Numéro spécial en l'honneur des 80 ans de Denis Bosq / Special issue in honour of Denis Bosq's 80th birthday

Autoregression Depth

Davy Paindaveine[†] **and Germain Van Bever**[‡]

Université libre de Bruxelles[†], *Université Toulouse 1 Capitole*[†] *and Université de Namur*[‡]

Abstract: We introduce a concept of autoregression depth that provides a robust ordering of autoregression parameter values according to their adequacy with respect to the underlying process. We derive a uniform strong consistency result for the corresponding sample autoregressive depth, which allows us to prove that the sample deepest parameter value is strongly consistent for its population version. Our depth concept finds applications in both point estimation and hypothesis testing: regarding point estimation, the deepest parameter value provides a robust estimator of the parameter of autoregressive processes, which we show to be strongly consistent by complementing the aforementioned consistency results with a Fisher-consistency result. Regarding hypothesis testing, the depth of the zero parameter value yields a natural test statistic to test for randomness. We investigate the AR(1) case in some details. Our results are illustrated with Monte Carlo exercises.

1. Introduction

Statistical depth is a device that allows measuring centrality of a d-vector z with respect to a probability measure P over $\mathbb{R}^d$. Many such depths are available in the literature; see [13]. Arguably, the most famous depth is the halfspace depth introduced in [10], that is defined as

$$HD(z,P) := \inf_{u \in \mathcal{S}^{d-1}} P[u'(Z-z) \geq 0],$$

where Z denotes a random d-vector with distribution P and $\mathcal{S}^{d-1} := \{z \in \mathbb{R}^d : \|z\|^2 = z'z = 1\}$ is the unit sphere of $\mathbb{R}^d$. Like any other depth, halfspace depth provides a center-outward ordering of the points in $\mathbb{R}^d$: if $HD(z_1,P) > HD(z_2,P)$, then z_1 is more central than z_2 with respect to P. This ordering is with respect to the central region $M_P := \{z \in \mathbb{R}^d : HD(z,P) = \sup_{y \in \mathbb{R}^d} HD(y,P)\}$, which is non-empty (see, e.g., [9]). If a unique center is needed, then it is traditional to use the *Tukey median*, that is defined as the barycentre of M_P and extends to the multivariate case the univariate concept of median. Like any other halfspace depth region $R_\alpha(P) := \{z \in \mathbb{R}^d : HD(z,P) \geq \alpha\}$, the most central region is a convex subset of $\mathbb{R}^d$, so that the Tukey median has itself maximal depth, hence constitutes a valid representative of M_P.

AMS 2000 subject classifications: Primary 62M10; secondary 62G20

Keywords and phrases: Autoregression processes, Runs tests, Statistical depth, Sign procedures, Tests of randomness

Sample versions of halfspace depth and Tukey median are readily obtained by plugging P_n, the empirical probability measure associated with a sample of observations $Z_1, \ldots, Z_n$, into the definitions above. [9] established the upper semicontinuity of the halfspace depth function and the compactness of the depth regions $R_\alpha(P)$, $\alpha > 0$. Jointly with the uniform strong consistency of the sample halfspace depth function ([3]), this allows showing that the sample Tukey median is strongly consistent for its population counterpart.

Depth notions have been extended to parametric settings other than location. [6] extended statistical depth to virtually any parametric setting by defining the *tangent depth* $D(\theta, P)$ of a d-dimensional parameter θ with respect to $P \in \mathcal{P}$, for $\mathcal{P}$ a parametric set of distributions with index set Θ. This depth notion quantifies the appropriateness of θ (as a parameter for P) by considering

$$D(\,\cdot\,, P) \;:\; \Theta \to \mathbb{R}^+ \;:\; \theta \mapsto D(\theta, P) = \inf_{u \in S^{d-1}} P[u' \nabla_\theta F_\theta(Z) \geq 0],$$

where $F_\theta(z)$ measures (lack of) fit of the parameter value θ for observation z. For example, if θ is a location parameter, then setting $F_\theta(z) = h(\|z-\theta\|)$, with $h : \mathbb{R}^+ \to \mathbb{R}^+$ smooth and monotone increasing, provides the halfspace depth $HD(\theta, P)$. Similarly, in the regression context where $z = (x, y)$ involves a d-dimensional covariate x and a scalar response y, taking $F_\theta(z) = h(|y - \theta' x|)$, with h as above, yields the concept of regression depth from [8]. Tangent depth therefore provides a turnkey notion of depth in any parametric space (by using the generic choice of measure of fit $F_\theta(z) = -\log L_\theta(z)$, for $L_\theta(z)$ the likelihood of z under parameter value θ).

Few general results about tangent depth are available, however, as the behaviour of the obtained depth depends crucially on the geometry of the parameter space and on the chosen measure of fit. Consequently, the properties of each parametric depth function need to be explored on a case by case basis. [7], for example, studied (a modification of) the tangent depth for shape parameters in multivariate distributions.

The present contribution introduces and studies a concept of autoregression depth. The concept, which is of a tangent depth nature, provides means for comparing the relevance of two such parameter values and, in the context of autoregressive processes, a robust estimate of the autoregressive parameter. The proposed autoregression depth also allows conducting hypothesis testing, in particular in the framework of testing for randomness, where it provides (generalized) runs tests of randomness.

