Giuseppe Cavaliere, Iliyan Georgiev

Exploiting Infinite Variance through Dummy Variables in Non-Stationary Autoregressions

Quaderni di Dipartimento

Serie Ricerche 2013, n. 1
ISSN 1973-9346

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche “Paolo Fortunati”

Exploiting Infinite Variance through Dummy Variables in Non-Stationary Autoregressions*

Giuseppe Cavaliere
Università di Bologna

Iliyan Georgiev
Universidade Nova de Lisboa

This version: December 2012

Abstract

We consider estimation and testing in finite-order autoregressive models with a (near) unit root and infinite-variance innovations. 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 preliminary parameter estimator and on the threshold which ensure that (i) the dummy-based estimator is consistent at higher rates than the OLS estimator, (ii) an asymptotically normal test statistic for the unit root hypothesis can be derived, and (iii) order of magnitude gains of local power are obtained.

1 Introduction

In this paper we study the problem of estimation and unit root [UR] testing in a finite-order autoregressions [AR] with infinite variance [IV] innovations.

Specifically, consider first the case where $\{y_t\}$ is the AR(1) process (the case of higher order processes will be discussed later)

$$\Delta y_t := y_t - y_{t-1} = \phi y_{t-1} + \varepsilon_t,\ (t = 1, ..., T) \tag{1.1}$$

initialized at some fixed value y_0. The innovations ε_t are infinite-variance i.i.d. and belong to the domain of attraction of an α-stable distribution, $\alpha \in (0, 2)$, and ϕ is either 0 (i.e., y_t is a random walk) or 'close' to 0 (i.e., y_t has an AR root near unity).

Estimation and inference on ϕ have been widely studied in the statistical and econometric literature, see Samarakoon and Knight (2009) and references therein. Typically, ϕ is estimated either using ordinary least squares [OLS] or robust M-estimation. In the former case, which

*We thank Pentti Saikkonen (Co-Editor), three anonymous referees and Keith Knight for their very useful comments. Support was provided by the Fundação para a Ciência e a Tecnologia, Portugal, through grants PTDC/EGE-ECO/108620/2008. Cavaliere also thanks the Danish Council for Independent Research, Sapere Aude | DFF Advanced Grant (Grant nr: 12-124980) for financial support. Address correspondence to: Giuseppe Cavaliere, Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, I-40126 Bologna, Italy; e-mail giuseppe.cavaliere@unibo.it.

is considered, e.g., in Chan and Tran (1989), Knight (1989) and Phillips (1990), inference is based on the estimator

$$\hat{\phi}_{OLS} := \frac{\sum_{t=1}^{T} y_{t-1}\Delta y_t}{\sum_{t=1}^{T} y_{t-1}^2} \ . \tag{1.2}$$

It is well known that under a unit root this estimator is consistent at the T rate for all $\alpha \in (0,2)$ and has a non-standard asymptotic distribution; see, among others, Phillips (1990).

In the latter case, see Knight (1989, 1991) and Samarakoon and Knight (2009), inference is based on the estimator

$$\tilde{\phi} := \arg\min_{\phi} \sum_{t=1}^{T} \rho\left(\Delta y_t - \phi y_{t-1}\right) \tag{1.3}$$

for some convex function ρ, or on a solution $\tilde{\phi}_M$ of the equation

$$\sum_{t=1}^{T} y_{t-1}\psi(\Delta y_t - \tilde{\phi}_M y_{t-1}) = 0 \tag{1.4}$$

for some function ψ (typically, $\psi = \rho'$). Knight (1989, 1991) and Samarakoon and Knight (2009) provide a set of sufficient conditions on ρ, ψ and ε_t ensuring that the M-estimators $\tilde{\phi}$, $\tilde{\phi}_M$ are consistent at a rate faster than the OLS rate and inference on ϕ is asymptotically Gaussian under the UR hypothesis.

In this paper, we take an alternative route by analysing estimators obtained by dummying out residuals (say, $\hat{\varepsilon}_t$) which are 'large', i.e., exceed a given threshold (say, $\hat{\theta}$).[1] That is, after an initial estimator of the parameters is obtained, observations with large residuals are discarded, and the model is re-estimated on the maintained observations only.[2] The approach is commonly implemented in applied econometric works by including a set of impulse dummies in the estimated equation.

Further, we study the iteration of the above procedure. Thus, given an initial estimator of ϕ, we analyse the iterative procedure consisting of (i) computing residuals, (ii) introducing dummy variables for those of them which exceed some threshold and (iii) reestimating ϕ (and, optionally, the threshold), possibly until convergence. The iteration can be related to the empirical strategy of re-examining the residuals and adjusting the set of dummy variables until the estimates stabilize, or become "robust" (insensitive) to the exclusion of further observations. In the paper we discuss the asymptotic properties of the iterated estimator, which is closely related to M-estimation based on (1.4) with $\psi()$ chosen as Huber's skip function. However, this ψ does not satisfy the smoothness hypotheses usually required for M estimation, see Knight (1989, 1991) and Samarakoon and Knight (2009).

Rather surprisingly, little is known about the asymptotic properties of dummy-based estimators, in spite of their rather simple computation and wide use in practice. In the finite-variance case, they have been recently analysed by Johansen and Nielsen (2011). Their

[1] The definition of our estimators involves dummying out the large time-series innovations, although they perfectly fit the maintained infinite-variance model and although their number is large. The idea of using dummy variables in number proportional to the sample size, solely as a means to construct estimators or test statistics, has been recently employed also by Hendry, Johansen and Santos (2008), and Johansen and Nielsen (2009, 2010).

[2] The resulting estimator is called the one-step Huber-skip estimator.

set up covers both autoregressions with a unit (or local-to-unit) root and stationary autoregressions, and large-sample properties are obtained under the assumption of finite fourth moments of the innovations. Near-UR autoregressions augmented with dummy variables have also been analysed in Cavaliere and Georgiev (2009), where it is shown that when finite-variance innovations are contaminated by infrequent, large outliers, the inclusion of dummy variables increases the efficiency of the AR parameter estimator (leaving the consistency rate unchanged) and gives rise to UR tests with significant power gains.

So far, no result is available for possibly non-stationary autoregressions with infinite-variance innovations, where large realizations are more likely to occur. We find that, with respect to the finite-variance case, dummy-based estimation under infinite-variance innovations has some additional attractive features. Due to the link with M-estimation, the iterated dummy-based estimator shares the two basic asymptotic properties of the M-estimators discussed by Knight (1989, 1991), though not belonging to their class. These are the properties of a fast consistency rate and Gaussian asymptotic (null) distribution of UR test statistics. As we will show, the iterated dummy-based estimator improves upon the consistency rate of the initial estimator, as long as the latter is reasonable (for $\alpha > 1$, the OLS rate suffices). At the same time, the dummy-based estimator is rather straightforward to compute, with its iterated version being no more demanding than a feasible GLS estimator. Hence, the desirable features of both least squares (simplicity) and M-estimation (asymptotic properties) are preserved.

A further, important feature of the dummy-based approach is that – as it will be shown in this paper – its asymptotics can be derived under fairly transparent conditions. It is mainly required that the innovations have symmetric density f and belong to the domain of attraction of a stable distribution (with index $\alpha \in (0, 2)$). This contrasts with the case of general robust estimators discussed by Knight (1991) and Samarakoon and Knight (2009): for instance, even in the case where ρ of (1.3) is convex and differentiable, further conditions involving the derivatives ρ' and ρ'' and their relations to ε_t are required; see e.g. conditions A2 and A3 in Samarakoon and Knight (2009).

Formally, under specification (1.1), our object of study is the estimator $\tilde{\phi}$ of ϕ defined by

$$\tilde{\phi}(\hat{\phi}, \hat{\theta}) := \frac{\sum_{t=1}^{T} y_{t-1} \Delta y_t \mathbb{I}_{\{|\hat{\varepsilon}_t| \leq \hat{\theta}\}}}{\sum_{t=1}^{T} y_{t-1}^2 \mathbb{I}_{\{|\hat{\varepsilon}_t| \leq \hat{\theta}\}}} , \tag{1.5}$$

where, with $\hat{\phi}$ a preliminary estimator of ϕ, $\hat{\varepsilon}_t = \Delta y_t - \hat{\phi} y_{t-1}$ $(t = 1, ..., T)$ are the residuals based on this preliminary estimator and $\hat{\theta}$ is a scale statistic (e.g., a quantile of the empirical distribution function of $|\hat{\varepsilon}_t|$). Further objects of study are the iterates of estimator (1.5), possibly augmented with iteration over $\hat{\theta}$, as well as modifications of $\tilde{\phi}$ suitable for AR(p) processes with $p > 1$. Notice that $\tilde{\phi}(\hat{\phi}, \hat{\theta})$ corresponds to the OLS estimator of ϕ from the augmented regression

$$\Delta y_t = \phi y_{t-1} + \varphi' \mathbf{D}_t + e_t, \; t = 1, ..., T$$

where $\mathbf{D}_t$ is a vector of impulse dummies, one for each t such that $|\hat{\varepsilon}_t|$ exceeds $\hat{\theta}$.

When the true ϕ is zero, a natural benchmark in terms of asymptotic properties is $\tilde{\phi}\,(0, \theta)$, with θ a positive constant. Under the assumptions we make in the next section, $\tilde{\phi}\,(0, \theta)$ vanishes at the same rate as the M-estimators studied by Knight (1989, 1991), and $(\sum_{t=1}^{T} y_{t-1}^2)^{1/2} \tilde{\phi}\,(0, \theta)$ has Gaussian limiting distribution like the UR test statistics of Knight

(1989, 1991) and the rank test of Hasan (2001). With respect to $\tilde{\phi}(\hat{\phi},\hat{\theta})$, as opposed to $\tilde{\phi}\,(0,\theta)$, there are two issues to tackle: (i) $\hat{\theta}$ is, generally, random and, (ii) $\{\varepsilon_t\}$ are estimated by $\{\hat{\varepsilon}_t\}$. To clarify ideas, we discuss the two issues (estimation of the threshold and ϕ) first separately, and then, in a joint setup. The main mathematical tools employed are weak convergence of weighted empirical processes and an implied asymptotic expansion of $\tilde{\phi}(\hat{\phi},\hat{\theta})$.

The paper is organized as follows. In section 2 we discuss the basic assumptions underlying the reference model. In sections 3 and 4 we analyse two special cases of estimator (1.5), respectively, with $\hat{\phi}$ fixed at 0 and with $\hat{\theta}$ fixed at some $\theta > 0$; iterates of estimator (1.5) with fixed $\hat{\theta}$ are also studied. In section 5 the general iteration over both ϕ and θ is considered. In section 6 we generalize our results to the case of higher order autoregressions. Section 7 contains simulation evidence, whereas section 8 concludes.

2 Model and assumptions

In this section we introduce and discuss the basic assumptions on the reference AR(1) process (1.1). Assumption $\mathcal{E}$ below summarizes the stochastic properties of the innovations ε_t, while Assumption $\mathcal{Y}$ determines the dynamic properties of the autoregression for y_t. The AR(1) assumption is relaxed in section 6.

Assumption $\mathcal{E}$. (i) $\{\varepsilon_t\}_{t=1}^{\infty}$ *is an i.i.d. sequence of random variables which have* $\mathrm{E}\varepsilon_1^2 = \infty$ *and belong to the domain of attraction of a stable distribution with index* $\alpha \in (0,2)$*.* (ii) ε_1 *has density* f *with respect to Lebesgue measure and* f *is a continuous even function, positive a.e., with* $\sup_{x\in\mathbb{R}} |xf\,(x)\,| < \infty$*.*

Some comments are due.

REMARK 2.1. Assumption $\mathcal{E}$(i) and the symmetry part of Assumption $\mathcal{E}$(ii) imply the existence of an α-stable process S in $D\,[0,1]$ and a normalizing sequence $a_T = T^{1/\alpha}\ell(T)$, with $\ell(\cdot)$ standing for a slowly varying function at ∞, such that $a_T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t \xrightarrow{w} S$ in $D[0,1]$ as $T\to\infty$ (Resnick and Greenwood, 1979).

REMARK 2.2. Under the assumption of a continuous f, symmetry of the distribution is equivalent to $\mathrm{E}\left(\varepsilon_1\mathbb{I}_{\{|\varepsilon_1|\le\theta\}}\right) = 0$ for all $\theta > 0$. With $\psi\,(x) := x\mathbb{I}_{[-1,1]}\,(x)$ this condition can be written as $\mathrm{E}\{\psi\left(\theta^{-1}\varepsilon_1\right)\} = 0$, which (for a different ψ) is used in Knight (1989) in the analysis of scale-parameter estimation. As long as estimated scale quantities satisfy $P(\hat{\theta} > \theta') \to 1$, the assumption can be relaxed to

$$\mathrm{E}(\varepsilon_1\mathbb{I}_{\{|\varepsilon_1|\le\theta\}}) = 0 \text{ for all } \theta \ge \theta' > 0 \tag{2.6}$$

without affecting the results. For a continuous f, the latter is equivalent to $\mathrm{E}(\varepsilon_1\mathbb{I}_{\{|\varepsilon_1|\le\theta'\}}) = 0$ and symmetry of the tails: $f\,(\theta) = f\,(-\theta)$ for $\theta > \theta'$. Although (2.6) is more general, there is a trade-off between higher generality (larger θ' increases the class of admissible distributions) and the need to determine θ' in practice (larger θ' are more difficult to determine empirically).

REMARK 2.3. The i.i.d. assumption and smoothness assumptions on f are common in the literature on empirical processes (see, e.g., Koul, 2002, and Engler and Nielsen, 2009). □

Assumption $\mathcal{Y}$. *The process* $\{y_t\}_{t=1}^{T}$ *satisfies* $\Delta y_t = \phi y_{t-1} + \varepsilon_t$ $(t = 1,...,T)$*, where* $\phi = -d_T^{-1}c$ *with* $d_T := T^{1/2}a_T$ *and* $c \in \mathbb{R}$*, and* y_0 *fixed.*

Remark 2.4. For $c = 0$, $\{y_t\}_{t=1}^T$ is a random walk with infinite-variance innovations. For $c \neq 0$ the process has a root near unity, in the sense that $\phi \to 0$ as $T \to \infty$. However, in contrast to the finite-variance case, where the choice $\phi = -c/T$ yields a non-trivial local power function of UR tests (see, e.g., Phillips, 1987), for our tests such a function is obtained under the faster shrinkage rate of d_T^{-1}, as postulated in Assumption $\mathcal{Y}$. Under infinite variance, the choice $\phi = -c/T$ defines a so-called moderate deviation from a UR (parametrized in the finite-variance case by $\phi = O(T^{-\beta})$ for $\beta \in (0, 1)$; see Phillips and Magdalinos, 2007a,b, and references therein); it is considered in Remarks 3.3 and 4.4. □

Throughout the paper, we use also the following notation related to the distribution of $\{\varepsilon_t\}$:

$$p_\theta(x) := \mathrm{E}\left(\mathbb{I}_{\{|\varepsilon_1 - x| \le \theta\}}\right), \ m_\theta(x) := \mathrm{E}\left(\varepsilon_1 \mathbb{I}_{\{|\varepsilon_1 - x| \le \theta\}}\right), \ V(\theta) := \mathrm{E}(\varepsilon_1^2 \mathbb{I}_{\{|\varepsilon_1| \le \theta\}})$$

and F for the cumulative distribution function of ε_1. Under assumption $\mathcal{E}$, $V()$ is strictly increasing on $[0, \infty)$.

