}|\leq M\}}(\mathbf{1}_{\{|\varepsilon_t|\leq\theta\}} - p_\theta\,(0)),$$

like earlier it suffices to show the tightness of $T^{-3/4}a_T\Phi_{T,3}^{\leq}(\cdot)$ for every fixed $M > 0$.

For $\theta_1 < \theta_m < \theta_2$ we find

$$F_{\theta_1,\theta_2,\theta_m} := \mathrm{E}\left[\left(\Phi_{T,3}^{\leq}(\theta_2) - \Phi_{T,1}^{\leq}(\theta_m)\right)^2\left(\Phi_{T,1}^{\leq}(\theta_m) - \Phi_{T,1}^{\leq}(\theta_1)\right)^2\right] = a_T^{-8}T^{-3}\sum_{t_1,t_2,t_3,t_4=1}^{T} F_{t_1t_2t_3t_4}$$

with $F_{t_1t_2t_3t_4} = \mathrm{E}\prod_{i=1}^4[y_{t_i-1}^2\mathbf{1}_{\{|a_T^{-1}y_{t_i-1}|\leq M\}}\Delta_{t_i}^{(i)}]$, $\Delta_t^{(1)} := \Delta_t^{(2)} := \mathbf{1}_{\{|\varepsilon_t|\leq\theta_2\}} - \mathbf{1}_{\{|\varepsilon_{t_i}|\leq\theta_m\}}$ and $\Delta_t^{(3)} := \Delta_t^{(4)} := \mathbf{1}_{\{|\varepsilon_t|\leq\theta_2\}} - \mathbf{1}_{\{|\varepsilon_{t_i}|\leq\theta_m\}}$. Whenever t_i $(i = 1, ..., 4)$ are distinct, $F_{t_1t_2t_3t_4} = 0$ by the independence and the symmetry of $\{\varepsilon_t\}$. For the remaining T^3 cases we use the evaluation $F_{t_1t_2t_3t_4} \leq M^8a_T^8 Var(\Delta_1^{(1)})Var(\Delta_2^{(2)})$ implied by the Cauchy-Schwartz inequality and the independence of $\{\varepsilon_t\}$. With $Var(\Delta_1^{(1)}) \leq p_{\theta_2}(0) - p_{\theta_m}(0)$ by a direct calculation, and similarly for $Var(\Delta_1^{(2)})$, we find

$$F_{t_1t_2t_3t_4} \leq M^8a_T^8 4^{-1}\{p_{\theta_2}(0) - p_{\theta_1}(0)\}^2 \leq a_T^8C(\theta_2 - \theta_1)^2$$

with $C = M^8\,\|f\|_\infty^2$, and further, $F_{\theta_1,\theta_2,\theta_m} \leq C(\theta_2-\theta_1)^2$, from where the tightness of $T^{-3/4}a_T^{-2}\Phi_{T,3}^{\leq}(\cdot)$ follows. ■

Proof of Proposition 1. As $p_a(0) > 0$ and $p_{(\cdot)}(0)$ is continuous under Assumption B, there exists an $\epsilon \in (0, a)$ such that $p_{a-\epsilon}(0) > 0$. Further, (i) $\varsigma_T := T^{1/2}a_T(\sum y_{t-1}^2)^{-1/2} \xrightarrow{w} (\int S^2)^{-1/2} > 0$ a.s. by the CMT (for a.s. positivity, see Chan and Tran, 1989, p.359) and (ii) for every $A > a - \epsilon$ it holds with probability approaching one that

$$T^{1/2}a_T \sup_{a-\epsilon\leq\theta\leq A}\left|(\sum y_{t-1}^2)^{-1/2} - \Phi_{T,1}(\theta)\right| \leq \frac{\varsigma_T^{-1}\sup_{a-\epsilon\leq\theta\leq A}\left|T^{-1}a_T^{-2}\Phi_{T,3}(\theta)\right|}{p_{a-\epsilon}(0)\,\varsigma_T^2 - \sup_{a-\epsilon\leq\theta\leq A}\left|T^{-1}a_T^{-2}\Phi_{T,3}(\theta)\right|} = o_P(1)$$

because $p_{a-\epsilon}(0) > 0$ and $T^{-1}a_T^{-2}\sup_{a-\epsilon\leq\theta\leq A}|\Phi_{T,3}(\theta)| = o_P(1)$; see Lemma 1(d). Therefore, it holds that $T^{1/2}a_T\Phi_{T,2}(\theta^* + T^{-\omega}(\cdot)) \xrightarrow{w} (\int S^2)^{-1/2}$ in $D[-A, A]$ for $\omega \geq 0$, and $T^{1/2}a_T\Phi_{T,2}(\cdot) \xrightarrow{w} (\int S^2)^{-1/2}$ in $D\,[a - \epsilon, \infty)$[1]. Combined with Lemma 1(b,c) these give, in view of the continuity of $p_{(\cdot)}(0)$,

$$T^{1/2}a_T\tilde{\phi}(0, \theta_* + T^{-\omega}(\cdot)) \xrightarrow{w} \frac{\{V(\theta^*)\}^{1/2}}{p_{\theta^*}(0)}\frac{B(\theta^*)}{(\int S^2)^{1/2}} \text{ and } T^{1/2}a_T\tilde{\phi}(0, \cdot) \xrightarrow{w} \frac{\{V(\cdot)\}^{1/2}}{p_{(\cdot)}(0)}\frac{B(\cdot)}{(\int S^2)^{1/2}}$$

respectively in $D[-A, A]$ for $\alpha \in (4/3, 2)$ and $\omega \geq 3/2-2/\alpha$, and in $D\,[a - \epsilon, \infty)$ for $\alpha \in (0, 4/3]$. The conclusions of the proposition are straightforward from here, in part (b) using the continuity of B. ■

[1] Understood as $T^{1/2}a_T\Phi_{T,2}(\cdot + a - \epsilon) \xrightarrow{w} (\int S^2)^{-1/2}$ in $D\,[0, \infty)$.