The outline of the paper is as follows. Section 2 defines the concept and establishes its Fisher-consistency in autoregressive models. Section 3 introduces the corresponding sample depth function and proves its uniform strong consistency. Section 4 further explores the concept in the particular case of AR(1) models, provides Bahadur representation results for the depth-based estimator of the parameter and a test for randomness in that setting. Finally, some final comments close the paper.

2. Autoregression depth

Let $\{Y_t : t \in \mathbb{Z}\}$ be an autoregressive process of order at most p, satisfying

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \varepsilon_t \tag{2.1}$$

for any integer t, where the ε_t's are mutually independent and admit the common density f. Here, we consider *any* parameter value $\phi = (\phi_1, \ldots, \phi_p) \in \mathbb{R}^p$, hence also those providing non-causal autoregressive processes or autoregressive processes with an order that is strictly smaller than p. Throughout, the density f will be assumed to belong to the collection $\mathcal{F}$ of densities having a unique median at zero (i.e., such that the corresponding cumulative distribution F takes value $1/2$ at zero only). Our goal in this section is to introduce a concept of *autoregression depth* that will allow us to measure how well an autoregressive model of this form fits a given stationary process.

To this end, consider a stationary process $\{Z_t : t \in \mathbb{Z}\}$ on the measure space $(\Omega, \mathcal{F}, P)$, which may or may not be an autoregressive process. For any $\phi \in \mathbb{R}^p$, we define the autoregressive depth of ϕ with respect to P as

$$ARD_p(\phi, P) := \inf_{u \in \mathcal{S}^{p-1}} P[u' \nabla_\phi F(Z_t, Z_{t-1}, \ldots, Z_{t-p}, \phi) \geq 0],$$

where $F(Z_t, Z_{t-1}, \ldots, Z_{t-p}, \phi) = h(|Z_t - \sum_{j=1}^{p} \phi_j Z_{t-j}|)$ involves an arbitrary function $h : \mathbb{R}^+ \to \mathbb{R}^+$ that is differentiable, monotone strictly increasing and is such that $h(0) = 0$. A direct computation shows that, irrespective of the function h adopted,

$$ARD_p(\phi, P) = \inf_{u \in \mathcal{S}^{p-1}} P[\varepsilon_t(\phi) u' \mathcal{Z}_{t-1} \geq 0],$$

where we let $\varepsilon_t(\phi) := Z_t - \sum_{j=1}^{p} \phi_j Z_{t-j}$ and $\mathcal{Z}_{t-1} := (Z_{t-1}, \ldots, Z_{t-p})'$. The larger the autoregressive depth of ϕ with respect to P, the better the corresponding autoregressive process fits the underlying stationary process.

Autoregression depth is Fisher-consistent in the sense that, if the underlying process $\{Z_t : t \in \mathbb{Z}\}$ is a stationary autoregressive process of order at most p, with parameter $\phi_0 = (\phi_{01}, \ldots, \phi_{0p})$ say, then the autoregression depth is uniquely maximized at $\phi = \phi_0$. Fisher-consistency thus also holds under over-identification, that is, when the order, q say, of the underlying autoregressive process is smaller than p. More precisely, we have the following result.

Theorem 2.1. *Let $\{Z_t : t \in \mathbb{Z}\}$ be a causal (hence, stationary) autoregressive process of order q with autoregressive parameter $\tilde{\phi}_0 = (\tilde{\phi}_{01}, \ldots, \tilde{\phi}_{0q})$ and innovation density $f \in \mathcal{F}$. Then, letting $\phi_0 = (\tilde{\phi}_0', 0, \ldots, 0)' \in \mathbb{R}^p$,*

$$ARD_p(\phi, P) \leq ARD_p(\phi_0, P) = \frac{1}{2}$$

for any $\phi \in \mathbb{R}^p$, and the equality holds if and only if $\phi = \phi_0$.

The proof requires the following preliminary result.

Lemma 2.1. *Let* $\{Z_t : t \in \mathbb{Z}\}$ *be a causal autoregressive process of order* q *with autoregressive parameter* $\tilde{\phi}_0 = (\tilde{\phi}_{01}, \ldots, \tilde{\phi}_{0q})$ *and innovation density* $f \in \mathcal{F}$. *Then,* $P[u'\mathcal{Z}_{t-1} = 0] = 0$ *for any* $u \in \mathbb{R}^p \setminus \{0\}$.

Proof. Fix $u \in \mathbb{R}^p \setminus \{0\}$ and let $r := \min\{j = 1, \ldots, p : u_j \neq 0\}$. Then $P[u'\mathcal{Z}_{t-1} = 0] = P[Z_{t-r} = -\sum_{j=r+1}^{p} v_j Z_{t-j}]$, where we let $v_j = u_j/u_r$ and where a sum over an empty collection of indices is defined as zero. Conditioning with respect to $\mathcal{F}_{t-r-1} := \sigma(Z_{t-r-1}, Z_{t-r-2}, \ldots)$ readily yields

$$
\begin{aligned}
P[u'\mathcal{Z}_{t-1} = 0] &= P[\varepsilon_{t-r} = -\textstyle\sum_{j=r+1}^{p} v_j Z_{t-j} - \sum_{j=1}^{q} \tilde{\phi}_{0j} Z_{t-r-j}] \\
&= \mathrm{E}[P[\varepsilon_{t-r} = -\textstyle\sum_{j=r+1}^{p} v_j Z_{t-j} - \sum_{j=1}^{q} \tilde{\phi}_{0j} Z_{t-r-j} | \mathcal{F}_{t-r-1}]] = 0,
\end{aligned}
$$

since the distribution of ε_{t-r} conditional on $\mathcal{F}_{t-r-1}$, which coincides with the unconditional distribution of ε_{t-r}, admits a density. □

We can now prove Theorem 2.1.

