ON THE BICKEL–ROSENBLATT TEST OF GOODNESS-OF-FIT FOR THE RESIDUALS OF AUTOREGRESSIVE PROCESSES

Agnès Lagnoux[1,*], Thi Mong Ngoc Nguyen[2] and Frédéric Proïa[3]

Abstract. We investigate in this paper a Bickel–Rosenblatt test of goodness-of-fit for the density of the noise in an autoregressive model. Since the seminal work of Bickel and Rosenblatt, it is well-known that the integrated squared error of the Parzen–Rosenblatt density estimator, once correctly renormalized, is asymptotically Gaussian for independent and identically distributed (i.i.d.) sequences. We show that the result still holds when the statistic is built from the residuals of general stable and explosive autoregressive processes. In the univariate unstable case, we prove that the result holds when the unit root is located at -1 whereas we give further results when the unit root is located at 1. In particular, we establish that except for some particular asymmetric kernels leading to a non-Gaussian limiting distribution and a slower convergence, the statistic has the same order of magnitude. We also study some common unstable cases, like the integrated seasonal process. Finally, we build a goodness-of-fit Bickel–Rosenblatt test for the true density of the noise together with its empirical properties on the basis of a simulation study.

1. Introduction and motivations

For independent and identically distributed (i.i.d.) sequences of random variables, there is a wide range of goodness-of-fit statistical procedures in connection with the underlying true distribution. Among many others, one can think about the Kolmogorov–Smirnov test, the Cramér–von Mises criterion, the Pearson's chi-squared test, or more specific ones like the whole class of normality tests. Most of them have become of frequent practical use and directly implemented on the set of residuals of regression models. For such applications the independence hypothesis is irrelevant, especially for time series where lagged dependent variables are included. Thus, the crucial issue that naturally arises consists in having an overview of their sensitivity facing some weakened assumptions. This paper focuses on such a generalization for the Bickel–Rosenblatt statistic, introduced by the eponymous statisticians [3] in 1973, who first established its asymptotic normality and gave their names to the associated testing procedure. The statistic is closely related to the L^2 distance on the real line between the

[1] Institut de Mathématiques de Toulouse, UMR 5219, CNRS UT2J, Université de Toulouse, F-31058 Toulouse, France.
[2] VNUHCM - University of Science, Ho Chi Minh City, Viet Nam.
[3] Laboratoire angevin de recherche en mathématiques, LAREMA, UMR 6093, CNRS, UNIV Angers, SFR MathSTIC, 2 Bd Lavoisier, 49045 Angers Cedex 01, France.

* Corresponding author: lagnoux@univ-tlse2.fr

Parzen–Rosenblatt kernel density estimator and a parametric distribution (or a smoothed version of it). Namely it takes the form of

$$\int_{\mathbb{R}} \left(\widehat{f}_n(x) - f(x) \right)^2 a(x)\, \mathrm{d}x$$

with notation that we will detail in the sequel. Some improvements are of interest for us. First, Takahata and Yoshihara [22] in 1987 and later Neumann and Paparoditis [15] in 2000 extended the result to weakly dependent sequences (with mixing or absolute regularity conditions). As they noticed, these assumptions are satisfied by several processes of the time series literature. Then, Lee and Na [13] showed in 2002 that it also holds for the residuals of an autoregressive process of order 1 as soon as it contains no unit root (we will explain precisely this fact in the paper). Such a study leads to a goodness-of-fit test for the distribution of the innovations of the process. Bachmann and Dette [1] went further in 2005 by putting the results obtained by Lee and Na into practice. Their study also enables to get an asymptotic normality of the correctly renormalized statistic under some fixed alternatives. Even if it is not directly related to our subject, we also mention the work of Horváth and Zitikis [11] in 2004, providing some results on goodness-of-fit tests in L^r norm (for any r greater than 1) for the residuals of first-order autoregressive processes.

The goal of this paper is two-fold. First and mainly, we generalize the results of Lee and Na to autoregressive processes of order p ($p \geqslant 1$) while refining the set of hypotheses and discussing on the effect of unit roots on the statistic of interest. Second, a goodness-of-fit test is derived as a corollary together with a short empirical study so as to get an overview of its capabilities. From a theoretical point of view, this is always a challenging target to extend well-known first-order results to more general cases, and it is even more difficult to deal with unit roots in a time series context. Moreover, autoregressive models are still widespread in applied fields like econometrics, mathematical finance, weather and energy forecasting, engineering, etc. Thus to provide some advances in the study of autoregressive processes, in terms of inference, prediction, statistical significance, or, in our case, goodness-of-fit, was our first motivation.

On finite samples, it has been observed that the Gaussian behavior is difficult to reach and that, instead, an asymmetry occurs for dependent frameworks (see, *e.g.*, Valeinis and Locmelis [23]). In the simulation study, we will use the configurations suggested in this previous paper, and that of Fan [9] and Ghosh and Huang [10] to try to minimize this effect. In the end of this section, we introduce the context of our study and present both notation and vocabulary used in the sequel. Moreover, we recall the well-known asymptotic behavior of the Bickel–Rosenblatt statistic for i.i.d. random variables. Section 3 is dedicated to our results that are proved in Appendix B. To sum up, we establish the asymptotic behavior of the Bickel–Rosenblatt statistic based on the residuals of stable and explosive autoregressive processes of order p ($p \geqslant 0$). In addition, we also prove results related to common unstable autoregressions, like the random walks and the seasonally integrated processes. Finally, we give some considerations when dealing with general unstable processes or mixed processes. For example, the unstable ARIMA($p-1$,1,0) process would deserve a particular attention due to its widespread use in the econometric field. In Section 4, we build a goodness-of-fit test and discuss some empirical bases. The Appendix A is devoted to an overview of existing results on the least-squares estimation of the autoregressive parameter, depending on the roots of its characteristic polynomial, since it is of crucial interest in our proofs.

To start with, let us consider an autoregressive process of order p (AR(p)) defined by

$$X_t = \theta_1\, X_{t-1} + \ldots + \theta_p\, X_{t-p} + \varepsilon_t \tag{1.1}$$

for any $t \geqslant 1$ or equivalently, in a compact form, by

$$X_t = \theta^T \Phi_{t-1} + \varepsilon_t$$

where $\theta = (\theta_1, \ldots, \theta_p)^T$ is a vector parameter, Φ_0 is an arbitrary initial random vector, $\Phi_t = (X_t, \ldots, X_{t-p+1})^T$ and (ε_t) is a strong white noise having a finite positive variance σ^2 and a marginal density f (positive on the

real line). The corresponding characteristic polynomial is defined, for all $z \in \mathbb{C}$, by

$$\Theta(z) = 1 - \theta_1 z - \ldots - \theta_p z^p \tag{1.2}$$

and the companion matrix associated with Θ (see, *e.g.*, [8], Sect. 4.1.2) is given by

$$C_\theta = \begin{pmatrix} \theta_1 & \theta_2 & \ldots & \theta_{p-1} & \theta_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}. \tag{1.3}$$

It follows that the process may also be written as

$$\Phi_t = C_\theta \, \Phi_{t-1} + E_t \tag{1.4}$$

where $E_t = (\varepsilon_t, 0, \ldots, 0)^T$ is a p-dimensional noise. It is well-known that the stability of this p-dimensional process is closely related to the eigenvalues of the companion matrix that we will denote and arrange like

$$\rho(C_\theta) = |\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_p|.$$

In particular, according to ([8], Def. 2.3.17), the process is said to be *stable* when $|\lambda_1| < 1$, *purely explosive* when $|\lambda_p| > 1$ and *purely unstable* when $|\lambda_1| = |\lambda_p| = 1$. Among the purely unstable processes of interest, let us mention the seasonal model admitting the complex s-th roots of unity as solutions of its autoregressive polynomial $\Theta(z) = 1 - z^s$ for a season $s \in \mathbb{N}\backslash\{0,1\}$. In the paper, this model will be shortened as *seasonal unstable of order $p = s$*, it satisfies $\theta_1 = \ldots = \theta_{s-1} = 0$ and $\theta_s = 1$. In a general way, it is easy to see that $\det(C_\theta) = (-1)^{p+1} \theta_p$ so that C_θ is invertible as soon as $\theta_p \neq 0$ (which will be one of our hypotheses when $p > 0$). In addition, a simple calculation shows that

$$\det(C_\theta - \lambda \, I_p) = (-\lambda)^p \, \Theta(\lambda^{-1})$$

which implies (since $\Theta(0) \neq 0$) that each zero of Θ is the inverse of an eigenvalue of C_θ. Consequently, the stability of the process may be expressed in the paper through the eigenvalues of C_θ as well as through the zeroes of Θ. We will also consider that

$$z^p \, \Theta(z^{-1}) = z^p - \theta_1 z^{p-1} - \ldots - \theta_p$$

is the minimal polynomial of C_θ (which, in the terminology of [8], means that the process is *regular*). Now, assume that $X_{-p+1}, \ldots, X_0, X_1, \ldots, X_n$ are observable (for $n \gg p$) and let

$$\widehat{\theta}_n = \left(\sum_{t=1}^n \Phi_{t-1} \Phi_{t-1}^T \right)^{-1} \sum_{t=1}^n \Phi_{t-1} \, X_t \tag{1.5}$$

be the least-squares estimator of θ (for $p > 0$). The associated residual process is

$$\widehat{\varepsilon}_t = X_t - \widehat{\theta}_n^T \, \Phi_{t-1} \tag{1.6}$$

for all $1 \leqslant t \leqslant n$, or simply $\widehat{\varepsilon}_t = X_t$ when $p = 0$. Hereafter, $\mathbb{K}$ is a kernel and (h_n) is a bandwidth. That is, $\mathbb{K}$ is a non-negative function satisfying

$$\int_{\mathbb{R}} \mathbb{K}(x)\,\mathrm{d}x = 1, \quad \int_{\mathbb{R}} \mathbb{K}^2(x)\,\mathrm{d}x < +\infty \quad \text{and} \quad \int_{\mathbb{R}} x^2\,\mathbb{K}(x)\,\mathrm{d}x < +\infty,$$

and (h_n) is a positive sequence decreasing to 0. The so-called Parzen–Rosenblatt estimator [16, 18] of the density f is given, for all $x \in \mathbb{R}$, by

$$\widehat{f}_n(x) = \frac{1}{n\,h_n} \sum_{t=1}^{n} \mathbb{K}\left(\frac{x - \widehat{\varepsilon}_t}{h_n}\right). \tag{1.7}$$

The local and global behaviors of this empirical density have been well studied in the literature. However, for a goodness-of-fit test, we focus on the global fitness of $\widehat{f}_n$ to f on the whole real line. From this viewpoint, we consider the Bickel–Rosenblatt statistic that we define as

$$\widehat{T}_n = n\,h_n \int_{\mathbb{R}} \left(\widehat{f}_n(x) - (\mathbb{K}_{h_n} * f)(x)\right)^2 a(x)\,\mathrm{d}x \tag{1.8}$$

where $\mathbb{K}_{h_n} = h_n^{-1}\,\mathbb{K}(\cdot/h_n)$, a is a positive piecewise continuous integrable function and $*$ denotes the convolution operator, *i.e.* $(g * h)(x) = \int_{\mathbb{R}} g(x - u)\,h(u)\,\mathrm{d}u$. A statistic of probably greater interest and easier to interpret is

$$\widetilde{T}_n = n\,h_n \int_{\mathbb{R}} \left(\widehat{f}_n(x) - f(x)\right)^2 a(x)\,\mathrm{d}x. \tag{1.9}$$

Bickel and Rosenblatt show in [3] that under appropriate conditions, if T_n is the statistic given in (1.8) built on the strong white noise (ε_t) instead of the residuals, then, as n tends to infinity,

$$\frac{T_n - \mu}{\sqrt{h_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2) \tag{1.10}$$

where the centering term is

$$\mu = \int_{\mathbb{R}} f(s)\,a(s)\,\mathrm{d}s \int_{\mathbb{R}} \mathbb{K}^2(s)\,\mathrm{d}s \tag{1.11}$$

and the asymptotic variance is given by

$$\tau^2 = 2 \int_{\mathbb{R}} f^{\,2}(s)\,a^2(s)\,\mathrm{d}s \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{K}(t)\,\mathbb{K}(t + s)\,\mathrm{d}t\right)^2 \mathrm{d}s. \tag{1.12}$$

The aforementioned conditions are summarized in ([10], Sect. 2), they come from the original work of Bickel and Rosenblatt later improved by Rosenblatt [19]. In addition to some technical assumptions that will be recalled in (H0) and similarly to the results in [13], notice that the results hold

- either if $\mathbb{K}$ is bounded and has a compact support and the bandwidth is given by $h_n = h_0\,n^{-\kappa}$ with $0 < \kappa < 1$;

- or if $\mathbb{K}$ is a continuous and positive kernel defined on $\mathbb{R}$, the bandwidth is given by $h_n = h_0\, n^{-\kappa}$ with $0 < \kappa < 1/4$ and

$$\int_{|x| \geqslant 3} |x|^{3/2}\, (\ln\ln|x|)^{1/2}\, |\mathbb{K}'(x)|\, \mathrm{d}x < +\infty \quad \text{and} \quad \int_{\mathbb{R}} (\mathbb{K}'(x))^2\, \mathrm{d}x < +\infty.$$

We can find in later references (like [22], [15] or [1]) some alternative proofs of the asymptotic normality (1.10) with $a(x) = 1$ and appropriate assumptions. However, in this paper, we keep a as an integrable function in order to follow the original framework of Bickel and Rosenblatt. Fortunately, it also makes the calculations easier and remains appropriate for applications since it is always possible to define a compact support including any mass of a given density f_0 to be tested, as it will be done in Section 3.