6.2 The approximations of $\tilde{\phi}$

Introduce

$$\begin{aligned} r_2(u) &= \sum y_{t-1}\left[\varepsilon_t \mathbf{1}_{\{|\varepsilon_t - uy_{t-1}|\leq\theta\}} - m_\theta(uy_{t-1})\right],\ r_4(u) = \sum y_{t-1}m_\theta(uy_{t-1}), \\ r_1(u) &= \mathbf{1}_{\{u\neq 0\}}\frac{1}{u}\left[\sum y_{t-1}m_\theta(uy_{t-1}) - uh_\theta \sum y_{t-1}^2 p_\theta(uy_{t-1})\right], \\ r_5(u) &= \sum y_{t-1}^2\left[\mathbf{1}_{\{|\varepsilon_t - uy_{t-1}|\leq\theta\}} - p_\theta(uy_{t-1})\right],\ r_3(u) = \sum y_{t-1}^2 p_\theta(uy_{t-1}), \end{aligned}$$

suppressing the dependence on θ. Then

$$\tilde{\phi}_\theta\,(u) = L_{T,\theta} + u(h_\theta + r_{T,\theta}\,(u)) + e_{T,\theta}(u), \tag{6}$$

where $L_{T,\theta} := r_2(0)/r_3(0)$, $r_{T,\theta}\,(u) := r_1(u)/\{r_3(u) + r_5(u)\}$ and

$$e_{T,\theta}(u) := \frac{r_2(u) - r_2(0) - L_{T,\theta}\left[r_3(u) - r_3(0) + r_5(u)\right] - uh_\theta r_5(u)}{r_3(u) + r_5(u)}; \tag{7}$$

also

$$\tilde{\phi}_\theta\,(u) = Q_{T,\theta}(u) + \frac{r_2\,(u)}{r_3(u) + r_5(u)} - \frac{r_4(u)}{r_3(u)}\frac{r_5(u)}{r_3(u) + r_5(u)}, \tag{8}$$

where $Q_{T,\theta}(u) := r_4(u)/r_3(u)$.

To prove Proposition 2, we study the stochastic magnitude orders of $r_i\,(u)$.

Lemma 2 *Let $\{b_T\}$ be a positive real sequence. Under Assumptions A-C it holds that:*

a. $\sup_{|u|\leq A}\left|r_1(b_T^{-1}u)\right| = o_P(Ta_T^2)$ *and* $\sup_{|u|\leq A}\left|r_3(b_T^{-1}u) - r_3(0)\right| = o_P(Ta_T^2)$ *if* $a_T = o(b_T)$.

b. $\sup_{|u|\leq A}\left|r_5(b_T^{-1}u)\right| = o_P(T^{3/4}a_T^{5/2}b_T^{-1/2})$ *if* $a_T = O(b_T)$ *and* $b_T = o(T^{1/2}a_T)$, *whereas* $\sup_{|u|\leq A}\left|r_5(b_T^{-1}u)\right| = O_P(T^{1/2}a_T^2)$ *if* $T^{1/2}a_T = O(b_T)$.

c. $\sup_{|u|\leq A}\left|r_2(b_T^{-1}u) - r_2(0)\right| = o_P(T^{3/4}a_T^{3/2}b_T^{-1/2})$ *if* $a_T = O(b_T)$.

PROOF. First, as $m_\theta(0) = 0$ by the symmetry of ε_t, from the Mean Value Theorem

$$r_1(u) = \mathbf{1}_{\{u\neq 0\}}\sum y_{t-1}^2\{m_\theta'(uw_{t-1}y_{t-1}) - h_\theta p_\theta(uy_{t-1})\},$$

where $w_t \in [0,1]$. Hence,

$$\sup_{|u|\leq A}\left|r_1(b_T^{-1}u)\right| \leq \sup_{|u|\leq b_T^{-1}A}\max_{t\leq T}\left|m_\theta'(uw_{t-1}y_{t-1}) - h_\theta p_\theta(uy_{t-1})\right|\sum y_{t-1}^2. \tag{9}$$

For a given $\epsilon \in (0,1)$, let M_ϵ be such that $P(\max_{t\leq T}|a_T^{-1}y_{t-1}| \leq M_\epsilon) > 1-\epsilon$ (M_ϵ exists since $\max_{t\leq T}|a_T^{-1}y_{t-1}| \xrightarrow{w} \max_{s\in[0,1]}|S\,(s)|$). For outcomes such that $\max_{t\leq T}|a_T^{-1}y_{t-1}| \leq M_\epsilon$, it holds that

$$\begin{aligned} &\sup_{|u|\leq b_T^{-1}A}\max_{t\leq T}\left|m_\theta'(uw_{t-1}y_{t-1}) - h_\theta p_\theta(uy_{t-1})\right| \\ \leq &\sup_{|u|\leq a_T b_T^{-1}AM_\epsilon}\left|m_\theta'\,(u) - m_\theta'\,(0)\right| + h_\theta \sup_{|u|\leq a_T b_T^{-1}AM_\epsilon}|p_\theta(u) - p_\theta(0)| \to 0, \end{aligned}$$

the inequality since $m'_\theta(0) = h_\theta p_\theta(0)$ and the convergence since (i) m'_θ and p_θ are continuous at 0 under Assumption B and (ii) $a_T = o(b_T)$. By the arbitrariness of ϵ and using (9), where $\sum y_{t-1}^2 = O_P(Ta_T^2)$, the first relation in part (a) obtains.