Proof. Lemma 2.1 ensures that $P[u'\mathcal{Z}_{t-1} = 0] = 0$ for any $u \in \mathcal{S}^{p-1}$. Therefore, causality implies that

$$
\begin{aligned}
P[\varepsilon_t(\phi_0)u'\mathcal{Z}_{t-1} \geq 0] \;&=\; P[\varepsilon_t(\phi_0) \leq 0]P[u'\mathcal{Z}_{t-1} < 0] + P[\varepsilon_t(\phi_0) \geq 0]P[u'\mathcal{Z}_{t-1} > 0] \\
&= \frac{1}{2}P[u'\mathcal{Z}_{t-1} < 0] + \frac{1}{2}P[u'\mathcal{Z}_{t-1} > 0] = \frac{1}{2}
\end{aligned}
$$

for any $u \in \mathcal{S}^{p-1}$, so that $ARD_p(\phi_0, P) = 1/2$.

Now, fix $\phi \in \mathbb{R}^p \setminus \{\phi_0\}$. Pick then $u_0 := (\phi - \phi_0)/\|\phi - \phi_0\| =: (\phi - \phi_0)/\lambda_0$ and write $\varepsilon_t(\phi) = \varepsilon_t(\phi_0) + (\phi_0 - \phi)'\mathcal{Z}_{t-1} = \varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1}$. We consider three situations.

(i) In the case $P[u_0'\mathcal{Z}_{t-1} < 0] > 0$ and $P[u_0'\mathcal{Z}_{t-1} > 0] > 0$, it holds

$$
\begin{aligned}
P[\varepsilon_t(\phi)u_0'\mathcal{Z}_{t-1} \geq 0] &= P[(\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1})u_0'\mathcal{Z}_{t-1} \geq 0] \\
&= P[(\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1})u_0'\mathcal{Z}_{t-1} \geq 0 | u_0'\mathcal{Z}_{t-1} < 0]P[u_0'\mathcal{Z}_{t-1} < 0] \\
&\qquad + P[(\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1})u_0'\mathcal{Z}_{t-1} \geq 0 | u_0'\mathcal{Z}_{t-1} > 0]P[u_0'\mathcal{Z}_{t-1} > 0] \\
&= P[\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1} \leq 0 | u_0'\mathcal{Z}_{t-1} < 0]P[u_0'\mathcal{Z}_{t-1} < 0] \\
&\qquad + P[\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1} \geq 0 | u_0'\mathcal{Z}_{t-1} > 0]P[u_0'\mathcal{Z}_{t-1} > 0]. \qquad (2.2)
\end{aligned}
$$

Note that $P[\varepsilon_t(\phi_0) - \lambda_0 u_0' \mathcal{Z}_{t-1} \leq 0 | u_0' \mathcal{Z}_{t-1} < 0]$

$$= \frac{1}{P[u_0'\mathcal{Z}_{t-1} < 0]} \int_{-\infty}^{0} P[\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1} \leq 0 | u_0'\mathcal{Z}_{t-1} = s] f^{u_0'\mathcal{Z}_{t-1}}(s)\, ds$$

$$= \frac{1}{P[u_0'\mathcal{Z}_{t-1} < 0]} \int_{-\infty}^{0} P[\varepsilon_t \leq \lambda_0 s | u_0'\mathcal{Z}_{t-1} = s] f^{u_0'\mathcal{Z}_{t-1}}(s)\, ds < \frac{1}{2}$$

and, similarly, $P[\varepsilon_t(\phi_0) - \lambda_0 u_0' \mathcal{Z}_{t-1} \geq 0 | u_0' \mathcal{Z}_{t-1} > 0]$

$$= \frac{1}{P[u_0'\mathcal{Z}_{t-1} > 0]} \int_{0}^{\infty} P[\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1} \geq 0 | u_0'\mathcal{Z}_{t-1} = s] f^{u_0'\mathcal{Z}_{t-1}}(s)\, ds$$

$$= \frac{1}{P[u_0'\mathcal{Z}_{t-1} > 0]} \int_{0}^{\infty} P[\varepsilon_t \geq \lambda_0 s | u_0'\mathcal{Z}_{t-1} = s] f^{u_0'\mathcal{Z}_{t-1}}(s)\, ds < \frac{1}{2},$$

so that (2.2) yields

$$ARD_p(\phi, P) \leq P[\varepsilon_t(\phi) u_0'\mathcal{Z}_{t-1} \geq 0] < \frac{1}{2} P[u_0'\mathcal{Z}_{t-1} > 0] + \frac{1}{2} P[u_0'\mathcal{Z}_{t-1} < 0] = \frac{1}{2}\cdot$$

(ii) In the case $P[u_0'\mathcal{Z}_{t-1} < 0] > 0$ and $P[u_0'\mathcal{Z}_{t-1} > 0] = 0$, Lemma 2.1 implies that $P[u_0'\mathcal{Z}_{t-1} < 0] = 1$, so that

$$\begin{aligned} ARD_p(\phi, P) &\leq P[\varepsilon_t(\phi) u_0'\mathcal{Z}_{t-1} \geq 0] = P[(\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1}) u_0'\mathcal{Z}_{t-1} \geq 0] \\ &= P[\varepsilon_t(\phi_0) - \lambda_0 u_0'\mathcal{Z}_{t-1} \leq 0] \\ &= \int_{-\infty}^{0} P[\varepsilon_t - \lambda_0 s \leq 0 | u_0'\mathcal{Z}_{t-1} = s] f^{u_0'\mathcal{Z}_{t-1}}(s)\, ds \\ &< \frac{1}{2}\cdot \end{aligned}$$

(iii) In the case $P[u_0'\mathcal{Z}_{t-1} < 0] = 0$ and $P[u_0'\mathcal{Z}_{t-1} > 0] > 0$, Lemma 2.1 implies that $P[u_0'\mathcal{Z}_{t-1} > 0] = 1$, and the same argument as in case (ii) shows that $ARD_p(\phi, P) < 1/2$.