Finally, the quantity $h_\theta := 2\theta f(\theta)/p_\theta(0)$ for $\theta > 0$ and $h_0 := 1$ (so that $h_{(\cdot)}$ is right-continuous at zero) will play a special role in the analysis of the iterative estimators.

Remark 2.5. It will turn out important whether h_θ is below or above unity. Under Assumption $\mathcal{E}$(ii) 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 $f(\theta) < f(\theta')$; hence, $h_\theta < 1$ 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 (see, e.g., Yamazoto, 1978). Moreover, even if f is plurimodal, for large θ it will necessarily hold that $h_\theta < 1$, because $h_\theta \to 0$ as $\theta \to \infty$. Nevertheless, distributions satisfying Assumption $\mathcal{E}$ and having $h_\theta > 1$ for *some* $\theta > 0$ do exist, see the example in section 7. □

3 A simple, benchmark estimator

In this section we consider a benchmark estimator of ϕ which is obtained by dummying out observations where $|\Delta y_t|$ exceeds an estimated threshold $\hat{\theta}$. Formally, this corresponds to the choice $\hat{\phi} = 0$ in (1.5)[3]:

$$\tilde{\phi}(0, \hat{\theta}) = \frac{\sum y_{t-1} \Delta y_t \mathbb{I}_{\{|\Delta y_t| \le \hat{\theta}\}}}{\sum y_{t-1}^2 \mathbb{I}_{\{|\Delta y_t| \le \hat{\theta}\}}} = \phi + \frac{\sum y_{t-1} \varepsilon_t \mathbb{I}_{\{|\Delta y_t| \le \hat{\theta}\}}}{\sum y_{t-1}^2 \mathbb{I}_{\{|\Delta y_t| \le \hat{\theta}\}}}. \tag{3.7}$$

The results for this estimator, besides their independent interest, are needed in the case where the preliminary estimator $\hat{\phi}$ depends on the data, since in that case we rely on an expansion of $\tilde{\phi}(\cdot, \hat{\theta})$ with leading term $\tilde{\phi}(0, \hat{\theta})$.

To formulate our first proposition, we make use of two limits implied by Assumptions $\mathcal{E}$ and $\mathcal{Y}$ as $T \to \infty$.[4] First, it holds that $a_T^{-1} y_{\lfloor T \cdot \rfloor} \xrightarrow{w} S$ in $D[0, 1]$ (cf. Remark 2.1). Second,

$$\left(\sum y_{t-1}^2\right)^{-1/2} \sum y_{t-1} \varepsilon_t \mathbb{I}_{\{|\varepsilon_t| \le (\cdot)\}} \xrightarrow{w} B(V(\cdot)) \tag{3.8}$$

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

[4] Due to the fast convergence of $\phi = -c d_T^{-1}$ to zero, the parameter c appears in none of the limits.

in $D[0,\infty)$, where B is a standard Brownian motion independent of S (see Lemma A.1(a) in the Appendix). It remains to specify how these convergences and the behaviour of $\hat{\theta}$ are related to each other.

Proposition 1 *Let Assumptions $\mathcal{E}$ and $\mathcal{Y}$ hold. If for some random variable Θ, a.s. positive, it holds that $(a_T^{-1}y_{\lfloor T\cdot\rfloor},\hat{\theta}) \xrightarrow{w} (S,\Theta)$ as random elements of the product space $D[0,1]\times\mathbb{R}$ as $T\to\infty$, and if B of (3.8) is independent of (S,Θ), then, for $d_T := T^{1/2}a_T$,*

$$d_T(\tilde{\phi}(0,\hat{\theta})-\phi) \xrightarrow{w} ch_\Theta + \frac{\{V(\Theta)\}^{1/2}}{p_\Theta(0)}\frac{N(0,1)}{(\int S^2)^{1/2}}.$$

Further, for any $\hat{\zeta}$ satisfying $(a_T^{-1}y_{\lfloor T\cdot\rfloor},\hat{\theta},\hat{\zeta}) \xrightarrow{w} \left(S,\Theta,\{V(\Theta)\}^{1/2}/p_\Theta(0)\right)$ it holds that

$$\frac{1}{\hat{\zeta}}\left(\sum y_{t-1}^2\right)^{1/2}(\tilde{\phi}(0,\hat{\theta})-\phi) \xrightarrow{w} c\frac{h_\Theta p_\Theta(0)}{\{V(\Theta)\}^{1/2}}\left(\int S^2\right)^{1/2}+N(0,1). \tag{3.9}$$

The standard Gaussian variable in both limits is independent of (S,Θ).

Some remarks are due.

Remark 3.1. For B and (S,Θ) to be independent, it is sufficient that Θ be $\sigma(S)$-measurable. A $\hat{\theta}$ that converges (in probability) to a constant, like a quantile of the sample distribution of $|\Delta y_t|$, is the simplest example. An example of a random Θ is obtained, e.g., for $\hat{\theta} = (\max_{t\le T}|y_t|)^{-1}[\sum(\Delta y_t)^2]^{1/2}$. In this case $(a_T^{-1}y_{\lfloor T\cdot\rfloor},\hat{\theta}) \xrightarrow{w} (S,\Theta)$ with $\Theta = (\sup_{[0,1]}|S|)^{-1}[S]_1^{1/2}$, $[S]_1$ being the quadratic variation of S at unity.

Remark 3.2. From (3.9) it can be seen that

$$\xi_T(0,\hat{\theta}) := \frac{1}{\hat{\zeta}}\left(\sum y_{t-1}^2\right)^{1/2}\tilde{\phi}(0,\hat{\theta}) \xrightarrow{w} -c\frac{(1-h_\Theta)p_\Theta(0)}{\{V(\Theta)\}^{1/2}}(\int S^2)^{1/2}+N(0,1). \tag{3.10}$$

In particular, under the UR null hypothesis $\phi=0$ the statistic $\xi_T(0,\hat{\theta})$ is asymptotically $N(0,1)$. In the limit, the power properties of UR tests based on $\xi_T(0,\hat{\theta})$ against the (local) alternative $\phi=-d_T^{-1}c$ ($c>0$) depend on h_Θ. In the typical case with $h_\Theta<1$ a.s., one-sided tests have non-trivial asymptotic power, whereas OLS-based UR tests are known to have asymptotic power equal to size. If $h_\Theta\ge 1$ a.s., asymptotic power does not exceed size also for tests based on $\xi_T(0,\hat{\theta})$.

Remark 3.3. From the argument in section A.1 of the appendix, it follows that, for $\alpha\in(1,2)$ and $h_\Theta<1$ a.s., a UR test based on $\xi_T(0,\hat{\theta})$ is consistent against any local alternative $\phi=-c/T$ ($c\neq 0$). This is in contrast with OLS based UR tests, which are never consistent against these alternatives. For $\alpha\in(0,1]$, consistency against $\phi=-c/T$ ($c\neq 0$) can also be conjectured to hold and the simulation evidence in section 7 confirms the conjecture.

Remark 3.4. The assumption that $\hat{\theta}$ has a weak limit implies, amongst other things, that it is stochastically bounded. This is crucial in order to obtain the d_T^{-1} convergence rate of $\tilde{\phi}(0,\hat{\theta})$. If we let the threshold grow at the rate of T^r ($r\in(0,1/\alpha)$), the convergence in (3.8) does not hold even pointwise. Instead, it could be shown that

$$T^{r(\alpha/2-1)}\mu_T(\sum y_{t-1}^2)^{-1/2}\sum y_{t-1}\varepsilon_t\mathbb{I}_{\{|\varepsilon_t|\le\theta T^r\}} \xrightarrow{w} B(\theta^{2-\alpha}) \tag{3.11}$$

for some slowly varying sequence μ_T and for every fixed $\theta > 0$. This leads to the inefficient estimator $\tilde{\phi}(0, \theta T^r)$, whose convergence rate under the UR null is $d_T^{-1} T^{r(1-\alpha/2)} \mu_T^{-1}$, slower than d_T^{-1}. Similarly, also the practice of dummying out residuals exceeding a fixed multiple of the residual standard deviation is likely to compromise the d_T^{-1} rate, since $\hat{\sigma}^2 = T^{-1} \sum (\Delta y_t)^2$ is not stochastically bounded.

Remark 3.5. Two examples of eligible estimators $\hat{\zeta}$ in (3.9) are

$$\hat{\zeta}_1 := T^{1/2} \frac{[\sum (\Delta y_t)^2 \mathbb{I}_{\{|\Delta y_t| \leq \hat{\theta}\}}]^{1/2}}{\sum \mathbb{I}_{\{|\Delta y_t| \leq \hat{\theta}\}}}, \quad \hat{\zeta}_2 := T^{-1/2} \frac{[\sum y_t^2][\sum (\Delta y_t)^2 \mathbb{I}_{\{|\Delta y_t| \leq \hat{\theta}\}}]^{1/2}}{\sum y_t^2 \mathbb{I}_{\{|\Delta y_t| \leq \hat{\theta}\}}}.$$

This follows from the fact that the pointwise convergences $T^{-1} \sum (\Delta y_t)^2 \mathbb{I}_{\{|\Delta y_t| \leq \theta\}} \xrightarrow{P} V(\theta)$ and $T^{-1} \sum \mathbb{I}_{\{|\Delta y_t| \leq \theta\}} \xrightarrow{P} p_\theta(0)$ are uniform on compacts, because the involved functions are non-decreasing in θ and the limits are continuous, whereas $(\sum y_t^2)^{-1} \sum y_t^2 \mathbb{I}_{\{|\Delta y_t| \leq \theta\}} = p_\theta(0) + o_P(1)$ again uniformly on compacts (see the proof of Lemma A.1(c)). $\square$

4 An iterative estimator with a fixed threshold

In the previous section we analysed a benchmark estimator of ϕ where impulse dummy variables based on Δy_t are used. We now turn to the case where dummy variables are based on general residuals of the form $\hat{\varepsilon}_t := \Delta y_t - \hat{\phi}^{(0)} y_{t-1}$ instead of Δy_t; here $\hat{\phi}^{(0)}$ is a preliminary estimator of ϕ, e.g., its OLS estimator. Thus, in this section we study the procedure consisting of (i) calculating residuals, given the estimator $\hat{\phi}^{(0)}$, (ii) introducing dummy variables for those of them which exceed a fixed threshold θ and (iii) reestimating ϕ. We are particularly interested in the iteration of these three steps. The condition that θ is fixed will be relaxed in section 5.

By letting, for every $u \in \mathbb{R}$,

$$\tilde{\phi}_\theta(u) := \tilde{\phi}(u, \theta) = \frac{\sum y_{t-1} \Delta y_t \mathbb{I}_{\{|\Delta y_t - u y_{t-1}| \leq \theta\}}}{\sum y_{t-1}^2 \mathbb{I}_{\{|\Delta y_t - u y_{t-1}| \leq \theta\}}}, \tag{4.12}$$

the estimator produced by steps (i)-(iii) can be written as $\tilde{\phi}_\theta(\hat{\phi}^{(0)})$, and its iterates as

$$\hat{\phi}^{(i)} := \tilde{\phi}_\theta(\hat{\phi}^{(i-1)}) \ , \, i = 1, 2, ... \tag{4.13}$$

We discuss under what conditions does the iteration conduct to the asymptotics found for the benchmark estimator of the previous section. To deal with the discontinuous sample paths of $\tilde{\phi}_\theta$, we replace the standard fixed-point property with an asymptotic approximation.

4.1 Near fixed points

The asymptotic properties of the iteration are formulated using the following concept.

Definition 1 (near fixed point) *Let $\{\upsilon_T\}$ and $\{\Phi_T\}$ be sequences of random variables and random maps $\mathbb{R} \to \mathbb{R}$, respectively. We call $\{\upsilon_T\}$ a sequence of non-zero near fixed points of $\{\Phi_T\}$ if $\Phi_T(\upsilon_T) = \upsilon_T + o_P(\upsilon_T)$, with $\upsilon_T = O_P(1)$ and $P(\upsilon_T \neq 0) \to 1$ as $T \to \infty$.*

In this section the relevant choice of Φ_T will be $\Phi_T = \tilde{\phi}_\theta$, see (4.12). The fixed points of $\tilde{\phi}_\theta$, if they exist, are solutions of $\sum y_{t-1}\psi(\Delta y_t - (\cdot)y_{t-1}) = 0$, where $\psi(x) := x\mathbb{I}_{[-\theta,\theta]}(x)$, $x \in \mathbb{R}$, is Huber's skip function. Thus, they are M-estimators, cf. (1.4). Given that $\tilde{\phi}_\theta$ has discontinuous sample paths, which makes the existence of fixed points problematic, we will discuss estimators that are near fixed points.

The near-fixed point property is closely related to the numerical convergence (i.e., convergence declared by a computational algorithm) of the iterates of Φ_T, as well as to the numerical solution of the fixed-point equation $\Phi_T(\upsilon) = \upsilon$. Specifically, for a non-zero near fixed point sequence υ_T and for every $\varepsilon > 0$, it holds that $P(|\Phi_T(\upsilon_T) - \upsilon_T|/|\upsilon_T| < \varepsilon) \to 1$ as $T \to \infty$. Therefore, for any desired precision and with probability approaching one as T grows, numerical algorithms aiming at solving the equation $\Phi_T(\upsilon) = \upsilon$ will regard υ_T as a fixed point of Φ_T with respect to the relative-error criterion.[5] More generally, for any normalization sequence $n_T = O_P(\upsilon_T^{-1})$ and any $\varepsilon > 0$ it holds that $P(n_T|\Phi_T(\upsilon_T) - \upsilon_T| < \varepsilon) \to 1$ as $T \to \infty$, so numerically $n_T\upsilon_T$ will be regarded as a fixed point of $n_T\Phi_T$ with respect to the absolute-error criterion with precision ε. Whenever $n_T\upsilon_T$ is bounded away from zero in probability, this criterion is meaningful.

4.2 Uniform approximations

The next proposition establishes two approximations of the map $\tilde{\phi}_\theta$. Both are related to the behavior of $\tilde{\phi}_\theta$ in neighborhoods of zero shrinking at some rate b_T, with b_T a deterministic, positive sequence. In section 4.3, the magnitude order of the sequence b_T will match that of the estimator $\hat{\phi}^{(0)}$ used to initialize iteration (4.13).

Recall that for a given $\hat{\phi}^{(0)}$, residuals are constructed as $\hat{\varepsilon}_t := \Delta y_t - \hat{\phi}^{(0)} y_{t-1} = \varepsilon_t - (\hat{\phi}^{(0)} - \phi)y_{t-1}$. Since $\phi = -d_T^{-1}c$ and $\max_{t=1,...,T}|y_t|$ is of magnitude order $a_T = T^{-1/2}d_T$, the difference between $\hat{\varepsilon}_t$ and ε_t is asymptotically negligible (uniformly in $t = 1, ..., T$) if and only if $\hat{\phi}^{(0)}$ convergences to zero at a rate faster than $T^{1/2}d_T^{-1}$ (all in probability). Hence, in the following we separate two cases.

(a) First, we let $T^{1/2}d_T^{-1}/b_T \to \infty$, which is useful in the analysis of situations where the convergence rate of the preliminary estimator $\hat{\phi}_0$ makes residuals and true innovations asymptotically indistinguishable. This happens, for instance, if (i) $\hat{\phi}^{(0)}$ is set to 0, or (ii) $\alpha > 1$ and $\hat{\phi}^{(0)}$ is the OLS estimator (so $b_T = T^{-1}$).

