Biometrika (2015), **xx**, x, *pp*. 1–10

Printed in Great Britain

Supplementary material for: Entropy testing for nonlinear serial dependence in time series.

BY SIMONE GIANNERINI

Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, 40126, Bologna, Italy.

simone.giannerini@unibo.it

ESFANDIAR MAASOUMI

Department of Economics, Emory University, 1602 Fishburne Dr, Economics, Atlanta, Georgia, USA.

esfandiar.maasoumi@emory.edu

AND ESTELA BEE DAGUM

Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, 40126, Bologna, Italy.

estela.beedagum@unibo.it

1. THE SEVEN AXIOMS OF RÉNYI

Let $H(\mathrm{X})$ be a measure of stochastic dependence between the random variables $\mathrm{X} = X_1, \ldots, X_n$. The seven axioms of Rényi (Rényi, 1959) define the desirable properties that an ideal measure of dependence should possess:

1. $H(\mathrm{X})$ is defined for any random vector $\mathrm{X} = (X_1, \ldots, X_n)$, being X_i a non-degenerate r.v. for all $i = 1, \ldots, n$.
2. For any permutation $(i_1, \ldots, i_n)$ of the indices $\{1, \ldots, n\}$, we have $H(\mathrm{X}) = H(X_1, \ldots, X_n) = H(X_{i1}, \ldots, X_{in})$.
3. $0 < H(\mathrm{X}) < a$, with $a \in \mathbb{R} \cup +\infty$
4. $H(\mathrm{X}) = 0$ if and only if the random variables $X_1, \ldots, X_n$ are independent.
5. $H(\mathrm{X}) = a$ if and only if there exists an exact measurable relationship between $X_1, \ldots, X_n$.
6. $H(\mathrm{X})$ is invariant w.r.t continuous and strictly increasing transformations.
7. $H(\mathrm{X})$ is in relation with the correlation coefficient in the Gaussian case.

In addition to these, Maasoumi (1993) defines three further properties for $H(\mathrm{X})$:

1. It is well defined for both continuous and discrete variables;
2. It is normalized to zero in case of independence and lies in $[0, 1]$;
3. It is a proper measure of distance;

As shown in Maasoumi (1993); Granger et al. (2004) the measure

$$S_\rho(k) = \frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\{f_{(X_t,X_{t+k})}(x_1,x_2)\}^{1/2} - \{f_{X_t}(x_1)f_{X_{t+k}}(x_2)\}^{1/2}\right]^2 dx_1 dx_2 \quad (1)$$

satisfies all the above mentioned properties. We have the following

PROPOSITION 1. *Let $(X_t, X_{t+k}) \sim N(0,1,\rho)$ be a standard Normal random vector with joint probability density function given by $f_{(X_t,X_{t+k})}(x_1,x_2,\rho)$, where ρ is the correlation coefficient at lag k. Then,*

$$S_\rho(k) = 1 - \frac{2\left(1-\rho^2\right)^{1/4}}{(4-\rho^2)^{1/2}} \quad (2)$$

Proof. Given a bivariate standard Gaussian density function: $f_{(X_t,X_{t+k})}(x_1,x_2)$ with marginal densities $f_{X_t}(x_1)$ and $f_{X_{t+k}}(x_2)$, we have:

$$\begin{aligned} S_\rho &= 1 - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{f_{X_t}(x_1)\, f_{X_{t+k}}(x_2)\, f_{(X_t,X_{t+k})}(x_1,x_2)\}^{1/2}\, dx_1\, dx_2 \\ &= 1 - \frac{1}{4\pi^2(1-\rho^2)^{1/2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x_1^2+x_2^2)(2-\rho^2)-2\rho x_1 x_2}{4(1-\rho^2)}\right\}^{1/2} dx_1\, dx_2\ . \end{aligned}$$

Now we use the polar coordinate transformation $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$ and obtain

$$\begin{aligned} S_\rho &= 1 - \frac{1}{2\pi(1-\rho^2)^{1/4}}\int_0^{2\pi}\int_0^{\infty}\exp\left\{-\frac{r^2\left(2-\rho^2-\rho\sin 2\theta\right)}{4\left(1-\rho^2\right)}\right\} d\theta\, dr \\ &= 1 - \frac{1}{\pi}\int_0^{2\pi}\frac{(1-\rho^2)^{\frac{3}{4}}}{(\rho^2-2+\rho\sin 2\theta)}d\theta = 1 - \frac{2\left(1-\rho^2\right)^{3/4}}{(4-5\rho^2+\rho^4)^{1/2}} = 1 - \frac{2\left(1-\rho^2\right)^{1/4}}{(4-\rho^2)^{1/2}}\ . \end{aligned}$$

In Figure 1 we depict the relationship we have found between S_ρ and the correlation coefficient ρ in the Gaussian case. The inspection of the plot reveals a remarkable steepness when ρ is close to $|1|$. In other words, S_ρ might be particularly sensitive for detecting and distinguishing between high levels of dependence. This property has proven useful for characterizing particular classes of processes.