Thus, $ARD_p(\phi, P) < 1/2$ for any $\phi \in \mathbb{R}^p \setminus \{\phi_0\}$, which establishes the result. □

We finish this section by showing that the function $\phi \mapsto ARD_p(\phi, P)$ is upper semicontinuous. To do so, first note that, for any fixed u, $P \mapsto P[\varepsilon_t(\phi) u'\mathcal{Z}_{t-1} \geq 0]$ is upper semicontinuous for weak convergence. As a consequence, by continuity of the function $\phi \mapsto \varepsilon_t(\phi) u'\mathcal{Z}_{t-1}$, the mapping $\phi \mapsto P[\varepsilon_t(\phi) u'\mathcal{Z}_{t-1} \geq 0]$ is also upper semicontinuous at each P. The result follows since $\phi \mapsto ARD_p(\phi, P)$ is then the infimum of a collection of upper semicontinuous functions.

3. Sample autoregression depth

If an observed series $Z_1, \ldots, Z_n$ is available, then the sample autoregressive depth of $\phi(\in \mathbb{R}^p)$ with respect to this series can be defined as

$$ARD_p(\phi, P_n) := \inf_{u \in \mathcal{S}^{p-1}} \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbb{I}[\varepsilon_t(\phi) u' \mathcal{Z}_{t-1} \geq 0],$$

still with $\varepsilon_t(\phi) = Z_t - \sum_{j=1}^{p} \phi_j Z_{t-j}$ and $\mathcal{Z}_{t-1} = (Z_{t-1}, \ldots, Z_{t-p})'$. Under mild ergodicity conditions on the underlying process, this sample autoregression depth is a strongly consistent estimator of its population analog. Recall that a stationary process X_t on the measure space $(\Omega, \mathcal{F}, P)$ is ergodic if, denoting $\mathcal{Z}_{s,t} := (Z_s, Z_{s+1}, \ldots, Z_t)'$, for any $k \geq 1$ and any $A, B \in \mathcal{F}^k$ (the usual product sigma-field on Ω^k),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} P[\mathcal{Z}_{1,k} \in A, \mathcal{Z}_{t+1,t+k} \in B] = P[\mathcal{Z}_{1,k} \in A] P[\mathcal{Z}_{1,k} \in B].$$

The condition above (see [1] for more details) is trivially verified for autoregressive processes. We then have the following result.

Theorem 3.1. *Let $\{Z_t : t \in \mathbb{Z}\}$ be a stationary ergodic process on the measure space $(\Omega, \mathcal{F}, P)$. Denote as $ARD_p(\phi, P_n)$ the sample autoregression depth of $\phi(\in \mathbb{R}^p)$ associated with a realization of length n from this process. Then,*

$$\sup_{\phi \in \mathbb{R}^p} |ARD_p(\phi, P_n) - ARD_p(\phi, P)| \to 0$$

almost surely as n diverges to infinity.

Proof. Consider the stationary process $\{V_t := W_t \otimes W_t : t \in \mathbb{Z}\}$, where we let $W_t := (Z_t, Z_{t-1}, \ldots, Z_{t-p})'$. With this notation,

$$ARD_p(\phi, P_n) = \inf_{u \in \mathcal{S}^{p-1}} \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbb{I}[V_t \in H_{\phi,u}]$$

and

$$ARD_p(\phi, P) = \inf_{u \in \mathcal{S}^{p-1}} \mathrm{P}[V_t \in H_{\phi,u}],$$

where $H_{\phi,u}$ belongs to the collection $\mathcal{H}$ of all closed halfspaces of $\mathbb{R}^{(p+1)^2}$. Now, fix $\phi \in \mathbb{R}^p$ and $\varepsilon > 0$. If $ARD_p(\phi, P_n) \leq ARD_p(\phi, P)$, then there exists $u_\varepsilon \in \mathcal{S}^{p-1}$

such that $(n-p)^{-1}\sum_{t=p+1}^{n}\mathbb{I}[V_t \in H_{\phi,u_\varepsilon}] \leq ARD_p(\phi, P_n)+\varepsilon$. Therefore,

$$
\begin{aligned}
|ARD_p(\phi,P_n) - ARD_p(\phi,P)|\mathbb{I}[ARD_p(\phi,P_n) \leq ARD_p(\phi,P)]\\
\leq \mathrm{P}[V_t \in H_{\phi,u_\varepsilon}] - \frac{1}{n-p}\sum_{t=p+1}^{n}\mathbb{I}[V_t \in H_{\phi,u_\varepsilon}] + \varepsilon\\
\leq \sup_{H\in\mathcal{H}}\left|\mathrm{P}[V_t \in H] - \frac{1}{n-p}\sum_{t=p+1}^{n}\mathbb{I}[V_t \in H]\right| + \varepsilon.
\end{aligned}
$$