For the second relation in (a), again by the Mean Value Theorem,

$$|r_3(u) - r_3(0)| \le |u| \sum |y_{t-1}^3 p'_\theta(uz_{t-1}y_{t-1})| \le |u| \, \|f\|_\infty \sum |y_{t-1}^3|,$$

where $z_t \in [0,1]$. Since $T^{-1}a_T^{-3} \sum |y_{t-1}^3| \xrightarrow{w} \int |S|^3$, the asserted relation follows.

To prove part (b), define $\iota_t(u) := \mathbf{1}_{\{|\varepsilon_t - uy_{t-1}| \le \theta\}} - p_\theta(uy_{t-1})$ and, for $M > 0$, also

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

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

$$P\left(\sup_{|u|\le A} \left|r_5(b_T^{-1}u)\right| > \eta_T\right) \le P\left(\sup_{|u|\le A} \left|r_5^M(b_T^{-1}u)\right| > \eta_T\right) + P\left(\max_{t\le T} |a_T^{-1}y_t| > M\right). \tag{10}$$

As $\max_{t\le T} \left|a_T^{-1}y_t\right|$ converges weakly to an a.s. finite random variable, M can be chosen large enough for $P(\max_{t\le T} \left|a_T^{-1}y_t\right| > M)$ to be as small as desired. So the sought relations for $r_5(b_T^{-1}(\cdot))$ will follow once we show that they hold for $r_5^M(b_T^{-1}(\cdot))$. To this aim we check below, first, that for $c_T := T^{3/4}(a_T^5 b_T^{-1})^{1/2}$ and for every fixed u, $T^{1/4}c_T^{-1}\{r_5^M(b_T^{-1}u) - r_5^M(0)\} = O_P(1)$, and hence, $c_T^{-1}\{r_5^M(b_T^{-1}u) - r_5^M(0)\} = o_P(1)$. Second, we check that $c_T^{-1}\{r_5^M(b_T^{-1}(\cdot)) - r_5^M(0)\}$ is tight. The two facts jointly imply that $c_T^{-1}\sup_{|u|\le A}\{r_5^M(b_T^{-1}(\cdot)) - r_5^M(0)\} = o_P(1)$. The proof is completed by noting that $T^{-1/2}a_T^{-1}r_5^M(0) = O_P(1)$ since it converges weakly to an a.s. finite random variable (see Lemma 1 of Knight (1989)).

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

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

the last inequality since $|\Delta p_\theta(u_2, u_1)| \le 1$. By applying the Mean Value Theorem,

$$\mathrm{E}\left[\{\iota_t(u_2) - \iota_t(u_1)\}^2 | \mathcal{F}_{t-1}\right] \le 6\,\|f\|_\infty (u_2 - u_1) y_{t-1} = 6\,\|f\|_\infty |u_2 - u_1||y_{t-1}|. \tag{11}$$

Now the fact that $T^{1/4}c_T^{-1}\{r_5^M(b_T^{-1}u) - r_5^M(0)\} = O_P(1)$ follows from Chebyshev's inequality: $\mathrm{E}\{r_5^M(b_T^{-1}u) - r_5^M(0)\} = 0$ because $\mathrm{E}\{\iota_t(b_T^{-1}u) - \iota_t(0) | \mathcal{F}_{t-1}\} = 0$ $(t = 1, ..., T)$, whereas

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

the first equality because $\mathrm{E}[\iota_{\max(t,s)}\left(b_T^{-1}u\right)|\mathcal{F}_{\max(t,s)-1}] = 0$ and the inequality from (11).

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 $r_5^M(b_T^{-1}(\cdot))$ are not càdlàg due to the terms $\mathbf{1}_{\{|\varepsilon_t-(\cdot)y_{t-1}|\leq\theta\}}$, which are not càdlàg. If we substitute them by

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

a càdlàg modified process, say $\tilde{r}_5^M(b_T^{-1}(\cdot))$, will obtain. The set of points at which the sample paths of $r_5^M(b_T^{-1}(\cdot))$ and $\tilde{r}_5^M(b_T^{-1}(\cdot))$ 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, with probability one at each of these points only one indicator is affected, so

$$\sup_{|u|\leq A} |r_5^M(b_T^{-1}(\cdot)) - \tilde{r}_5^M(b_T^{-1}(\cdot))| \leq \max_{t\leq T} y_t^2 = O_P(a_T^2) = o_P(c_T).$$

It is therefore enough to establish the tightness of $c_T^{-1}\{\tilde{r}_5^M(b_T^{-1}(\cdot)) - \tilde{r}_5^M(0)\}$. Since we argue in terms of expectations, which are unaffected by the change from $\mathbf{1}_{\{|\varepsilon_t-(\cdot)y_{t-1}|\leq\theta\}}$ to $\mathbf{1}_{\{|\varepsilon_t-(\cdot)y_{t-1}|\lhd\theta\}}$, we continue writing in terms of $r_5^M(b_T^{-1}(\cdot))$.