(b) The second case is $b_T = T^{1/2}d_T^{-1}$. A convergence rate of $a_T^{-1} = T^{1/2}d_T^{-1}$ for $\hat{\phi}^{(0)}$ is the bridge towards less satisfactory preliminary estimators, since it implies that the difference between residuals and true innovations, though uniformly bounded in probability, is not asymptotically negligible. This is the case, e.g., if $\hat{\phi}^{(0)}$ is the OLS estimator (so $b_T = T^{-1}$) and $\{\varepsilon_t\}$ are Cauchy distributed.[6]

The following proposition contains the approximations of $\tilde{\phi}_\theta$ under (a) and (b).

[5] The requirement $P(\upsilon_T \neq 0) \to 1$ is included to ensure the good definition of the relative error; it is not restrictive if one thinks of υ_T as a statistic with a non-degenerate limiting distribution.

[6] We focus away from the possibility $T^{1/2}d_T^{-1}/b_T \to 0$, since for $\hat{\phi}^{(0)}$ vanishing at such slow b_T-rates the uniform distance between residuals and true innovations becomes unbounded and too many periods with large y_{t-1} are dummied out, compromising the desired d_T^{-1} convergence rate of the iterated estimator.

Proposition 2 *Let Assumptions $\mathcal{E}$ and $\mathcal{Y}$ hold, and let b_T be a positive real sequence. For every fixed $A, \theta > 0$ it holds that:*

a. If $T^{1/2}d_T^{-1}/b_T \to \infty$, then

$$\tilde{\phi}_\theta(u) = \tilde{\phi}(0,\theta) + u(h_\theta + o_P(1)) + \begin{cases} o_P(d_T^{-1}), \text{ if } d_T b_T = O(1) \\ o_P(d_T^{-1/2} b_T^{1/2}), \text{ if } d_T b_T \to \infty \end{cases}$$

uniformly over $|u| \leq b_T A$.

b. If $b_T = T^{1/2}d_T^{-1}$, then

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

uniformly over $|u| \leq b_T A$, where $Q_{T,\theta}$ is a random process such that, as $T \to \infty$,

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

in $D[-A, A]$. If f is strictly decreasing on $(0,\infty)$, then there exist random variables $H_{T,\theta} \in [0,1)$ a.s. such that $\sup_{|u|\leq b_T A} |\mathbb{I}_{\{u\neq 0\}} u^{-1} Q_{T,\theta}(u)| \leq H_{T,\theta}$ and $H_{T,\theta}$ converges weakly as $T \to \infty$ to a random variable $H_\theta < 1$ a.s.

An immediate corollary of Propositions 1 and 2 is the existence of near fixed points of $\tilde{\phi}_\theta$.

Corollary 3 *Let Assumptions $\mathcal{E}$ and $\mathcal{Y}$ hold, and $\theta > 0$ be such that $h_\theta \neq 1$. Then $\upsilon_T := (1-h_\theta)^{-1}\tilde{\phi}(0,\theta)$ defines a sequence of non-zero near fixed points of $\{\tilde{\phi}_\theta\}$.*

In the next subsection we shall study the proximity of the iterates $\hat{\phi}^{(i)}$ to $(1-h_\theta)^{-1}\tilde{\phi}(0,\theta)$ using the two approximations in Proposition 2. The approximation in part (a) (which is formally similar to the one obtained by Johansen and Nielsen, 2009, for $b_T = T^{-1/2}$ and $\{\varepsilon_t\}$ with finite fourth moment) 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. The non-linear approximation in part (b) is less tractable and our study of this borderline case will be less complete.

4.3 Asymptotic properties

We are now able to discuss the asymptotic properties of the iterative estimator (4.12)-(4.13). To this aim, two kinds of limits for i, T are considered. The first is a sequential limit, where $i \to \infty$ followed by $T \to \infty$. The second is a path-wise limit where $T, i \to \infty$ simultaneously and i is given as a function of T; we consider the path $i = \psi(T) := \lfloor T^\nu \rfloor$ for some $\nu > 0$.[7] The difference turns out to be that, in the sequential case, the number of iterations until numerical convergence can be chosen independently of T, whereas in the path-wise case, which provides results for a wider class of preliminary estimators, it increases with T. This is of limited relevance from practical point of view.

Let the preliminary estimator $\hat{\phi}^{(0)}$ of ϕ satisfy $b_T^{-1}\hat{\phi}^{(0)} = O_P(1)$. Moreover, let us initially assume that $T^{1/2}d_T^{-1}/b_T \to \infty$, such that the uniform distance between residuals and true

[7] Our conclusions hold for any natural-valued ψ such that $(h_\theta + \omega)^{\{\psi(T)\}^{1/2}} = o(d_T^{-1})$ for some $\omega > 0$.

innovations shrinks as T increases; see section 4.2. Proposition 2(a) suggests that in this case the behaviour of the sequence of iterates $\hat{\phi}^{(i)}$ depends on the quantity h_θ. If $h_\theta < 1$, iterations improve upon the preliminary estimator, leading to the concentration of probability mass around the near fixed point $\tilde{\phi}(0,\theta)/(1-h_\theta)$, whose asymptotics follow from Proposition 1. In contrast, for $h_\theta > 1$ the iteration is unstable and deteriorates the properties of the preliminary estimator.[8]

The following theorem makes these observations more precise.

Theorem 4 *Let $h_\theta < 1$ and b_T be a real sequence such that $b_T^{-1}\hat{\phi}^{(0)} = O_P(1)$ and $T^{1/2}d_T^{-1}/b_T \to \infty$. Let also $\hat{\phi}^{(i)} := \tilde{\phi}_\theta(\hat{\phi}^{(i-1)})$, $i \in \mathbb{N}$. Under Assumptions $\mathcal{E}$ and $\mathcal{Y}$, as $T \to \infty$:*

a. If $d_T b_T = O(1)$, then $\limsup_{i\to\infty} |\hat{\phi}^{(i)} - (1-h_\theta)^{-1}\tilde{\phi}\,(0,\theta)\,| = o_P(d_T^{-1})$.

b. If $\psi(T) := \lfloor T^\nu \rfloor$ with $\nu > 0$, then $\hat{\phi}^{(\psi(T))} - (1-h_\theta)^{-1}\tilde{\phi}\,(0,\theta) = o_P(d_T^{-1})$.

The conclusion in part (a) is established for preliminary estimators which are of magnitude order at least d_T^{-1}. The iteration initialized at $\hat{\phi}^{(0)} = 0$ is an example. In part (b) any initial magnitude order lower than $T^{1/2}d_T^{-1}$ is covered. For instance, the T^{-1} magnitude order of the OLS estimator under a (near) unit root makes it an admissible $\hat{\phi}^{(0)}$ in part (b) for any $\alpha \in (1,2)$.

Some further remarks are due, all assuming that the conditions of Theorem 4 hold.

Remark 4.1. Typical numerical criteria would declare $\hat{\phi}^{(i)}$ to converge if, given some $\varepsilon > 0$, they find that $|\phi^{(i-1)} - \phi^{(i)}| < \varepsilon$ or $|\phi^{(i-1)} - \phi^{(i)}|/|\phi^{(i-1)}| < \varepsilon$ for some i. Under the hypotheses of Theorem 4, the iteration will be declared to converge with probability approaching one, with respect to both criteria.

Remark 4.2. Theorem 4 justifies asymptotically Gaussian inference based on the iterated estimator $\hat{\phi}^{(i)}$ due to its proximity to the near fixed point $(1-h_\theta)^{-1}\tilde{\phi}\,(0,\theta)$ of ϕ_θ. If, as in the previous remark, a numerical convergence criterion is used, let the iteration be halted at step N. Then, for large T, the 't'-statistics

$$\xi_T^{(x)}(\hat{\phi}^{(0)},\theta) := \frac{1-\hat{h}_\theta}{\hat{\zeta}}\left(\sum y_{t-1}^2\right)^{1/2}\hat{\phi}^{(x)}, \quad x \in \{N,\ \psi(T)\},$$

are approximately standard Gaussian under the UR null, if $\hat{h}_\theta$ and $\hat{\zeta}$ are consistent estimators of h_θ and $\{V\,(\theta)\}^{1/2}p_\theta\,(0)^{-1}$, respectively. To estimate h_θ, we can compute residuals $\hat{\varepsilon}_t$ using $\hat{\phi}^{(N)}$ (resp. $\hat{\phi}^{(\psi(T))}$), and use them to obtain an empirical distribution function and a kernel estimate of $f(\theta)$:

$$\hat{h}_\theta := 2\theta\frac{w_T^{-1}\sum K(w_T^{-1}(\theta-\hat{\varepsilon}_t))}{\sum \mathbb{I}_{\{|\hat{\varepsilon}_t|\le\theta\}}} \tag{4.14}$$

for some bandwidth sequence w_T and positive kernel K integrating to unity (see Ling, 2005, eq. (2.1) for a similar approach). If K is chosen as the standard Gaussian or the logistic kernel, this estimator can be seen to be consistent if (i) $w_T \to 0$, $Tw_T \to \infty$ and

[8] A similar phenomenon occurs if $b_T = T^{1/2}d_T^{-1}$, though in this case it is not possible to determine the outcome of the iteration by looking at a one-dimensional quantity like h_θ only.

(ii) $w_T^{-2}a_T\hat{\phi}^{(N+1)} = o_P(1)$ (resp. $w_T^{-2}a_T\hat{\phi}^{(\psi(T))} = o_P(1)$). While conditions (i) are standard consistency requirement for i.i.d. data, see e.g. Theorem 2.6 of Pagan and Ullah (1999), condition (ii) is needed in order to control the order of magnitude of $\hat{\varepsilon}_t - \varepsilon_t$. Eligible choices of w_T are $w_T = aT^{-\nu}$ with $a > 0$ and $\nu \in \left(0, \frac{1}{4}\right)$.[9] Similarly to Remark 3.5,

$$\hat{\zeta}_1 := T^{1/2}\frac{[\sum \hat{\varepsilon}_t^2 \mathbb{I}_{\{|\hat{\varepsilon}_t|\leq\theta\}}]^{1/2}}{\sum \mathbb{I}_{\{|\hat{\varepsilon}_t|\leq\theta\}}}, \quad \hat{\zeta}_2 := T^{-1/2}\frac{[\sum y_t^2][\sum \hat{\varepsilon}_t^2 \mathbb{I}_{\{|\hat{\varepsilon}_t|\leq\theta\}}]^{1/2}}{\sum y_t^2 \mathbb{I}_{\{|\hat{\varepsilon}_t|\leq\theta\}}} \tag{4.15}$$

can be used to estimate $\{V(\theta)\}^{1/2} p_\theta(0)^{-1}$ consistently.

Remark 4.3. Non-trivial local power against the UR null is obtained for alternatives of the form $\phi = -d_T^{-1}c$ $(c \neq 0)$, under which as $T \to \infty$

$$d_T\hat{\phi}^{(\psi(T))} \xrightarrow{w} -c + \frac{\{V(\theta)\}^{1/2}}{(1-h_\theta)p_\theta(0)} \frac{N(0,1)}{(\int S^2)^{1/2}} \tag{4.16}$$

and

$$\xi_T^{(\psi(T))}(\hat{\phi}^0, \theta) \xrightarrow{w} -c\frac{(1-h_\theta)p_\theta(0)}{\{V(\theta)\}^{1/2}}(\int S^2)^{1/2} + N(0,1), \tag{4.17}$$

with $\hat{h}_\theta$ and $\hat{\zeta}$ as in Remark 4.2, and the Gaussian variable independent of S. From the argument in section A.3 of the Appendix it follows further that pathwise consistency holds $(d_T\hat{\phi}^{(\psi(T))} \xrightarrow{P} -\infty)$ against the usual local alternatives $\phi = -c/T$ $(c \neq 0)$, and if additionally $d_T b_T = O(1)$, then $P(d_T \liminf_{i\to\infty} \hat{\phi}^{(i)} < -K) \to 1$ as $T \to \infty$, for every $K > 0$. □

Let us now consider the borderline case where $b_T = T^{1/2}d_T^{-1}$. The following proposition shows that a result similar to Theorem 4(b) holds if f is unimodal. It covers, in particular, the case of Cauchy errors and $\hat{\phi}^{(0)}$ chosen equal to the OLS estimator.

Proposition 5 *Let $d_T T^{-1/2}\hat{\phi}^{(0)} = O_P(1)$ and the density f be strictly monotone on $(-\infty, 0)$ and $(0,\infty)$. Under Assumptions $\mathcal{E}$ and $\mathcal{Y}$, $\hat{\phi}^{(\psi(T))} - (1-h_\theta)^{-1}\tilde{\phi}(0,\theta) = o_P(d_T^{-1})$, with $\psi(T)$ defined as in Theorem 4(b).*

We conclude this section by stating how an inappropriate choice of θ can deteriorate the properties of a good preliminary estimator, even if it is the true value of ϕ. Due to the counter-exemplary function of the result, we consider the case $\phi = 0$ only.

Proposition 6 *Let $h_\theta > 1$, and b_T be a real sequence such that $b_T^{-1}\hat{\phi}^{(0)} = O_P(1)$ and $T^{1/2}d_T^{-1}/b_T \to \infty$ as $T \to \infty$. Under Assumptions $\mathcal{E}$,$\mathcal{Y}$ with $\phi = 0$:*

a. If $d_T b_T \to 0$, then for every $A > 0$

$$P\left(|d_T\hat{\phi}^{(i)}| < A \text{ for all } i \in \mathbb{N}\right) \to 0 \quad \text{as } T \to \infty,$$

i.e., the sequence $\{d_T\hat{\phi}^{(i)}\}_{i=0}^{\infty}$ is unbounded with probability approaching one.

b. If either (i) $d_T b_T \to \infty$ and $b_T^{-1}\hat{\phi}^{(0)}$ is bounded away from zero in probability or (ii), $b_T = d_T^{-1}$ and $d_T\{\hat{\phi}^{(0)} - (1-h_\theta)^{-1}\tilde{\phi}(0,\theta)\}$ is bounded away from zero in probability, then for every $A > 0$

$$P\left(|b_T^{-1}\phi^{(i)}| < A \text{ for all } i \in \mathbb{N}\right) \to 0 \quad \text{as } T \to \infty.$$

[9] In applications it could be useful to check if $\hat{h}_\theta < 1$, and otherwise, consider a larger threshold.

5 An iterative estimator with estimated threshold

In this section we combine the results of sections 3 and 4.3 in a joint estimation setup. In particular, we study the following iterative procedure: (i) given an initial estimate of ϕ and the associated residuals, determine a threshold $\hat{\theta}$ as a function of the residuals; (ii) fix the threshold $\hat{\theta}$ and reestimate ϕ; considering the updated estimate of ϕ as a new initial estimate, repeat steps (i) and (ii) until possible convergence. In what follows we provide conditions for this iterative procedure to produce a d_T-consistent estimator which is asymptotically standard Gaussian under the UR null hypothesis.

To formalize the iteration, for every $T \in \mathbb{N}$ let $\sigma_T : \mathbb{R} \to \mathbb{R}$ be a (possibly random) nonnegative function; e.g., $\sigma_T(u) = a_T^{-2} \sum (\Delta y_t - u y_{t-1})^2$. If, given an initial estimate $\hat{\phi}^{(0)}$, we use $\sigma(\hat{\phi}^{(0)})$ as the threshold $\hat{\theta}$ in the reestimation of ϕ (suppressing the dependence of σ on T), and then iterate this procedure, we obtain a sequence of iterates $\hat{\phi}^{(i)} := \phi(\hat{\phi}^{(i-1)}, \sigma(\hat{\phi}^{(i-1)}))$, $i \in \mathbb{N}$, as estimates of ϕ. The next proposition relates this sequence to a non-zero near fixed point of $\phi(\cdot, \sigma(\cdot))$.