Since this holds for any $\varepsilon > 0$, we obtain

$$
\begin{aligned}
|ARD_p(\phi,P_n) - ARD_p(\phi,P)|\mathbb{I}[ARD_p(\phi,P_n) \leq ARD_p(\phi,P)]\\
\leq \sup_{H\in\mathcal{H}}\left|\mathrm{P}[V_t \in H] - \frac{1}{n-p}\sum_{t=p+1}^{n}\mathbb{I}[V_t \in H]\right|.
\end{aligned} \tag{3.1}
$$

Working similarly, the inequality still holds after replacing $\mathbb{I}[ARD_p(\phi,P_n) \leq ARD_p(\phi,P)]$ with $\mathbb{I}[ARD_p(\phi,P_n) > ARD_p(\phi,P)]$ in (3.1). Adding up both inequalities then yields

$$
|ARD_p(\phi,P_n) - ARD_p(\phi,P)| \leq 2\sup_{H\in\mathcal{H}}\left|\mathrm{P}[V_t \in H] - \frac{1}{n-p}\sum_{t=p+1}^{n}\mathbb{I}[V_t \in H]\right|,
$$

hence

$$
\sup_{\phi\in\mathbb{R}^p}|ARD_p(\phi,P_n) - ARD_p(\phi,P)| \leq 2\sup_{H\in\mathcal{H}}\left|\mathrm{P}[V_t \in H] - \frac{1}{n-p}\sum_{t=p+1}^{n}\mathbb{I}[V_t \in H]\right|, \tag{3.2}
$$

Since $\mathcal{H}$ is a Vapnik–Chervonenkis class (see, e.g., page 152 of [11]), the result follows directly from the Glivenko-Cantelli results for ergodic sequences; see [1]. □

The following result, that shows that the sample deepest parameter value is strongly consistent for its population analog, is then a rather direct corollary.

Theorem 3.2. *Let $\{Z_t : t \in \mathbb{Z}\}$ be a stationary ergodic process on the measure space $(\Omega, \mathcal{F}, P)$. Assume that $ARD_p(\phi, P)$ admits a unique maximizer ϕ. Denote as $ARD_p(\phi, P_n)$ the sample autoregression depth of $\phi(\in \mathbb{R}^p)$ associated with a realization of length n from this process and let $\hat{\phi}_n$ be an arbitrary maximizer of $ARD_p(\phi, P_n)$. Then, $\hat{\phi}_n \to \phi$ almost surely as n diverges to infinity.*

Proof. In view of the upper semicontinuity of $\phi \mapsto ARD_p(\phi, P)$ and of the uniform consistency result in Theorem 3.1, the result readily follows from Theorem 2.12(ii) and Lemma 14.3 in Kosorok (2008). □

We close this section with the following illustration of Theorem 3.2 in the framework of AR(p) models (note that Theorem 2.1 guarantees that the unique maximization assumption in Theorem 3.2 is always met when the underlying process is

AR(p), so that Theorem 3.2 applies for such processes). We generated $n = 1{,}000$ observations from four AR processes after an initial burn-in period of 2,000 observations starting from $Z_{-1} = Z_0 = 0$. The autoregression parameter ϕ was taken as either $\phi = \phi_a = (0,0)'$ or $\phi = \phi_b = (0.25, -0.375)'$, while the innovation density was taken either standard Gaussian or Cauchy. Figure 1 displays the sample depth values $ARD_p(\phi, P_n)$ for each of the four resulting series. Clearly, the proximity between the true parameter value ϕ and estimated parameter value $\hat{\phi}_n$ supports the consistency result of Theorem 3.2 in all four cases.

4. The AR(1) particular case

We assume, as in the previous section, that we observe $Z_1, \ldots, Z_n$, but we restrict now to autoregressive depth of order $p = 1$. For any ϕ, we then have

$$ARD_1(\phi, P_n) := \frac{1}{n-1} \min\Bigg(\sum_{t=2}^{n} \mathbb{I}[\varepsilon_t(\phi) Z_{t-1} \geq 0], \sum_{t=2}^{n} \mathbb{I}[\varepsilon_t(\phi) Z_{t-1} \leq 0] \Bigg), \tag{4.1}$$

which, since Z_{t-1} is different from zero with probability one (Lemma 2.1), almost surely rewrites

$$ARD_1(\phi, P_n) = \frac{1}{n-1} \min\Bigg(\sum_{t=2}^{n} \mathbb{I}[Z_t/Z_{t-1} \geq \phi], \sum_{t=2}^{n} \mathbb{I}[Z_t/Z_{t-1} \leq \phi] \Bigg).$$

The sample deepest parameter ϕ is then a median of $Z_2/Z_1, \ldots, Z_n/Z_{n-1}$. If a unique representative is needed, then, in the same spirit as what is done for the Tukey median, the barycentre, $\hat{\phi}_n$ say, of the set of medians can be used. If the underlying process is an AR(1) process with parameter ϕ, then $\hat{\phi}_n$ is a natural estimator for ϕ, which we expect to have nice robustness properties. This is confirmed by the following result (for a proof, we refer to [2], where this estimator was first considered).