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

$$G_{u_2,u_1,u_m} := \mathrm{E}\left(\{r_5^M(b_T^{-1}u_2) - r_5^M(b_T^{-1}u_m)\}^2\{r_5^M(b_T^{-1}u_m) - r_5^M(b_T^{-1}u_1)\}^2\right) = \sum_{t_1,t_2,t_3,t_4} G_{t_1,t_2,t_3,t_4}$$

with $G_{t_1,t_2,t_3,t_4} := \mathrm{E}\prod_{i=1}^4[y_{t_i-1}^2\mathbf{1}_{\{|a_T^{-1}y_{t_i-1}|\leq M\}}\Delta_{t_i}^{(i)}]$, $\Delta_t^{(1)} := \Delta_t^{(2)} := \iota_t(b_T^{-1}u_2) - \iota_t(b_T^{-1}u_m)$ and $\Delta_t^{(3)} := \Delta_t^{(4)} := \iota_t(b_T^{-1}u_m) - \iota_t(b_T^{-1}u_1)$. If all t_i $(i = 1, ..., 4)$ are distinct, $G_{t_1,t_2,t_3,t_4} = 0$ 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}, each of which can be evaluated, using the Cauchy-Schwartz inequality and (11), as

$$\begin{aligned}G_{t_1,t_2,t_3,t_4} &\leq a_T^8 M^8 \prod_{i=1}^4 \left(\mathrm{E}[\mathbf{1}_{\{|a_T^{-1}y_{t_i-1}|\leq M\}}\{\mathrm{E}(\Delta_{t_i}^{(i)})^2|\mathcal{F}_{t_i-1}\}]\right)^{1/2} \\ &\leq (a_T^{10}b_T^{-2})36M^{10}\|f\|_\infty^2|u_2-u_m||u_1-u_m| \leq (a_T^{10}b_T^{-2})9M^{10}\|f\|_\infty^2(u_2-u_1)^2;\end{aligned}$$

hence

$$G_{u_2,u_1,u_m} \leq (T^3a_T^{10}b_T^{-2})9M^{10}\|f\|_\infty^2(u_2-u_1)^2. \tag{12}$$

Since $c_T^{-1}\{r_5^M(b_T^{-1}u) - r_5^M(0)\} = o_P(1)$ for fixed u, from (12) and Theorem 15.6 of Billingsley (1968) it follows that $c_T^{-1}\sup_{|u|\leq A}|r_5^M(b_T^{-1}u) - r_5^M(0)| = o_P(1)$. 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 (12), with $\upsilon_t(u) := \varepsilon_t\mathbf{1}_{\{|\varepsilon_t-uy_{t-1}|\leq\theta\}} - m_\theta(uy_{t-1})$ instead of $\iota_t(u)$. Now, for $|u_1|, |u_2| \leq A$,

$$\begin{aligned}&\mathrm{E}\left[\{\upsilon_t(u_2) - \upsilon_t(u_1)\}^2|\mathcal{F}_{t-1}\right] \leq \{m_\theta(u_1y_{t-1}) - m_\theta(u_2y_{t-1})\}^2 \\ &+3(\theta+A|y_{t-1}|)^2\{|F(-\theta+u_2y_{t-1}) - F(-\theta+u_1y_{t-1})| + |F(\theta+u_2y_{t-1}) - F(\theta+u_1y_{t-1})|\},\end{aligned}$$

where the first term after the inequality sign corresponds to integration over values of ε_t such that $\mathbf{1}_{\{|\varepsilon_t-u_1y_{t-1}|\leq\theta\}} = \mathbf{1}_{\{|\varepsilon_t-u_2y_{t-1}|\leq\theta\}}$, and the second term – over the remaining values of ε_t. As $|m_\theta'(x)| \leq 2\|f\|_\infty(\theta+|x|)$, from the Mean Value Theorem

$$\mathrm{E}\left[\{\upsilon_t(u_2)-\upsilon_t(u_1)\}^2|\mathcal{F}_{t-1}\right] \leq (\theta+A|y_{t-1}|)^2\{4\|f\|_\infty^2(u_2-u_1)^2y_{t-1}^2+6\|f\|_\infty|u_2-u_1||y_{t-1}|\},$$

so, with $C=\{\theta+AM\sup_T(a_Tb_T^{-1})\}^2\{8\|f\|_\infty^2M^2A\sup_T(a_Tb_T^{-1})+6\|f\|_\infty M\}$, it holds that

$$\mathrm{E}\left[\mathbf{1}_{\{|a_T^{-1}y_{t-1}|\leq M\}}\{\upsilon_t(b_T^{-1}u_2)-\upsilon_t(b_T^{-1}u_1)\}^2|\mathcal{F}_{t-1}\right] \leq Ca_Tb_T^{-1}|u_2-u_1|. \tag{13}$$

Introducing

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

by an argument like for the process $r_5^M(\cdot)$ it follows from (13) and the independence of $\{\varepsilon_t\}$ that

$$\mathrm{E}\left(\{r_2^M(b_T^{-1}u_2)-r_2^M(b_T^{-1}u_m)\}^2\{r_2^M(b_T^{-1}u_m)-r_2^M(b_T^{-1}u_1)\}^2\right) \leq (T^3a_T^6b_T^{-2})M^4C^2(u_2-u_1)^2.$$

By Theorem 15.6 of Billingsley (1968), $T^{-3/4}(a_T^{-3}b_T)^{1/2}\{r_2^M(b_T^{-1}(\cdot))-r_2^M(0)\}$ is tight in $D[-A,A]$ for every fixed M (more precisely, the process can be modified like $r_5^M(b_T^{-1}(\cdot))$ earlier so that a tight cadlag sequence obtains). Since $T^{-3/4}(a_T^{-3}b_T)^{1/2}\{r_2^M(b_T^{-1}u)-r_2^M(0)\}=o_P(1)$ for every fixed u (as $\mathrm{E}\{r_2^M(b_T^{-1}u)-r_2^M(0)\}=0$, $\mathrm{E}\{r_2^M(b_T^{-1}u)-r_2^M(0)\}^2\leq Ta_T^3b_T^{-1}M^2C|u|$ using (13)), by tightness the convergence is uniform on $[-A,A]$, as asserted in part (b). ■

Proof of Proposition 2. Let sup and inf be taken over $|u|\leq b_T^{-1}A$.

