Available online at www.sciencedirect.com

SCIENCE DIRECT®

Statistics & Probability Letters 74 (2005) 221–234

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

A note on the Bickel–Rosenblatt test in autoregressive time series

Dirk Bachmann, Holger Dette*

Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany

Received 10 October 2004; received in revised form 11 February 2005
Available online 26 April 2005

Abstract

In a recent paper Lee and Na [2002. Statist. Probab. Lett. 56(1), 23–25] introduced a test for the parametric form of the distribution of the innovations in autoregressive models, which is based on the integrated squared error of the nonparametric density estimate from the residuals and a smoothed version of the parametric fit of the density. They derived the asymptotic distribution under the null-hypothesis, which is the same as for the classical Bickel–Rosenblatt [1973. Ann. Statist. 1, 1071–1095] test for the distribution of i.i.d. observations. In this note we first extend the results of Bickel and Rosenblatt to the case of fixed alternatives, for which asymptotic normality is still true but with a different rate of convergence. As a by-product we also provide an alternative proof of the Bickel and Rosenblatt result under substantially weaker assumptions on the kernel density estimate. As a further application we derive the asymptotic behaviour of Lee and Na's statistic in autoregressive models under fixed alternatives. The results can be used for the calculation of the probability of the type II error if the Bickel–Rosenblatt test is used to check the parametric form of the error distribution or to test interval hypotheses in this context.

MSC: primary 62M07

Keywords: Autoregressive process; Goodness-of-fit test; Nonparametric density estimation; Asymptotic distribution under fixed alternatives

*Corresponding author. Fax: +49 2 34 32 14 559.
E-mail addresses: dirk.bachmann@ruhr-uni-bochum.de (D. Bachmann), holger.dette@ruhr-uni-bochum.de (H. Dette).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.04.003

1. Introduction

The goodness-of-fit testing problem for the distribution of the innovations is of particular importance in time series analysis. In particular the hypothesis of Gaussian errors is of interest. Under this additional assumption inference simplifies substantially and many statistical procedures in time series are based on the assumption of normality (see e.g. Brockwell and Davis (1991) or Fan and Yao (2003)). In a recent paper Lee and Na (2002) considered the problem of testing the hypothesis

$$H_0 : f = f_0, \quad H_1 : f \neq f_0 \tag{1.1}$$

in the first-order-autoregressive process

$$X_j = \varphi X_{j-1} + Z_j, \tag{1.2}$$

where f_0 is a given density, Z_j are i.i.d. random variables with density f, mean 0 and variance $\sigma^2 > 0$. Their work was motivated by the fact that for the more general hypothesis of a location-scale family the limit distribution of tests based on functionals of the empirical process of the residuals $\hat{Z}_j = X_j - \hat{\varphi} X_{j-1}$ depends on the parameter estimates involved in the empirical process and is no longer a functional of the standard Brownian bridge (see e.g. Boldin (1982), Koul (1991, 2002), Koul and Levental (1989)). Lee and Na (2002) proposed to use the Bickel–Rosenblatt test based on the residuals $\hat{Z}_1, \ldots, \hat{Z}_n$ for the hypotheses (1.1) and proved asymptotic normality of the corresponding test statistic under the null hypothesis $H_0 : f = f_0$. They also generalized this result to the problem of testing for a location-scale family.

It is the purpose of the present paper to provide a more refined analysis of the Bickel–Rosenblatt test by a discussion of the asymptotic behaviour of the test statistic under fixed alternatives of the form

$$d(f, f_0) = \int (f - f_0)^2(x)\, \mathrm{d}x > 0. \tag{1.3}$$

In Section 2 we show that under the alternative (1.3) a standardized version of the statistic of Bickel and Rosenblatt (1973) based on i.i.d. observations is still asymptotically normal distributed but with a different rate of convergence.

This result allows a simple calculation of the probability of the type II error of the Bickel–Rosenblatt test. It is therefore of particular importance if the null hypothesis cannot be rejected (see Berger and Delampady (1987) or Sellke et al. (2001)). The asymptotic distribution of the test statistic under fixed alternatives can also be used for the calculation of critical values in the problem of testing precise hypotheses of the form

$$H_0 : d(f, f_0) > \pi, \quad H_1 : d(f, f_0) \leqslant \pi. \tag{1.4}$$

Here π is a given bound in which the experimenter would denote deviations from the assumed density f_0 as not relevant. Note that the formulation of the hypotheses (1.4) allows the experimenter to test that the density f is approximately equal to f_0 (i.e. $d(f, f_0) \leqslant \pi$) at a controlled type I error.