Theorem 7 *Let Assumptions $\mathcal{E}$ and $\mathcal{Y}$ hold. If*

(i) $\sigma(u) = \sigma(0) + o_P(1)$ uniformly on compacts $|u| \le d_T^{-1} C$ ($C > 0$),

(ii) $\sigma(0)$ is bounded and bounded away from zero in probability, and

(iii) $h(\sigma(0)) \neq 1$ a.s.,

then

$$\upsilon_T := \frac{\tilde{\phi}(0, \sigma(0))}{1 - h(\sigma(0))}$$

defines a sequence of non-zero near fixed points of $\{\phi(\cdot, \sigma(\cdot))\}$. If further

(iv) $\limsup_T h(\sigma(0)) < 1$ a.s.,

(v) $\hat{\phi}^{(0)} = O_P(d_T^{-1})$,

then $\limsup_{i\to\infty} |\hat{\phi}^{(i)} - \upsilon_T| = o_P(d_T^{-1})$, so for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that, as $T \to \infty$, $P(d_T \sup_{i>N} |\hat{\phi}^{(i)} - \upsilon_T| < \varepsilon) \to 1$.

Note that the near-fixed point υ_T is computable. The choice $\hat{\phi}^{(0)} = 0$ can be used as a trivial preliminary estimator satisfying $\hat{\phi}^{(0)} = O_P(d_T^{-1})$. Alternatively, under the conditions of Theorem 4, for $\alpha \in (1, 2)$ another preliminary estimator with magnitude order d_T^{-1} is $\hat{\phi}^{(\psi(T))}$, obtained by iterating $\tilde{\phi}_\theta$, for fixed θ, starting from the OLS estimator. Any of these preliminary estimators can be used to initialize the iteration in Theorem 7.

Some remarks follow.

Remark 5.1. Possible choices of σ include the empirical quantile

$$\sigma(u) = \inf\{x > 0 : T^{-1} \sum \mathbb{I}_{\{|\varepsilon_t - u y_{t-1}| \le x\}} > \tau\} \tag{5.18}$$

for fixed $\tau \in (0, 1)$. It is shown in the Appendix that it satisfies hypothesis (i), and since it holds that $\sigma(0) \xrightarrow{P} q_\tau$, where the τth quantile q_τ of the distribution of $|\varepsilon_1|$ is positive under Assumption $\mathcal{E}$(ii), it satisfies hypothesis (ii) as well. In this case, if $h(q_\tau) \neq 1$, also $(1 - h(q_\tau))^{-1} \tilde{\phi}(0, q_\tau)$ is a near-fixed point of $\phi(\cdot, \sigma(\cdot))$ and is $o_P(d_T^{-1})$-close to υ_T.

An eligible choice of σ which remains random in the limit is

$$\sigma(u) = \frac{\{\sum(\Delta y_t - u y_{t-1})^2\}^{1/2}}{\max_{t=1,...,T}|\Delta y_t - u y_{t-1}|} \text{ with } \sigma(0) = \frac{\{\sum(\Delta y_t)^2\}^{1/2}}{\max_{t=1,...,T}|\Delta y_t|} \xrightarrow{w} \frac{[S]_1^{1/2}}{\sup_{[0,1]}|\Delta S|} \tag{5.19}$$

as $T \to \infty$, under Assumptions $\mathcal{E}, \mathcal{Y}$ (here, for any $u \in (0,1]$, $\Delta S(u) := S(u) - S(u-)$ and $\Delta S(0) = 0$). The verification of hypothesis (i) is straightforward.[10]

The condition $\limsup_T h(\sigma(0)) < 1$ a.s. is trivially satisfied when f is unimodal, and in particular, when ε_t are α-stable.

Remark 5.2. The last convergence in Theorem 7 implies that Remark 4.1 applies to the iterated estimator $\hat{\phi}^{(i)}$. Asymptotic Gaussian inference under the UR null holds as in Remark 4.2, with θ replaced by $\sigma(\hat{\phi}^{(N)})$ or $\sigma(\hat{\phi}^{(\psi(T))})$ in (4.14) and (4.15), assuming additionally that $Tw_T^2 \to \infty$ as $T \to \infty$ and the density function of ε_1 is uniformly continuous on $\mathbb{R}$ (cf. Theorem 2.8 of Pagan and Ullah, 1999). Asymptotically non-trivial power against the local alternatives $\phi = -cd_T^{-1}$ ($c \neq 0$) is obtained like in Remark 4.3. However, in order to write an analogue of (4.17), we need $\sigma(0)$ to have a limit as $T \to \infty$. Namely, if together with the hypotheses of Theorem 7 it holds that $(a_T^{-1} y_{\lfloor T\cdot \rfloor}, \sigma(0), \hat{\zeta}, \hat{h}_\theta) \xrightarrow{w} \big(S, \Theta, \{V(\Theta)\}^{1/2}/p_\Theta(0), h(\Theta)\big)$ with (S, Θ) independent of the Brownian motion B from Proposition 1, then

$$\xi_T^{(\psi(T))}(\hat{\phi}^0, \hat{\theta}) := \frac{1 - \hat{h}_\theta}{\hat{\zeta}} \left(\sum y_{t-1}^2\right)^{1/2} \hat{\phi}^{(\psi(T))} \xrightarrow{w} -c\frac{(1 - h_\Theta)p_\Theta(0)}{\{V(\Theta)\}^{1/2}}\left(\int S^2\right)^{1/2} + N(0,1),$$

a hybrid version of (3.10) and (4.17). All the choices of σ suggested above satisfy the extra assumption, with $\Theta = q_\tau$ for the empirical quantile functions and $\Theta = (\sup_{[0,1]}|\Delta S|)^{-1}[S]_1^{1/2}$ for σ in (5.19). □

6 Extension to higher-order autoregressions

The goal of this section is to show that the preceding results are not specific to first-order autoregressions but can be extended to higher-order processes. Specifically, instead of Assumption $\mathcal{Y}$, consider the following one.

Assumption $\mathcal{Y}(k)$. *The process $\{y_t\}_{t=1}^T$ satisfies $\Delta y_t = \phi y_{t-1} + \partial' \Delta \mathbf{y}_{t-1} + \varepsilon_t$ ($t = 1, ..., T$), where $\Delta \mathbf{y}_{t-1} := (\Delta y_{t-1}, ..., \Delta y_{t-k})'$, $\phi = -d_T^{-1} c$ with $c \in \mathbb{R}$, $\partial := (\partial_1, ..., \partial_k)' \in \mathbb{R}^k$ is such that $1 - \sum_{i=1}^k \partial_i z^i \neq 0$ on the closed unit complex disk, and y_0 and $\Delta \mathbf{y}_0$ are fixed.*

We discuss two solutions of the problem of inference on ϕ under Assumptions $\mathcal{E}$ and $\mathcal{Y}(k)$. First, we show that a dummy-based estimator of ϕ with asymptotics as in Theorem 7 can be obtained by using a preliminary estimator of ∂ and iteration over the estimators of ϕ and θ alone. Second, we discuss how iteration over the estimators of all ϕ, ∂ and θ can be analysed.

Let an estimator $\hat{\partial}$ of ∂ be available. It can be used to 'de-lag' Δy_t (i.e., to replace Δy_t by $\Delta \hat{y}_t := \Delta y_t - \hat{\partial}' \Delta \mathbf{y}_{t-1}$) and then apply the estimation methods of sections 3 and 5 to $\Delta \hat{y}_t$

[10] Notice that for this choice of σ, if $\mathrm{E}\varepsilon_1^2 < \infty$, then $\sigma(u) \xrightarrow{P} \infty$ uniformly on compacts $|u| \leq T^{-1}C$. Hence, a large estimated σ could indicate a failure of Assumption $\mathcal{E}$, with consequent inapplicability of our theory.

and y_{t-1} instead of Δy_t and y_{t-1}. Accordingly, instead of (4.12), this requires us to consider its modification

$$\Phi\left(u,\theta\right)=\frac{\sum y_{t-1}\Delta\hat{y}_t\mathbb{I}_{\{|\Delta\hat{y}_t-uy_{t-1}|\leq\theta\}}}{\sum y_{t-1}^2\mathbb{I}_{\{|\Delta\hat{y}_t-uy_{t-1}|\leq\theta\}}}. \tag{6.20}$$

We prove in the Appendix, Section A.5, that Theorem 7 remains valid for $\Phi(\cdot,\sigma\left(\cdot\right))$ in place of $\phi(\cdot,\sigma\left(\cdot\right))$ if $\hat{\partial}-\partial=O_P(b_T^{-1})$ for a real sequence b_T such that $T^{-1/2}b_T\to\infty$ and $T^{1/4}a_T^{-1}b_T\to\infty$. Specifically, the near-fixed point and its asymptotics are as in the first-order autoregressive case, except for a standard long-run impact coefficient which now appears in the limits. As to the choice of $\hat{\partial}$, it could be the OLS estimator from a regression of Δy_t on $\Delta\mathbf{y}_{t-1}$ and a constant (then $\hat{\partial}$ has the required consistency rate for all α under the unit root null, and for $\alpha>2/3$ also under local alternatives), or from a regression of Δy_t on y_{t-1}, $\Delta\mathbf{y}_{t-1}$ and a constant (then the consistency rate of $\hat{\partial}$ is sufficiently fast for $\alpha>4/5$, under the null and under local alternatives); in both cases, for $\alpha>1$ the constant can be dispensed with. This follows from the asymptotics for correlations in Davis and Resnick (1985).

In spite of the formal and computational similarity with the AR(1) case, the study of $\Phi(\cdot,\sigma\left(\cdot\right))$ requires the development of some mathematics under new conditions. Results in the previous section were based on the properties of weighted empirical processes constructed from residuals whose distance from the true innovations is infinitesimal uniformly in t, in probability. This is analogous to the setup of Koul and Ossiander (1994). Here, however, under our hypotheses on $\hat{\partial}-\partial$, it need not hold that $\max_{t=1,\dots,T}|\Delta\mathbf{y}_{t-1}|=o\left(b_T\right)$, so $\max_{t=1,\dots,T}|\Delta\hat{y}_t-uy_{t-1}-\varepsilon_t|$ need not be $o_P\left(1\right)$ even if the true value $u=-d_T^{-1}c$ is inserted (in fact, it is not $o_P(1)$ if $\hat{\partial}$ is the OLS estimator). Thus, we extend the empirical processes results to cover this situation too. Formulations are in the Appendix (Propositions A.1 and A.2), and proofs in the supplement Cavaliere and Georgiev (2012).

We finally turn to the possibility of iterating over the estimator of ∂, besides those of ϕ and θ. In particular, iteration of the OLS estimator of ∂ from the dummy-augmented regression would require to redefine Φ as (using partial regression format):

$$\Phi\left(u,s,\theta\right)=\frac{\sum y_{t-1}(\Delta y_t-x(u,s,\theta)'\Delta\mathbf{y}_{t-1})\mathbb{I}_{\{|\Delta y_t-uy_{t-1}-s'\Delta\mathbf{y}_{t-1}|\leq\theta\}}}{\sum y_{t-1}(y_{t-1}-w(u,s,\theta)'\Delta\mathbf{y}_{t-1})\mathbb{I}_{\{|\Delta y_t-uy_{t-1}-s'\Delta\mathbf{y}_{t-1}|\leq\theta\}}},$$

where x and w are the estimators from the dummy-augmented regressions of, respectively, Δy_t and y_{t-1}, on $\Delta\mathbf{y}_{t-1}$; that is, $x(u,s,\theta):=S_{\Delta\Delta}^{-1}S_{\Delta 0}$ and $w(u,s,\theta):=S_{\Delta\Delta}^{-1}S_{\Delta 1}$, with $S_{\Delta\Delta}:=\sum\Delta\mathbf{y}_{t-1}\Delta\mathbf{y}_{t-1}'\mathbb{I}_{\{|\Delta y_t-uy_{t-1}-s'\Delta\mathbf{y}_{t-1}|\leq\theta\}}$, $S_{\Delta 0}:=\sum\Delta\mathbf{y}_{t-1}\Delta y_t\mathbb{I}_{\{|\Delta y_t-uy_{t-1}-s'\Delta\mathbf{y}_{t-1}|\leq\theta\}}$ and $S_{\Delta 1}:=\sum\Delta\mathbf{y}_{t-1}y_{t-1}\mathbb{I}_{\{|\Delta y_t-uy_{t-1}-s'\Delta\mathbf{y}_{t-1}|\leq\theta\}}$. Given initial estimators $(\hat{\phi}^{(0)},\hat{\partial}^{(0)})$, the iteration could be formalized as

$$\begin{aligned}\hat{\sigma}^{(i+1)}&=\sigma(\hat{\phi}^{(i)},\hat{\partial}^{(i)}),\ \hat{\phi}^{(i+1)}=\Phi(\hat{\phi}^{(i)},\hat{\partial}^{(i)},\hat{\sigma}^{(i+1)}),\\ \hat{\partial}^{(i+1)}&=x(\hat{\phi}^{(i)},\hat{\partial}^{(i)},\hat{\sigma}^{(i+1)})-\hat{\phi}^{(i+1)}w(\hat{\phi}^{(i)},\hat{\partial}^{(i)},\hat{\sigma}^{(i+1)})\ .\end{aligned}$$

For the iterates $\hat{\phi}^{(i+1)}$ to have the same asymptotics as with 'de-lagging' ($\hat{\partial}$ independent of i), we need to bound $\hat{\partial}^{(i+1)}$ for all i to a neighbourhood of ∂ shrinking at the rate of b_T^{-1}, with b_T as before. As can be seen from the expression for Φ, this can be achieved relying on results for two kinds of weighted empirical processes: (i) results uniform over (u,s,θ) for processes with weights y_{t-1} and y_{t-1}^2 (these are provided in Proposition A.2); (ii) for the discussion of x and w, results for processes with weights depending on $\{\Delta\mathbf{y}_{t-1}\}$. Such processes differ substantially from the class we study as their weights are not uniformly asymptotically negligible ($\max_{1\le t\le T}\|\Delta\mathbf{y}_{t-1}\|$ and $\sum\Delta\mathbf{y}_{t-1}$ are of the same magnitude order under IV), contrary to one of the main hypotheses in Propositions A.1 and A.2. Thus, we do not undertake the generalization here.

7 Simulation evidence

The asymptotic results in the previous sections show that the large-sample properties of the iterated dummy-variable estimator have two main determinants: the convergence 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 function h_θ (see Remarks 2.5 and 3.2). Our first goal in this section is to illustrate the importance of these two determinants numerically. The second goal is to investigate the finite-sample precision of the Gaussian approximation to the null distribution of the studied test statistics, as well as the finite-sample relevance of the theoretically predicted asymptotic power gains.

We generate data according to the autoregression (1.1), initialized at $y_0 = 0$. Four distributions for the innovations ε_t are considered. In the first three cases, ε_t are drawn from a symmetric stable distribution with tail index α, for $\alpha = 1/2$, 1 (Cauchy distribution) and 3/2. In all the three cases, the condition $h_\theta < 1$ holds for all $\theta > 0$. The fourth case is a *bimodal* distribution. This distribution is in the domain of attraction of the Cauchy distribution ($\alpha = 1$) but for some $\theta > 0$ it violates the condition $h_\theta < 1$; specifically, for $\theta \in (0, 1]$ it holds that $h_\theta = 3$.[11]

The autoregressive parameter ϕ is set either to 0 (y_t has a unit root) or to $-7/d_T$ (local alternative), where $d_T = T^{1/2+1/\alpha}$ is determined according to the distribution of the innovations.

For the size analysis, samples of size $T = 100$ and 500 are considered. For the local power analysis, we also report results for $T = 10,000$ and, in addition, we report asymptotic power ($T = \infty$) based on a simulation of the limiting distributions in (3.10) and (4.17), with S discretised over a grid of $500,000$ points.