For part (a), consider eq. (6) and the related definitions. The convergence of $T^{1/2}a_TL_{T,\theta}$ follows from its representation as

$$T^{1/2}a_TL_{T,\theta}=\Phi_{T,1}(\theta)\frac{(T^{-1}a_T^{-2}\sum y_{t-1}^2)^{1/2}}{T^{-1}a_T^{-2}r_3(0)},$$

where $\Phi_{T,1}(\theta)\xrightarrow{w}B(V(\theta))$ by Lemma 2(a), $T^{-1}a_T^{-2}\sum y_{t-1}^2\xrightarrow{w}\int S^2$ and $T^{-1}a_T^{-2}r_3(0)\xrightarrow{w}p_\theta(0)\int S^2$ by the CMT, and the convergences are joint by the independence of B and S.

To evaluate $r_{T,\theta}(u)$, note that $T^{-1}a_T^{-2}(\inf|r_3(u)|-\sup|r_5(u)|)$ is bounded away from $(-\infty,0]$ in probability because

$$T^{-1}a_T^{-2}\inf|r_3(u)|\geq\left||T^{-1}a_T^{-2}r_3(0)|-T^{-1}a_T^{-2}\sup|r_3(u)-r_3(0)|\right|\xrightarrow{w}p_\theta(0)\int S^2$$

by the CMT and Lemma 2(a), with $p_\theta(0)>0$ for $\theta\geq a$ and $\int S^2>0$ a.s., whereas $T^{-1}a_T^{-2}\sup|r_5(u)|=o_P(1)$ by Lemma 2(b). Hence, $\sup|r_{T,\theta}(\theta)|\leq\sup|r_1(u)|/\{\inf|r_3(u)|-\sup|r_5(u)|\}$ with probability approaching one, and using also Lemma 2(a), it obtains that $\sup|r_{T,\theta}(u)|=o_P(1)$.

Finally, from eq. (7) it is seen that does not exceed

$$\sup|e_{T,\theta}(u)|\leq\frac{\sup|r_2(u)-r_2(0)|+|L_{T,\theta}|\{\sup|r_3(u)-r_3(0)|+\sup|r_5(u)|\}+|u|\sup|r_5(u)|}{\inf|r_3(u)|-\sup|r_5(u)|},$$

where the term $L_{T,\theta}$ and the denominator were discussed above. Using Lemma 2, in the case $T^{1/2}a_T = O(b_T)$ the upper bound is seen to be

$$\left[o_P(T^{3/4}a_T^{3/2}b_T^{-1/2}) + O_P(T^{-1/2}a_T^{-1})\{o_P(Ta_T^2) + O_P(T^{1/2}a_T^2)\} + O(b_T^{-1})O_P(T^{1/2}a_T^2)\right] O_P(T^{-1}a_T^{-2})$$

for $|u| \leq b_T^{-1}A$ and reduces to $o_P(T^{-1/2}a_T^{-1})$, whereas in the case where $a_T = o(b_T)$ and $b_T = o(T^{1/2}a_T)$ it is

$$\left[o_P(T^{3/4}a_T^{3/2}b_T^{-1/2}) + O_P(T^{-1/2}a_T^{-1})\{o_P(Ta_T^2) + o_P(T^{3/4}a_T^{5/2}b_T^{-1/2})\} + O(b_T^{-1})o_P(T^{3/4}a_T^{5/2}b_T^{-1/2})\right] O_P(T^{-1}a_T^{-2})$$

and reduces to $o_P(T^{-1/4}a_T^{-1/2}b_T^{-1/2})$.

For part (b), consider eq. (8). The CMT implies that $T^{-1}a_T^{-2}r_3\left(a_T^{-1}(\cdot)\right) \xrightarrow{w} \int S^2 p_\theta\left((\cdot)S\right)$ and $T^{-1}a_T^{-1}r_4\left(a_T^{-1}(\cdot)\right) \xrightarrow{w} \int S m_\theta\left((\cdot)S\right)$ jointly on $D\left[-A, A\right]$, from where the convergence of $a_T Q_{T,\theta}\left(a_T^{-1}(\cdot)\right)$ follows. It also follows that $(\inf|r_3(u)|)^{-1} = O_P(T^{-1}a_T^{-2})$ and $\sup|r_4(u)| = O_P(Ta_T)$, which jointly with Lemma 2 and the weak convergence of $T^{-1/2}a_T^{-1}r_2(0)$ give

$$\begin{aligned}
\sup|\tilde{\phi}_\theta\left(u\right) - Q_{T,\theta}(u)| &\leq \frac{|r_2(0)| + \sup|r_2\left(u\right) - r_2(0)| + \sup|r_4(u)|\sup|r_5(u)|\left(\inf|r_3(u)|\right)^{-1}}{\inf|r_3(u)| - \sup|r_5(u)|} \\
&= \left[O_P(T^{1/2}a_T) + o_P(T^{3/4}a_T) + O_P(Ta_T)o_P(T^{3/4}a_T^2)O_P(T^{-1}a_T^{-2})\right] O_P(T^{-1}a_T^{-2}),
\end{aligned}$$

which simplifies to $o_P(T^{-1/4}a_T^{-1})$, as asserted.