Notation. It is convenient to have short expressions for terms that converge in probability to zero. In the whole study, the notation $o_{\mathbb{P}}(1)$ (resp. $O_{\mathbb{P}}(1)$) stands for a sequence of random variables that converges to zero in probability (resp. is bounded in probability) as $n \to \infty$.

2. The Bickel–Rosenblatt statistic

In this section, we derive the limiting distribution of the test statistics $\widehat{T}_n$ and $\widetilde{T}_n$ given by (1.8) and (1.9), based on the residuals in the stable, some unstable and purely explosive cases.

2.1. Assumptions

For the whole study, we make the following assumptions.

(H0) The strong white noise (ε_t) has a bounded density f which is positive, twice differentiable, and the second derivative f'' is itself bounded. The weighting function a is positive, piecewise continuous and integrable. The kernel $\mathbb{K}$ is bounded, continuous on its support. The kernel $\mathbb{K}$ and the bandwidth (h_n) are chosen so that
- either (a) $\mathbb{K}$ is bounded and has a compact support and the bandwidth is given by $h_n = h_0\, n^{-\kappa}$ with $0 < \kappa < 1$;
- or (b) $\mathbb{K}$ is continuous and positive kernel defined on $\mathbb{R}$, the bandwidth is given by $h_n = h_0\, n^{-\kappa}$ with $0 < \kappa < 1/4$ and

$$\int_{|x| \geqslant 3} |x|^{3/2}\, (\ln\ln|x|)^{1/2}\, |\mathbb{K}'(x)|\, \mathrm{d}x < +\infty \quad \text{and} \quad \int_{\mathbb{R}} (\mathbb{K}'(x))^2\, \mathrm{d}x < +\infty.$$

Some additional hypotheses are given below, not simultaneously needed.

(H1) The kernel $\mathbb{K}$ is such that $\mathbb{K}'''$ exists and is bounded on its support,

$$\int_{\mathbb{R}} \left\{ |\mathbb{K}'(x)| + |\mathbb{K}''(x)| \right\} \mathrm{d}x < +\infty, \quad \int_{\mathbb{R}} (\mathbb{K}''(x))^2\, \mathrm{d}x < +\infty,$$

$$\int_{\mathbb{R}} |x\, \mathbb{K}'(x)|\, \mathrm{d}x < +\infty \quad \text{and} \quad \int_{\mathbb{R}} x^2\, |\mathbb{K}'(x)|\, \mathrm{d}x < +\infty.$$

(H2) The kernel $\mathbb{K}$ satisfies

$$\int_{\mathbb{R}} |\mathbb{K}(x + \delta) - \mathbb{K}(x)|\, \mathrm{d}x \leqslant B\, \delta$$

for some $B > 0$ and all $\delta > 0$.

(H3) For some $\alpha > 0$, the bandwidth (h_n) satisfies

$$\lim_{n \to +\infty} n\, h_n^\alpha = +\infty.$$

(H4) For some $\beta > 0$, the bandwidth (h_n) satisfies

$$\lim_{n \to +\infty} n\, h_n^\beta = 0.$$

(H5) The noise (ε_t) and the initial vector Φ_0 have a finite moment of order $\nu > 0$.

Notice that

$$\int_{\mathbb{R}} |\mathbb{K}'(x)|\, \mathrm{d}x < +\infty$$

is a sufficient condition to get (H2).

2.2. Asymptotic behaviors

Theorem 2.1 (Stable case)**.** *In the stable case (*$|\lambda_1| < 1$*), assume that (H0), (H3) with $\alpha = 4$, and (H5) with $\nu = 4$ hold. Then, as n tends to infinity,*

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$$

where μ and τ^2 are given in (1.11) and (1.12), respectively. In addition, the result is still valid for $\widetilde{T}_n$ if (H4) with $\beta = 9/2$ holds.

Proof. See Section B.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Remark 2.2. Theorem 2.1 holds for $h_n = h_0\, n^{-\kappa}$ as soon as $2/9 < \kappa < 1/4$. A standard choice may be $h_n = h_0\, n^{-1/4+\epsilon}$ for a small $\epsilon > 0$.

Theorem 2.3 (Purely explosive case)**.** *In the purely explosive case (*$|\lambda_p| > 1$*), assume that (H0), (H2), (H3) with $\alpha = 1$ and (H5) with $\nu = 2 + \gamma$ hold, for some $\gamma > 0$. Then, as n tends to infinity,*

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$$

where μ and τ^2 are given in (1.11) and (1.12), respectively. In addition, the result is still valid for $\widetilde{T}_n$ if (H4) with $\beta = 9/2$ holds.

Proof. See Section B.2. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Remark 2.4. Hence by (H0), Theorem 2.3 holds for $h_n = h_0\, n^{-\kappa}$ as soon as $2/9 < \kappa < 1$ if $\mathbb{K}$ has a compact support (the usual bandwidth $h_n = h_0\, n^{-1/4}$ is thus appropriate) or as soon as $2/9 < \kappa < 1/4$ for a kernel positive on $\mathbb{R}$. In that case, a standard choice may be $h_n = h_0\, n^{-1/4+\epsilon}$ for a small $\epsilon > 0$.

We now investigate the asymptotic behavior of the statistics for some unstable cases. We establish in particular that the analysis of [13] is only partially true.

Proposition 2.5 (Purely unstable case)**.** *In the unstable case for $p = 1$ ($\lambda = \pm 1$), assume that (H0), (H1), (H3) with $\alpha = 4$, and (H5) with $\nu = 4$ hold. If $\lambda = -1$ then, as n tends to infinity,*

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$$

where μ and τ^2 are given in (1.11) and (1.12), respectively. If $\lambda = 1$ and $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s = 0$ then

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} = O_{\mathbb{P}}(1).$$

The results are still valid for $\widetilde{T}_n$ if (H4) with $\beta = 9/2$ holds. Finally, if $\lambda = 1$ and $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s \neq 0$ then, as n tends to infinity,

$$h_n(\widehat{T}_n - \mu) \xrightarrow{\mathcal{D}} \sigma^2 \left(\frac{\frac{1}{2}\left(W^2(1) - 1\right)}{\int_0^1 W^2(u)\,\mathrm{d}u} \int_0^1 W(u)\,\mathrm{d}u \right)^2 \int_{\mathbb{R}} f^2(s)\,a(s)\,\mathrm{d}s \left(\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s \right)^2$$

where $(W(t),\, t \in [0,1])$ is a standard Wiener process.

In the seasonal case with $p = s$, we also reach

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} = O_{\mathbb{P}}(1)$$

or

$$h_n(\widehat{T}_n - \mu) \xrightarrow{\mathcal{D}} \sigma^2 \left(S(W_s)^T H(W_s) \right)^2 \int_{\mathbb{R}} f^2(s)\,a(s)\,\mathrm{d}s \left(\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s \right)^2$$

depending on whether $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s = 0$ or $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s \neq 0$, respectively, where $S(W_s)$ is defined in Proposition A.6, $H(W_s)$ will be clarified in the proof and $(W_s(t),\, t \in [0,1])$ is a standard Wiener process of dimension s.

Proof. See Section B.3. □

Remark 2.6. This last result needs some observations.

- Proposition 2.5 holds for $h_n = h_0\, n^{-\kappa}$ as soon as $2/9 < \kappa < 1/4$ and $\lambda = -1$. A standard choice may be $h_n = h_0\, n^{-1/4+\epsilon}$ for a small $\epsilon > 0$.
- One can observe that the value of $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s$ is crucial to deal with the unstable case. In fact, all usual kernels are even, leading to $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s = 0$. On simulations, the unstable case gives results similar to the stable and explosive ones, except for some very particular asymmetric kernels having different bound values that we can build to violate the aforementioned natural condition. In that case, the convergence is slower as we can see from the result of Proposition 2.5, and the limiting distribution is not Gaussian.
- At the end of the associated proof, we show that, for $\lambda = 1$ and a choice of kernel such that $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s \neq 0$, we also need $\int_{\mathbb{R}} s\,\mathbb{K}(s)\,\mathrm{d}s = 0$ to translate the result from $\widehat{T}_n$ to $\widetilde{T}_n$ under (H4) with $\beta = 6$. In this case, the result holds for $h_n = h_0\, n^{-\kappa}$ as soon as $1/6 < \kappa < 1/4$.
- In ([14], Thm. 3.2), Lee and Wei had already noticed that only the unit roots located at 1 affect the residual empirical process defined, for $0 \leqslant u \leqslant 1$, by

$$\widehat{H}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\mathbb{1}_{\{F(\widehat{\varepsilon}_t) \leqslant u\}} - u \right),$$

where F is the cumulative distribution function of the noise. Our result is consistent with that fact since one may rewrite $\widehat{T}_n$ in terms of $\widehat{H}_n$ as

$$\widehat{T}_n = n\, h_n \int_{\mathbb{R}} \left(\frac{1}{h_n\, \sqrt{n}} \int_{\mathbb{R}} \mathbb{K}\left(\frac{x-s}{h_n} \right) \mathrm{d}\widehat{H}_n(F(s)) \right)^2 a(x)\, \mathrm{d}x.$$

Thus, $\widehat{H}_n$ appearing as the only stochastic part of the statistic, it is not surprising that, similarly, only the unit roots located at 1 affect our results.

Remark 2.7 (Beyond the results of Lee and Na). We conclude this section by a short comparison with the results of Lee and Na [13]. The set of hypotheses is refined but in the stable and explosive cases, the proofs follow similar strategies to extend their results from first-order ($p = 1$) to general autoregressive processes ($p \geqslant 1$) and to the more natural unsmoothed statistic (1.9). However a substantial improvement lies in the unstable case, since this work is not only a matter of generalization of the known results, but also a matter of correction of them. In particular, we have established that the negative unit root, although behind instability, does not play any role. In addition, for a positive unit root, we have shown that the rate h_n^{-1} is not systematic but only reached at the cost of the unrealistic hypotheses $\int_{\mathbb{R}} \mathbb{K}'(s)\, \mathrm{d}s = 0$, and in this case we have provided the limiting distribution. Otherwise the order of magnitude of the statistic remains $h_n^{1/2}$ even if the limiting distribution (if it exists) is not still established.

The authors of [11] mentioned that the statistic has a rate of $n^{1/2}\, h_n^{(1-1/r)/2}$ when built on the residuals of first-order autoregressive processes, for a goodness-of-fit test based on the L^r norm (which indeed corresponds to Bickel–Rosenblatt for $r = 2$). Even if it is hardly comparable since it relies on the sample cumulative distribution and not on any density estimation, let us also recall that the Kolmogorov–Smirnov test, probably the most famous and popular non-parametric test of goodness-of-fit, is associated with an L^∞ norm and has a rate of $\sqrt{n}$. The generalization of this work to these norms could also be valuable, by way of comparison.

To conclude this part, we draw the reader's attention to the fact that the purely unstable case is not fully treated. The general results may be a challenging study due to the phenomenon of compensation arising through unit roots different from 1. Lemma B.8 at the end of the Appendix B is not used as part of this paper but may be a trail for future studies. It illustrates the compensation *via* the fact that $(\sum_{k=1}^{t} X_k)$ is of the same order as (X_t) in a purely unstable process having no unit root located at 1. Mixed models ($|\lambda_1| \geqslant 1$ and $|\lambda_p| \leqslant 1$) should lead to similar reasonings, they also have to be handled to definitively lift the veil on the Bickel–Rosenblatt statistic for the residuals of autoregressive processes. As a priority, it seems that unstable ARIMA($p-1$,1,0) processes would deserve close investigations due to their widespread use in the econometric field. The difficulty arising here is that the estimator converges at the slow rate of stability while the process grows at the fast rate of instability: a compensation will be needed. Another technical improvement that would require an adjustment of the proofs, and probably a strengthening of the hypotheses on $\mathbb{K}$ and f, could be the extension of the results to the weighting function $a(x) = 1$. The goodness-of-fit test could then be implemented on the whole real line and not only on a part of it, however big it is, as it will be the case in the next section.