In Section 3 we consider the statistic of Lee and Na (2002) under alternative (1.3). We show that it has the same asymptotic behaviour as Bickel and Rosenblatt's statistic in the i.i.d. case which was derived in Section 2. It is also demonstrated that this result holds for composite hypotheses

$$H_0 : f \in \mathscr{F}, \quad H_1 : f \notin \mathscr{F}, \tag{1.5}$$

where

$$\mathscr{F} = \left\{ \frac{1}{\sigma} f_0\left(\frac{\cdot - \mu}{\sigma}\right) \mid \mu \in \mathbb{R}; \sigma > 0 \right\} \tag{1.6}$$

is a location-scale family and f_0 is a given density. In Section 4 we investigate the finite sample properties of a bootstrap version of the proposed test and compare its properties with the Kolmogorov–Smirnov test. Finally, some of the proofs are given in an appendix in Section 5.

2. The test of Bickel and Rosenblatt revisited

Let $Z_1, Z_2, \ldots, Z_n$ denote iid random variables with two times continuously differentiable density f with bounded second derivative and $K : \mathbb{R} \to \mathbb{R}$ be a continuous bounded symmetric kernel with compact support satisfying

$$\int K(x)\,\mathrm{d}x = 1, \quad \int x^2 K(x)\,\mathrm{d}x < \infty, \quad \int K^2(x)\,\mathrm{d}x < \infty. \tag{2.1}$$

We consider the kernel estimator

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - Z_i), \tag{2.2}$$

where $K_h(\cdot) = (1/h)K(\frac{\cdot}{h})$ is the scaled kernel and $h > 0$ denotes a bandwidth satisfying

$$nh^2 \to \infty, \quad h \to 0 \tag{2.3}$$

if $n \to \infty$. For the problem of testing the hypothesis (1.1) Bickel and Rosenblatt (1973) proposed to reject the null-hypothesis for large values of the statistic

$$T_n = \int [f_n - K_h * f_0]^2(x)\,\mathrm{d}x, \tag{2.4}$$

where $f_1 * f_2$ denotes the convolution of the functions f_1 and f_2. Under the null hypothesis $H_0 : f = f_0$ these authors showed asymptotic normality of T_n, namely

$$n\sqrt{h}\left\{ T_n - \frac{1}{nh} \int K^2(t)\,\mathrm{d}t \right\} \xrightarrow{\mathscr{D}} \mathscr{N}(0, \tau^2), \tag{2.5}$$

where

$$\tau^2 = 2 \int f_0^2(x)\,\mathrm{d}x \int (K * K)^2(x)\,\mathrm{d}x. \tag{2.6}$$

The following result now establishes asymptotic normality of an appropriately standardized version of T_n under fixed alternatives.

Theorem 2.1. *If the assumptions* (2.1)–(2.3) *are valid and the alternative* $H_1: f \neq f_0$ *is satisfied in the sense of* (1.3) *we have*

$$\sqrt{n}\left[T_n - \int (K_h * (f - f_0))^2(x)\,\mathrm{d}x\right] \xrightarrow{\mathscr{D}} \mathcal{N}(0, 4\varrho^2), \tag{2.7}$$

where the asymptotic variance is given by

$$\varrho^2 = \mathrm{Var}[(f - f_0)(Z_i)]. \tag{2.8}$$

In the appendix we provide an alternative proof of the statement (2.5) based on a central limit theorem for degenerate U-statistics, which is of its own interest and particularly helpful to identify the limit distribution in the proof of Theorem 2.1. Moreover, with this technique the statement (2.5) can be proved under substantially weaker assumptions than imposed by Bickel and Rosenblatt (1973). These authors derived this result using an approximation of the normalized and centered sample distribution function by an appropriate Brownian process on a convenient probability space.

It is also interesting to note that the centered version of T_n is of different order under the null hypothesis and alternative, namely

$$T_n - E[T_n] \overset{H_0}{\sim} \mathrm{O_p}\left(\frac{1}{n\sqrt{h}}\right),$$
$$T_n - E[T_n] \overset{H_1}{\sim} \mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right). \tag{2.9}$$

A detailed proof of these properties is given in the appendix. For a heuristic explanation note that

$$T_n = \int [f_n - K_h * f]^2(x)\,\mathrm{d}x + \int [K_h * (f - f_0)]^2(x)\,\mathrm{d}x$$
$$- 2\int (f_n - K_h * f)[K_h * (f - f_0)](x)\,\mathrm{d}x$$

The first term is essentially the integrated mean-squared error of a kernel density estimate, which is known to be of order

$$\mathrm{O_p}\left(\frac{1}{n\sqrt{h}}\right)$$

(see e.g. Bickel and Rosenblatt (1973) or Hall (1984)). The second term corresponds approximately to $E[T_n]$ and vanishes under the null hypothesis. Finally, the third term also vanishes under the null hypothesis. However, under the alternative, it is a sum of i.i.d. random variables, which can be shown to be of order $\mathrm{O_p}(1/\sqrt{n})$ (see our proof in the appendix).