Further, define

$$\chi_\theta(u) := \frac{m_\theta(u)}{up_\theta(u)} \text{ for } u \in \mathbb{R}\setminus\{0\} \text{ and } \chi_\theta(0) := h_\theta = \lim_{u\to 0}\chi(u).$$

If the density f is strictly increasing (resp. decreasing) on $(-\infty, 0)$ (resp. $(0, \infty)$), then $|\chi(u)| < 1$ for every $u \in \mathbb{R}$. Indeed,

$$m_\theta(u) - up_\theta\left(u\right) = \int_{u-\theta}^{u}(x-u)f\left(x\right)dx + \int_{u}^{u+\theta}(x-u)f\left(x\right)dx$$

and for $u > 0$ it follows that

$$m_\theta(u) - up_\theta\left(u\right) < f\left(u\right)\{\int_{u-\theta}^{u}(x-u)dx + \int_{u}^{u+\theta}(x-u)dx\} = 0,$$

whereas for $u < 0$ the opposite inequality holds (the case $u = 0$ was discussed in Remark ...). Using also the continuity of $\chi\left(\cdot\right)$, we can conclude that $H(U) = \max_{|u|\leq U}|\chi(u)| < 1$ for every $U \geq 0$ and $H(\cdot)$ is continuous. Since

$$\mathbf{1}_{\{u\neq 0\}}u^{-1}Q_{T,u} = \mathbf{1}_{\{u\neq 0\}}\frac{\sum y_{t-1}^2 p_\theta(uy_{t-1})\chi_\theta(uy_{t-1})}{\sum y_{t-1}^2 p_\theta(uy_{t-1})},$$

it obtains that

$$\sup_{|u|\leq a_T^{-1}A}|\mathbf{1}_{\{u\neq 0\}}u^{-1}Q_{T,u}| \leq H(A\max_{t\leq T}|a_T^{-1}y_{t-1}|) \xrightarrow{w} H(A\sup_{s\in[0,1]}|S(s)|),$$

the convergence from the CMT. The proof is completed by defining $H_{T,\theta} := H(A \max_{t\le T} |a_T^{-1} y_{t-1}|)$, $H_\theta := H(A \sup_{s\in[0,1]} |S(s)|)$ and observing that $H_\theta < 1$ a.s. because $\sup_{s\in[0,1]} |S(s)| < \infty$ a.s. ■

Proof of Proposition 3. For $h_\theta < 1$ and $b_T \hat{\phi}^{(0)} = O_P(1)$, it follows from Proposition 2(a), by recursive substitution, that $\sup_{i\in\mathbb{N}} |\gamma_T \hat{\phi}^{(i)}| = O_P(1)$, where $\gamma_T := \min\{T^{1/2} a_T, b_T\}$. Therefore, again from Proposition 2(a),

$$\hat{\phi}^{(i)} = L_{T,\theta} + \hat{\phi}^{(i-1)}(h_\theta + R_1^{(i)}) + R_2^{(i)},$$

where $\sup_{i\in\mathbb{N}} |R_1^{(i)}| = o_P(1)$ and $\sup_{i\in\mathbb{N}} |R_2^{(i)}| = o_P(\delta_T)$ with $\delta_T := \max\{T^{-1/2}a_T^{-1}, [(T^{-1/2}a_T^{-1})(b_T^{-1})]^{1/2}\}$. Solving for $\hat{\phi}^{(i)}$ gives $\hat{\phi}^{(i)} = \hat{\phi}_1^{(i)} + \hat{\phi}_2^{(i)}$, where

$$\hat{\phi}_1^{(i)} = \hat{\phi}^{(0)} \prod_{j=1}^{i} (h_\theta + R_1^{(i)}) + L_{T,\theta} \sum_{j=1}^{i} \prod_{k=j+1}^{i} (h_\theta + R_1^{(k)}), \quad \hat{\phi}_2^{(i)} = \sum_{j=1}^{i} R_2^{(j)} \prod_{k=j+1}^{i} (h_\theta + R_1^{(k)}).$$

For the terms in $\hat{\phi}_1^{(i)}$ we find that

$$P\left(\hat{\phi}^{(0)} \prod_{j=1}^{i} (h_\theta + R_1^{(i)}) \to 0 \text{ as } i \to \infty\right) \le P\left(\sup_{i\in\mathbb{N}} |R_1^{(i)}| < \tfrac{1}{2}(1-h_\theta)\right) \to 1$$

as $T \to \infty$, and for every $\eta \in (0, 1-h_\theta)$,

$$P\left(\exists \lim_{i\to\infty} \prod_{j=1}^{i} (h_\theta + R_1^{(i)}) \text{ and } (1-h_\theta+\eta)^{-1} \le \lim_{i\to\infty} \prod_{j=1}^{i} (h_\theta + R_1^{(i)}) \le (1-h_\theta-\eta)^{-1}\right)$$
$$\le P\left(\sup_{i\in\mathbb{N}} |R_1^{(i)}| < \eta\right) \to 1$$

as $T \to \infty$. So if we define $\hat{\phi}_1^{(\infty)} := \mathbf{1}\{\exists \lim_{i\to\infty} \prod_{j=1}^{i} (h_\theta + R_1^{(i)})\} \lim_{i\to\infty} \prod_{j=1}^{i} (h_\theta + R_1^{(i)}) - (1-h_\theta)^{-1}$, it holds that $\hat{\phi}_1^{(\infty)} = o_P(1)$ and

$$P\left(\hat{\phi}_1^{(i)} \to (1-h_\theta)^{-1} L_{T,\theta} + \hat{\phi}_1^{(\infty)} L_{T,\theta} \text{ as } i \to \infty\right) \to 1$$

as $T \to \infty$. From here

$$P\left(\text{limsup}_{i\to\infty} |\hat{\phi}_1^{(i)} - (1-h_\theta)^{-1} L_{T,\theta}| \le |\hat{\phi}_1^{(\infty)}||L_{T,\theta}| + \sup_{i\in\mathbb{N}} |\hat{\phi}_2^{(i)}|\right) \to 1.$$