Theorem 4.1. *Denote as $\dot{\mathcal{F}}$ the collection of densities $\mathcal{F}$ that are bounded, are positive at* 0*, satisfy a Lipschitz condition at* 0*, and admit finite moments of order* $1+\delta$ *for some* $\delta > 0$*. Assume that* $\{Z_t : t \in \mathbb{Z}\}$ *is an autoregressive process of order* $p = 1$ *with autoregressive parameter* $\phi \in (-1, 1)$ *and an innovation density* $f \in \dot{\mathcal{F}}$*. Then, as* $n \to \infty$*,*

$$\sqrt{n}(\hat{\phi}_n - \phi) = \frac{1}{2f(0)\mathrm{E}[|Z_t|]\sqrt{n}} \sum_{t=2}^{n} \mathrm{Sign}(\varepsilon_t)\mathrm{Sign}(Z_{t-1}) + o_{\mathrm{P}}(1), \tag{4.2}$$

where $\mathrm{Sign}(x)$ *stands for the sign of* x*, so that* $\sqrt{n}(\hat{\phi}_n - \phi)$ *is asymptotically normal with mean zero and variance* $1/\{4f^2(0)\mathrm{E}^2[|Z_t|]\}$*.*

As usual, the Bahadur representation result in (4.2) readily implies that the influence function of $\hat{\phi}_n$ at $(z_{t-1}, z_t)'$ is given by

$$\frac{1}{2f(0)\mathrm{E}[|Z_t|]} \mathrm{Sign}(z_t - \phi z_{t-1})\mathrm{Sign}(z_{t-1}).$$

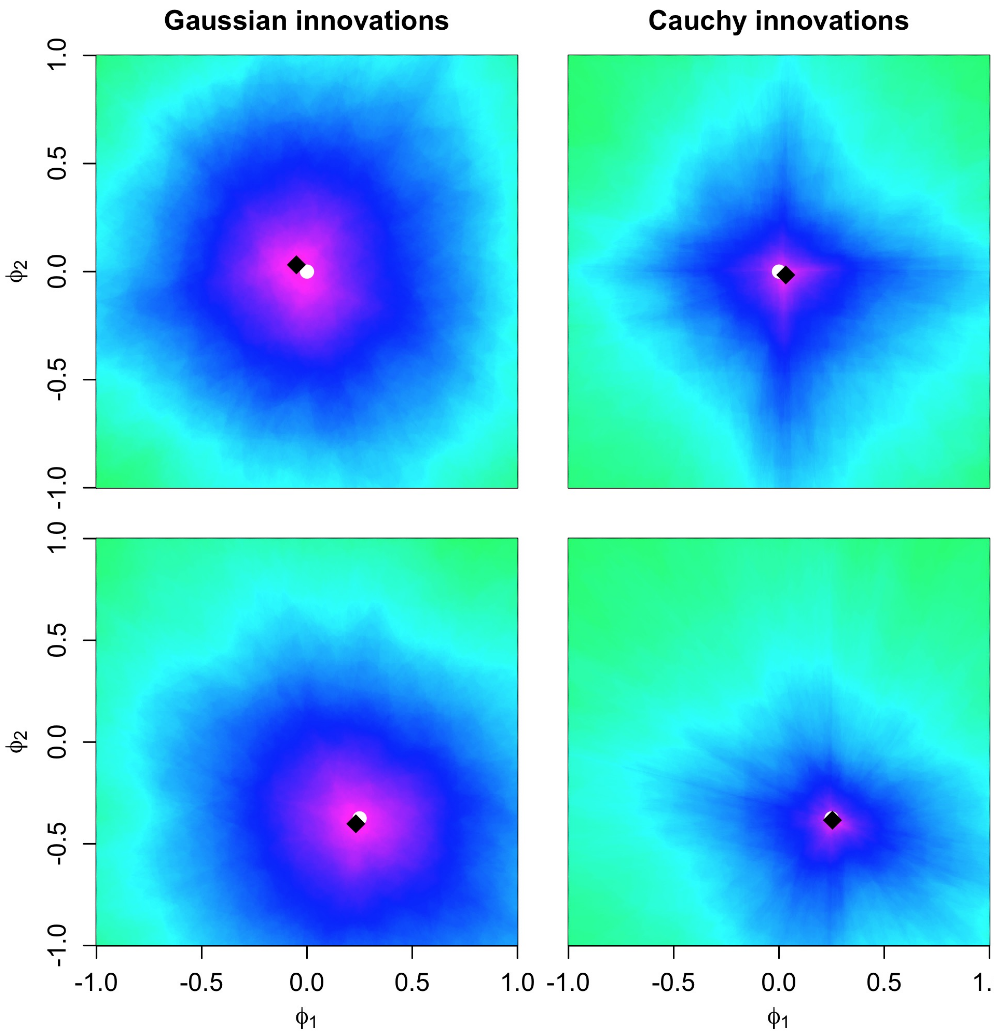

FIGURE 1. *Heat plots of the sample depth values $ARD_2(\phi, P_n)$ for finite realizations of length $n =$ 1,000 from four AR processes. These processes involve parameter values $\phi = \phi_a = (0, 0)'$ (top) or $\phi = \phi_b = (0.25, -0.375)'$ (bottom), and an innovation density that is standard Gaussian (left) or Cauchy (right). All panels display the true parameter value ϕ (white disk) and the sample deepest parameter value $\hat{\phi}_n$ (black diamond); see Section 3 for details.*