We consider several (left-sided) dummy-based tests for the UR null hypothesis $\phi = 0$, all of them run at the 5% nominal asymptotic significance level. First, we consider tests based on a fixed threshold θ. Specifically, we consider the benchmark statistics based on $\tilde{\phi}_\theta(0)$ of section 3, denoted by $\xi_T(0,\theta)$ in the following. Moreover, we consider the statistics based on the iterated estimator of section 4, initialized at $\hat{\phi}^{(0)} = 0$ and at the OLS estimator of ϕ. These statistics are denoted by $\xi_T^{(N)}(0,\theta)$ and $\xi_T^{(\sqrt{T})}(\hat{\phi}_{OLS},\theta)$, respectively. The iteration initialized at zero is halted at the smallest iterate N such that the difference between the Nth and the $(N-1)$th iterate of the test statistics is smaller than 10^{-6}; under the conditions of Theorem 4(a), such an N exists with probability tending to one.[12] The iteration initialized at the OLS estimator is halted at the $\lfloor\sqrt{T}\rfloor$th iterate, in agreement with Theorem 4(b). We

[11] Its density $\tilde{f}$ is given, for every $x \in \mathbb{R}$ and with f denoting the standard Cauchy density, as

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

[12] In fact, for $T = 500$ our convergence criterion failed to be satisfied in at most 2 out of $50,000$ replications.

set $\theta = F^{-1}(0.875)$ in what follows, corresponding to the 75th percentile of the distribution of $|\varepsilon_1|$ (apart from the case $\alpha = 1$, the values of θ and the associated $p_\theta(0)$ and $V(\theta)$ were found by simulation in advance). Results are reported for normalization with the statistics $\hat{\zeta}_1$ from Remarks 3.5 and 4.2 (results for $\hat{\zeta}_2$ are similar and omitted for brevity), using a Gaussian kernel and the bandwidth $w_T = T^{-9/40}$.

Second, we consider the tests of section 5, where also the threshold is determined iteratively. We work with three different initial estimators of ϕ: $\hat{\phi}^{(0)} = 0$, $\hat{\phi}^{(0)} = \tilde{\phi}_\theta^{(N)}(0)$ and $\hat{\phi}^{(0)} = \tilde{\phi}_\theta^{(\sqrt{T})}(\hat{\phi}_{OLS})$. The function σ from (5.19) is employed, and the convergence criterion is as for $\xi_T^{(N)}(0, \theta)$, following Theorem 7. As for the tests with a fixed threshold, normalization by $\hat{\zeta}_1$ is employed.

Third, we also present some results for the well-known Dickey-Fuller UR test based on the t-statistics. The asymptotic critical value employed is the 5th percentile of the Dickey-Fuller distribution. Although under infinite-variance innovations this critical value is not justified, it can be used to provide a clear illustration of the fact that, asymptotically, the Dickey-Fuller test cannot distinguish between the null and the postulated local alternative.

TABLE 1 ABOUT HERE

Let us discuss the cases $\alpha = 3/2$ and $\alpha = 1$ (Cauchy distribution) first. According to Theorems 4 and 7, all the dummy-based test statistics should be asymptotically $N(0,1)$ distributed under the UR null hypothesis. The Monte Carlo results support this result. Specifically, tests performed using a fixed threshold are seen to be only slightly oversized, with the distortions decreasing slowly as T grows. Size distortions of similar magnitude are observed, now in both directions, for tests with iteration over the threshold.

Local power for the dummy-based tests is high, though convergence to the asymptotic power is slow. Nevertheless, even for small T the superiority of dummy-based tests over the standard OLS t-based UR test is obvious, with the local power of the latter decreasing towards its size as T grows. We have also simulated power against $\phi = -7/T$ (not reported), where the asymptotic power of the OLS t-based UR test is non-trivial but bounded away from one, and dummy-based tests are again superior, as their power approaches one.

Results for the case $\alpha = 1/2$ clearly show that the choice of a preliminary estimator is indeed crucial, and suggest that the conditions on $\hat{\phi}^{(0)}$ in Theorems 4 and 7 are not only sufficient, but also necessary. For $\alpha = 1/2$ the OLS preliminary estimator (which converges at the T^{-1} rate) does not have the convergence rate required by Theorem 4 and Proposition 5 (i.e., no slower than T^{-2}) for the iteration with a fixed threshold. The consequences are (i) the severe size distortions of the $\xi_T^{(\sqrt{T})}(\hat{\phi}_{OLS}, \theta)$ test, indicating that a Gaussian approximation to its null distribution is inappropriate, and (ii) similar distortions of the corresponding full-estimation test, indicating that $\tilde{\phi}_\theta^{(\sqrt{T})}(\hat{\phi}_{OLS})$ is not d_T-consistent as required by Theorem 7(v).

For the stable distributions with $\alpha = 1/2, 1$ (Cauchy) and $3/2$ considered so far, the condition $h_\theta < 1$ is satisfied in the experiments with fixed $\hat{\theta}$, and h_θ evaluated at $(\sup_{[0,1]} |\Delta S|)^{-1}[S]_1^{1/2}$ is (almost surely) smaller than 1 in the experiments with iterated $\hat{\theta}$. While the former fact holds also for the bimodal density, the latter one does not, implying a violation of hypothesis

(iv) in Theorem 7. This allows us to see the importance of the threshold and the related quantity h_θ by examining the results in Table 1 for the bimodal distribution. Contrary to the tests with fixed θ, which follow the pattern observed for $\alpha = 3/2$ and 1, for all tests where the threshold is determined iteratively rejection frequencies under the alternative are very close to those under the null, indicating an inability of the latter tests to distinguish between the examined hypotheses.

8 Concluding remarks

In this paper we considered estimation and testing in autoregressive models characterized by infinite-variance innovations and a unit or near-to-unit AR root. We analysed the large-sample properties of robust estimators (one-step and iterated), obtained by following the much used practice of dummying out 'large' residuals, i.e. exceeding some (given or estimated) threshold. Our results provide a statistical justification for this approach: specifically, we proved that it guarantees (i) a convergence rate faster than the T rate of standard OLS estimators of the (near) unit root, (ii) asymptotically Gaussian UR test statistics under the UR (null) hypothesis and (iii) massive local power improvements, together with (iv) easy computability.

Our asymptotic and finite sample results show that the choice of an initial estimator for the iteration plays a key role. Specifically, a fast consistency rate of the dummy-based estimator is achieved only for initial estimators sufficiently close to the true value of the autoregressive parameters. This conclusion is likely to extend to the iterative computation of the generic M-estimator defined by (1.4), when it is not simultaneously a minimizer in (1.3). We provide sufficient conditions on the consistency rate of the initial estimator, easy to satisfy, given that it is chosen at the discretion of the econometrician. In the empirically most relevant case of $\alpha > 1$, a corollary of our results is that the OLS initial estimator works if the threshold is fixed (at least in a first round of iterations), thus supporting the practice of dummying out large OLS residuals.

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. A further, related issue is to assess whether iterative estimation of the threshold enhances, diminishes or does not affect these gains. Unreported simulations (available in Section S3 of the supplement, Cavaliere and Georgiev, 2012) show that for $\alpha > 1$ iterations of the threshold seem to make little practical difference. Moreover, among the two thresholds we consider, a self-normalized residual standard deviation and a residual quantile, the quantile appears to have a slight advantage in terms of both size and power.

Although we have not discussed the case of deterministic terms in the autoregressive equation, the theory of the Appendix suffices to work out an extension along the lines of section 6. The effect of deterministic components could be removed by 'demeaning' and/or 'detrending', given a preliminary estimator of the associated parameters that makes the contribution of the deterministics to the detrended Δy_t asymptotically negligible, uniformly in t. Under infinite innovation variance, this estimator may need to be a robust one.

Two extensions, both related to the presence of stationary regressors, are left to further research. One concerns the properties of dummy-variable iteration for stationary AR processes, and the other one, iteration over the estimator of the short-run parameters in autoregressions with a (near) unit root. In the stationary case, the literature has only studied M-estimators of the type 1.3; see Davis, Knight and Liu (1992). The analysis of Huber's skip-estimator would require the consideration of empirical processes with weights which are not uniformly asymptotically infinitesimal, calling for a novel treatment.

A Appendix

Uniform asymptotic approximations of the map $\tilde{\phi}$ are our main tools in the Appendix. A joint approximation in (u, θ) follows via the approach of Koul and Ossiander (1994) and Koul (2002). However, it requires that u be restricted to a compact around zero which shrinks at two fast a rate, not permitting the study of dummy-variable iteration initialized at an estimator whose consistency rate is slower than $T^{1/2}a_T$. As improving on the rate of the initial estimator is the rationale of dummy-variable iteration, we derive also an alternative approximation which allows for slower shrinkage rates, the price being that it is univariate, with θ hold fixed.

In order to cover also the local alternatives $\phi = -c/T$ ($c \neq 0$), we derive some results under the following generalization of Assumption $\mathcal{Y}$.

Assumption $\mathcal{Y}$'. The process $\{y_t\}_{t=1}^T$ satisfies $\Delta y_t = -cd_T^{-1}y_{t-1}+\varepsilon_t$ $(t = 1, ..., T)$, where $d_T = T^{\kappa}(T^{1/2}a_T)^{1-\kappa}$ with $\kappa \in [0, 1]$ such that $a_T/d_T \to 0$ as $T \to \infty$, and y_0 is fixed.

For $\kappa = 0$ this is Assumption $\mathcal{Y}$, whereas for $\kappa > 0$ and $c \neq 0$ moderate deviations from UR are obtained; see Remark 2.4. For $\kappa = 1$ (and $\alpha \in (1, 2)$ to ensure $a_T/d_T \to 0$) these are the alternatives $\phi = -c/T$ ($c \neq 0$). In the Appendix the meaning of d_T is as in Assumption $\mathcal{Y}'$; in the special case $\kappa = 0$ we write explicitly $T^{1/2}a_T$ instead of d_T. Under Assumptions $\mathcal{E}$ and $\mathcal{Y}'$, $a_T^{-1}y_{\lfloor T\cdot\rfloor} \xrightarrow{w} S_c^{\kappa}$ in $D[0, 1]$, where $S_c^1 = \int^{(\cdot)} e^{-c(\cdot-s)}dS(s)$ and $S_c^{\kappa} = S$ for $\kappa \in [0, 1)$.

A.1 The benchmark estimator with estimated scale

We need first a result on weighted empirical processes. Let $\varepsilon_t^+ := \varepsilon_t \vee 0$ and $\varepsilon_t^- := (-\varepsilon_t) \vee 0$, with $\vee$ denoting binary maximum, and similarly for other r.v.'s. 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 $\mathbf{\Delta y}_0$ of section 6, are assumed fixed constants). Consider an array $\{\eta_t, \xi_{Tt}, \psi_{Tt}, \gamma_{Tt}\}_{t=1}^T$ of r.v.'s such that $\eta_t \in \{1, \varepsilon_t, \varepsilon_t^+, \varepsilon_t^-\}$ (the same choice for all t, T), and $(\xi_{Tt}, \psi_{Tt}, \gamma_{Tt})$ is $\mathcal{F}_{t-1}$-measurable and a.s. finite for all t, T. Define $m(\theta, z) := E(\eta_1\mathbb{I}_{\{|\varepsilon_1-z|\leq\theta\}})$,

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

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

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

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

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

PROOF. The proof mimics that of Theorem 1.1 in Koul and Ossiander (1994), and is given in the supplement Cavaliere and Georgiev (2012) [hereafter CG12]. ■

Let next $\Psi_T(\theta) := p_\theta(0)\{cd_T^{-1}+\tilde{\phi}(0,\theta)\}$ for $\theta>0$ and let it be written as $\Psi_T(\theta) = \{\Psi_{T,1}(\theta)+\Psi_{T,2}(\theta)\}\Psi_{T,3}(\theta)$, where

$$\begin{aligned}
\Psi_{T,1}(\theta) &:= \left(\sum y_{t-1}^2\right)^{-1/2}\sum y_{t-1}\varepsilon_t\mathbb{I}_{\{|\varepsilon_t|\le\theta\}},\\
\Psi_{T,2}(\theta) &:= \left(\sum y_{t-1}^2\right)^{-1/2}\sum y_{t-1}\varepsilon_t(\mathbb{I}_{\{|\varepsilon_t-d_T^{-1}cy_{t-1}|\le\theta\}}-\mathbb{I}_{\{|\varepsilon_t|\le\theta\}}),\\
\Psi_{T,3}(\theta) &:= p_\theta(0)\left(\sum y_{t-1}^2\mathbb{I}_{\{|\varepsilon_t-d_T^{-1}cy_{t-1}|\le\theta\}}\right)^{-1}\left(\sum y_{t-1}^2\right)^{1/2}.
\end{aligned}$$

We establish some asymptotic properties of $\Psi_{T,i}$ as processes in θ.

Lemma A.1 *Under Assumptions $\mathcal{E}$, $\mathcal{Y}$', as $T\to\infty$ it holds that:*

a. $\Psi_{T,1}\xrightarrow{w}B(V)$ in $D[0,\infty)$, where B is a standard Brownian motion independent of S_c^κ.

b. $d_Ta_T^{-1}T^{-1/2}\Psi_{T,2}\xrightarrow{w}2c(\cdot)f\{\int(S_c^\kappa)^2\}^{1/2}$ in $D[0,\infty)$, where $(\cdot)$ is the identity function of $[0,\infty)$.

c. $T^{1/2}a_T\Psi_{T,3}\xrightarrow{w}\{\int(S^\kappa)^2\}^{-1/2}$ in $D[a,\infty)$ for all $a>0$.

PROOF. We start the proof of part (a) by noting that $\Psi_{T,1}$ is tight in $D[0,A]$ for every $A>0$. This follows from Proposition A.1 with $\gamma_{Tt}=a_T^{-1}y_{t-1}$, $\eta_t=\varepsilon_t$, $\xi_{Tt},\psi_{Tt}=0$, and from the weak convergence of $T^{-1}\sum\gamma_{Tt}^2=T^{-1}\sum(a_T^{-1}y_{t-1})^2$ to $\int(S_c^\kappa)^2\in(0,\infty)$ a.s.

Next we turn to convergence of the finite-dimensional distributions, and for notational ease we discuss the bivariate ones, the generalization being straightforward. For given $\theta_1<\theta_2$, $\varepsilon_t\mathbb{I}_{\{|\varepsilon_t|\le\theta_1\}}$ and $\varepsilon_t\mathbb{I}_{\{\theta_1<|\varepsilon_t|\le\theta_2\}}$ are uncorrelated under the assumption that the distribution of ε_t is symmetric. Thus, using also the continuity of the distribution of $\{\varepsilon_t\}$, from Donsker's invariance principle it is seen that

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

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

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

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

$$\begin{aligned}
(\Psi_{T,1}(\theta_1),\Psi_{T,1}(\theta_2)-\Psi_{T,1}(\theta_1)) &\xrightarrow{w}\left(\{\frac{V(\theta_1)}{\int(S_c^\kappa)^2}\}^{1/2}\int S_c^\kappa dB_1,\{\frac{V(\theta_2)-V(\theta_1)}{\int(S_c^\kappa)^2}\}^{1/2}\int S_c^\kappa dB_2\right)\\
&\overset{d}{=}(B(V(\theta_1)),B(V(\theta_2))-B(V(\theta_1)))
\end{aligned}$$

with B independent of S_c^κ, the distributional equality by the independence of B_1, B_2 and S_c^κ, and by the independence of the increments of a Brownian motion. Hence, $(\Psi_{T,1}(\theta_1), \Psi_{T,1}(\theta_2)) \xrightarrow{w} (B\left(V\left(\theta_1\right)\right), B\left(V\left(\theta_2\right)\right))$.