As $L_{T,\theta} = T^{-1/2} a_T^{-1}$ and, in view of the magnitude orders of $R_1^{(i)}$ and $R_2^{(i)}$, $\sup_{i\in\mathbb{N}} |\hat{\phi}_2^{(i)}| = o_P(\delta_T)$, it follows that $\limsup_i |\delta_T^{-1} \hat{\phi}^{(i)} - \delta_T^{-1}(1-h_\theta)^{-1} L_{T,\theta}| = o_P(1)$ as $T \to \infty$, which implies the conclusions in part (a) about iterated limits.

Consider next limits along a path $i = \psi(T)$ and define $\nu_T = \lfloor \{\psi(T)\}^{1/2} \rfloor$. Note that $\nu_T \to \infty$ as $T \to \infty$. Similarly to the previous argument, by the hypothesis on $\psi(T)$ and since $\sup_{i\in\mathbb{N}} |R_1^{(i)}| = o_P(1)$ it holds that

$$|\hat{\phi}^{(0)}| \left| \prod_{j=1}^{\nu_T} (h_\theta + R_1^{(i)}) \right| \le |\hat{\phi}^{(0)}| (h_\theta + \sup_{i\in\mathbb{N}} |R_1^{(i)}|)^{\nu_T} = o_P(T^{-1/2} a_T^{-1})$$

and $\sum_{j=1}^{\nu_T}\prod_{k=j+1}^{\nu_T}(h_\theta+R_1^{(k)}) \xrightarrow{P} (1-h_\theta)^{-1}$ as $T\to\infty$. So $\hat{\phi}_1^{(\nu_T)} = (1-h_\theta)^{-1}L_{T,\theta}+o_P(T^{-1/2}a_T^{-1}) = O_P(T^{-1/2}a_T^{-1})$, and as $\hat{\phi}_2^{(\nu_T)} = o_P(\delta_T^{(1)})$, we find that $\hat{\phi}^{(\nu_T)} = O_P(\delta_T^{(1)})$, $\delta_T^{(1)} := \delta_T = \max\{T^{-1/2}a_T^{-1}, [(T^{-1/2}a_T^{-1})(b_T^{-1})]^{1/2}\}$. We can think of $\hat{\phi}^{(2v_T)}$ as obtained from an iteration with initial value $\hat{\phi}_1^{(\nu_T)}$ and with initial magnitude order $(\delta_T^{(1)})^{-1}$ instead of b_T. By the same argument as for $\hat{\phi}^{(v_T)}$, we can conclude that $\hat{\phi}_1^{(2\nu_T)} = (1-h_\theta)^{-1}L_{T,\theta} + o_P(T^{-1/2}a_T^{-1})$, $\hat{\phi}_2^{(2\nu_T)} = o_P(\delta_T^{(2)})$ and $\hat{\phi}^{(2\nu_T)} = O_P(\delta_T^{(2)})$, $\delta_T^{(2)} := \max\{T^{-1/2}a_T^{-1}, [(T^{-1/2}a_T^{-1})^3(b_T^{-1})]^{1/4}\}$. Continuing by induction, $\hat{\phi}_1^{(\nu_T^2)} = (1-h_\theta)^{-1}L_{T,\theta}+o_P(T^{-1/2}a_T^{-1})$ and $\hat{\phi}_2^{(\nu_T^2)} = o_P(\delta_T^{(T)})$, $\delta_T^{(T)} := \max\{T^{-1/2}a_T^{-1}, [(T^{-1/2}a_T^{-1})^{2^T-1}(b_T^{-1})]^{2^{-T}}\}$. For the b_T considered in part (b), $\delta_T^{(T)} = O(T^{-1/2}a_T^{-1})$ and thus, $\hat{\phi}^{(\nu_T^2)} = (1-h_\theta)^{-1}L_{T,\theta} + o_P(T^{-1/2}a_T^{-1})$. Finally, iterating from $\hat{\phi}^{(\nu_T^2)}$ to $\hat{\phi}^{(\psi(T))}$ yields $\hat{\phi}^{(\psi(T))} = (1-h_\theta)^{-1}L_{T,\theta}+o_P(T^{-1/2}a_T^{-1})$, from where the convergence in part (b) obtains. ■

Proof of Proposition 4. The proof is similar to that of Proposition 3(b). In view of Proposition 2(b) we can write

$$|\hat{\phi}^{(\lfloor T^{\nu/2}\rfloor)}| \le |\hat{\phi}^{(0)}| H_{T,\theta}^{\lfloor T^{\nu/2}\rfloor} + o_P(T^{-1/4}a_T^{-1}) = o_P(T^{-1/4}a_T^{-1}),$$

since $H_{T,\theta}^{\lfloor T^{\nu/2}\rfloor} \xrightarrow{P} 0$ exponentially fast. Then the desired convergences follow by applying Proposition 3(b) with $b_T = T^{1/4}a_T$ and $\psi(T) = \lfloor T^\nu\rfloor - \lfloor T^{\nu/2}\rfloor$ to the iteration started at $\hat{\phi}^{(\lfloor T^{\nu/2}\rfloor)}$. ■