In the following we briefly indicate two potential applications of Theorem 2.1. Note that the weak convergence in (2.7) can be used for the calculation of the probability of the type II error of the test, which rejects the null hypothesis $H_0: f = f_0$, whenever

$$n\sqrt{h}\left\{T_n - \frac{1}{nh}\int K^2(t)\,\mathrm{d}t\right\} > \tau u_{1-\alpha}. \tag{2.10}$$

Here $u_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution. A straightforward calculation gives under the alternative (1.3) for the probability of rejection the approximation

$$P(\text{“ rejection” }) \approx \Phi\left(\frac{\sqrt{n}}{2\varrho} d(f,f_0) - \frac{\tau}{2\varrho}\frac{u_{1-\alpha}}{\sqrt{nh}}\right) \approx \Phi\left(\frac{\sqrt{n}}{2\varrho} d(f,f_0)\right).$$

A further application of Theorem 2.1 consists in the calculation of critical values of the test for the precise hypotheses defined in (1.4). Here the null hypothesis is rejected for small values of the statistic T_n, namely

$$\sqrt{n}\frac{T_n - \pi}{2\hat{\varrho}} \leqslant u_\alpha, \tag{2.11}$$

where $\hat{\varrho}$ is an appropriate estimator of the asymptotic variance ϱ in Theorem 2.1. Note that the test of the form (2.11) decides in favour of the alternative $H_1 : d(f,f_0) \leqslant \pi$ at a controlled type I error of size α. In other words if we decide that the “true” density is approximately equal to f_0, the probability of a possible error is approximately α. We finally note that it is important to control this probability because subsequent data analysis will be performed under the assumption $f = f_0$ if the null hypothesis in (1.4) is rejected.

3. A goodness-of-fit test in autoregressive models

Consider the first-order autoregressive model, where we are interested in testing the hypothesis (1.1) for the distribution of the innovations Z_i. Because these values are unobservable, we replace them by the residuals $\hat{Z}_i = X_i - \hat{\varphi}X_{i-1}$, where $\hat{\varphi}$ is a $\sqrt{n}$-consistent estimator of the parameter φ. Let

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - \hat{Z}_i) \tag{3.1}$$

denote the kernel density estimate based on the residuals $\hat{Z}_1, \dots, \hat{Z}_n$ and define the statistic $\hat{T}_n$ as the analogue of T_n, where the random variable f_n defined in (2.2) is replaced by $\hat{f}_n$, i.e.

$$\hat{T}_n = \int [\hat{f}_n - K_h * f_0]^2(x)\,\mathrm{d}x. \tag{3.2}$$

Lee and Na (2002) made the following assumptions:

$$K^{(3)} \text{ exists, } K^{(2)} \text{ is bounded} \tag{3.3}$$

$$\int |K^{(j)}(x)|\,\mathrm{d}x < \infty, \ j = 1,2,3, \quad \int |K^{(j)}(x)|^2\,\mathrm{d}x < \infty, \ j = 1,2 \tag{3.4}$$

$$nh^4 \to \infty. \tag{3.5}$$

They showed that the statistics T_n and $\hat{T}_n$ are asymptotically equivalent, i.e.

$$n\sqrt{h}[\hat{T}_n - T_n] = o_P(1), \tag{3.6}$$

and derived as a consequence the asymptotic normality of $\hat{T}_n$ from the corresponding result of Bickel and Rosenblatt (1973). The following results show that statements of this form remain true under fixed alternatives.

Theorem 3.1. *Assume that* $|\varphi|<1$. *If the assumptions* (2.1)–(2.3), (3.3)–(3.5) *are satisfied and alternative* (1.3) *is valid, then*

$$\sqrt{n}\left[\hat{T}_n - \int (K_h * (f-f_0))^2(x)\,\mathrm{d}x\right] \xrightarrow{\mathscr{D}} N(0,4\varrho^2), \tag{3.7}$$

where ϱ^2 *is given in* (2.8).

Theorem 3.2. *Assume that* $|\varphi|>1$ *and that the assumptions* (2.1)–(2.3) *and* (3.5) *are satisfied. If additionally the kernel* K *in the density estimate* (3.1) *is bounded such that there exists a constant* $B>0$ *with*

$$\int |K(x+\delta)-K(x)|\,\mathrm{d}x \leqslant B\delta \tag{3.8}$$

for all $\delta>0$, *then assertion* (3.7) *holds.*

Remark 3.3. Theorems 3.1 and 3.2 are also valid for testing the composite hypothesis (1.6) of a location-scale family. To be precise consider the first-order autoregressive model

$$X_t = \mu + \rho X_{t-1} + Z_t. \tag{3.9}$$

We are interested in the problem of testing the hypothesis

$$H_0 : M(f,f_0) = 0 \quad H_1 : M(f,f_0) > 0 \tag{3.10}$$

or the corresponding precise hypotheses of the form (1.4). Here

$$M(f,f_0) = \min_{\sigma>0} \int \left(f(x) - \frac{1}{\sigma} f_0\left(\frac{x}{\sigma}\right)\right)^2 \mathrm{d}x \tag{3.11}$$

is the L^2-distance of the best approximation of the density f by elements from the scale family

$$\mathscr{F} = \left\{\frac{1}{\sigma} f_0\left(\frac{\cdot}{\sigma}\right) \mid \sigma>0\right\}.$$