3. A GOODNESS-OF-FIT BICKEL–ROSENBLATT TEST

Our objective is now to derive a goodness-of-fit testing procedure from the results established in the previous section. First, one can notice that the choice of

$$a(x) = \begin{cases} 1/f(x) & \text{for } x \in [-\delta\,;\,\delta] \\ 0 & \text{for } x \notin [-\delta\,;\,\delta] \end{cases}$$

for any $\delta > 0$ leads to the simplifications

$$\widetilde{T}_n = n\, h_n \int_{-\delta}^{\delta} \frac{\big(\widehat{f}_n(x) - f(x)\big)^2}{f(x)}\, \mathrm{d}x, \qquad \mu = 2\,\delta \int_{\mathbb{R}} \mathbb{K}^2(s)\, \mathrm{d}s$$

and

$$\tau^2 = 4\,\delta \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{K}(t)\, \mathbb{K}(t+s)\, \mathrm{d}t \right)^2 \mathrm{d}s.$$

Bickel and Rosenblatt [3] suggested a similar choice for the weight function a, with $[0\,;1]$ for compact support. Nevertheless, it seems more reasonable to work on a symmetric interval in order to test for the density of a random noise. In addition, μ and τ^2 become independent of f, which will be useful to build a statistical procedure based on f. For a couple of densities f and f_0 such that f_0 does not cancel on $[-\delta\,;\delta]$, let us define

$$\Delta_\delta(f, f_0) = \int_{-\delta}^{\delta} \frac{\big(f(x) - f_0(x)\big)^2}{f_0(x)}\, \mathrm{d}x. \tag{3.1}$$

Hence, $\Delta_\delta(f, f_0) = 0$ means that f and f_0 coincide almost everywhere on $[-\delta\,;\delta]$, and everywhere under our usual continuity hypotheses on the densities. On the contrary, $\Delta_\delta(f, f_0) > 0$ means that there exists an interval $I \subseteq [-\delta\,;\delta]$ with non-empty interior on which f and f_0 differ. Accordingly, let

$$\mathcal{H}_0 : \text{``}\Delta_\delta(f, f_0) = 0\text{''} \quad \textit{vs.} \quad \mathcal{H}_1 : \text{``}\Delta_\delta(f, f_0) > 0\text{''}.$$

The natural test statistic is therefore given by

$$\widetilde{Z}_n^0 = \frac{\widetilde{T}_n^0 - \mu}{\tau\, \sqrt{h_n}} \tag{3.2}$$

where $\widetilde{T}_n^0$ is the statistic $\widetilde{T}_n$ reminded above built using f_0 instead of f.

Proposition 3.1 (A goodness-of-fit test). *Consider the set of residuals from one of the following AR(p) models:*

- *a stable process with $p \geqslant 0$, under the hypotheses of Theorem 2.1,*
- *an explosive process with $p \geqslant 1$, under the hypotheses of Theorem 2.3,*
- *an unstable process with $p = 1$ and $\lambda = -1$, under the hypotheses of Proposition 2.5.*

Then, under $\mathcal{H}_0 : \text{``}\Delta_\delta(f, f_0) = 0\text{''}$ where f_0 is a continuous density which does not cancel on $[-\delta\,;\delta]$ for some $\delta > 0$,

$$\widetilde{Z}_n^0 \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

In addition, under $\mathcal{H}_1 : \text{``}\Delta_\delta(f, f_0) > 0\text{''}$,

$$\widetilde{Z}_n^0 \xrightarrow{\mathbb{P}} +\infty.$$

Proof. The proof is immediate using our previous results. The consistency under $\mathcal{H}_1$ is reached using the fact that

$$\widetilde{Z}_n^0 = \frac{\widetilde{T}_n - \mu}{\tau\, \sqrt{h_n}} + \frac{\widetilde{T}_n^0 - \widetilde{T}_n}{\tau\, \sqrt{h_n}}.$$

$\square$

TABLE 1. Empirical level of the test under $\mathcal{H}_0$, for the configurations described above. We used $n \in \{50, 100, 500\}$ and 1000 replications.

$\mathbb{K}$	$\mathcal{N}(0,1)$			$\mathcal{U}([-1\,;1])$		
n	50	100	500	50	100	500
h_0	0.10	0.14	0.14	0.20	0.25	0.32
M_0	0.051	0.049	0.049	0.052	0.048	0.048
M_1	0.059	0.051	0.047	0.058	0.050	0.050
M_2	0.055	0.046	0.051	0.044	0.046	0.051
M_3	0.054	0.054	0.047	0.059	0.051	0.048
M_4	0.048	0.051	0.047	0.050	0.045	0.050
M_5	0.049	0.047	0.058^*	0.046	0.049	0.054^*

* Simulations that needed more than one numerical trial, due to the explosive nature of the process and the large value of n.

For any level $0 < \alpha < 1$, we reject $\mathcal{H}_0$ as soon as

$$\widetilde{Z}_n^0 > u_{1-\alpha}$$

where $u_{1-\alpha}$ stands for the $(1-\alpha)$-quantile of the $\mathcal{N}(0,1)$ distribution. For our simulations, we focus on a normality test (probably the most useful in regression, for goodness-of-fit). We have trusted the observations of [10], [9] or [23]. In particular, only the $\mathcal{N}(0,1)$ and the $\mathcal{U}([-1\,;1])$ kernels are used, with obviously $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s = 0$. The bandwidth is $h_n = h_0\,n^{-1/4+\epsilon}$ for $\epsilon = 10^{-3}$ where h_0 is calibrated to reach an empirical level close to $\alpha = 5\%$ for the neutral model ($p = 0$), and the true distribution of the noise is $\mathcal{N}(0,1)$. We shall note at this point that, despite appearances, the choice of κ is not crucial for our study. Indeed, for the moderate values of n that we consider, h_0 plays a more important role. That is the reason why we choose κ in the restricted area of validity ($2/9 < \kappa < 1/4$) and why we focus first on h_0. The selection of the hyperparameter δ will be described thereafter. We only give an overview of the results in Table 1 for some typical models:

- M_0 – neutral model ($p = 0$),
- M_1 – stable model ($p = 3$, $\theta_1 = -1/12$, $\theta_2 = 5/24$, $\theta_3 = 1/24$),
- M_2 – stable but almost unstable model ($p = 1$, $\theta_1 = 99/100$),
- M_3 – unstable model with negative unit root ($p = 1$, $\theta_1 = -1$),
- M_4 – unstable model with positive unit root ($p = 1$, $\theta_1 = 1$),
- M_5 – explosive model ($p = 2$, $\theta_1 = 0$, $\theta_2 = 121/100$).

For model M_4, it is important to note that Proposition 3.1 may not hold. Nevertheless, we know by virtue of Proposition 2.5 that the statistic has the same order of magnitude, thus it seemed interesting to look at its empirical behavior in comparison with the other models (and we observe that it reacts as well). Now we turn to the frequency of rejection of $\mathcal{H}_0$ when f_0 is not the true distribution, for the configuration $n = 100$ and the $\mathcal{N}(0,1)$ kernel. First of all, we are going to discuss on the hyperparameter δ that has to be non-arbitrary. A natural choice relies on the quantiles of the distribution that we test to guarantee that a sufficient mass is covered by the statistic. As the simulations below will emphasize, at least 95% seems relevant and accordingly, we let

$$\delta = \max(-q_0(0.025), q_0(0.975))$$

where q_0 is the quantile function of f_0, which takes the simplified form $\delta = q_0(0.975)$ whenever the distribution is both centered and symmetrical. The latter situation is generally preferred since we deal with the noise of a regression model, but for example in our simulations we also need the general one for $\mathcal{N}(m,1)$. Among the alternatives $\mathcal{N}(m,1)$ and $\mathcal{N}(0,\sigma^2)$, we pick two control models that will be $\mathcal{N}(0.5,1)$ and $\mathcal{N}(0,0.3)$, chosen to

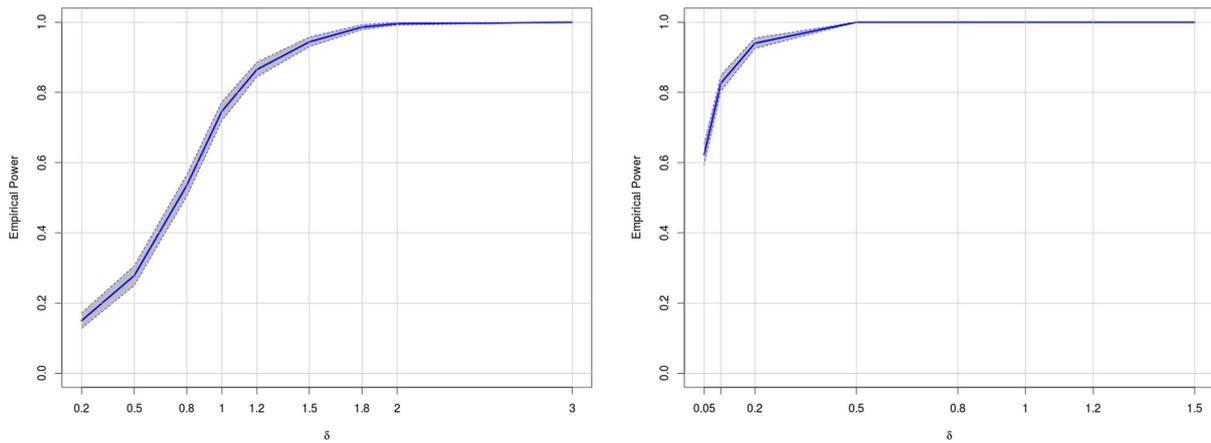FIGURE 1. Frequency of rejection of the test for $n = 100$, $\mathbb{K} = \mathcal{N}(0,1)$, $h_0 = 0.14$ and 1000 replications, depending on δ on the abscissa. The darkened areas are the 95% confidence intervals. The alternatives are $\mathcal{N}(0.5, 1)$ on the left, and $\mathcal{N}(0, 0.3)$ on the right.
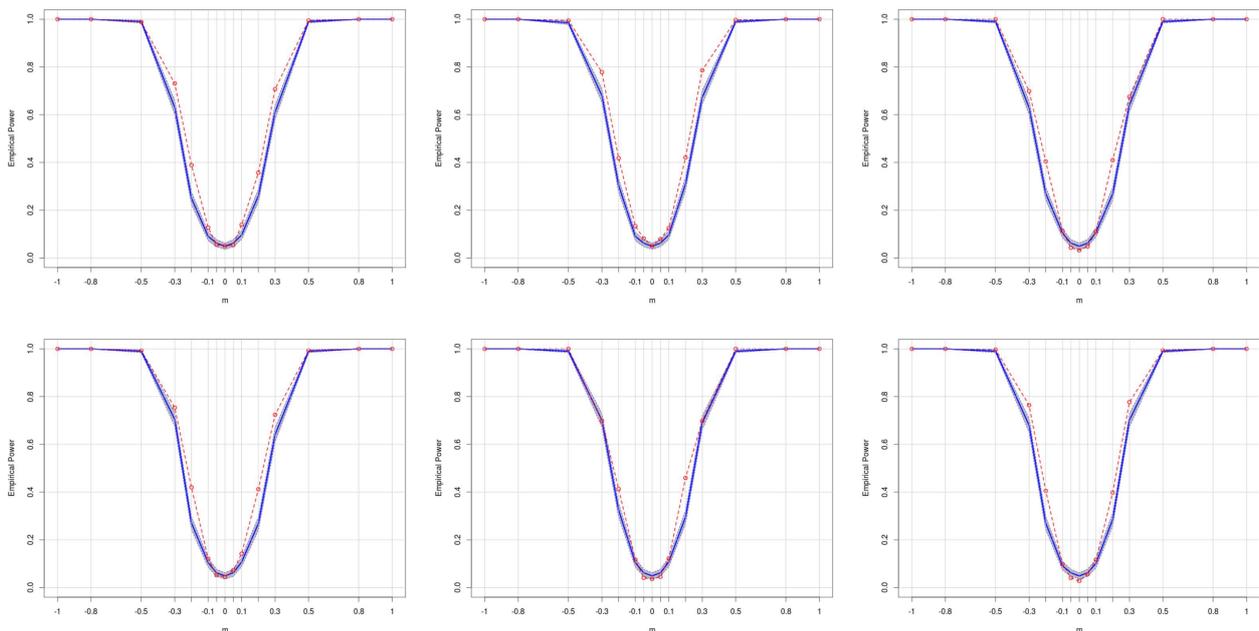