2. ALGORITHM FOR GENERATING SURROGATE DATA THROUGH SIMULATED ANNEALING

The simulated annealing is a useful tool for combinatorial minimization of functions with false minima (see e.g. Vidal, 1993). The cost function C is interpreted as an energy in a thermodynamic system. Minimizing C is then equivalent to finding the ground state of a system. The annealing is a procedure used to bring a glassy solid close to the optimal state by first heating it and then cooling it. The simulation of this tempering procedure exploits the fact that in thermodynamic equilibrium at some finite temperature T, the possible configurations of the system are visited with a probability according to the Boltzmann distribution of the canonical ensemble $p = \exp\{-C/T\}$. In the simulation, this is attained by accepting changes of the configuration with a probability $p = 1$ if the energy is decreased ($\Delta C < 0$) and $p = \exp\{-\Delta C/T\}$ if the energy is increased ($\Delta C \geqslant 0$). This rule is often referred to as the Metropolis step. In the minimization problem, the temperature is the parameter in the Boltzmann distribution that determines the probability of accepting the unfavorable changes that are needed to get out of false minima.

Let $\mathrm{x} = (x_1,\ldots,x_n)^T$ be an observed series of length n being $\hat{\rho}_k$ its sample autocorrelation function at lag k and let x^* be the candidate surrogate with autocorrelation function at lag k

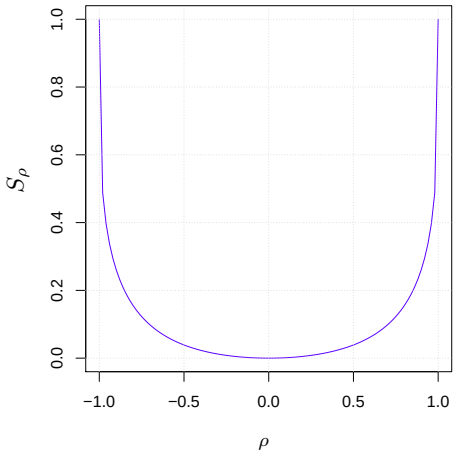

Fig. 1. Relationship between S_ρ and the correlation coefficient ρ in the Gaussian case, see Eq. (2).

denoted by $\hat{\rho}_k^*$. The cost function we have adopted here is

$$C(\mathrm{x}, \mathrm{x}^*) = \max_{k=1}^{3k_{\max}} |1.05\hat{\rho}_k - \hat{\rho}_k^*|. \tag{3}$$

Other cost functions based on a discrepancy between two correlation functions can be employed. First, denote with x^* a random permutation of the original series x; also, start with a temperature T. For each temperature value T, the algorithm is the following:

1. swap two observations of x^* and obtain the series $\mathrm{x}^{*(s)}$;
2. compute $\Delta C = \left\{C(\mathrm{x}, \mathrm{x}^{*(s)}) - C(\mathrm{x}, \mathrm{x}^*)\right\}$;
3. if $\Delta C < 0$ accept the swap, that is, $\mathrm{x}^* = \mathrm{x}^{*(s)}$
   if $\Delta C \geqslant 0$ accept the swap with probability $p = \exp\{-\Delta C/T\}$;
4. repeat step (1) – (3) until either the number of accepted swaps or the number of trials reach specified thresholds;
5. lower the temperature T, for instance by setting $T = \alpha T$ where $\alpha < 1$;
6. repeat the whole procedure until the cost function reaches a specified threshold *eps*;

In general, the choice of the parameters for the algorithm is problem-specific and a certain amount of experimentation and tuning is expected in order to obtain good results. In our instance, we have found parameters' settings that can be employed almost automatically as follows:

parameter	value	description
T	0.001	initial temperature
R_T	0.9	reduction factor for T
eps	0.02	threshold
nsuccmax	30	T is decreased after *nsuccmax*$\times n$ successes
nmax	300	T is decreased after *nmax*$\times n$ trials
che	1e+05	after *che*$\times 2n$ global iterations the algorithm starts again

The accuracy of the results is a user defined parameter, in our case we have set 0.02 as the threshold for the cost function C, namely, for the first 15 lags, the maximum discrepancy between the ACF of the original series and that of the surrogates does not exceed 0.02. While this requirement appears reasonable for our purposes, more stringent values can be chosen if needed

at the cost of longer computing times. Clearly, the great flexibility of this method is that different hypotheses can be tested simply by changing the cost function. However, the computational cost is a problem that has to be taken into account. For a more detailed discussion on different cooling schemes, cost functions and computational issues related to the algorithm see Schreiber & Schmitz (2000). In Schreiber & Schmitz (1996) it is shown that the constrained randomization approach leads to tests with better power and size as compared to those derived through the phase randomization.