We assume that the minimum in (3.11) exists and is attained at a unique point, say $\sigma_0>0$. Assume that $\hat{\mu}, \hat{\varphi}, \hat{\sigma}$ are $\sqrt{n}$-consistent estimates of μ, φ, σ, respectively and that $\hat{f}_n$ is the density estimate (3.1) from the residuals $\hat{Z}_i = X_i - \hat{\mu} - \hat{\varphi} X_{i-1}$. Lee and Na (2002) showed for the statistic

$$\bar{T}_n = \int \left\{\hat{f}_n(x) - \left(K_h * \frac{1}{\hat{\sigma}} f\left(\frac{\cdot}{\hat{\sigma}}\right)\right)\right\}^2 (x)\,\mathrm{d}x \tag{3.12}$$

the asymptotic normality

$$n\sqrt{h}\left(\bar{T}_n - \frac{1}{nh}\int K^2(x)\,\mathrm{d}x\right) \xrightarrow{\mathscr{D}} \mathscr{N}(0,\tau^2) \tag{3.13}$$

under the null hypothesis (3.10), where τ^2 is defined in (2.6). Combining these arguments with the arguments given for the proof of Theorems 2.1, 3.1 and 3.2 it can be shown that under any fixed alternative $M(f, f_0) > 0$ it follows

$$\sqrt{n}(\bar{T}_n - M(f, f_0)) \xrightarrow{\mathscr{D}} \mathcal{N}(0, 4\bar{\rho}^2), \tag{3.14}$$

where

$$\bar{\rho} = \mathrm{Var}\left(\left\{f(Z_i) - \frac{1}{\sigma_0} f_0\left(\frac{Z_i}{\sigma_0}\right)\right\}^2\right)$$

and σ_0 is the unique minimizer in (3.11). The details are omitted for the sake of brevity.

Remark 3.4. It is well known (see Boldin (1982) or Koul (1991)) that the Kolmogorov–Smirnov test, based on the residual empirical process, is asymptotically distribution free for testing an error distribution with zero mean in the autoregressive model (1.2). However, this fact is neither true in the case $\int x f_0(x)\,\mathrm{d}x \neq 0$ nor in the case where a location scale family has to be tested. On the other hand the Bickel–Rosenblatt test is always asymptotically distribution free. Note also that in the case $\int x f_0(x)\,\mathrm{d}x \neq 0$ the consistent estimation of φ in model (1.2) is not possible.

Moreover, while the Kolmogorov–Smirnov test is more powerful with respect to Pitman alternatives than the proposed test based on density estimation, the opposite may be true for the power with respect to local alternatives of the form

$$k_n(x) = f_0(x) + \alpha_n w((x - c)\gamma_n^{-1}),$$

where $\alpha_n, \gamma_n \to 0$. Here we assume that the function w is two times continuously differentiable and square integrable such that for sufficiently large n the function $k_n(x)$ is nonnegative (note that the condition $\int k_n(x)\,\mathrm{d}x = 1$ implies $\int w(x)\,\mathrm{d}x = 0$, which means that w must have negative values). It can be shown by similar arguments as given in the appendix, that for alternatives of this type with

$$\alpha_n = \frac{1}{n^{1/2} h^{2/3}}, \quad \gamma_n = h^{5/6}$$

the statistic on the left-hand side of (2.5) is also asymptotically normal with variance τ^2 and mean $\int w^2(x)\,\mathrm{d}x$. Because the size of the integral $\int_{-\infty}^{c} \alpha_n w((x - c)\gamma_n^{-1})\,\mathrm{d}x$ is of order $\alpha_n \gamma_n = h^{1/6} n^{-1/2}$ the Bickel–Rosenblatt test has greater power against such local alternatives than tests based on the deviation between the sample and the true distribution function (see Rosenblatt (1975) or Gosh and Huang (1991) for more details).

4. A finite sample comparison

In this section we briefly investigate the finite sample properties of the test based on the L^2-distance between the densities and the Kolmogorov–Smirnov test which compares the distribution functions directly. For this purpose we consider the problem of testing the hypothesis of a scale family

$$H_0 : f \in \mathscr{F} = \left\{\frac{1}{\sigma} f_0\left(\frac{\cdot}{\sigma}\right) \mid \sigma > 0\right\} \tag{3.15}$$

for the error distribution in the first-order autoregressive model (1.2). The test statistic for the L^2-distance is given by (3.12), where $\hat{f}_n$ is the density estimate from the residuals $\hat{Z}_t = X_t - \hat{\varphi}X_{t-1}$, $\hat{\varphi}$ and $\hat{\sigma}^2$ are the Yule–Walker estimates of the parameter φ and the variance of the innovations, respectively. For the Kolmogorov–Smirnov test we used the statistic

$$U_n = \sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F_0\left(\frac{x}{\hat{\sigma}}\right)\right|, \tag{3.16}$$