FIGURE 2. Frequency of rejection of the test for $n = 100$, $\mathbb{K} = \mathcal{N}(0,1)$, $h_0 = 0.14$ and 1000 replications. The alternative f_0 are $\mathcal{N}(m, 1)$ for some $m \in [-1\,;\,1]$ with true value $m = 0$. Results obtained from M_0, M_1 and M_2 are in the top (from left to right), results obtained from M_3, M_4 and M_5 are in the bottom (from left to right). The dotted line in red corresponds to the Kolmogorov–Smirnov test and the darkened areas are the 95% confidence intervals.

be the closest distributions for which our procedure is able to reach a maximal frequency of rejection under this configuration (see Figs. 2 and 3). Figure 1 is an illustration of the percentage of rejection (together with its 95% confidence interval) of $\mathcal{H}_0$ while we make δ increase, for the neutral model M_0. The corresponding values $\delta \approx 2.46$ and $\delta \approx 1.07$ that we suggest to use respectively for these examples seem therefore suitable.
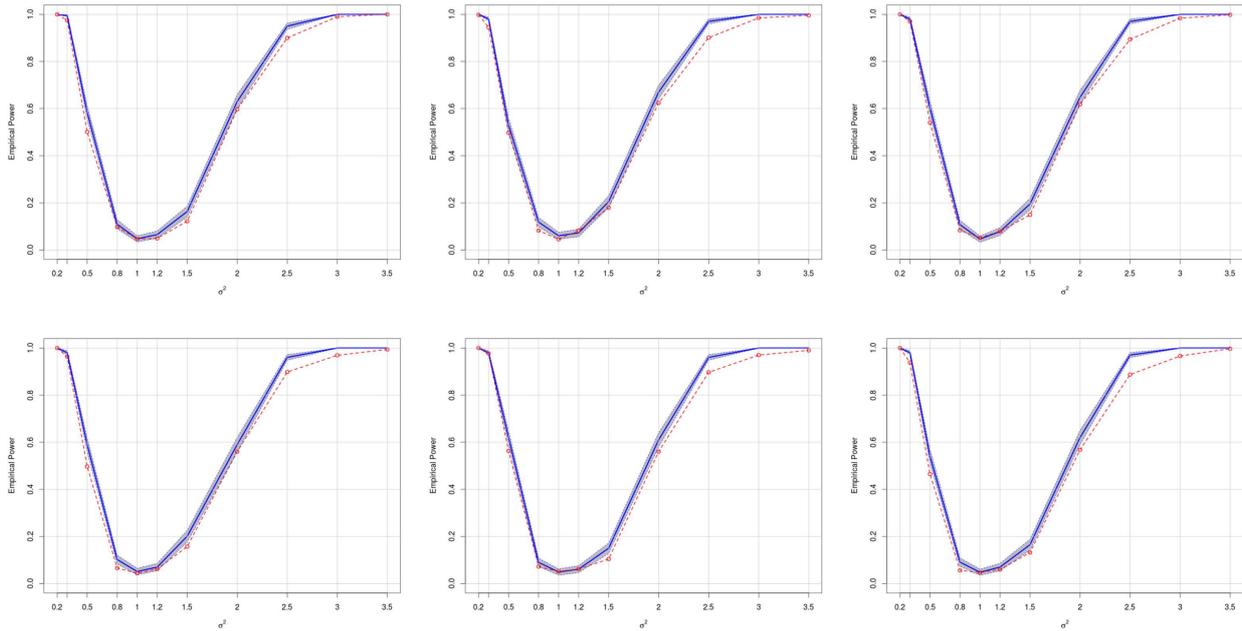
FIGURE 3. Frequency of rejection of the test for $n = 100$, $\mathbb{K} = \mathcal{N}(0,1)$, $h_0 = 0.14$ and 1000 replications. The alternative f_0 are $\mathcal{N}(0, \sigma^2)$ for some $\sigma^2 \in [0.2\,;3.5]$ with true value $\sigma^2 = 1$. Results obtained from M_0, M_1 and M_2 are in the top (from left to right), results obtained from M_3, M_4 and M_5 are in the bottom (from left to right). The dotted line in red corresponds to the Kolmogorov–Smirnov test and the darkened areas are the 95% confidence intervals.

Using this choice of δ, we represent in Figures 2 and 3 below the percentage of rejection of $\mathcal{H}_0$ against the aforementioned $\mathcal{N}(m, 1)$ and $\mathcal{N}(0, \sigma^2)$ alternatives for f_0, for different values of the parameters, to investigate the sensitivity towards location and scale. We also make experiments in Figure 4 with different distributions as alternatives (choosing f_0 as the Student, uniform, Laplace and Cauchy distributions, respectively). First of all, the main observation is that all models give very similar results (all curves are almost superimposed) even if n is not so large. That corroborates the results of the paper: residuals from stable, explosive or some (univariate) unstable models satisfy the Bickel–Rosenblatt original convergence. Our procedure is roughly equivalent to the Kolmogorov–Smirnov one to test for location or scale in the Gaussian family (in fact it seems to be less powerful for location and slightly more powerful for scale). However, it appears that our procedure better managed to discriminate some alternatives with close but different distributions. The objective of the paper is mainly theoretical and of course, a much more extensive study is needed to give any permanent conclusion about the comparison (sensitivity in n, h_0, κ, $\mathbb{K}$, δ, role of the true distribution of (ε_t), etc.). Also, we must not underestimate the difficulty to proceed to the numerical integrations required for the statistic, and the very likely improvements that may stem from in-depth applied studies.

Remark 3.2. In the unstable case with $p = 1$ and a positive unit root (namely, the random walk), and a kernel satisfying $\int_{\mathbb{R}} \mathbb{K}'(s) \, ds \neq 0$, even if it is of lesser statistical interest, it is also possible to exploit Proposition 2.5 to derive a statistical procedure. Indeed, let

$$\sigma_0^2 = \int_{\mathbb{R}} s^2 \, f_0(s) \, ds \quad \text{and} \quad F_0 = \int_{-\delta}^{\delta} f_0(s) \, ds$$
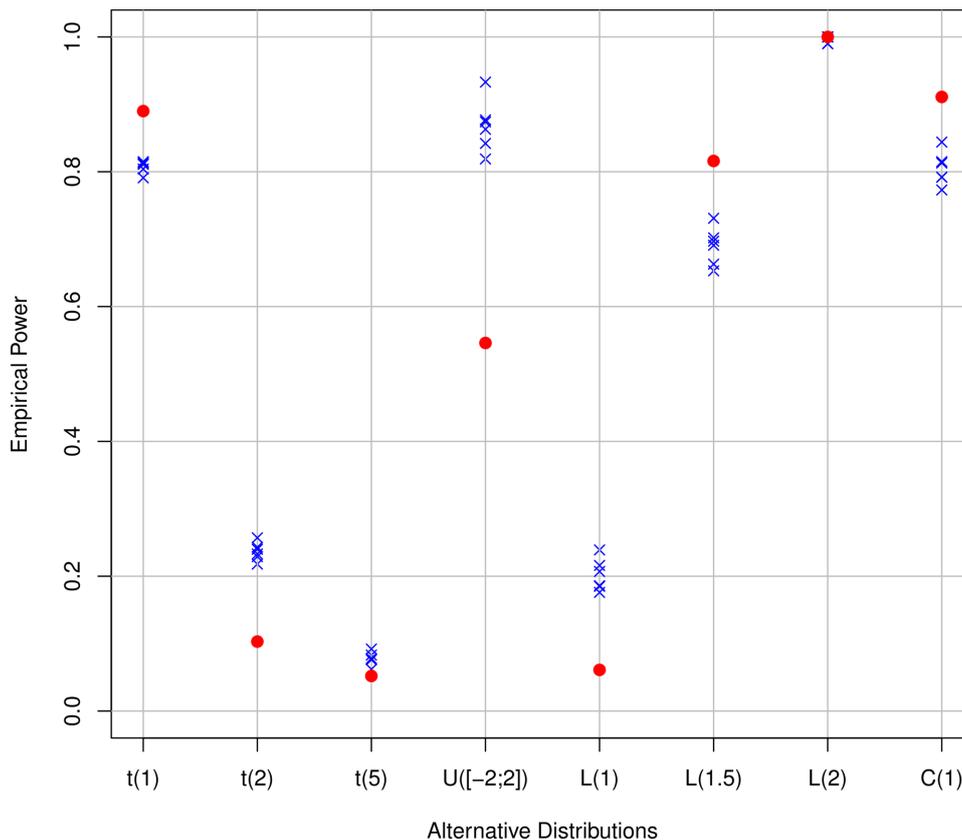
FIGURE 4. Frequency of rejection of the test for $n = 100$, $\mathbb{K} = \mathcal{N}(0, 1)$, $h_0 = 0.14$ and 1000 replications. The alternative f_0 are from left to right the Student $t(1)$, $t(2)$, $t(5)$ distributions, the uniform $\mathcal{U}([-2 \, ; \, 2])$ distribution, the centered Laplace $\mathcal{L}(1)$, $\mathcal{L}(1.5)$ and $\mathcal{L}(2)$ distributions and the centered Cauchy $\mathcal{C}(1)$ distribution. Results obtained from $M_0, \ldots, M_5$ are superimposed as blue crosses. The red circles correspond to the Kolmogorov–Smirnov test for M_0.

with an adjustment of σ_0^2 if f_0 is not centered. Then, we can choose

$$\widetilde{Z}_n^0 = \frac{h_n \big(\widetilde{T}_n^0 - \mu \big)}{\sigma_0 \, \sqrt{F_0} \int_{\mathbb{R}} \mathbb{K}'(s) \, \mathrm{d}s}$$

and compare it with the quantiles associated with the distribution of

$$Z = \left(\frac{\frac{1}{2} \left(W^2(1) - 1 \right)}{\int_0^1 W^2(u) \, \mathrm{d}u} \int_0^1 W(u) \, \mathrm{d}u \right)^2 .$$

APPENDIX A. INTERMEDIATE RESULTS

We recall some (already known) preliminary results related to the behavior of (X_t), depending on the eigenvalues of C_θ and, in each case, the asymptotic behavior of the least-squares estimator. In the sequel, we assume that Φ_0 shares the same assumptions of moments as (ε_t). We only focus on results that will be useful in our proofs.

A.1 Asymptotic behavior of the process

Proposition A.1 (Stable case). *Assume that (ε_t) is a strong white noise such that, for some $\nu \geqslant 1$, $\mathbb{E}[|\varepsilon_1|^\nu]$ is finite. If (X_t) satisfies (1.1) and is stable (that is $|\lambda_1| < 1$ or, equivalently, $\Theta(z) \neq 0$ for all $|z| \leqslant 1$), then as n tends to infinity,*

$$\sum_{t=1}^{n} |X_t|^\nu = O(n) \quad \text{a.s.} \quad and \quad \sup_{1 \leqslant t \leqslant n} |X_t| = o(n^{1/\nu}) \quad \text{a.s.}$$

In addition if $\nu \geqslant 2$, then

$$\sum_{t=1}^{n} \Phi_t = O_{\mathbb{P}}(\sqrt{n}).$$

Proof. See ([2], Lem. A.2) and ([17], Lem. B.2). The last result is deduced from the invariance principle of ([7], Thm. 1), since (X_t) is (asymptotically) stationary in this case. Indeed, we clearly have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Phi_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Phi_t^* + \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{n} C_\theta^t \right) (\Phi_0 - \Phi_0^*)$$

where (Φ_t^*) is the (second-order) stationary version of the process for $\nu \geqslant 2$. We conclude that (Φ_t) and (Φ_t^*) share the same invariance principle (with rate $\sqrt{n}$), since $\rho(C_\theta) < 1$. □

Proposition A.2 (Purely explosive case). *Assume that (ε_t) is a strong white noise having a finite variance σ^2. If (X_t) satisfies (1.1)–(1.4) and is purely explosive (that is $|\lambda_p| > 1$ or, equivalently, $\Theta(z) \neq 0$ for all $|z| \geqslant 1$), then*

$$\lim_{n \to +\infty} C_\theta^{-n} \Phi_n = \Phi_0 + \sum_{k=1}^{\infty} C_\theta^{-k} E_k = Z \quad \text{a.s.}$$

and

$$\lim_{n \to +\infty} C_\theta^{-n} \left(\sum_{t=1}^{n} \Phi_{t-1} \Phi_{t-1}^T \right) (C_\theta^{-n})^T = \sum_{k=1}^{\infty} C_\theta^{-k} Z Z^T (C_\theta^{-k})^T = G \quad \text{a.s.}$$

In addition, G is (a.s.) positive definite.