For part (b), from Proposition A.1 with $\gamma_{Tt} = a_T^{-1} y_{t-1}$, $\eta_t = \varepsilon_t$, $\xi_{Tt} = cd_T^{-1} y_{t-1}$ and $\psi_{Tt} = 0$ we have, for any fixed c and $A > 0$,

$$
\begin{aligned}
& a_T^{-1} T^{-1/2} \sum y_{t-1} \varepsilon_t \left[\mathbb{I}_{\{|\varepsilon_t - d_T^{-1} c y_{t-1}| \le \theta\}} - \mathbb{I}_{\{|\varepsilon_t| \le \theta\}}\right] \\
= \ & a_T^{-1} T^{-1/2} \sum y_{t-1} \left[m(\theta, d_T^{-1} c y_{t-1}) - m\left(\theta, 0\right)\right] + o_P(1) \\
= \ & (d_T^{-1} a_T T^{1/2}) 2c\theta f\left(\theta\right) a_T^{-2} T^{-1} \sum y_{t-1}^2 + o_P(d_T^{-1} a_T T^{1/2})
\end{aligned}
$$

uniformly over $\theta \in [0, A]$, the last equality by the Mean Value Theorem [MVT] and the uniform continuity of the partial derivative $m_2'(\theta, \cdot) = (\cdot + \theta) f\left(\cdot + \theta\right) - (\cdot - \theta) f(\cdot - \theta)$ on compacts. As $a_T^{-2} T^{-1} \sum y_{t-1}^2 \xrightarrow{w} \int (S_c^\kappa)^2 > 0$ a.s., we obtain part (b).

Let next $T a_T^2 \Psi_{T,4}(\theta) := \sum y_{t-1}^2 (\mathbb{I}_{\{|\varepsilon_t - d_T^{-1} c y_{t-1}| \le \theta\}} - p_\theta\left(0\right)) = \sum y_{t-1}^2 (\mathbb{I}_{\{|\varepsilon_t - d_T^{-1} c y_{t-1}| \le \theta\}} - p_\theta(d_T^{-1} c y_{t-1})) + \sum y_{t-1}^2 (p_\theta(d_T^{-1} c y_{t-1}) - p_\theta\left(0\right))$. The first of the latter summations is $O_P(T^{1/2} a_T^2)$ uniformly over $\theta \in [0, A]$ for any $A > 0$, by Proposition A.1 with $\gamma_{Tt} = a_T^{-2} y_{t-1}^2$, $\eta_t = 1$, $\xi_{Tt} = cd_T^{-1} y_{t-1}$, $\psi_{Tt} = 0$ and $\gamma = \int (S_c^\kappa)^4$. The second summation is bounded by $d_T^{-1} 2|c| \|f\|_\infty \sum |y_{t-1}^3| = O_P(d_T^{-1} a_T^3 T) = o_P(a_T^2 T)$ by the MVT and the CMT (with $\|f\|_\infty = \sup_{\mathbb{R}} |f| < \infty$ under Assumption $\mathcal{E}$(ii)), so $\sup_{\theta \in [0,A]} |\Psi_{T,4}\left(\theta\right)| = o_P(1)$. As further $\varsigma_T := a_T^{-1} T^{-1/2} (\sum y_{t-1}^2)^{1/2} \xrightarrow{w} \{\int (S_c^\kappa)^2\}^{1/2} > 0$ a.s., and for any $A > a$ it holds that $p_a\left(0\right) > 0$ by Assumption $\mathcal{E}$(ii), it follows that $p_a\left(0\right) \varsigma_T^2 - \sup_{\theta \in [0,A]} |\Psi_{T,4}\left(\theta\right)| > 0$ with probability approaching one. With the same probability

$$
\begin{aligned}
\sup_{\theta \in [a,A]} \left| T^{1/2} a_T \Psi_{T,3}\left(\theta\right) - \varsigma_T^{-1} \right| &= \sup_{\theta \in [a,A]} \frac{\varsigma_T^{-1} \left|\Psi_{T,4}\left(\theta\right)\right|}{\left|p_\theta\left(0\right) \varsigma_T^2 - \Psi_{T,4}\left(\theta\right)\right|} \\
&\le \frac{\varsigma_T^{-1} \sup_{\theta \in [0,A]} \left|\Psi_{T,4}\left(\theta\right)\right|}{p_a\left(0\right) \varsigma_T^2 - \sup_{\theta \in [0,A]} \left|\Psi_{T,4}\left(\theta\right)\right|} = o_P(1),
\end{aligned}
$$

using the fact that $p_\theta\left(0\right)$ is increasing in θ. Therefore, $T^{1/2} a_T \Psi_{T,3} \xrightarrow{w} \{\int (S_c^\kappa)^2\}^{-1/2}$ in $D[a, A]$, and by the arbitrariness of A, in $D\left[a, \infty\right)$. ■

Proof of Proposition 1. The proof is given under Assumptions $\mathcal{E}$ and $\mathcal{Y}$'. Lemma A.1 implies, in view of the continuity of $p_{(\cdot)}(0)$ and $h_{(\cdot)}$, that in $D\left[a, \infty\right)$, for any $a > 0$,

$$
d_T\{cd_T^{-1} + \tilde{\phi}(0, \cdot)\} = d_T \frac{\Psi_T(\cdot)}{p_{(\cdot)}(0)} \xrightarrow{w} ch_{(\cdot)} + \frac{B(V(\cdot))}{p_{(\cdot)}(0) (\int S^2)^{1/2}} \mathbb{I}_{\{\kappa = 0\}}. \tag{A.1}
$$

Let now $(a_T^{-1} y_{\lfloor T \cdot \rfloor}, \hat{\theta}) \xrightarrow{w} (S_c^\kappa, \Theta)$ hold in $D[0, \infty) \times \mathbb{R}$. From the CMT and the independence of (S_c^κ, Θ) and B, $(\Psi_{T,1}, a_T^{-1} y_{\lfloor T \cdot \rfloor}, \hat{\theta}, \hat{\theta} \vee a) \xrightarrow{w} (B(V), S_c^\kappa, \Theta, \Theta \vee a)$ in $D^2[0, \infty) \times \mathbb{R} \times [a, \infty)$ for any $a > 0$. Using (A.1) and the continuity of V, B, $p_{(\cdot)}(0)$ and $h_{(\cdot)}$, we can conclude that

$$
\begin{aligned}
d_T\{cd_T^{-1} + \tilde{\phi}(0, \hat{\theta} \vee a)\} &\xrightarrow{w} ch_{\Theta \vee a} + \frac{B(V(\Theta \vee a))}{p_{\Theta \vee a}(0) (\int S^2)^{1/2}} \mathbb{I}_{\{\kappa = 0\}} \text{ (as } T \to \infty) \\
&\xrightarrow{a.s.} ch_\Theta + \frac{B(V(\Theta))}{p_\Theta(0) (\int S^2)^{1/2}} \mathbb{I}_{\{\kappa = 0\}} \text{ (as } a \downarrow 0)
\end{aligned} \tag{A.2}
$$

because $\Theta > 0$ a.s. At the same time, for any $a > 0$ there is a $b \in [1, 2]$ such that the distribution function of Θ is continuous at ab, so for every $\epsilon > 0$,

$$\begin{aligned} P\left(d_T|\tilde{\phi}(0,\hat{\theta}) - \tilde{\phi}(0,\hat{\theta}\vee a)| \geq \epsilon\right) &\leq P(\hat{\theta} < a) \leq P(\hat{\theta} \leq ab) \to P(\Theta \leq ab) \text{ (as } T \to \infty) \\ &\to P(\Theta \leq 0) = 0 \text{ (as } a \downarrow 0), \end{aligned}$$

by hypothesis, so

$$\lim_{a\to 0} \operatorname*{limsup}_{T\to\infty} P\left(d_T|\tilde{\phi}(0,\hat{\theta}) - \tilde{\phi}(0,\hat{\theta}\vee a)| > \epsilon\right) = 0.$$

Jointly with (A.2), this implies that

$$\begin{aligned} d_T\{cd_T^{-1} + \tilde{\phi}(0,\hat{\theta})\} &\xrightarrow{w} ch_\Theta + \frac{B(V(\Theta))}{p_\Theta(0)(\int S^2)^{1/2}}\mathbb{I}_{\{\kappa=0\}} \text{ (as } T \to \infty) \qquad \text{(A.3)} \\ &\overset{d}{=} ch_\Theta + \frac{\{V(\Theta)\}^{1/2}}{p_\Theta(0)} \frac{B(1)}{(\int S^2)^{1/2}}\mathbb{I}_{\{\kappa=0\}}, \end{aligned}$$

the convergence according to Billingsley (1968, Th. 4.2), and the equality in distribution by the independence of B and (S, Θ). Hence, for $\kappa = 0$, the proposition. ■

Notice that from the preceding display with $d_T = T$ (i.e., $\kappa = 1$) and $\alpha \in (1, 2)$, the consistency statement of Remark 3.3 regarding the local alternatives $\phi = -c/T$ follows.

A.2 Univariate approximations of $\tilde{\phi}_\theta$

Let in this section $\xi_{Tt}(u) := (u + d_T^{-1}c)y_{t-1}$. Introduce

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

suppressing the dependence on θ which is fixed. Then

$$\tilde{\phi}_\theta(u) = -d_T^{-1}c + \{d_T^{-1}c + \tilde{\phi}(0,\theta)\}\phi_1(u) + \phi_2(u) + \phi_3(u), \qquad \text{(A.4)}$$

with

$$\begin{aligned} \phi_1(u) &:= 1 - \{r_3(u) + r_5(u)\}^{-1}\{r_5(u) - r_5(0) + r_3(u) - r_3(0)\}, \\ \phi_2(u) &:= uh_\theta[1 + \{r_3(u) + r_5(u)\}^{-1} r_1(u)], \\ \phi_3(u) &:= \{r_3(u) + r_5(u)\}^{-1}\{r_2(u) - r_2(0) - uh_\theta r_5(u)\}, \end{aligned}$$

and also

$$\tilde{\phi}_\theta(u) = -d_T^{-1}c + Q_{T,\theta}(u) + \phi_4, \qquad \text{(A.5)}$$

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

$$\phi_4(u) := \frac{r_2(u)}{r_3(u) + r_5(u)} - \frac{r_4(u)}{r_3(u)}\frac{r_5(u)}{r_3(u) + r_5(u)}.$$

To prove Proposition 2, we use the identities (A.4) and (A.5), respectively for parts (a) and (b). The stochastic magnitude orders of $r_i(u)$ are studied first.

Lemma A.2 *Let $\{b_T\}$ be a positive real sequence. Under Assumptions $\mathcal{E}$ and $\mathcal{Y}$' it holds that:*

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

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

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

Proof. First, from the MVT,

$$r_1(u) = \mathbb{I}_{\{u\neq 0\}}\sum y_{t-1}^2\{m_\theta'(\xi_{Tt}(w_t u)) - h_\theta p_\theta(\xi_{Tt}(u))\}$$

for some $w_t \in [0,1]$. Hence,

$$\sup_{|u|\leq A}|r_1(b_T u)| \leq \sup_{|u|\leq b_T A}\max_{t\leq T}\left|m_\theta'(\xi_{Tt}(w_t u)) - h_\theta p_\theta(\xi_{Tt}(u))\right|\sum y_{t-1}^2. \tag{A.6}$$

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 $\max_{t\leq T}|a_T^{-1}y_{t-1}| \xrightarrow{w} \max_{[0,1]}|S_c^\kappa|$). Let $k_T = b_T \vee d_T^{-1}$. For outcomes such that $\max_{t\leq T}|a_T^{-1}y_{t-1}| \leq M_\epsilon$, it holds that

$$\begin{aligned}
&\sup_{|u|\leq b_T A}\max_{t\leq T}\left|m_\theta'(\xi_{Tt}(w_t u)) - h_\theta p_\theta(\xi_{Tt}(u))\right| \\
\leq &\sup_{|u|\leq a_T k_T(A+|c|)M_\epsilon}\left(\left|m_\theta'(u) - m_\theta'(0)\right| + h_\theta\left|p_\theta(u) - p_\theta(0)\right|\right) \to 0 \text{ as } T\to\infty,
\end{aligned}$$

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 $\mathcal{E}$(ii) and (ii) $a_T k_T = o(1)$. By the arbitrariness of ϵ and using (A.6), where $\sum y_{t-1}^2 = O_P(Ta_T^2)$, the first relation in part (a) is obtained.

For the second relation in (a), again by the MVT,

$$|r_3(u) - r_3(0)| \leq |u|\sum|y_{t-1}^3 p_\theta'(\xi_{Tt}(z_t u))| \leq 2|u|\,\|f\|_\infty\sum|y_{t-1}^3|$$

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

To prove part (b), for $M > 0$ define

$$r_5^M(u) := \sum y_{t-1}^2\mathbb{I}_{\{|a_T^{-1}y_{t-1}|\leq M\}}\left[\mathbb{I}_{\{|\varepsilon_t - \xi_{Tt}(u)|\leq\theta\}} - p_\theta(\xi_{Tt}(u))\right].$$

Then, for any $M, K > 0$,

$$P\Big(\sup_{|u|\leq A}|r_5(b_T u)| > K\Big) \leq P\Big(\sup_{|u|\leq A}|r_5^M(b_T u)| > K\Big) + P\Big(\max_{t\leq T}|a_T^{-1}y_t| > M\Big).$$

As $\max_{t\leq T}|a_T^{-1}y_t| \xrightarrow{w} \max_{[0,1]}|S_c^\kappa| < \infty$ a.s., M can be chosen such that $P(\max_{t\leq T}|a_T^{-1}y_t| > M)$ 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, in the supplement CG12, first, that for $c_T := T^{3/4}(a_T^5 b_T)^{1/2}$ and for every fixed u,

$$\mathrm{E}\{r_5^M(b_T u) - r_5^M(0)\}^2 \leq T^{-1/2}c_T^2 M^5 6\,\|f\|_\infty|u| < \infty,$$

so $c_T^{-1}\{r_5^M(b_T u) - r_5^M(0)\} = o_P(1)$. Second, we check that $c_T^{-1}\{r_5^M(b_T(\cdot)) - r_5^M(0)\}$ is tight by applying 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 $\mathbb{I}_{\{|\varepsilon_t - (\cdot)y_{t-1}| \leq \theta\}}$, which are not càdlàg. If we substitute them by

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

a càdlàg modified process, say $\tilde{r}_5^M$, will be 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 a.s.

$$\sup_{|u| \leq A} |r_5^M(b_T(\cdot)) - \tilde{r}_5^M(b_T(\cdot))| \leq \max_{t \leq T} y_t^2 = O_P(a_T^2) = o_P(c_T). \tag{A.7}$$

It is enough, therefore, to establish the tightness of $c_T^{-1}\{\tilde{r}_5^M(b_T(\cdot)) - \tilde{r}_5^M(0)\}$ in $D[-A, A]$. We show in the supplement CG12 that for a fixed M and $u_2 > u_m > u_1 \geq 0$,

$$\begin{aligned}&\mathrm{E}\left(\{\tilde{r}_5^M(b_T u_2) - \tilde{r}_5^M(b_T u_m)\}^2\{\tilde{r}_5^M(b_T u_m) - \tilde{r}_5^M(b_T u_1)\}^2\right) \qquad\qquad (\text{A.8})\\ \leq\ &(T^3 a_T^{10} b_T^2) 4M^{10}\left[16\,\|f\|_\infty (2\theta)^{-1} + 9\,\|f\|_\infty^2\right](u_2 - u_1)^2.\end{aligned}$$

Since additionally, for fixed u, $c_T^{-1}\{\tilde{r}_5^M(b_T u) - \tilde{r}_5^M(0)\} = c_T^{-1}\{r_5^M(b_T u) - r_5^M(0)\} + o_P(1) = o_P(1)$ by (A.7) and the earlier argument, from (A.8) and Theorem 15.6 of Billingsley (1968) it follows that $c_T^{-1}\sup_{|u| \leq A}|\tilde{r}_5^M(b_T u) - \tilde{r}_5^M(0)| = o_P(1)$. This and (A.7) yield $c_T^{-1}\sup_{|u| \leq A}\{r_5^M(b_T(\cdot)) - r_5^M(0)\} = o_P(1)$. The proof is completed by noting that $T^{-1/2}a_T^{-1}r_5^M(0) = O_P(1)$ since it equals $T^{-1/2}a_T^{-1}\sum y_{t-1}^2\mathbb{I}_{\{|a_T^{-1}y_{t-1}| \leq M\}}\left[\mathbb{I}_{\{|\varepsilon_t| \leq \theta\}} - p_\theta(0)\right] + o_P(1)$ (by Proposition A.1), where the normalized summation converges weakly to an a.s. finite random variable (see Lemma 1 of Knight, 1989).