where $\hat{F}_n(x) = (1/n)\sum_{t=0}^{n} I\{\hat{Z}_t \leqslant x\}$ is the empirical distribution function of the residuals. The implementation of the statistic (3.12) requires the specification of a bandwidth, and we used

$$h = \left(\frac{\hat{\sigma}^2}{n}\right)^{1/5} \tag{3.17}$$

for this purpose. We consider the problem of testing the distribution of the innovations for a centered normal and a double exponential distribution with unknown variance. Because the Kolmogorov–Smirnov test is not asymptotically distribution free in this case and the approximation of the distribution of the standardized statistic $\bar{T}_n$ in (3.12) by the normal distribution is not too accurate, we implemented a bootstrap version of both tests. For this we adapted a resampling scheme which was recently proposed by Neumann and Kreiss (1998) in the more general context of first-order nonparametric autoregressive models. To be precise, we determined the Yule–Walker estimates $\hat{\varphi}$ and $\hat{\sigma}^2$ of the parameters in the model (1.2) and generated bootstrap observations as follows:

$$Y_i^* = \hat{\varphi}X_{i-1} + \hat{\sigma}\varepsilon_i^*, \quad i = 1, \ldots, n, \tag{3.18}$$

where the ε_i^* are i.i.d. random variables with a standardized distribution corresponding to the null hypothesis (that is a standard normal distribution, if we are testing for normality or a double exponential distribution with mean 0 and variance 1 if we are testing for the double exponential distribution). The statistics $\bar{T}_n$ defined in (3.12) and U_n defined in (3.16) are now calculated for the bootstrap sample and denoted by $\bar{T}_n^*$ and U_n^*, respectively.

If $\bar{T}_{n(1)}^*, \ldots, \bar{T}_{n(B)}^*$ (or $U_{n(1)}^*, \ldots, U_{n(B)}^*$) denote the order statistics obtained from B bootstrap replications the null hypothesis of a scale family with density f_0 is rejected if

$$\bar{T}_n > \bar{T}^*_{n\lfloor B(1-\alpha)\rfloor} \quad (U_n > U^*_{n\lfloor B(1-\alpha)\rfloor}), \tag{3.19}$$

where the level α is 2.5%, 5% and 10%. In Table 1 we show the rejection probabilities of the two tests for the null hypothesis of a centered normal distribution and the alternatives

$$\begin{aligned} &\mathcal{N}(0.5, 1), \\ &\mathcal{N}(1, 1), \\ &(\chi_k^2 - k)/\sqrt{2k}, \quad k = 1, 2, 3, \end{aligned} \tag{3.20}$$

where the symbol χ_k^2 denotes a χ^2 distribution with k degress of freedom. The sample sizes are $n = 25, 50$ and we used $B = 200$ bootstrap replications and 1000 simulation runs for the calculation of the rejection probabilities. Data was generated according to the first-order autoregressive model (1.2) with $\varphi = 0.1$ and a standard normal distribution for the errors. We observe a reasonable approximation of the nominal level under the null hypothesis for both tests

Table 1
Simulated rejection probabilities of the Kolmogorov–Smirnov test based on the statistic U_n in (3.16) and the test based on the statistic $\bar{T}_n$ defined in (3.12) in the problem of testing for a scale family

	$n=25$						$n=50$					
	$\bar{T}_n$			U_n			$\bar{T}_n$			U_n		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
$\mathcal{N}(0,1)$	0.027	0.048	0.107	0.030	0.049	0.105	0.032	0.057	0.109	0.043	0.063	0.106
$\mathcal{N}(0.5,1)$	0.180	0.216	0.301	0.308	0.386	0.556	0.434	0.533	0.650	0.681	0.767	0.880
$\mathcal{N}(1,1)$	0.291	0.316	0.426	0.543	0.640	0.817	0.496	0.584	0.672	0.939	0.962	0.987
df_1	0.748	0.805	0.872	0.476	0.594	0.740	0.986	0.992	0.996	0.913	0.955	0.981
df_2	0.445	0.525	0.633	0.255	0.319	0.449	0.803	0.834	0.902	0.530	0.660	0.798
df_3	0.311	0.365	0.448	0.171	0.226	0.318	0.576	0.643	0.738	0.404	0.491	0.677

Six error distributions are considered: a standard normal distribution (corresponding to the null hypothesis), a $\mathcal{N}(0.5,1)$, a $\mathcal{N}(1,1)$ distribution and a standardized χ^2_k-distribution with mean 0 and variance 1 denoted by the symbol df_k $(k=1,2,3)$.