Proof of Proposition 5. We use the notation introduced in the proof of Proposition 3. Fix an $A>0$ and let $R_{1,T} := \sup_{|u|\le A}\left|r_{T,\theta}(\gamma_T^{-1}u)\right| = o_P(1)$ and $R_{2,T} := \sup_{|u|\le A}\left|e_{T,\theta}(\gamma_T^{-1}u)\right| = o_P(\delta_T)$, where $r_{T,\theta}$ and $e_{T,\theta}$ are as in Proposition 2(a). By the same proposition, the event $\mathcal{A} := \{|\gamma_T\hat{\phi}^{(i)}| < A, \forall i\in\mathbb{N}\}$ is contained in

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

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

$$\begin{aligned} \hat{\phi}_1^{(i)} &= \hat{\Delta}_T\prod_{j=1}^{i}(h_\theta + R_1^{(i)}) + L_{T,\theta}[(1-h_\theta)^{-1} + \lambda_T^{(i)}], \qquad (14)\\ \lambda_T^{(i)} :&= \sum_{j=1}^{i}\Big[\prod_{k=j+1}^{i}(h_\theta + R_1^{(k)}) - h_\theta^{i-j}\Big] + (1-h_\theta)^{-1}\Big[\prod_{j=1}^{i}(h_\theta + R_1^{(i)}) - h_\theta^i\Big]. \end{aligned}$$

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

$$\left|\prod_{j=1}^{i}(h_\theta + R_1^{(i)})\right| \ge \left(\frac{h_\theta+1}{2}\right)^i \quad \text{and} \quad \sum_{j=1}^{i}\prod_{k=j+1}^{i}|h_\theta + R_1^{(k)}| < \frac{(3h_\theta-1)^i}{2^i(h_\theta-1)}. \qquad (15)$$

Define

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

Then $I_T = O_P(1)$ because, by hypothesis, $|\gamma_T \hat{\Delta}_T|$ is bounded away from zero in probability, whereas $\gamma_T L_{T,\theta} = O_P(1)$ by Proposition (a). Further, for outcomes in $\mathcal{A}_1 \cap \mathcal{A}_2$

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

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

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

As $I_T = O_P(1)$ and $R_{1,T} = o_P(1)$, it follows that $\lambda_T^{(I_T)} = o_P(1)$, and since $\gamma_T L_{T,\theta} = O_P(1)$, $P(\mathcal{A}_3) \to 1$ for $\mathcal{A}_3 := \{|\gamma_T L_{T,\theta} \hat{\lambda}_T^{(I_T)}| < A\}$. Recalling eq. (14) and the definition of I_T, we can conclude that for outcomes in $\cap_{i=1}^3 \mathcal{A}_i$, $|\gamma_T \hat{\phi}_1^{(I_T)}| > 2A$. Recalling also (15), for such outcomes $|\gamma_T \hat{\phi}_2^{(I_T)}| \leq \gamma_T R_{2,T} \{2^{-1}(3h_\theta - 1)\}^{I_T} (h_\theta - 1)^{-1} =: \kappa_T$, and since $P(\mathcal{A}_4) \to 1$ for $\mathcal{A}_4 := \{\kappa_T < A\}$, we finally obtain that, for outcomes in $\cap_{i=1}^4 \mathcal{A}_i$, $|\gamma_T \hat{\phi}^{(I_T)}| \geq |\gamma_T \hat{\phi}_1^{(I_T)}| - |\gamma_T \hat{\phi}_2^{(I_T)}| > A$. Therefore, $\mathcal{A} \subset (\cap_{i=1}^4 \mathcal{A}_i)^c$ (with c denoting complement). Recalling that $\mathcal{A} \subset \mathcal{A}_1$, we find that $\mathcal{A} \subset (\cap_{i=2}^4 \mathcal{A}_i)^c$, where $P(\cap_{i=2}^4 \mathcal{A}_i) \to 1$. This proves that $P(\mathcal{A}) \to 0$. ■

7 References

Billingsley P. (1968). Convergence of Probability Measures, John Wiley and Sons

Chan, N.H. and L.T. Tran (1989). On the first order autoregressive process with infinite variance. Econometric Theory 5: 354-362

Hendry D., C. Santos (2006). Automatic tests for super exogeneity. Unpublished Paper, University of Oxford

Engler E. and B. Nielsen (2007). The empirical process of autoregressive residuals. Unpublished paper, University of Oxford.

Johansen S. and B. Nielsen (2008). An analysis of the indicator saturation estimator as a robust regression estimator. To appear in Castle, J. and Shephard, N. The Methodology and Practice of Econometrics: Festschrift for David F. Hendry. Oxford: Oxford University Press.

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

Knight K. (1991). Limit theory for M-estimates in an integrated infinite variance process. Econometric Theory 7: 200-212.

Lee S. and C.Z. Wei (1999). On residual empirical processes of stochastic regression models with applications to time series. Annals of Statistics 27: 237-261

Resnick S. and P. Greenwood (1979). A bivariate stable characterization and domains of attraction. Journal of Multivariate Analysis 9: 206-221

Santos C., Hendry D. and S. Johansen (2008). Selecting a regression saturated by indicators. Computational Statistics 23, 317-335 and Erratum 337-339