The boundedness of this influence function in both z_{t-1} and z_t confirms the robustness of the depth-based estimator $\hat{\phi}_n$. For the sake of comparison, recall that the classical least-squares estimator $\hat{\phi}_n^{\mathrm{LS}}$ and the LAD estimator $\hat{\phi}_n^{\mathrm{LAD}}$, which are defined as

$$\hat{\phi}_n^{\mathrm{LS}} := \arg\min_{\phi \in (-1,1)} \sum_{t=2}^{n} (Z_t - \phi Z_{t-1})^2 \quad \text{and} \quad \hat{\phi}_n^{\mathrm{LAD}} := \arg\min_{\phi \in (-1,1)} \sum_{t=2}^{n} |Z_t - \phi Z_{t-1}|.$$

If $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{E}[\varepsilon_t^2] < \infty$, then the least-squares estimator satisfies

$$\sqrt{n}(\hat{\phi}_n^{\mathrm{LS}} - \phi) = \frac{1-\phi^2}{\mathrm{E}[\varepsilon_t^2]\sqrt{n}} \sum_{t=2}^{n} \varepsilon_t Z_{t-1} + o_{\mathrm{P}}(1) \tag{4.3}$$

whereas, if $f \in \dot{\mathcal{F}}$, $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{E}[|\varepsilon|^{2+\delta}] < \infty$ for some $\delta > 0$, then the LAD estimator satisfies

$$\sqrt{n}(\hat{\phi}_n^{\mathrm{LAD}} - \phi) = \frac{1-\phi^2}{2f(0)\mathrm{E}[\varepsilon_t^2]\sqrt{n}} \sum_{t=2}^{n} \mathrm{Sign}(\varepsilon_t) Z_{t-1} + o_{\mathrm{P}}(1), \tag{4.4}$$

so that the corresponding influence functions are unbounded both in z_{t-1} and z_t for $\hat{\phi}_n^{\mathrm{LS}}$ and unbounded in z_t for $\hat{\phi}_n^{\mathrm{LAD}}$ (we refer to [2] for proofs of the Bahadur representation results in (4.3)–(4.4)).

While this settles the case for robustness, efficiency is of course also of primary interest. Obviously, (4.3)–(4.4) ensures that, under the assumptions stated there, $\hat{\phi}_n^{\mathrm{LS}}$ is asymptotically normal with mean zero and variance $1-\phi^2$, while $\hat{\phi}_n^{\mathrm{LAD}}$ is asymptotically normal with mean zero and variance $(1-\phi^2)/(4f^2(0)\mathrm{E}[\varepsilon_t^2])$. Figure 2 plots the resulting MSEs for each of the three estimators considered, more precisely the limiting values of $\mathrm{E}[\{\sqrt{n}(\hat{\phi} - \phi)\}^2]$, for $\phi = 0.6$ and for t innovations with ν degrees of freedom, $\nu = 1, 1.2, \ldots, 12$. The figure also reports Monte Carlo estimates of the same MSEs evaluated from $M = 100{,}000$ independent samples of length $n = 400$ (obtained after an initial burn-in period of 1,000 observations starting from $Z_0 = 0$). The same figure further provides the empirical MSEs obtained when 20 observations, chosen randomly among the $n = 400$ observations used in each replication, are multiplied by a factor 10. Clearly, without contamination, the depth-based estimator competes well with the LS and LAD ones for heavy tails only, but it outperforms its competitors with contamination.

Turning to hypothesis testing, depth-based tests can be defined for any null hypothesis of the form $\mathcal{H}_0 : \phi = \phi_0$, where $\phi_0 \in (-1, 1)$ is fixed. Typically, such a test will reject the null hypothesis when the value of $ARD_p(\phi_0, P_n)$ (see (4.1)) is too small, since this will indicate that the AR(1) with parameter ϕ_0 does not provide a suitable fit for the observed series. In the important particular case of testing for randomness (that is obtained with $\phi_0 = 0$), the test statistic rewrites (with probability one)

$$\begin{aligned} ARD_p(\phi, P_n) &= \frac{1}{n-1} \min\Big(\sum_{t=2}^{n} \mathbb{I}[Z_t Z_{t-1} > 0], \sum_{t=2}^{n} \mathbb{I}[Z_t Z_{t-1} < 0] \Big) \\ &= \frac{1}{n-1} \min\big(n - R_n, R_n - 1\big), \end{aligned}$$

where R_n is the number of runs in the series $\mathrm{Sign}(Z_1), \ldots, \mathrm{Sign}(Z_n)$, where a run is a maximal sequence of consecutive equal signs. Consequently, the depth-based test of randomness coincides, in the AR(1) case, with the classical runs test of randomness; see [5] and the references therein.