Table 2
Simulated rejection probabilities of the Kolmogorov–Smirnov test based on the statistic U_n in (3.16) and the test based on the statistic $\bar{T}_n$ defined in (3.12) in the problem of testing for a scale family

	$n=25$						$n=50$					
	$\bar{T}_n$			U_n			$\bar{T}_n$			U_n		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
$DE(0)$	0.021	0.042	0.089	0.033	0.053	0.105	0.029	0.047	0.089	0.034	0.052	0.115
$DE(0.5)$	0.236	0.303	0.439	0.374	0.466	0.609	0.599	0.688	0.791	0.878	0.932	0.967
$DE(1)$	0.392	0.463	0.589	0.652	0.733	0.863	0.663	0.750	0.888	0.972	0.991	1.000
df_1	0.761	0.824	0.886	0.537	0.619	0.750	0.993	0.998	1.000	0.928	0.965	0.990
df_2	0.567	0.651	0.766	0.263	0.318	0.426	0.908	0.932	0.966	0.592	0.702	0.864
df_3	0.435	0.511	0.640	0.176	0.220	0.304	0.821	0.863	0.922	0.367	0.452	0.648

Six error distributions are considered: a double exponential distribution with variance 1 (corresponding to the null hypothesis), a $DE(0.5)$, a $DE(1)$ distribution (where the symbol $DE(\mu)$ denotes a double exponential distribution with mean μ and variance 1) and a standardized χ^2_k-distribution with mean 0 and variance 1 denoted by the symbol df_k $(k=1,2,3)$.

(see the first row in Table 1). For shift alternatives the Kolmogorov–Smirnov test is substantially more powerful than the Bickel–Rosenblatt test (see the second and third line of Table 1). On the other hand if χ^2 distributions appear as alternatives, the test based on the L^2-distance of the two densities yields remarkably larger power (see row 4–6 of Table 1).

In Table 2 we show the corresponding results for the problem of testing the hypothesis (3.15) for a double exponential distribution with density

$$f_0(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right), \quad x \in \mathbb{R}$$

(note that $\int xf_0(x)\,\mathrm{d}x = 0$, $\int x^2 f_0(x)\,\mathrm{d}x = 1$ for this choice). The alternatives are a double exponential distribution with variance 1 and mean 0.5, 1 and the standardized χ^2 distributions specified by (3.20). We observe a similar performance of the two tests as described for the problem of testing for a normal distribution. The level is approximated with reasonable accuracy under the null hypothesis (see the first row in Table 2). Under shift alternatives the Kolmogorov–Smirnov test outperforms the test based on the L^2-distance of the densities (see row 2,3 of Table 2), while the opposite behaviour can be observed, if χ^2 distributions are considered as alternatives (see row 4–6 of Table 2).

Acknowledgements

The support of the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt B1) is gratefully acknowledged. Parts of the paper were written during a visit of the second author at the Institute of Statistics in Louvain-la-Neuve and this author would like to thank the institute for its hospitality. The authors are also grateful to two unknown referees for their constructive comments on an earlier version of this paper and to Isolde Gottschlich, who typed parts of this paper with considerable technical expertise.

Appendix. Proofs

Proof of (2.5) and Theorem 2.1. Note that we will establish asymptotic normality under the null hypothesis $f = f_0$ <u>and</u> under fixed alternatives $f \neq f_0$ with different rates of convergence in both cases. The weak convergence (2.5) under the null hypothesis has already been established by Bickel and Rosenblatt (1973), but the argument presented here is more direct and requires weaker assumptions.

Let f denote the "true" density of the random variables Z_i. Recalling the definition of the statistic T_n and the density estimate f_n we obtain the following decomposition:

$$\begin{aligned}
T_n &= \int [f_n - K_h * f_0]^2(x)\,\mathrm{d}x \\
&= \int [f_n - K_h * f]^2(x)\,\mathrm{d}x + 2\int [f_n - K_h * f](x) g_h(x)\,\mathrm{d}x + \int g_h^2(x)\,\mathrm{d}x \\
&= \frac{2}{n^2}\sum_{i<j} \int [K_h(x - Z_i) - e_h(x)][K_h(x - Z_j) - e_h(x)]\,\mathrm{d}x \\
&\quad + \frac{2}{n}\sum_{i=1}^{n} [(K_h * g_h)(Z_i) - E[(K_h * g_h)(Z_i)]] + \frac{1}{n^2}\sum_{i=1}^{n} \int [K_h(x - Z_i) - e_h(x)]^2\,\mathrm{d}x \\
&\quad + \int g_h^2(x)\,\mathrm{d}x, \qquad \text{(A.1)}
\end{aligned}$$

where the functions e_h and g_h are defined by $e_h := K_h * f$ and $g_h := K_h * (f - f_0)$, respectively. A straightforward calculation shows

$$\frac{1}{n^2}\sum_{i=1}^{n}\int [K_h(x - Z_i) - e_h(x)]^2\,\mathrm{d}x = \frac{1}{nh}\int K^2(x)\,\mathrm{d}x + \mathrm{O_P}\left(\frac{1}{n}\right). \tag{A.2}$$