For part (c), we first derive an inequality analogous to (A.8). Introducing

$$r_2^M(u) = \sum y_{t-1}\mathbb{I}_{\{|a_T^{-1}y_{t-1}| \leq M\}}\left[\varepsilon_t \mathbb{I}_{\{|\varepsilon_t - \xi_{Tt}(u)| \leq \theta\}} - m_\theta(\xi_{Tt}(u))\right],$$

it follows by an argument like for the process $\tilde{r}_5^M$ 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 $\mathrm{E}\{r_2^M(b_T u) - r_2^M(0)\}^2 \leq T a_T^3 b_T M^2 C|u|$ for some $C > 0$, see the supplement CG12), by tightness the convergence is uniform on $[-A, A]$, as asserted in part (c). ∎

Proof of Proposition 2. For part (a) we employ equality (A.4). Using the proof of Lemma A.1(c) and Lemma A.2(a,b), we find that

$$\begin{aligned}r_3(u) + r_5(u) &= p_\theta(0)\sum y_{t-1}^2 + \Psi_{T4} + \{r_5(u) - r_5(0)\} + \{r_3(u) - r_3(0)\} \\ &= p_\theta(0)\sum y_{t-1}^2 + o_P(T a_T^2)\end{aligned}$$

uniformly over $|b_T u| \leq A$. Thus, $T^{-1}a_T^{-2}\{r_3(u) + r_5(u)\}$ is uniformly bounded away from zero, since $T^{-1}a_T^{-2}\sum y_{t-1}^2$ is such and $p_\theta(0) > 0$. Together with Lemma A.2(b,c) this implies that

$$\phi_1 = 1 + o_P(1),\ \phi_2 = u(h_\theta + o_P(1)),\ \phi_3 = o_P(T^{-1/4}a_T^{-1/2}b_T^{1/2})$$

uniformly over $|b_T u| \leq A$. Plugging these into (A.4) and using $d_T^{-1}c + \tilde{\phi}(0, \theta) = O_P(d_T^{-1})$, see (A.3), proves part (a).

For part (b), consider eq. (A.5). The CMT yields that $T^{-1}a_T^{-2}r_3(a_T^{-1}(\cdot)) \xrightarrow{w} \int S^2 p_\theta\left((\cdot)S\right)$ and $T^{-1}a_T^{-1}r_4(a_T^{-1}(\cdot)) \xrightarrow{w} \int S m_\theta\left((\cdot)S\right)$ jointly on $D\left[-A, A\right]$, from where the convergence of $a_T Q_{T,\theta}(a_T^{-1}(\cdot))$ follows. It also follows that $\inf_{|u|\leq A}\left|r_3(a_T^{-1}u)\right| \propto Ta_T^2$ in probability (so, by Lemma A.2(b), $\inf_{|u|\leq A}\left|r_3(a_T^{-1}u) - r_5(a_T^{-1}u)\right| \propto Ta_T^2$ in probability) and $\sup_{|u|\leq A}|r_4(a_T^{-1}u)| = O_P(Ta_T)$. Jointly with Lemma A.2(b) and the weak convergence of $T^{-1/2}a_T^{-1}r_2(0)$ these give

$$\begin{aligned}\sup|\phi_4(u)| &\leq \frac{|r_2(0)| + \sup|r_2\left(u\right) - r_2(0)|}{\inf|r_3(u) - r_5(u)|} + \frac{\sup|r_4(u)|}{\inf|r_3(u)|}\frac{\sup|r_4(u)|\sup|r_5(u)|}{\inf|r_3(u) - r_5(u)|}\\ &= o_P(T^{-1/4}a_T^{-1}),\end{aligned}$$

with sup and inf taken over $|u| \leq a_T^{-1}A$.

Further, define

$$\chi_\theta(u) := \{up_\theta(u)\}^{-1}m_\theta(u) \text{ for } u \in \mathbb{R}\setminus\{0\} \text{ and } \chi_\theta(0) := h_\theta = \lim_{u\to 0}\chi_\theta(u).$$

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

$$m_\theta(u) - up_\theta\left(u\right) = \int_u^{u+\theta}(x - u)\{f\left(x\right) - f(2u - x)\}dx < 0$$

by the strict piece-wise monotonicity and symmetry of f, whereas for $u < 0$ the opposite inequality holds (the case $u = 0$ was discussed in Remark 2.5). Using also the continuity of χ_θ, we can conclude that $H(U) = \max_{|u|\leq U}|\chi_\theta(u)| < 1$ for every $U \geq 0$ and H is continuous on $[0, \infty)$. Since

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

where $\xi_{Tt}(u) = (u + d_T^{-1}c)y_{t-1}$, we obtain that

$$\sup_{|u|\leq A}|\mathbb{I}_{\{u\neq 0\}}u^{-1}Q_{T,\theta}(a_T^{-1}u)| \leq H_{T,\theta} := H(A\max_{t\leq T}|\xi_{Tt}(a_T^{-1}u)|) \xrightarrow{w} H(A\sup_{[0,1]}|S_c^\kappa|) =: H_\theta,$$

the convergence using the CMT and the assumption $d_T = o(a_T)$. The proof is completed by observing that $H_\theta < 1$ a.s. because $\sup_{[0,1]}|S_c^\kappa| < \infty$ a.s. ■

A.3 The iterated estimator with fixed scale

PROOF OF THEOREM 4. For $h_\theta < 1$ and $b_T^{-1}\hat{\phi}^{(0)} = O_P(1)$, it follows from Proposition 2(a), by recursive substitution, that $\sup_{i\in\mathbb{N}} |n_T^{-1}\hat{\phi}^{(i)}| = O_P(1)$, where $n_T^{-1} := \min\{b_T^{-1}, T^{1/2}a_T, |c|^{-1}d_T\}$ with $0^{-1} := \infty$. Therefore, again from Proposition 2(a),

$$\hat{\phi}^{(i)} = \tilde{\phi}(0,\theta) + \hat{\phi}^{(i-1)}(h_\theta + R_1^{(i)}) + R_2^{(i)},$$

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

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

and $\rho^{(i)} := \sum_{j=1}^{i}\prod_{k=j+1}^{i}(h_\theta + R_1^{(k)})$. For the terms in $\hat{\phi}_1^{(i)}$ we find that

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

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

$$P\left(\exists \lim_{i\to\infty} \rho^{(i)} \in \left[(1-h_\theta+\eta)^{-1}, (1-h_\theta-\eta)^{-1}\right]\right) \geq P\left(\sup_{i\in\mathbb{N}} |R_1^{(i)}| < \eta\right) \to 1$$

as $T\to\infty$. So if we define $\rho^{(\infty)} := \mathbb{I}\{\exists \lim_{i\to\infty}\rho^{(i)}\}\lim_{i\to\infty}\rho^{(i)} - (1-h_\theta)^{-1}$, it holds that $\rho^{(\infty)} = o_P(1)$ as $T\to\infty$, and

$$P\left(\hat{\phi}_1^{(i)} \to (1-h_\theta)^{-1}\tilde{\phi}(0,\theta) + \rho^{(\infty)}\tilde{\phi}(0,\theta) \text{ as } i\to\infty\right) \to 1$$

as $T\to\infty$. From here

$$P\left(\text{limsup}_{i\to\infty}|\hat{\phi}^{(i)} - (1-h_\theta)^{-1}\tilde{\phi}(0,\theta)| \leq |\rho^{(\infty)}||\tilde{\phi}(0,\theta)| + \sup_{i\in\mathbb{N}}|\hat{\phi}_2^{(i)}|\right) \to 1.$$

As $\tilde{\phi}(0,\theta) = O_P(T^{-1/2}a_T^{-1} \vee |c|d_T^{-1})$ and, in view of the magnitude orders of $R_1^{(i)}$and $R_2^{(i)}$, $\sup_{i\in\mathbb{N}} |\hat{\phi}_2^{(i)}| = o_P(\delta_T^{-1})$, it follows that $\limsup_i |\hat{\phi}^{(i)} - (1-h_\theta)^{-1}\tilde{\phi}(0,\theta)| = o_P(\delta_T^{-1})$ as $T\to\infty$, which implies the conclusions in part (a) about iterated limits (for $\kappa = 0$) and, jointly with (A.3), those of Remark 4.4 about consistency against traditional local alternatives (for $\kappa = 1$).

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

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

and $\rho^{(\nu_T)} \xrightarrow{P} (1-h_\theta)^{-1}$ as $T\to\infty$. So $\hat{\phi}_1^{(\nu_T)} = (1-h_\theta)^{-1}\tilde{\phi}(0,\theta) + o_P(T^{-1/2}a_T^{-1}) = O_P(T^{-1/2}a_T^{-1} \vee |c|d_T^{-1})$, and as $\hat{\phi}_2^{(\nu_T)} = o_P(\delta_T^{(1)})$ with $\delta_T^{(1)} := \delta_T^{-1}$, we find that $\hat{\phi}^{(\nu_T)} =$

$O_P(\delta_T^{(1)})$. We can treat $\hat{\phi}^{(2\nu_T)}$ as obtained by iteration with initial value $\hat{\phi}_1^{(\nu_T)}$ and with initial magnitude order $\delta_T^{(1)}$ instead of b_T. By the same argument as for $\hat{\phi}^{(\nu_T)}$, we can conclude that $\hat{\phi}_1^{(2\nu_T)} = (1-h_\theta)^{-1}\tilde{\phi}(0,\theta) + o_P(T^{-1/2}a_T^{-1})$, $\hat{\phi}_2^{(2\nu_T)} = o_P(\delta_T^{(2)})$ and $\hat{\phi}^{(2\nu_T)} = O_P(\delta_T^{(2)})$, $\delta_T^{(2)} := \max\{T^{-1/2}a_T^{-1}, [(T^{-1/2}a_T^{-1})^3 b_T]^{1/4}, |c|d_T^{-1}\}$. Further, by induction, $\hat{\phi}_1^{(\nu_T^2)} = (1-h_\theta)^{-1}\tilde{\phi}(0,\theta) + o_P(T^{-1/2}a_T^{-1})$ and $\hat{\phi}_2^{(\nu_T^2)} = o_P(\delta_T^{(\nu_T)})$, $\delta_T^{(\nu_T)} := \max\{T^{-1/2}a_T^{-1}, [(T^{-1/2}a_T^{-1})^{2^{\nu_T}-1}b_T]^{2^{-\nu_T}}, |c|d_T^{-1}\}$. For the b_T considered in part (b), $\delta_T^{(\nu_T)} = O(T^{-1/2}a_T^{-1} \vee |c|d_T^{-1})$ and thus, $\hat{\phi}^{(\nu_T^2)} = (1-h_\theta)^{-1}\tilde{\phi}(0,\theta) + o_P(\delta_T^{(\nu_T)})$. Finally, iterating from $\hat{\phi}^{(\nu_T^2)}$ to $\hat{\phi}^{(\psi(T))}$ (if $\nu_T^2 < \psi(T)$) maintains $\hat{\phi}^{(\psi(T))} = (1-h_\theta)^{-1}\tilde{\phi}(0,\theta) + o_P(\delta_T^{(\nu_T)})$, from where part (b) and the consistency part of Remark 4.4 are obtained. ■

Proof of Proposition 5. The proof is similar to that of Theorem 4(b). In view of Proposition 2(b) we can write

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

since $H_{T,\theta}^{\lfloor T^{\nu/2}\rfloor} \xrightarrow{P} 0$ exponentially fast. Then the desired convergences follow by applying Proposition 4(b) with $b_T = (T^{-1/4}a_T^{-1}) \vee (|c|d_T^{-1})$ 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 6. The proof is, in a sense, reciprocal to that of Theorem 4, and is relegated to the supplement CG12.■

A.4 The iterated estimator with iteration over the scale

Here and in the next section, let $\xi_{Tt}(u) := (u + a_T^{-1}T^{-1/2}c)y_{t-1}$ and $\psi_{Tt}(s) := s'\Delta\mathbf{y}_{t-1}$ for $s \in \mathbb{R}^k$ and $\Delta\mathbf{y}_{t-1} := (\Delta y_{t-1}, ..., \Delta y_{t-k})$. For the proof of Theorem 7 we need a version of Proposition A.1 with this $\xi_{Tt}(u)$, $\psi_{Tt}\,(0)$ for arbitrary k and uniformity over (u,θ), whereas for section 6 we need $\xi_{Tt}(u)$, $\psi_{Tt}(s)$ with k as in Assumption $\mathcal{Y}(k)$ and uniformity over (u,s,θ). Dimensions other than k could be considered if interest is in estimated autoregressions with order possibly different from the true one. Let finally

$$U_T(u,s,\theta) := T^{-1/2}\sum \gamma_{Tt}\left[\eta_t \mathbb{I}_{\{|\varepsilon_t - \xi_{Tt}(u) - \psi_{Tt}(s)| \le \theta\}} - m(\theta, \xi_{Tt}(u) + \psi_{Tt}(s))\right],$$

the rest of the notation being as in Proposition A.1.