Proof. See ([12], Thm. 2). □

Proposition A.3 (Purely unstable case). *Assume that (ε_t) is a strong white noise having a finite variance σ^2. If (X_t) satisfies (1.1)–(1.4) with $p = 1$ and is unstable (that is $|\lambda| = 1$ or, equivalently, $\Theta(z) \neq 0$ for all $|z| \neq 1$), then for $k \in \mathbb{N}$,*

$$\sum_{t=1}^{n} X_t^{2k} = O_{\mathbb{P}}(n^{k+1}) \quad and \quad |X_n| = O_{\mathbb{P}}(n^{1/2}). \tag{A.1}$$

In addition,

$$\sum_{t=1}^{n} X_t = \begin{cases} O_{\mathbb{P}}(n^{3/2}) & for \ \lambda = 1 \\ O_{\mathbb{P}}(n^{1/2}) & for \ \lambda = -1. \end{cases} \tag{A.2}$$

If (X_t) satisfies *(1.1)–(1.4)* with $p = s$ and is seasonal unstable (that is $\theta_1 = \ldots = \theta_{s-1} = 0$ and $\theta_s = 1$), then *(A.1)* and *(A.2)* hold with $\lambda = \theta_s = 1$.

Proof. These results are detailed in Lemma B.4 and then proved. □

A.2 Asymptotic behavior of the least-squares estimator

In this subsection, (ε_t) is supposed to have a finite moment of order $2 + \gamma$, for some $\gamma > 0$. The consistency results have been established in ([12], Thm. 1), while the weak convergences are proved in [5], ([21], Thm. 2) and ([6], Thm. 3.5.1) respectively.

Proposition A.4 (Stable case)**.** *In the stable case ($|\lambda_1| < 1$), the least-squares estimator $\widehat{\theta}_n$ of θ is strongly consistent. In addition,*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\, \Gamma_p^{-1}\right)$$

where Γ_p is the $p \times p$ asymptotic covariance matrix of (X_t).

Proposition A.5 (Purely explosive case)**.** *In the purely explosive case ($|\lambda_p| > 1$), the least-squares estimator $\widehat{\theta}_n$ of θ is strongly consistent. In addition,*

$$C_\theta^n\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} U$$

where U is a nondegenerate random vector given in (1.8)–(1.9)–(1.10) of [21].

Proposition A.6 (Purely unstable case)**.** *In the purely unstable case ($|\lambda_1| = |\lambda_p| = 1$), the least-squares estimator $\widehat{\theta}_n$ of θ is strongly consistent. In the univariate case ($p = 1$), we have in addition*

$$n\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} \mathrm{sgn}(\theta)\, \frac{\frac{1}{2}\left(W^2(1) - 1\right)}{\int_0^1 W^2(u)\, \mathrm{d}u}$$

where $(W(t),\, t \in [0,1])$ is a standard Wiener process and $\mathrm{sgn}(\theta)$ stands for the sign of θ. In the seasonal case ($p = s$), since $\Theta(z) = 0$ is an equation admitting the complex s–th roots of unity as solutions,

$$n\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} S(W_s)$$

where $S(W_s)$ is a functional of a standard Wiener process $(W_s(t),\, t \in [0,1])$ of dimension s that can be explicitly determined following ([6], Thm. 3.5.1).

APPENDIX B. PROOFS OF THE MAIN RESULTS

In this section, we prove our results.

B.1 Stable case – Proof of Theorem 2.1

We consider the case where the kernel satisfies hypothesis (H1). The difference between the statistics can be expressed like

$$\widehat{T}_n - T_n = n\, h_n \int_{\mathbb{R}} \left(\widehat{f}_n(x) - f_n(x)\right)^2 a(x)\, \mathrm{d}x + R_n \tag{B.1}$$

with

$$R_n = 2\,n\,h_n \int_{\mathbb{R}} \left(\widehat{f}_n(x) - f_n(x)\right)\left(f_n(x) - (\mathbb{K}_{h_n} * f)(x)\right) a(x)\,\mathrm{d}x$$

and f_n is the analog of $\widehat{f}_n$ defined in (1.7) where the residuals $\widehat{\varepsilon}$ have been replaced by the strong white noise ε. Recall that T_n has been defined before (1.10). Now we follow an idea illustrated in the proof of ([11], Thm. 1.1) and in that of ([13], Thm. 2.1), which consists in using Taylor expansion to get, for all $1 \leqslant t \leqslant n$,

$$\mathbb{K}\left(\frac{x - \widehat{\varepsilon}_t}{h_n}\right) - \mathbb{K}\left(\frac{x - \varepsilon_t}{h_n}\right) = \frac{\varepsilon_t - \widehat{\varepsilon}_t}{h_n} \mathbb{K}'\left(\frac{x - \varepsilon_t}{h_n}\right) + \frac{(\varepsilon_t - \widehat{\varepsilon}_t)^2}{2\,h_n^2} \mathbb{K}''(\Delta_{tx})$$

where

$$\Delta_{tx} = \frac{x - \varepsilon_t + \zeta(\varepsilon_t - \widehat{\varepsilon}_t)}{h_n}$$

for some $0 < \zeta < 1$. Note that

$$\varepsilon_t - \widehat{\varepsilon}_t = \left(\widehat{\theta}_n - \theta\right)^T \Phi_{t-1} = \left\langle\widehat{\theta}_n - \theta, \Phi_{t-1}\right\rangle$$

and obviously that

$$\left|\left\langle\widehat{\theta}_n - \theta, \Phi_{t-1}\right\rangle\right| \leqslant \|\widehat{\theta}_n - \theta\|\,\|\Phi_{t-1}\|.$$

On the one hand, we consider the first term (say, I_n) of the right-hand side of (B.1) that can be bounded like

$$I_n = n\,h_n \int_{\mathbb{R}} \left\{\frac{1}{n\,h_n} \sum_{t=1}^{n}\left(\mathbb{K}\left(\frac{x - \widehat{\varepsilon}_t}{h_n}\right) - \mathbb{K}\left(\frac{x - \varepsilon_t}{h_n}\right)\right)\right\}^2 a(x)\,\mathrm{d}x$$

$$\leqslant \frac{2}{n\,h_n^3} \int_{\mathbb{R}} \left\{\sum_{t=1}^{n}\langle\widehat{\theta}_n - \theta, \Phi_{t-1}\rangle\,\mathbb{K}'\left(\frac{x - \varepsilon_t}{h_n}\right)\right\}^2 a(x)\,\mathrm{d}x$$

$$+ \frac{1}{2\,n\,h_n^5} \int_{\mathbb{R}} \left\{\sum_{t=1}^{n}\langle\widehat{\theta}_n - \theta, \Phi_{t-1}\rangle^2\,\mathbb{K}''(\Delta_{tx})\right\}^2 a(x)\,\mathrm{d}x. \tag{B.2}$$

At this step, we need two technical lemmas.

Lemma B.1. *We have*

$$I_{1,n} = \int_{\mathbb{R}} \left\{\sum_{t=1}^{n}\langle\widehat{\theta}_n - \theta, \Phi_{t-1}\rangle\,\mathbb{K}'\left(\frac{x - \varepsilon_t}{h_n}\right)\right\}^2 a(x)\,\mathrm{d}x = O_{\mathbb{P}}(h_n).$$

Proof. One clearly has

$$I_{1,n} \leqslant 2 \left\{\int_{\mathbb{R}} \left(\sum_{t=1}^{n}\langle\widehat{\theta}_n - \theta, \Phi_{t-1}\rangle\,v_t(x)\right)^2 a(x)\,\mathrm{d}x\right.$$

$$+ \int_{\mathbb{R}} \left(\sum_{t=1}^{n} \langle \widehat{\theta}_n - \theta, \Phi_{t-1} \rangle \, e(x) \right)^2 a(x) \, dx \right\} \; = \; 2 \left(J_{1,n} + J_{2,n} \right) \tag{B.3}$$

where

$$v_t(x) = \mathbb{K}' \left(\frac{x - \varepsilon_t}{h_n} \right) - \mathbb{E} \left[\mathbb{K}' \left(\frac{x - \varepsilon_1}{h_n} \right) \right] \quad \text{and} \quad e(x) = \mathbb{E} \left[\mathbb{K}' \left(\frac{x - \varepsilon_1}{h_n} \right) \right].$$

Now let us consider $J_{1,n}$ and $J_{2,n}$. First,

$$J_{1,n} \leqslant \| \widehat{\theta}_n - \theta \|^2 \int_{\mathbb{R}} \left\| \sum_{t=1}^{n} \Phi_{t-1} \, v_t(x) \right\|^2 a(x) \, dx$$

$$= \| \widehat{\theta}_n - \theta \|^2 \sum_{i=1}^{p} \int_{\mathbb{R}} \left(\sum_{t=1}^{n} X_{t-i} \, v_t(x) \right)^2 a(x) \, dx. \tag{B.4}$$

Let $i \in \{1, \ldots, p\}$. Then,

$$\mathbb{E} \left[\int_{\mathbb{R}} \left(\sum_{t=1}^{n} X_{t-i} \, v_t(x) \right)^2 a(x) \, dx \right] = \int_{\mathbb{R}} \mathbb{E} \left[\left(\sum_{t=1}^{n} X_{t-i} \, v_t(x) \right)^2 \right] a(x) \, dx$$

$$= \int_{\mathbb{R}} \sum_{t=1}^{n} \mathbb{E} \left[X_{t-i}^2 \right] \mathbb{E} \left[v_t^2(x) \right] a(x) \, dx$$

$$= \sum_{t=1}^{n} \mathbb{E} \left[X_{t-i}^2 \right] \int_{\mathbb{R}} \mathbb{E} \left[v_t^2(x) \right] a(x) \, dx.$$

But, under our assumptions, we recall that $\mathbb{E}[X_n^2] = O(1)$ from the asymptotic stationarity of the process and

$$\int_{\mathbb{R}} \mathbb{E} \left[v_t^2(x) \right] a(x) \, dx \leqslant \int_{\mathbb{R}} \mathbb{E} \left[\left(\mathbb{K}' \left(\frac{x - \varepsilon_1}{h_n} \right) \right)^2 \right] a(x) \, dx$$

$$= h_n \int_{\mathbb{R}} (\mathbb{K}'(z))^2 \int_{\mathbb{R}} f(x - h_n \, z) \, a(x) \, dx \, dz \; = \; O(h_n)$$

since $\mathrm{Var}(Z) \leqslant \mathbb{E}[Z^2]$ for some random variable Z and by (H0) and (H1). Thus $J_{1,n} = O_{\mathbb{P}}(h_n)$ *via* Proposition A.4. Then, by a direct calculation,

$$J_{2,n} = \left(\left(\widehat{\theta}_n - \theta \right)^T \sum_{t=1}^{n} \Phi_{t-1} \right)^2 \int_{\mathbb{R}} e^2(x) \, a(x) \, dx$$

$$= h_n^2 \left(\left(\widehat{\theta}_n - \theta \right)^T \sum_{t=1}^{n} \Phi_{t-1} \right)^2 \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{K}'(z) \, f(x - h_n \, z) \, dz \right)^2 a(x) \, dx. \tag{B.5}$$

Using Proposition A.1, it follows that $J_{2,n} = O_{\mathbb{P}}(h_n^2)$. $\qquad \square$

Lemma B.2. *We have*

$$I_{2,n} = \int_{\mathbb{R}} \left\{ \sum_{t=1}^{n} \left\langle \widehat{\theta}_n - \theta, \Phi_{t-1} \right\rangle^2 \mathbb{K}'' \left(\Delta_{tx} \right) \right\}^2 a(x)\, \mathrm{d}x = O_{\mathbb{P}} \left(h_n^2 + \frac{1}{n\,h_n^2} \right).$$

Proof. We directly get

$$I_{2,n} \leqslant \|\widehat{\theta}_n - \theta\|^4 \int_{\mathbb{R}} \left\{ \sum_{t=1}^{n} \|\Phi_{t-1}\|^2 \, \mathbb{K}'' \left(\Delta_{tx} \right) \right\}^2 a(x)\, \mathrm{d}x$$

$$\leqslant 3 \|\widehat{\theta}_n - \theta\|^4 \left(\int_{\mathbb{R}} \left\{ \sum_{t=1}^{n} \|\Phi_{t-1}\|^2 \left(\mathbb{K}'' \left(\Delta_{tx} \right) - \mathbb{K}'' \left(\frac{x - \varepsilon_t}{h_n} \right) \right) \right\}^2 a(x)\, \mathrm{d}x \right.$$

$$+ \int_{\mathbb{R}} \left\{ \sum_{t=1}^{n} \|\Phi_{t-1}\|^2 \left(\mathbb{K}'' \left(\frac{x - \varepsilon_t}{h_n} \right) - \mathbb{E} \left[\mathbb{K}'' \left(\frac{x - \varepsilon_1}{h_n} \right) \right] \right) \right\}^2 a(x)\, \mathrm{d}x$$

$$+ \left. \int_{\mathbb{R}} \left\{ \sum_{t=1}^{n} \|\Phi_{t-1}\|^2 \, \mathbb{E} \left[\mathbb{K}'' \left(\frac{x - \varepsilon_1}{h_n} \right) \right] \right\}^2 a(x)\, \mathrm{d}x \right)$$

$$= 3 \|\widehat{\theta}_n - \theta\|^4 \left(K_{1,n} + K_{2,n} + K_{3,n} \right). \tag{B.6}$$