Consequently we obtain the stochastic expansion

$$T_n - \frac{1}{nh}\int K^2(x)\,\mathrm{d}x - \int [K_h * (f - f_0)]^2(x)\,\mathrm{d}x = \frac{2}{n^2}\sum_{i<j} H_n(Z_i, Z_j) + \frac{2}{n}\sum_{i=1}^{n} Y_i + \mathrm{O_P}\left(\frac{1}{n}\right), \tag{A.3}$$

where the random variables $H_n(Z_i, Z_j)$ and Y_i are defined by

$$H_n(Z_i, Z_j) = \int [K_h(x - Z_i) - e_h(x)][K_h(x - Z_j) - e_h(x)]\,\mathrm{d}x, \tag{A.4}$$

$$Y_i = (K_h * g_h)(Z_i) - E[K_h * g_h(Z_i)], \tag{A.5}$$

respectively. Define the first term in this decomposition as

$$U_n = \frac{2}{n^2}\sum_{i<j} H_n(Z_i, Z_j) \tag{A.6}$$

and note that U_n does not depend on the density f_0 specified by the null hypothesis. As a consequence any asymptotic property of U_n holds independently if the null hypothesis is satisfied or not.

In the following we will establish the weak convergence of this statistic. For this we apply a central limit theorem for degenerate U-statistics proved by Hall (1984) (see Theorem 2.1 in this reference). Obviously, H_n is symmetric, $E[H_n(Z_1, Z_2) \mid Z_1] = 0$, and $E[H_n^2(Z_1, Z_2)] < \infty$ for each $n \in \mathbb{N}$. Moreover, a straightforward but tedious calculation shows

$$\lim_{n\to\infty} \mathrm{Var}\left(\sqrt{h}H_n(Z_i, Z_j)\right) = \lim_{n\to\infty} E[hH_n^2(Z_i, Z_j)] = \int (K * K)^2(x)\,\mathrm{d}x \int f_0^2(x)\,\mathrm{d}x. \tag{A.7}$$

This gives for the variance of $n\sqrt{h}U_n$

$$\begin{aligned}\mathrm{Var}\left(n\sqrt{h}U_n\right) &= E\left[\frac{4h}{n^2}\sum_{\substack{i<j\\ i'<j'}} H_n(Z_i, Z_j)H_n(Z_{i'}, Z_{j'})\right]\\ &= E\left[2h\frac{n-1}{n}H_n^2(Z_i, Z_j)\right] = \tau^2 + \mathrm{o}(1),\end{aligned}$$

where τ^2 is defined in (2.6). The final condition (2.1) of Hall's (1984) Theorem 2.1 is more difficult to check. First note that it follows from (A.7) that $E[H_n^2(Z_i, Z_j)] = \mathrm{O}(\frac{1}{h})$. A similar calculation gives

$$E[H_n^4(Z_i, Z_j)] = \frac{1}{h^3}\int f_0^2(x)\,\mathrm{d}x \int (K*K)^4(x)\,\mathrm{d}x + \mathrm{O}\left(\frac{1}{h^2}\right) = \mathrm{O}\left(\frac{1}{h^3}\right). \tag{A.8}$$

Finally, we have to consider the quantity

$$G_n(Z_1, Z_2) = E[H_n(Z_1, Z_3)H_n(Z_3, Z_2) \mid Z_1, Z_2],$$

and obtain

$$\begin{aligned} E[G_n^2(Z_i, Z_j)] &= \frac{1}{h^2}E\left[\left\{\int (K*K)(w)(K*K)\left(w - \frac{Z_i - Z_j}{h}\right)f(Z_i)\,\mathrm{d}w\right\}^2\right] + \mathrm{O}\left(\frac{1}{h}\right) \\ &= \frac{1}{h}\int\int \{(K*K)*(K*K)\}^2(s)\,\mathrm{d}s f^4(v)\,\mathrm{d}v + \mathrm{O}\left(\frac{1}{h}\right) \\ &= \mathrm{O}\left(\frac{1}{h}\right). \end{aligned}$$

This gives

$$\frac{E[G_n^2(Z_i, Z_j)] + \frac{1}{n}E[H_n^4(Z_i, Z_j)]}{(E[H_n^2(Z_i, Z_j)])^2} = \mathrm{O}\left(h + \frac{1}{nh}\right) = \mathrm{o}(1),$$

and establishes condition (2.1) of Hall's (1984) Theorem 2.1. We therefore obtain the weak convergence

$$n\sqrt{h}U_n \xrightarrow{\mathscr{D}} \mathscr{N}(0, \tau^2). \tag{A.9}$$

We are now in a position to prove the assertions (2.5) and (2.7) of Section 2.

Proof of (2.5). Weak convergence under the null hypothesis $f = f_0$.

Under the null hypothesis $H_0 : f = f_0$ we have $Y_i \equiv 0$ and obtain from (A.3) the stochastic expansion

$$T_n - \frac{1}{nh}\int K^2(x)\,\mathrm{d}x = \frac{2}{n^2}\sum_{i<j} H_n(Z_i, Z_j) + \mathrm{O_P}\left(\frac{1}{n}\right) = U_n + \mathrm{O_P}\left(\frac{1}{n}\right), \tag{A.10}$$

where the statistic U_n is defined in (A.6). The asymptotic normality of the statistic

$$n\sqrt{h}\left\{T_n - \frac{1}{nh}\int K^2(x)\,\mathrm{d}x\right\}$$

now follows from the corresponding statement for the random variable $n\sqrt{h}U_n$ in (A.9).