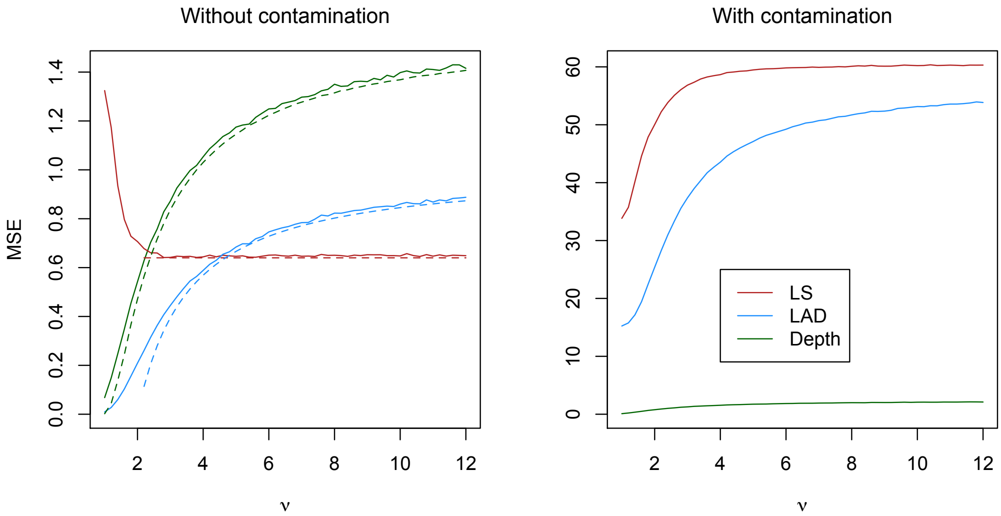

FIGURE 2. *(Left:) For the depth-based, LS and LAD estimators, plots of the corresponding limiting values of* $\mathrm{E}[\{\sqrt{n}(\hat{\phi}-\phi)\}^2]$ *(dashed lines) when the underlying AR(1) process is based on* $\phi = 0.6$ *and* t *innovations with* $\nu = 1, 1.2, \ldots, 12$ *degrees of freedom. Finite-sample versions of these MSEs, obtained from* $M = 100{,}000$ *independent realizations of length* $n = 400$*, are also provided (solid lines); see Section 4 for details. (Right:) The corresponding finite-sample MSEs when obtained from contaminated samples where 20 randomly selected observations out of the 400 used in each replication are multiplied by a factor 10.*

5. Final comments

In this paper, we introduced a concept of autoregression depth and derived various consistency results. A Bahadur representation for the estimator $\hat{\phi}_n$ of the deepest parameter value was also provided in the AR(1) case. In the general AR(p) case, it seems much more challenging to investigate the asymptotic behavior of this estimator. While we considered depth-based tests for randomness in the AR(1) case only, it is clear how to define them in the general AR(p) case, too: such tests would simply reject the null hypothesis of randomness for small values of $T_n = ARD_p(0, P_n)$. The resulting tests are of a *generalized runs* nature; see, e.g., [4] or [12] (the latter providing simplicial-based runs tests in the case $p = 2$). Exact rejection rules, however, are beyond the scope of the present paper since, contrary to the case $p = 1$, T is not distribution-free under the null hypothesis for $p \geq 2$, even with symmetric innovation densities.

Acknowledgements

Davy Paindaveine's research is supported by a research fellowship from the Francqui Foundation and by the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles. Germain Van Bever's research is supported a Crédit de Recherche grant of the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique.

References

[1] Adams, T. M. ; Nobel, A. B.: Uniform convergence of Vapnik-Chervonenkis classes under ergodic sampling. In: *Ann. Probab.* 38 (2010), Nr. 4, S. 1345–1367

[2] Boldin, M.V. ; Simonova, G. I. ; Tyurin, Yu. N.: *Translations of Mathematical Monographs*. Bd. 62: *Sign-based Methods in Linear Statistical Models*. American Mathematical Society, 1997

[3] Donoho, David L. ; Gasko, Miriam: Breakdown properties of location estimates based on halfspace depth and projected outlyingness. In: *Ann. Statist.* 20 (1992), Nr. 4, S. 1803–1827

[4] Dufour, J.-M. ; Hallin, M. ; Mizera, I.: Generalized runs tests for heteroskedastic time series. In: *J. Nonparametr. Stat.* 9 (1998), S. 39–86

[5] Ley, C. ; Paindaveine, D.: Runs tests. In: El-Shaarawi, A. H. (Hrsg.) ; Piegorsch, W. (Hrsg.): *Encyclopedia of Environmetrics*, John Wiley & Sons Ltd, Chichester, UK, 2012, S. 2474–2481

[6] Mizera, I.: On depth and deep points: a calculus. In: *Ann. Statist.* 30 (2002), Nr. 6, S. 1681–1736

[7] Paindaveine, D. ; Van Bever, G.: Tyler shape depth. In: *Biometrika* 4 (2019), S. 913–927

[8] Rousseeuw, P. J. ; Hubert, M.: Regression depth (with discussion). In: *J. Amer. Statist. Assoc.* 94 (1999), Nr. 446, S. 388–433

[9] Rousseeuw, P. J. ; Ruts, I.: The depth function of a population distribution. In: *Metrika* 49 (1999), Nr. 3, S. 213–244. – ISSN 0026-1335

[10] Tukey, J. W.: Mathematics and the picturing of data. In: *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, Canad. Math. Congress, Montreal, Que., 1975, S. 523–531

[11] Vaart, A. W. Van der ; Wellner, J. A.: *Weak Convergence and Empirical Processes.* New York : Springer, 1996 (Springer Series in Statistics)

[12] Van Bever, G.: Bivariate simplicial tests for randomness. In: *Statist. Probab. Lett.* 12 (2016), S. 20–25

[13] Zuo, Y. ; Serfling, R.: General notions of statistical depth function. In: *Ann. Statist.* 28 (2000), Nr. 2, S. 461–482

Université libre de Bruxelles
ECARES and Department of Mathematics
Avenue F.D. Roosevelt, 50, CP 114/04
B-1050 Brussels, Belgium
e-mail: dpaindav@ulb.ac.be
url: http://homepages.ulb.ac.be/ dpaindav

Department of Mathematics
and Namur Institute for Complex Systems
Université de Namur
Rempart de la Vierge, 8
5000, Namur, Belgium
e-mail: germain.vanbever@unamur.be
url: https://sites.google.com/site/germainvanbever