From the mean value theorem, under our hypotheses,

$$\mathbb{K}'' \left(\Delta_{tx} \right) - \mathbb{K}'' \left(\frac{x - \varepsilon_t}{h_n} \right) = \mathbb{K}'' \left(\frac{x - \varepsilon_t + \zeta(\varepsilon_t - \widehat{\varepsilon}_t)}{h_n} \right) - \mathbb{K}'' \left(\frac{x - \varepsilon_t}{h_n} \right)$$

$$= \frac{\zeta(\varepsilon_t - \widehat{\varepsilon}_t)}{h_n} \mathbb{K}''' \left(\frac{x - \varepsilon_t + \xi\,\zeta(\varepsilon_t - \widehat{\varepsilon}_t)}{h_n} \right)$$

for some $0 < \zeta, \xi < 1$. We deduce that

$$\left| \mathbb{K}'' \left(\Delta_{tx} \right) - \mathbb{K}'' \left(\frac{x - \varepsilon_t}{h_n} \right) \right| \leqslant \frac{|\varepsilon_t - \widehat{\varepsilon}_t|}{h_n} \left| \mathbb{K}''' \left(\frac{x - \varepsilon_t + \xi\,\zeta(\varepsilon_t - \widehat{\varepsilon}_t)}{h_n} \right) \right|.$$

Consequently, since $\mathbb{K}'''$ is bounded,

$$K_{1,n} \leqslant \frac{C}{h_n^2} \left(\sum_{t=1}^{n} \|\Phi_{t-1}\|^2 \, |\varepsilon_t - \widehat{\varepsilon}_t| \right)^2 \leqslant \frac{C}{h_n^2} \sum_{t=1}^{n} \|\Phi_{t-1}\|^4 \sum_{t=1}^{n} (\varepsilon_t - \widehat{\varepsilon}_t)^2$$

$$\leqslant \frac{C}{h_n^2} \|\widehat{\theta}_n - \theta\|^2 \sum_{t=1}^{n} \|\Phi_{t-1}\|^4 \sum_{t=1}^{n} \|\Phi_{t-1}\|^2 \tag{B.7}$$

for some constants. Then, $K_{1,n} = O_{\mathbb{P}}(n\,h_n^{-2})$ as soon as we suppose that (ε_t) has a finite moment of order $\nu = 4$ (by virtue of Prop. A.1). Now we proceed as for $J_{1,n}$ to get

$$\mathbb{E}[K_{2,n}] \leqslant \sum_{t=1}^{n} \mathbb{E}\left[\|\Phi_{t-1}\|^4\right] \int_{\mathbb{R}} \mathbb{E} \left[\mathbb{K}'' \left(\frac{x - \varepsilon_1}{h_n} \right)^2 \right] a(x)\, \mathrm{d}x$$

$$= h_n \sum_{t=1}^{n} \mathbb{E}\big[\|\Phi_{t-1}\|^4\big] \int_{\mathbb{R}} (\mathbb{K}''(z))^2 \int_{\mathbb{R}} f(x - h_n\, z)\, a(x)\, \mathrm{d}x\, \mathrm{d}z \;=\; O(n\, h_n) \qquad (B.8)$$

which shows $K_{2,n} = O_\mathbb{P}(n\, h_n)$, since $\mathbb{E}\big[X_n^4\big] = O(1)$ under the hypotheses of stability and fourth-order moments. Finally,

$$K_{3,n} \leqslant h_n^2 \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |\mathbb{K}''(z)|\, f(x - h_n\, z)\, \mathrm{d}z \right)^2 a(x)\, \mathrm{d}x\, \left(\sum_{t=1}^{n} \|\Phi_{t-1}\|^2 \right)^2$$

$$\leqslant C\, h_n^2 \left(\sum_{t=1}^{n} \|\Phi_{t-1}\|^2 \right)^2 \;=\; O_\mathbb{P}(n^2\, h_n^2) \qquad (B.9)$$

for a constant C. Whence we deduce that $I_{2,n} = O_\mathbb{P}(h_n^2 + n^{-1} h_n^{-2})$. $\qquad\square$

We are now ready to conclude the proof of Theorem 2.1. If we come back to I_n in (B.2), then the combination of Lemmas B.1 and B.2 leads to

$$\frac{I_n}{\sqrt{h_n}} = O_\mathbb{P}\left(\frac{1}{n\, h_n^{5/2}} + \frac{1}{n\, h_n^{7/2}} + \frac{1}{n^2\, h_n^{15/2}} \right) = o_\mathbb{P}(1) \qquad (B.10)$$

as soon as $n\, h_n^{15/4} \to +\infty$. Now by Cauchy–Schwarz inequality,

$$\frac{|R_n|}{\sqrt{h_n}} \leqslant 2 \sqrt{\frac{I_n\, T_n}{h_n}} = O_\mathbb{P}\left(\sqrt{\frac{I_n}{h_n}} \right) = o_\mathbb{P}(1) \qquad (B.11)$$

from the previous reasoning, the asymptotic normality (1.10) and assumption (H3) with $\alpha = 4$. It follows from (B.1), (B.10) and (B.11) that

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} = \frac{T_n - \mu}{\sqrt{h_n}} + o_\mathbb{P}(1)$$

which ends the first part of the proof. The second part makes use of a result of Bickel and Rosenblatt [3]. Indeed, denote by $\bar{T}_n$ the statistic given in (1.9) built on the strong white noise (ε_t) instead of the residuals. They show that

$$\frac{|T_n - \bar{T}_n|}{\sqrt{h_n}} = o_\mathbb{P}(1) \qquad (B.12)$$

as soon as $n\, h_n^{9/2} \to 0$. A similar calculation leads to

$$\widetilde{T}_n = I_n + \bar{T}_n + \bar{R}_n$$

where

$$\bar{R}_n = 2\, n\, h_n \int_{\mathbb{R}} \left(\widehat{f}_n(x) - f_n(x) \right) \left(f_n(x) - f(x) \right) a(x)\, \mathrm{d}x.$$

We deduce that

$$\frac{|\widetilde{T}_n - \bar{T}_n|}{\sqrt{h_n}} \leqslant \frac{I_n}{\sqrt{h_n}} + \frac{|\bar{R}_n|}{\sqrt{h_n}} = O_{\mathbb{P}}\left(\sqrt{\frac{I_n}{h_n}}\right) = o_{\mathbb{P}}(1)$$

as soon as $\bar{T}_n$ satisfies the original Bickel–Rosenblatt convergence and (B.10) and (B.11) hold, that is $n\, h_n^{9/2} \to 0$ and $n\, h_n^4 \to +\infty$. It only remains to note that

$$\frac{|\widehat{T}_n - \widetilde{T}_n|}{\sqrt{h_n}} \leqslant \frac{|\widehat{T}_n - T_n|}{\sqrt{h_n}} + \frac{|T_n - \bar{T}_n|}{\sqrt{h_n}} + \frac{|\bar{T}_n - \widetilde{T}_n|}{\sqrt{h_n}},$$

each term being $o_{\mathbb{P}}(1)$. The proof of Theorem 2.1 is now complete.

B.2 Purely explosive case – Proof of Theorem 2.3

Let I_n be the first term of the right-hand side of (B.1), like in the previous proof, and note that

$$I_n = n\, h_n \int_{\mathbb{R}} \left(\widehat{f}_n(x) - f_n(x)\right)^2 a(x)\, \mathrm{d}x \leqslant 2\,(I_{n,1} + I_{n,2}), \tag{B.13}$$

where for an arbitrary rate that we set to $v_n^2 = \ln(n\, h_n)$,

$$I_{n,1} = \frac{1}{n\, h_n} \int_{\mathbb{R}} \left\{ \sum_{t=1}^{n-[v_n]} \left(\mathbb{K}\left(\frac{x - \widehat{\varepsilon}_t}{h_n}\right) - \mathbb{K}\left(\frac{x - \varepsilon_t}{h_n}\right)\right) \right\}^2 a(x)\, \mathrm{d}x,$$

$$I_{n,2} = \frac{1}{n\, h_n} \int_{\mathbb{R}} \left\{ \sum_{t=n-[v_n]+1}^{n} \left(\mathbb{K}\left(\frac{x - \widehat{\varepsilon}_t}{h_n}\right) - \mathbb{K}\left(\frac{x - \varepsilon_t}{h_n}\right)\right) \right\}^2 a(x)\, \mathrm{d}x.$$

Under our hypotheses, this choice of (v_n) ensures

$$\lim_{n \to +\infty} v_n = +\infty \quad \text{and} \quad \lim_{n \to +\infty} \frac{v_n^2}{n\sqrt{h_n}} = 0. \tag{B.14}$$

Let us look at $I_{n,1}$. By Cauchy–Schwarz,

$$\begin{aligned}
I_{n,1} &\leqslant \frac{n - [v_n]}{n\, h_n} \int_{\mathbb{R}} \sum_{t=1}^{n-[v_n]} \left[\mathbb{K}\left(\frac{x - \widehat{\varepsilon}_t}{h_n}\right) - \mathbb{K}\left(\frac{x - \varepsilon_t}{h_n}\right)\right]^2 a(x)\, \mathrm{d}x \\
&\leqslant \frac{2\,\|\mathbb{K}\|_\infty\,(n - [v_n])}{n\, h_n} \int_{\mathbb{R}} \sum_{t=1}^{n-[v_n]} \left|\mathbb{K}\left(\frac{x - \widehat{\varepsilon}_t}{h_n}\right) - \mathbb{K}\left(\frac{x - \varepsilon_t}{h_n}\right)\right| a(x)\, \mathrm{d}x \\
&\leqslant \frac{2\,\|a\|_\infty\,\|\mathbb{K}\|_\infty\,(n - [v_n])}{n} \int_{\mathbb{R}} \sum_{t=1}^{n-[v_n]} \left|\mathbb{K}\left(u + \frac{\varepsilon_t - \widehat{\varepsilon}_t}{h_n}\right) - \mathbb{K}(u)\right| \mathrm{d}u \\
&\leqslant \frac{2\,B\,\|a\|_\infty\,\|\mathbb{K}\|_\infty\,(n - [v_n])}{n\, h_n} \sum_{t=1}^{n-[v_n]} |\varepsilon_t - \widehat{\varepsilon}_t|
\end{aligned}$$

$$= \frac{2\,B\,\|a\|_\infty\,\|\mathbb{K}\|_\infty\,(n-[v_n])}{n\,h_n} \sum_{t=1}^{n-[v_n]} \left|(\widehat{\theta}_n-\theta)^T\,\Phi_{t-1}\right|$$

$$= \frac{2\,B\,\|a\|_\infty\,\|\mathbb{K}\|_\infty\,(n-[v_n])}{n\,h_n} \sum_{t=1}^{n-[v_n]} \left|(\widehat{\theta}_n-\theta)^T\,C_\theta^n\,C_\theta^{-[v_n]}\,C_\theta^{-n+[v_n]}\,\Phi_{t-1}\right|.$$

Here, we recall that $|\lambda_p| > 1$. Consequently, $\rho(C_\theta^{-1}) = 1/|\lambda_p| < 1$. It follows that there exists a matrix norm $\|\cdot\|_* = \sup(|\cdot u|_*\,;\,u \in \mathbb{C}^p,\,|u|_* = 1)$ satisfying $\|C_\theta^{-1}\|_* < 1$ (see *e.g.* [8], Prop. 2.3.15), and a constant k^* such that, with $C^* = 2\,B\,k^*\,\|a\|_\infty\,\|\mathbb{K}\|_\infty$, we obtain

$$I_{n,1} \leqslant \frac{C^*\,(n-[v_n])}{n\,h_n} \sum_{t=1}^{n-[v_n]} \left|(\widehat{\theta}_n-\theta)^T\,C_\theta^n\right|_* \left|C_\theta^{-[v_n]}\,C_\theta^{-n+[v_n]}\,\Phi_{t-1}\right|_*$$

$$\leqslant \frac{C^*\,(n-[v_n])}{n\,h_n} \left|(\widehat{\theta}_n-\theta)^T\,C_\theta^n\right|_* \|C_\theta^{-1}\|_*^{[v_n]} \sum_{t=1}^{n-[v_n]} \left|C_\theta^{-n+[v_n]}\,\Phi_{t-1}\right|_*$$

$$\leqslant \frac{C^*\,(n-[v_n])}{n\,h_n} \left|(\widehat{\theta}_n-\theta)^T\,C_\theta^n\right|_* \|C_\theta^{-1}\|_*^{[v_n]} \sup_{0\leqslant t\leqslant n-[v_n]-1} \left|C_\theta^{-t}\,\Phi_t\right|_* \sum_{\ell=1}^{n-[v_n]} \|C_\theta^{-1}\|_*^\ell.$$