Proof of (2.7). Weak convergence under a fixed alternative $f \neq f_0$.

For a proof of asymptotic normality of the statistic T_n under a fixed alternative satisfying (1.3) we note that it follows from (A.9) that

$$U_n = \frac{2}{n^2}\sum_{i<j} H_n(Z_i, Z_j) = \mathrm{O_p}\left(\frac{1}{n\sqrt{h}}\right).$$

From (2.3) and (A.3) we obtain

$$T_n - \int [K_h * (f - f_0)]^2(x)\,\mathrm{d}x = \frac{2}{n}\sum_{i=1}^{n} Y_i + \mathrm{o_p}\left(\frac{1}{\sqrt{n}}\right),$$

where the random variables Y_i are defined by (A.5). A straightforward but tedious calculation shows

$$\mathrm{Var}(Y_i) = \mathrm{Var}((f - f_0)(Z_i)) + \mathrm{O}(h^2) = \varrho^2 + \mathrm{O}(h^2).$$

Consequently we have

$$\mathrm{Var}\left(\frac{2}{\sqrt{n}}\sum_{i=1}^{n} Y_i\right) = 4\varrho^2 + \mathrm{o}(1),$$

while

$$E[Y_i^4] = \mathrm{O}(1),$$

uniformly with respect to $i = 1, \ldots, n$. The asymptotic normality in Theorem 2.1 now follows from Lindeberg–Feller's theorem, which completes the proof of this Theorem. □

Proof of Theorems 3.1 and 3.2. We only consider the case $|\varphi| < 1$, the proof of Theorem 3.2 can be obtained by similar arguments. Obviously, the assertion follows from the estimate

$$\sqrt{n}(\hat{T}_n - T_n) = \mathrm{o_p}(1) \tag{A.11}$$

For a proof of this estimate we will proceed as in Lee and Na (2002) who obtained the estimate

$$\int (\hat{f}_n - f_n)^2(x)\,\mathrm{d}x = \mathrm{O_P}(n^{-2}h^{-4}). \tag{A.12}$$

On the other hand Theorem 2.1 shows that under a fixed alternative

$$\int (f_n - K_h * f_0)^2(x)\,\mathrm{d}x = \mathrm{O_P}(1), \tag{A.13}$$

and a straightforward calculation [using condition (3.5)] gives

$$\begin{aligned} |\hat{T}_n - T_n| &\leqslant \int (\hat{f}_n - f_n)^2(x)\,\mathrm{d}x + 2\left[\int (\hat{f}_n - f_n)^2(x)\,\mathrm{d}x\right]^{1/2}\left[\int (f_n - K_h * f_0)^2(x)\,\mathrm{d}x\right]^{1/2} \\ &= \mathrm{O_P}\left(\frac{1}{nh^2}\right) = \mathrm{o_P}\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

which proves the assertion of Theorem 3.1. □

References

Berger, J.O., Delampady, M., 1987. Testing precise hypotheses. With comments and a rejoinder by the authors. Stat. Sci. 2 (3), 317–352.

Bickel, P.J., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Ann. Statist. 1, 1071–1095.

Boldin, M.V., 1982. Estimation of the distribution of noise in an autoregression scheme. Theory Probab. Appl. 27, 866–871.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, second ed., Springer Series in Statistics. Springer, New York.

Fan, J., Yao, Q., 2003. Nonlinear Time Series. Springer Series in Statistics. Springer, New York, Berlin, Heidelberg.

Gosh, B.K., Huang, W.-M., 1991. The power and optimal kernel of the Bickel–Rosenblatt test for goodness of fit. Ann. Statist. 19, 999–1009.

Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. J. Multivariate Anal. 14, 1–16.

Koul, H.L., 1991. A weak convergence result useful in robust autoregression. J. Statist. Plann. Inference 29, 291–308.

Koul, H.L., 2002. Weighted empirical processes in dynamic nonlinear models. Lecture Notes in Statistics, vol. 166, Springer, New York.

Koul, H.L., Levental, S., 1989. Weak convergence of the residual empirical process in explosive autoregression. Ann. Statist. 17 (4), 1784–1794.

Lee, S., Na, S., 2002. On the Bickel–Rosenblatt test for first-order autoregressive models. Statist. Probab. Lett. 56 (1), 23–35.

Neumann, M.H., Kreiss, J.-P., 1998. Regression type inference in nonparametric autoregression. Ann. Statist. 26, 1570–1613.

Rosenblatt, M., 1975. A quadratic measure of deviation of two dimensional density estimate and a test of independence. Ann. Statist. 3, 1–14.

Sellke, T., Bayarri, M.J., Berger, J.O., 2001. Calibration of p values for testing precise null hypotheses. Amer. Statist. 55 (1), 62–71.