Proposition A.2 *Under Assumptions $\mathcal{E}$, and $\mathcal{Y}$ or $\mathcal{Y}(k)$, and under the hypotheses that*

$$\max_{1\le t\le T}|\gamma_{Tt}| = O_P(1),\ T^{-1}\sum \gamma_{Tt}^2 \xrightarrow{w} \gamma < \infty \ a.s.,$$

it holds as $T \to \infty$ that

$$\sup_{(u,s,\theta)\in K_T} |U_T(u,s,\theta) - U_T(0,0,\theta)| = o_P(1) \tag{A.9}$$

with $K_T := \{(u,s,\theta) : |u| \le a_T^{-1}T^{-1/2}C, \|s\| \le b_T^{-1}C, \theta \in [0,A]\}$ for a real sequence b_T with $T^{-1/2}b_T \to \infty$ and $T^{1/4}b_T/a_T \to \infty$. Furthermore,

$$\sup_{(u,s,\theta)\in K_T}\left|\sum \gamma_{Tt}\left\{\varepsilon_t[\mathbb{I}_{\{|\varepsilon_t - \xi_{Tt}(u) - \psi_{Tt}(s)|\le\theta\}} - \mathbb{I}_{\{|\varepsilon_t - \xi_{Tt}(0)|\le\theta\}}] - 2u\theta f(\theta)y_{t-1}\right\}\right| = o_P(T^{1/2}).$$

Proof. The assertions follow from Proposition A.1 by adapting the compactness and monotonicity arguments of Koul (2002) for his Theorem 7.2.1. In the supplement CG12 we discuss the necessary modifications. ■

Proof of Theorem 7. From Proposition A.2 under Assumptions $\mathcal{E}$ and $\mathcal{Y}$ it follows that

$$\sum y_{t-1}^{i}\varepsilon_t^{2-i}\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)|\leq\theta\}}=\sum y_{t-1}^{i}\varepsilon_t^{2-i}\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(0)|\leq\theta\}}+2u\theta^{2-i}f(\theta)\sum y_{t-1}^{i+1}+o_P(T^{1/2}a_T^{i}) \tag{A.10}$$

for $i=1,2$, uniformly on K_T. Since, in view of the CMT, $\sum y_{t-1}^3=O_P(Ta_T^3)$, and, from the proof of Lemma A.1(c), $\sum y_{t-1}^2\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(0)|\leq\theta\}}=p_\theta\,(0)\sum y_{t-1}^2+o_P(Ta_T^2)$ uniformly over $\theta\in[0,A]$, where $p_\theta\,(0)\,T^{-1}a_T^{-2}\sum y_{t-1}^2$ is uniformly bounded away from zero in probability over $\theta\in[a,A]$ $(0<a<A)$, we find that

$$\begin{aligned}
\tilde{\phi}(u,\theta) &= -T^{-1/2}a_T^{-1}c+\frac{\sum y_{t-1}\varepsilon_t\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)|\leq\theta\}}}{\sum y_{t-1}^2\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)|\leq\theta\}}}=\\
&= -T^{-1/2}a_T^{-1}c+\frac{\sum y_{t-1}\varepsilon_t\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(0)|\leq\theta\}}+2\theta uf(\theta)\sum y_{t-1}^2}{\sum y_{t-1}^2\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(0)|\leq\theta\}}+O_P(T^{1/2}a_T^2)}+o_P(T^{-1/2}a_T^{-1})\\
&= T^{-1/2}a_T^{-1}c+\tilde{\phi}(0,\theta)+uh_\theta(1+o_P(1))+o_P(T^{-1/2}a_T^{-1}),
\end{aligned}$$

uniformly over $\theta\in[a,A]$, $u\in K_{Tu}(C):=[-T^{-1/2}a_T^{-1}C,T^{-1/2}a_T^{-1}C]$.

By hypotheses (i) and (ii), $\sigma(u)=O_P(1)$ and $\sigma(u)^{-1}=O_P(1)$ uniformly over $K_{Tu}(C)$ for all $C>0$. Thus, from the previous results,

$$\tilde{\phi}(u,\sigma(u))=\tilde{\phi}(0,\sigma(u))+uh(\sigma(u))+o_P(T^{-1/2}a_T^{-1})$$

uniformly on such compacts. Still uniformly, $\sigma(u)=\sigma(0)+o_P(1)$ by hypothesis (i), $\tilde{\phi}(0,\sigma(u))-\tilde{\phi}(0,\sigma(0))=o_P(T^{-1/2}a_T^{-1})$ by the tightness of $\tilde{\phi}(0,\cdot)$ in the uniform metric (see (A.1), where the limiting process is continuous), and $h(\sigma(u))=h(\sigma(0))+o_P(1)$ by uniform continuity of h on compacts. Hence,

$$\tilde{\phi}(u,\sigma(u))=\tilde{\phi}(0,\sigma(0))+uh(\sigma(0))+o_P(T^{-1/2}a_T^{-1}) \tag{A.11}$$

uniformly on compacts $K_{Tu}(C)$.

By Proposition 1, $\upsilon_T=O_P(T^{-1/2}a_T^{-1})$, so (A.11) is valid for υ_T in place of u, and since by the same proposition $T^{1/2}a_T\upsilon_T$ is bounded away from zero in probability, υ_T is a non-zero near-fixed point of $\phi(\cdot,\sigma(\cdot))$. Numerical convergence of the iterates of $\phi(\cdot,\sigma(\cdot))$ follows from (A.11) as in the proof of Theorem 4. ■

Finally, we show that the quantile functions (5.18) satisfy hypothesis (i) of Theorem 7. Since $\sigma\,(0)-q_\tau=o_P(1)$, it is enough to show that $\sup_{u\in K_{Tu}(C)}|\sigma\,(u)-q_\tau|=o_P\,(1)$. In fact, in the sense of inclusion of events, $\{\sigma(u)<a\}\subset\{F_T\,(u,a)\geq\tau\}$ and $\{\sigma(u)>a\}\subset\{F_T\,(u,a)<\tau\}$, where $F_T\,(u,a):=T^{-1}\sum_{t=1}^{T}\mathbb{I}_{\{|\varepsilon_t-uy_{t-1}|\leq a\}}$. Thus, for every $\varepsilon>0$,

$$\begin{aligned}
\{|\sigma\,(u)-q_\tau|>\varepsilon\} &\subset \{F_T\,(u,q_\tau-\varepsilon)\geq\tau\}\cap\{F_T\,(u,q_\tau+\varepsilon)<\tau\}\\
&\subset \{F_T\,(0,q_\tau-\varepsilon)+|F_T\,(u,q_\tau-\varepsilon)-F_T\,(0,q_\tau-\varepsilon)\,|\geq\tau\}\\
&\qquad \cap\{F_T\,(0,q_\tau+\varepsilon)-|F_T\,(u,q_\tau+\varepsilon)-F_T\,(0,q_\tau+\varepsilon)\,|<\tau\}
\end{aligned}$$

so

$$\left\{\sup_{u\in K_{Tu}(C)}|\sigma(u)-q_\tau|>\varepsilon\right\}\subset\left\{F_T(0,q_\tau-\varepsilon)+\sup_{u\in K_{Tu}(C)}|F_T(u,q_\tau-\varepsilon)-F_T(0,q_\tau-\varepsilon)|\geq\tau\right\}$$
$$\cap\left\{F_T(0,q_\tau+\varepsilon)-\sup_{u\in K_{Tu}(C)}|F_T(u,q_\tau+\varepsilon)-F_T(0,q_\tau+\varepsilon)|<\tau\right\}.$$

The two suprema on the right-hand side are $o_P(1)$ by Proposition A.2, whereas $F_T(0,q_\tau\mp\varepsilon)\xrightarrow{P} F(q_\tau\mp\varepsilon)\lesseqgtr\tau$ under Assumption $\mathcal{E}$(ii), so $\sup_{u\in K_{Tu}(C)}|\sigma(u)-q_\tau|=o_P(1)$.

A.5 Higher-order autoregressions

Asymptotics for the iterates of Φ defined in (6.20) follow as in the proof of Theorem 7.

We note first that under Assumption $\mathcal{Y}(k)$, 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 $Q=1-\sum_{i=1}^k\partial_i\neq0$ and $\{q_i\}_{i=0}^\infty$ decrease exponentially. It can be used to show that $a_T^{-2i}T^{-1}\sum y_{t-1}^{2i}\xrightarrow{w}Q^{2i}\int S^{2i}\in(0,\infty)$ a.s. $(i=1,2)$. Thus, by Proposition A.2 for $\gamma_{Tt}=a_T^{-1}y_{t-1}$ and $\gamma_{Tt}=a_T^{-2}y_{t-1}^2$, a version of (A.10) holds with $\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)|\leq\theta\}}$ replaced by $\mathbb{I}_{\{|\Delta\hat{y}_t-uy_{t-1}|\leq\theta\}}=\mathbb{I}_{\{|\varepsilon_t-\xi_{Tt}(u)-\psi_{Tt}|\leq\theta\}}$, $\psi_{Tt}:=(\hat\partial-\partial)'\Delta\mathbf{y}_{t-1}$. As in the proof of Theorem 7, this implies for $\tilde\Phi(u,\sigma(u)):=-cT^{-1/2}a_T^{-1}+(\sum y_{t-1}^2\mathbb{I}_{\{|\Delta\hat{y}_t-uy_{t-1}|\leq\sigma(u)\}})^{-1}\sum y_{t-1}\varepsilon_t\mathbb{I}_{\{|\Delta\hat{y}_t-uy_{t-1}|\leq\sigma(u)\}}$ the expansion

$$\tilde\Phi(u,\sigma(u))=\tilde\phi(0,\sigma(0))+uh(\sigma(0))+o_P(T^{-1/2}a_T^{-1})\tag{A.12}$$

similar to (A.11) and uniform on $\theta\in[0,A]$, $u\in K_{Tu}(C)$. As further

$$\begin{aligned}|\Phi(u,\sigma(u))-\tilde\Phi(u,\sigma(u))| &\leq (\sum y_{t-1}^2\mathbb{I}_{\{|\Delta\hat{y}_t-uy_{t-1}|\leq\theta\}})^{-1}\|\hat\partial-\partial\|\|\sum y_{t-1}\Delta\mathbf{y}_{t-1}\mathbb{I}_{\{|\Delta\hat{y}_t-uy_{t-1}|\leq\theta\}}\|\\ &\leq O_P(T^{-1}a_T^{-1}b_T^{-1})\max_{1\leq t\leq T}|a_T^{-1}y_{t-1}|\sum\|\Delta\mathbf{y}_{t-1}\|\end{aligned}$$

by the updated version of (A.10), and $\sum\|\Delta\mathbf{y}_{t-1}\|=O_P(\max\{T,a_TT^\varepsilon\})$ with $\varepsilon>0$ arbitrarily small by Markov's inequality:

$$\mathrm{E}(\sum\|\Delta\mathbf{y}_{t-1}\|)^{\min(1,\alpha-\varepsilon)}\leq\sum\mathrm{E}\|\Delta\mathbf{y}_{t-1}\|^{\min(1,\alpha-\varepsilon)}=O(T),$$

for $T^{-1/2}b_T\to\infty$ and $T^{1/4}b_T/a_T\to\infty$ it holds that $|\Phi(u,\sigma(u))-\tilde\Phi(u,\sigma(u))|=o_P(T^{-1/2}a_T^{-1})$. Hence, also $\Phi(u,\sigma(u))$ has expansion (A.12), from where the properties of its iterates follow as in the proof of Theorem 4. ■

References

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

Cavaliere G. and I. Georgiev (2009). Robust inference in autoregressions with multiple outliers. *Econometric Theory* 25: 1625-1661

Cavaliere G. and I. Georgiev (2012). Supplement to "Exploiting Infinite Variance through Dummy Variables in Non-Stationary Autoregressions", http://docentes.fe.unl.pt/~igrg/dusup.pdf

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

Davis R., Knight K., Liu J. (1992). M-estimation for autoregressions with infinite variance, *Stochastic Processes and their Applications* 40: 145-180

Davis R. and S. Resnick (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities, *The Annals of Probability* 13: 179-195.

Engler E. and B. Nielsen (2009). The empirical process of autoregressive residuals. *Econometrics Journal* 12: 367-381

Hasan, M. N. (2001). Rank tests of unit root hypothesis with infinite variance errors. *Journal of Econometrics* 104: 49-65.

Hendry D., Johansen S. and C. Santos (2008). Automatic selection of indicators in a fully saturated regression. Computational Statistics 23: 317-335.

Johansen S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, J. and Shephard, N. eds. *The Methodology and Practice of Econometrics: a Festschrift in Honour of David F. Hendry*. Oxford: Oxford University Press.

Johansen S. and B. Nielsen (2010). Discussion: The forward search: Theory and data analysis. *Journal of the Korean Statistical Society* 39: 137-145.

Johansen S. and B. Nielsen (2011). Asymptotic theory for iterated one-step Huber-skip estimators. Department of Economics. University of Copenhagen

Koul H. and Ossiander M. (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*. Berlin: Springer.

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.

Ling S. (2005), Self-weighted least absolute deviation estimation for infinite variance autoregressive models, *Journal of the Royal Statistical Society B* 67, 381–393.

Pagan A. and Ullah A. (1999). *Nonparametric Econometrics*. CUP: Cambridge

Phillips P.C.B. (1987), Time Series Regression with a Unit Root, *Econometrica* 55, 277–301.

Phillips P.C.B. (1990). Time series regression with a unit root and infinite-variance errors. *Econometric Theory* 6: 44-62.

Phillips, P.C.B. & T. Magdalinos (2007a) Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.

Phillips, P.C.B. & T. Magdalinos (2007b) Limit theory for moderate deviations from a unit root under weak dependence. In G.D.A. Phillips & E. Tzavalis (eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, pp. 123–162. Cambridge University Press.

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

Samarakoon M. and K. Knight (2009), A note on unit root tests with infinite variance noise. *Econometric Reviews* 28: 314-334

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

Table 1. Empirical Size and Power of the dummy-based Tests

	Estimation with fixed θ			Full estimation initialized at			
T	$\xi_T(0,\theta)$	$\xi_T^{(N)}(0,\theta)$	$\xi_T^{(\sqrt{T})}(\phi_{OLS},\theta)$	0	$\tilde{\phi}_\theta^{(N)}(0)$	$\tilde{\phi}_\theta^{(\sqrt{T})}(LS)$	t_{OLS}
				Empirical Size			
	$\alpha = 3/2$						
100	6.1	5.2	6.4	5.4	6.2	6.6	3.9
500	6.0	5.2	5.8	5.4	5.7	5.8	3.9
	$\alpha = 1$ (Cauchy)						
100	5.4	5.2	6.8	3.8	4.6	5.4	2.9
500	5.3	5.1	5.8	4.1	4.4	4.7	2.9
	$\alpha = 1$ (Bimodal)						
100	5.6	5.3	7.8	*23.1*	*26.1*	*27.0*	2.9
500	5.3	4.7	5.7	*34.0*	*35.8*	*36.0*	2.9
	$\alpha = 1/2$						
100	5.3	5.4	*28.4*	3.3	4.8	*26.1*	1.9
500	5.1	5.2	*38.0*	4.1	4.8	*36.4*	1.7
			Empirical rejection frequencies for $\phi = -7/d_T$				
	$\alpha = 3/2$						
100	49.8	44.2	49.9	37.6	43.9	46.1	15.4
500	55.1	51.6	54.8	47.9	51.2	51.7	11.1
10^4	61.9	60.5	61.2	59.1	59.7	59.8	7.6
∞	67.1	67.1	67.1	70.0	70.0	70.0	
	$\alpha = 1$ (Cauchy)						
100	66.8	65.4	69.4	52.3	62.1	64.2	10.7
500	71.0	69.9	71.5	64.2	67.7	68.3	7.7
10^4	72.9	72.6	72.9	70.4	70.9	70.9	5.2
∞	74.2	74.2	74.2	74.8	74.8	74.8	
	$\alpha = 1$ (Bimodal)						
100	39.6	38.9	43.9	*11.5*	*25.4*	*27.1*	11.2
500	45.2	43.7	45.5	*23.2*	*33.8*	*34.3*	8.0
10^4	47.9	47.1	47.5	*35.4*	*43.1*	*43.2*	5.5
∞	48.9	48.9	48.9	*n.a.*	*n.a.*	*n.a.*	
	$\alpha = 1/2$						
100	51.8	51.6	*47.1*	54.6	68.3	*56.0*	5.7
500	54.4	54.4	*49.1*	64.0	71.5	*54.9*	4.0
10^4	57.6	57.0	*52.0*	71.7	73.9	*53.2*	3.0
∞	63.0	63.0	*n.a.*	78.3	78.3	*n.a.*	

Notes. Monte Carlo results based on 50,000 replications. Experiments where the conditions of Theorem 4 or 7 are not satisfied are in italics.