But Proposition A.2 ensures that $\sup_t |C_\theta^{-t}\,\Phi_t|_*$ is a.s. bounded for a sufficiently large n, and we also have

$$(\widehat{\theta}_n-\theta)^T\,C_\theta^n = \left(\sum_{t=1}^{n}\Phi_{t-1}^T\,\varepsilon_t\right)(C_\theta^{-n})^T\,(C_\theta^n)^T \left(\sum_{t=1}^{n}\Phi_{t-1}\Phi_{t-1}^T\right)^{-1} C_\theta^n$$

$$= \sum_{t=1}^{n}\left(C_\theta^{-n}\,\Phi_{t-1}\,\varepsilon_t\right)^T G_n^{-1}$$

where

$$G_n = C_\theta^{-n}\left(\sum_{t=1}^{n}\Phi_{t-1}\Phi_{t-1}^T\right)(C_\theta^{-n})^T.$$

We deduce that, for some constant k^*,

$$\left|(\widehat{\theta}_n-\theta)^T\,C_\theta^n\right|_* \leqslant k^*\,\varepsilon_n^\sharp\,\|G_n^{-1}\|_* \sup_{0\leqslant t\leqslant n-1} \left|(C_\theta^{-t}\,\Phi_t)^T\right|_* \sum_{\ell=1}^{n} \|C_\theta^{-1}\|_*^\ell$$

where ([8], Cor. 1.3.21) shows that $\varepsilon_n^\sharp = \sup_t |\varepsilon_t| = o(\sqrt{n})$ a.s. under our conditions of moments on (ε_t). Proposition A.2 and the fact that $\|C_\theta^{-1}\|_* < 1$ lead to

$$I_{n,1} = o\left(\frac{n-[v_n]}{\sqrt{n}\,h_n}\,R^{[v_n]}\right) \quad \text{a.s.} \tag{B.15}$$

for some $0 < R < 1$. Let us now turn to $I_{n,2}$ for which the same strategy gives

$$
\begin{aligned}
I_{2,n} &\leqslant \frac{[v_n]}{n\, h_n} \sum_{t=n-[v_n]+1}^{n} \int_{\mathbb{R}} \left[\mathbb{K}\left(\frac{x-\widehat{\varepsilon}_t}{h_n}\right) - \mathbb{K}\left(\frac{x-\varepsilon_t}{h_n}\right) \right]^2 a(x)\, \mathrm{d}x \\
&\leqslant \frac{2\,[v_n]}{n\, h_n} \sum_{t=n-[v_n]+1}^{n} \int_{\mathbb{R}} \left[\mathbb{K}^2\left(\frac{x-\widehat{\varepsilon}_t}{h_n}\right) + \mathbb{K}^2\left(\frac{x-\varepsilon_t}{h_n}\right) \right] a(x)\, \mathrm{d}x \\
&\leqslant \frac{4\,\|a\|_\infty\,[v_n]}{n} \sum_{t=n-[v_n]+1}^{n} \int_{\mathbb{R}} \mathbb{K}^2(u)\, \mathrm{d}u \\
&= \frac{4\,\|a\|_\infty\,[v_n]^2}{n} \int_{\mathbb{R}} \mathbb{K}^2(u)\, \mathrm{d}u.
\end{aligned}
$$

Thus,

$$
I_{n,2} = O\left(\frac{[v_n]^2}{n}\right) \quad \text{a.s.} \tag{B.16}
$$

Finally, from (B.1), (B.15) and (B.16), we have

$$
I_n = O\left(\frac{n-[v_n]}{\sqrt{n}\, h_n}\, R^{[v_n]} + \frac{[v_n]^2}{n}\right) = o\left(\sqrt{h_n}\right) \quad \text{a.s.} \tag{B.17}
$$

for some $0 < R < 1$, as a consequence of the properties of (v_n) in (B.14). The cross term R_n is treated in the same way as in the proof of Theorem (2.1). Indeed, since our choice of (v_n) also ensures that $I_n = o(h_n)$ a.s., the same reasoning leads to the conclusion. It follows from (B.13) and (B.17) that

$$
\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} = \frac{T_n - \mu}{\sqrt{h_n}} + o(1) \quad \text{a.s.}
$$

which ends the first part of the proof. The second part merely consists in noting that (B.12) still holds. The proof of Theorem 2.3 is now complete.

B.3 Unstable case – Proof of Propositions A.3 and 2.5

In this proof, the notation $\Longrightarrow$ refers to the weak convergence of sequences of random elements in $D([0,1])$, the space of right continuous functions on $[0,1]$ having left-hand limits, equipped with the Skorokhod topology (see Billingsley [4]). One can find in Thm. 2.7 of the same reference the statement and the proof of the continuous mapping theorem. In Lemmas B.4 and B.8 below, $(Y_n)_{n\geqslant 1}$ is a sequence of random variables satisfying, for some $\delta > 0$,

$$
\frac{1}{n^\delta} \sum_{t \in \mathcal{P}_{[n\cdot]}} Y_t \Longrightarrow L_{\mathcal{P}}(\cdot) \quad \text{and} \quad \frac{1}{n^\delta} \sum_{t=1}^{[n\cdot]} (-1)^{[n\cdot]-t}\, Y_t \Longrightarrow \Lambda(\cdot) \tag{B.18}
$$

where $L_{\mathcal{P}}$ and Λ are random paths in $D([0,1])$ and where $\mathcal{P}_{[n\cdot]}$ stands either for $\{1, \dots, [n\cdot]\}$ or for the sets of even/odd integers in $\{1, \dots, [n\cdot]\}$. We use the convention $\sum_{\varnothing} = 0$ and the notation $[x]$ to designate the integer part of any $x \geqslant 0$.

Remark B.3. The first convergence obviously holds for a strong white noise having a finite variance and $\delta = 1/2$, *via* Donsker theorem ([4], Thm. 8.2). In this case, $L_{\mathcal{P}}$ is a Wiener process and $\mathcal{P}_{[n\cdot]}$ changes its variance (half as many terms leads to a doubled variance). If the distribution is symmetric, then $\Lambda \equiv L_{\mathcal{P}}$ for $\mathcal{P}_{[n\cdot]} = \{1, \dots, [n\cdot]\}$, whereas an invariance principle for stationary processes (see, *e.g.*, ([7], Thm. 1) enables to identify Λ in the nonsymmetric case.

Lemma B.4 (Univariate unstable case). *Let $X_0 = 0$ and, for $1 \leqslant t \leqslant n$, consider*

$$X_t = \theta\, X_{t-1} + Y_t, \qquad \theta = \pm 1,$$

where (Y_t) satisfies (B.18). Then,

$$\sum_{t=1}^{n} X_t = \begin{cases} O_{\mathbb{P}}(n^{\delta+1}) & \text{for } \theta = 1 \\ O_{\mathbb{P}}(n^{\delta}) & \text{for } \theta = -1 \end{cases}$$

and, in both cases, for $k \in \mathbb{N}$,

$$\sum_{t=1}^{n} X_t^{2k} = O_{\mathbb{P}}(n^{2k\delta+1}).$$

Proof. Let $\theta = 1$. Then,

$$\frac{1}{n^{\delta+1}} \sum_{t=1}^{n} X_t = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{n^{\delta}} \sum_{s=1}^{t} Y_s$$

$$= \sum_{t=1}^{n} \int_{\frac{t}{n}}^{\frac{t+1}{n}} \frac{1}{n^{\delta}} \sum_{s=1}^{[nu]} Y_s \,\mathrm{d}u \quad \xrightarrow{\mathcal{D}} \quad \int_0^1 L(u)\,\mathrm{d}u$$

from the continuous mapping theorem, where $L \equiv L_{\mathcal{P}}$ for $\mathcal{P}_{[n\cdot]} = \{1, \dots, [n\cdot]\}$. Following the same lines,

$$\frac{1}{n^{2k\delta+1}} \sum_{t=1}^{n} X_t^{2k} = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{n^{\delta}} \sum_{s=1}^{t} Y_s \right)^{2k}$$

$$= \sum_{t=1}^{n} \int_{\frac{t}{n}}^{\frac{t+1}{n}} \left(\frac{1}{n^{\delta}} \sum_{s=1}^{[nu]} Y_s \right)^{2k} \mathrm{d}u \quad \xrightarrow{\mathcal{D}} \quad \int_0^1 L^{2k}(u)\,\mathrm{d}u.$$

Now let $\theta = -1$. Then,

$$\sum_{t=1}^{n} X_t = \sum_{t \in \mathcal{P}_n} Y_t$$

where $\mathcal{P}_n \subset \{1, \dots, n\}$ is the set of even integers (resp. odd integers) between 1 and n, for an even (resp. odd) n. Then we conclude to the first result. Finally,

$$\frac{1}{n^{2k\delta+1}} \sum_{t=1}^{n} X_t^{2k} = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{n^{\delta}} \sum_{s=1}^{t} (-1)^{t-s}\, Y_s \right)^{2k}$$

$$= \sum_{t=1}^{n} \int_{\frac{t}{n}}^{\frac{t+1}{n}} \left(\frac{1}{n^{\delta}} \sum_{s=1}^{[nu]} (-1)^{[nu]-s} Y_s \right)^{2k} \mathrm{d}u \xrightarrow{\mathcal{D}} \int_0^1 \Lambda^{2k}(u)\, \mathrm{d}u.$$

$\square$

Proof of Proposition A.3 – Seasonal case. Let $X_{-s+1} = \ldots = X_0 = 0$ and $\mathcal{S}_k = \{\ell \in \{1, \ldots, n\}, \ell[s] = k\}$, for $k \in \{0, \ldots, s-1\}$, the notation $\ell[s]$ referring to the remainder of the Euclidean division of ℓ by s. Then the subsets $\mathcal{S}_k$ form a partition of $\{1, \ldots, n\}$, the path $(X_1, \ldots, X_n)$ is made of s uncorrelated random walks and Donsker theorem directly gives the $O_{\mathbb{P}}(\sqrt{n})$ rate for $|X_n|$. In addition,

$$\sum_{t=1}^{n} X_t^a = \sum_{k=0}^{s-1} \sum_{t \in \mathcal{S}_k} X_t^a$$

so that, for $a \in \{1, 2, 4\}$, the results follow from Lemma B.4. $\square$

We now turn to the proof of Proposition 2.5. Following the idea of [13], we make once again the decomposition

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} = \frac{I_n}{\sqrt{h_n}} + \frac{R_n}{\sqrt{h_n}} + \frac{T_n - \mu}{\sqrt{h_n}} \tag{B.19}$$

where μ is the centering term (1.11) of the statistic, I_n is given in (B.2), T_n is the original Bickel-Rosenblatt statistic and R_n is the cross term in (B.1) such that

$$\frac{|R_n|}{\sqrt{h_n}} = O_{\mathbb{P}}\left(\sqrt{\frac{I_n}{h_n}} \right) \tag{B.20}$$

as we have seen earlier. Going on with the decomposition of I_n in (B.2) and using the same notation, we establish the two following lemmas. In all the sequel, the case $p = s$ refers to the seasonal process $\theta_1 = \ldots = \theta_{s-1} = 0$ and $\theta_s = 1$.

Lemma B.5. *When $p = 1$ or $p = s$, Lemma B.2 still holds in the unstable case.*

Proof. First, with $p = 1$ and $\theta = \pm 1$, we note that

$$\mathbb{E}[X_n^2] = \sum_{1 \leqslant t, s \leqslant n} \theta^{2n-t-s}\, \mathbb{E}[\varepsilon_t\, \varepsilon_s] = O(n)$$

and

$$\mathbb{E}[X_n^4] = \sum_{1 \leqslant t,s,u,v \leqslant n} \theta^{4n-t-s-u-v}\, \mathbb{E}[\varepsilon_t\, \varepsilon_s\, \varepsilon_u\, \varepsilon_v] = O(n^2)$$

under our assumptions on the noise, and the same results obviously hold for $p = s$ where X_n still has a random walk behavior. Thus, $K_{2,n}$ in (B.8) is $O_{\mathbb{P}}(n^3\, h_n)$. Lemma B.4 and the second part of Proposition A.3 also give

$$\sum_{t=1}^{n} X_t^2 = O_{\mathbb{P}}(n^2) \quad \text{and} \quad \sum_{t=1}^{n} X_t^4 = O_{\mathbb{P}}(n^3)$$

and, accordingly with Proposition A.6, $K_{1,n}$ in (B.7) is $O_{\mathbb{P}}(n^3 h_n^{-2})$ and $K_{3,n}$ in (B.9) is $O_{\mathbb{P}}(n^4 h_n^2)$. Finally, (B.6) concludes the proof. $\qquad\square$

Lemma B.6. *Let $p = 1$. Then, Lemma B.1 holds for $\lambda = \theta = -1$. On the contrary, for $\lambda = \theta = 1$ and $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s \neq 0$,*

$$I_{1,n} = O_{\mathbb{P}}(n\, h_n^2)$$

whereas for $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s = 0$,

$$I_{1,n} = O_{\mathbb{P}}(n\, h_n^4).$$

The last two results also hold for $p = s$.

Proof. On the one hand, let $\theta = -1$. The reasoning of (B.4) shows that $J_{1,n}$ is still $O_{\mathbb{P}}(h_n)$, using Proposition A.6 and $\mathbb{E}[X_n^2] = O(n)$. From Lemma B.4, we know that

$$\sum_{t=1}^{n} X_t = O_{\mathbb{P}}(\sqrt{n}).$$

Then, $J_{2,n}$ in (B.5) must be $O_{\mathbb{P}}(n^{-1} h_n^2)$ and $I_{1,n} = O_{\mathbb{P}}(h_n)$. On the other hand, we have to investigate the case where $\theta = 1$. Since $\mathbb{E}[X_n^2]$ does not depend on θ, $J_{1,n}$ has exactly the same behavior. For a better readability, let

$$L_n = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{K}'(z)\, f(x - h_n\, z)\,\mathrm{d}z \right)^2 a(x)\,\mathrm{d}x.$$

Under our assumptions, it is easy to see that

$$\lim_{n \to +\infty} L_n = \left(\int_{\mathbb{R}} f^2(s)\, a(s)\,\mathrm{d}s \right) \left(\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s \right)^2 \tag{B.21}$$

as soon as $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s \neq 0$, and that

$$\lim_{n \to +\infty} \frac{L_n}{h_n^2} = \left(\int_{\mathbb{R}} f'(s)^2\, a(s)\,\mathrm{d}s \right) \left(\int_{\mathbb{R}} s\, \mathbb{K}'(s)\,\mathrm{d}s \right)^2 \tag{B.22}$$

as soon as $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s = 0$. Lemma B.4 also gives

$$\sum_{t=1}^{n} X_t = O_{\mathbb{P}}(n^{3/2})$$

which shows that $J_{2,n}$ is $O_{\mathbb{P}}(n\, h_n^2)$ or $O_{\mathbb{P}}(n\, h_n^4)$, depending on $\int_{\mathbb{R}} \mathbb{K}'(s)\,\mathrm{d}s$. The latter reasoning still applies for $p = s$, *via* the second part of Proposition A.3. The proof is achieved using (B.3) and the hypotheses of Proposition 2.5. $\qquad\square$

The combination of Lemmas B.5 and B.6 is sufficient to establish that

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$$

holds for $\theta = -1$, despite the instability, and replacing $\widehat{T}_n$ by $\widetilde{T}_n$ is possible without disturbing the asymptotic normality, as a consequence of (B.12) for $n\, h_n^{9/2} \to 0$. When $\theta = 1$ or $p = s$ and $\int_{\mathbb{R}} \mathbb{K}'(s)\, \mathrm{d}s = 0$, Lemmas B.5 and B.6 show that $I_n = O_{\mathbb{P}}(h_n) = o_{\mathbb{P}}(\sqrt{h_n})$. It follows from (B.19) and (B.20) that

$$\frac{|R_n|}{\sqrt{h_n}} = O_{\mathbb{P}}(1)$$

which unfortunately prevents us from concluding to the Bickel–Rosenblatt convergence in this case. However, we still have the order of magnitude

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} = O_{\mathbb{P}}(1) \quad \text{and} \quad \frac{\widetilde{T}_n - \mu}{\sqrt{h_n}} = O_{\mathbb{P}}(1).$$

Finally for $\int_{\mathbb{R}} \mathbb{K}'(s)\, \mathrm{d}s \neq 0$, the same lines also imply, together with (B.19), Lemmas B.5 and B.6 and the asymptotic normality (1.10), that

$$h_n(\widehat{T}_n - \mu) = O_{\mathbb{P}}(1).$$

It remains to study the asymptotic behavior of the correctly renormalized statistic using the explicit expression of $J_{2,n}$ in (B.5). From the proof of Lemma B.4, Proposition A.6 and the continuous mapping theorem, we get for the univariate unstable case with $\theta = 1$,

$$\left(n\,(\widehat{\theta}_n - \theta) \right)^2 \left(\frac{1}{n^{3/2}} \sum_{t=1}^{n} X_{t-1} \right)^2 \xrightarrow{\mathcal{D}} \sigma^2 \left(\frac{\frac{1}{2}(W^2(1) - 1)}{\int_0^1 W^2(u)\, \mathrm{d}u} \int_0^1 W(u)\, \mathrm{d}u \right)^2$$

where $(W(t),\, t \in [0,1])$ is a standard Wiener process. Now for the seasonal case, reusing the same notation as above,

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} X_t = \sum_{k=0}^{s-1} \frac{1}{n^{3/2}} \sum_{t \in \mathcal{S}_k} X_t \xrightarrow{\mathcal{D}} \sigma \sum_{k=0}^{s-1} \int_0^{\frac{1}{s}} W_s^{(k+1)}(u)\, \mathrm{d}u \tag{B.23}$$

where $W_s^{(i)}$ stands for the i-th component of the standard Wiener process $(W_s(t),\, t \in [0,1])$ of dimension s, since the path can be seen as the concatenation of s uncorrelated random walks. Using the Cramèr–Wold and the continuous mapping theorems, it follows that

$$\left(n\left(\widehat{\theta}_n - \theta\right)^T \frac{1}{n^{3/2}} \sum_{t=1}^{n} \Phi_{t-1} \right)^2 \xrightarrow{\mathcal{D}} \sigma^2 \left(S(W_s)^T H(W_s) \right)^2$$

with the notation of Proposition A.6, where $H(W_s)$ is a random vector of size s containing the limit variable (B.23) on each component. The proof is then ended using the limiting behavior of L_n. The proof of Theorem 2.5 is now complete.

Remark B.7. To go beyond, note that for $\theta = 1$ and $\int_{\mathbb{R}} \mathbb{K}'(s)\, \mathrm{d}s \neq 0$, we get

$$h_n\, |\widehat{T}_n - \widetilde{T}_n| \leqslant 2\, h_n\, \sqrt{\widehat{T}_n\, U_n} + h_n\, U_n$$

where

$$U_n = n\,h_n \int_{\mathbb{R}} \left((\mathbb{K}_{h_n} * f)(x) - f(x) \right)^2 a(x)\,\mathrm{d}x.$$

A straightforward calculation shows that $U_n = O(n\,h_n^3)$ if $\int_{\mathbb{R}} s\,\mathbb{K}(s)\,\mathrm{d}s \neq 0$ whereas $U_n = O(n\,h_n^5)$ if $\int_{\mathbb{R}} s\,\mathbb{K}(s)\,\mathrm{d}s = 0$. In the second case, $\widehat{T}_n$ may be replaced by $\widetilde{T}_n$ as soon as $n\,h_n^6 \to 0$. But it the first case, one needs $n\,h_n^4 \to 0$ which contradicts (H3) with $\alpha = 4$.

This last lemma is not used as part of this paper, but it may be a trail for future studies. It is related to the conclusion of Section 2 and illustrates the phenomenon of compensation in purely unstable processes.

Lemma B.8. *Let $X_{-1} = X_0 = 0$ and, for $1 \leqslant t \leqslant n$, consider*

$$X_t = 2\cos(\omega)\,X_{t-1} - X_{t-2} + Y_t, \qquad 0 \leqslant \omega \leqslant \pi,$$

where (Y_t) satisfies (B.18). Then,

$$\sum_{t=1}^{n} X_t = \begin{cases} O_{\mathbb{P}}(n^{\delta+2}) & \text{for } \omega = 0 \\ O_{\mathbb{P}}(n^{\delta}) & \text{for } 0 < \omega \leqslant \pi. \end{cases}$$

Moreover, if the process is generated by

$$X_t = X_{t-2} + Y_t,$$

then

$$\sum_{t=1}^{n} X_t = O_{\mathbb{P}}(n^{\delta+1}).$$

Proof. For $\omega = 0$, let $Z_t = X_t - X_{t-1}$. Then, $Z_t - Z_{t-1} = Y_t$ and from Lemma B.4,

$$\sum_{t=1}^{n} Z_t = O_{\mathbb{P}}(n^{\delta+1}).$$

A second application of Lemma B.4 gives the result. If $0 < \omega \leqslant \pi$, we obtain

$$\frac{2\left(1 - \cos(\omega)\right)}{n^{\delta}} \sum_{t=1}^{n} X_t = \frac{X_{n-1} + \left(1 - 2\cos(\omega)\right) X_n}{n^{\delta}} + \frac{1}{n^{\delta}} \sum_{t=1}^{n} Y_t.$$

In the last case, let $Z_t = X_t - X_{t-1}$ and note that $Z_t = -Z_{t-1} + Y_t$. Then, Lemma B.4 gives

$$\sum_{t=1}^{n} Z_t = O_{\mathbb{P}}(n^{\delta}).$$

We conclude by applying again Lemma B.4 to $X_t = X_{t-1} + Z_t$. $\qquad\square$

Remark B.9. In the previous lemma, $\omega = 0$ corresponds to the unit roots $\{1, 1\}$ in an AR(2) process generated by (Y_t) whereas $0 < \omega \leqslant \pi$ corresponds to the unit roots $\{\mathrm{e}^{\mathrm{i}\omega}, \mathrm{e}^{-\mathrm{i}\omega}\}$, including $\{-1, -1\}$ for $\omega = \pi$. The last case corresponds to the unit roots $\{-1, 1\}$, in other words this is the seasonal model for $s = 2$.

References

[1] D. Bachmann and H. Dette, A note on the the Bickel-Rosenblatt test in autoregressive time series. *Stat. Probab. Lett.* **74** (2005) 221–234.

[2] B. Bercu and F. Proïa, A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process. *ESAIM: PS* **17** (2013) 500–530.

[3] P. J. Bickel and M. Rosenblatt, On some global measures of the deviations of density function estimates. *Ann. Stat.* **1** (1973) 1071–1095.

[4] P. Billingsley, Convergence of Probability Measures. *Wiley Series in Probability and Statistics: Probability and Statistics*. 2nd edn. John Wiley & Sons Inc., New York (1999).

[5] P.J. Brockwell and R.A. Davis, Time Series: Theory and Methods. *Springer Series in Statistics*. Springer, New York (2006).

[6] N. H. Chan and C.Z. Wei, Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Stat.* **16** (1988) 367–401.

[7] J. Dedecker and E. Rio. On the functional central limit theorem for stationary processes. *Ann. Inst. Henri Poincaré, B.* **36** (2000) 1–34.

[8] M. Duflo, Random iterative models, in Vol. 34 of Applications of Mathematics. Springer-Verlag, Berlin, New York (1997).

[9] Y. Fan, Testing the goodness of fit of a parametric density function by kernel method. *Econom. Theory* **10** (1994) 316–356.

[10] B.K. Ghosh and W.M. Huang, The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit. *Ann. Stat.* **19** (1991) 999–1009.

[11] L. Horváth and R. Zitikis, Asymptotics of the L_p-norms of density estimators in the first-order autoregressive models. *Stat. Probab. Lett.* **66** (2004) 91–103.

[12] T.L. Lai and C.Z. Wei, Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.* **13** (1983) 1–23.

[13] S. Lee and S. Na, On the Bickel-Rosenblatt test for first-order autoregressive models. *Stat. Probab. Lett.* **56** (2002) 23–35.

[14] S. Lee and C.Z. Wei, On residual empirical processes of stochastic regression models with applications to time series. *Ann. Stat.* **27** (1999) 237–261.

[15] M.H. Neumann and E. Paparoditis, On bootstrapping L_2-type statistics in density testing. *Stat. Probab. Lett.* **50** (2000) 137–147.

[16] E. Parzen, On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** (1962) 1065–1076.

[17] F. Proïa, Further results on the H-Test of Durbin for stable autoregressive processes. *J. Multivariate Anal.* **118** (2013) 77–101.

[18] M. Rosenblatt, Remark on some nonparametric estimates of a density function. *Ann. Math. Stat.* **27** (1956) 832–837.

[19] M. Rosenblatt, A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Stat.* **3** (1975) 1–14.

[20] A.V. Skorokhod, Limit theorems for stochastic processes. *Theory Probab. Appl.* **1** (1956) 261–290.

[21] B.P. Stigum, Asymptotic properties of dynamic stochastic parameter estimates (III). *J. Multivariate Anal.* **4** (1974) 351–381.

[22] H. Takahata and K.I. Yoshihara, Central limit theorems for integrated square error of nonparametric density estimators based on absolutely regular random sequences. *Yokohama Math. J.* **35** (1987) 95–111.

[23] J. Valeinis and A. Locmelis, Bickel-Rosenblatt test for weakly dependent data. *Math. Model. Anal.* **17** (2012) 383–395.