3. THE SMOOTHED SIEVE BOOTSTRAP ALGORITHM

The main idea of the sieve bootstrap relies upon the Wold decomposition of a stationary stochastic process. A stationary stochastic process X_t can be decomposed into the sum of two uncorrelated processes: $X_t = Z_t + V_t$ where Z_t is the MA(∞) component:

$$Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \ , \ \psi_0 = 1 \quad \text{with} \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty \ , \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2). \tag{4}$$

V_t is a deterministic component $V_t = \mu + \sum_{j=1}^{\infty} \{\alpha_j \sin(\omega_j t) + \beta_j \cos(\omega_j t)\}$. If $\{V_t\} = 0$ (purely non-deterministic process) X_t admits an infinite moving average representation. This fact was at the basis of the extension of classical residual based bootstrap methods to linear processes (see e.g. Berkowitz & Kilian, 2000; Politis, 2003). Such methods can be adapted as follows in order to obtain surrogates:

1. Given a time series $(x_1, \ldots, x_n)$ fit an AR(p) model
$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_t$$
and obtain the estimates $\{\hat{a}_{1,n}(p), \ldots, \hat{a}_{p,n}(p)\}$
2. Obtain the residuals.
$$\hat{\varepsilon}_{t,n} = x_t - \sum_{i=1}^{p} \hat{a}_{i,n}(p) x_{t-i} \qquad t = p+1, \ldots, n$$
3. Compute a kernel density estimate for $f_\varepsilon(\cdot)$:
$$\hat{f}_\varepsilon(x) = \frac{1}{n-p} h^{-1} \sum_{t=p+1}^{n} K\left(\frac{x - \hat{\varepsilon}_{t,n}}{h}\right)$$
where $h = h(n)$ is a bandwidth such that $h(n) \to 0$ and $h(n)^{-1} = o(n)$
4. Resample the residuals from the kernel density estimate
$$e_t^* \sim \text{i.i.d.} \hat{f}_\varepsilon(x)dx$$
denote the sample obtained with $(e^*_{-Q}, \ldots, e^*_{-n})$, $Q \in \mathbb{N}$.
5. Obtain the bootstrapped time series $\mathrm{X}^* = (x_1^*, \ldots, x_n^*)$ according to:
$$x_t^* = \sum_{i=1}^{p} \hat{a}_{i,n}(p) x_{t-i}^* + e_t^* \quad t = -Q, \ldots, n$$
where the initial values are $x^*_{-Q-1} = \cdots = x^*_{-Q-p} = 0$.

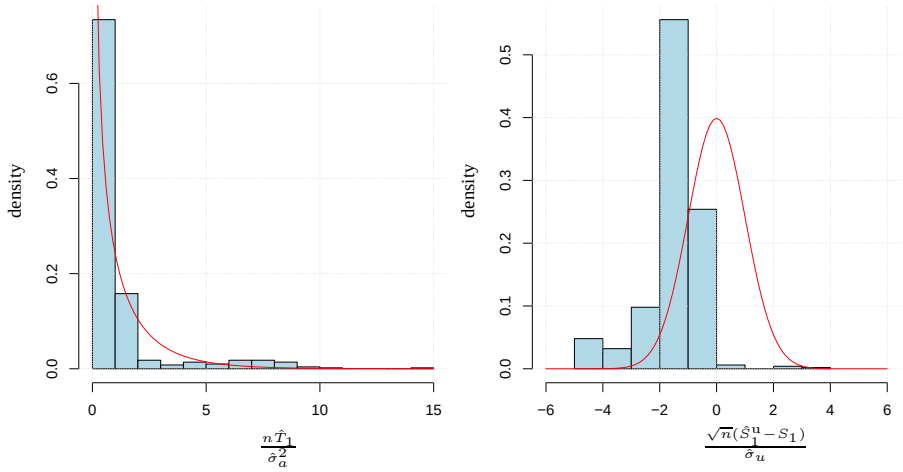

Fig. 2. Left: histogram of $n\hat{T}_1/\hat{\sigma}_a^2$ for 500 realizations of a moving average process of order 1. The length of the series is $n = 2000$. The solid red line is a χ_1^2 density. Right: the same but for $\sqrt{n}(\hat{S}_1^{\mathrm{u}} - S_1)/\hat{\sigma}_u$. The solid red line is a standard normal density.

6. Repeat steps (3) and (4) B times.

The method we have chosen for obtaining the transient set is not the only possible solution. In our instance we do not condition every surrogate realization to the same initial values so that we avoid problems related to such a choice, see also Berkowitz & Kilian (2000) for a discussion. One of the main issues related to this method is the choice of the order of the fitted process p. In practice a finite order must be chosen, nevertheless, several authors have stressed that if we require $p = p(n) \to \infty$ as $n \to \infty$ but $p(n)^k/n \to 0$, where k is a small integer, then such method becomes non parametric in that the procedure is valid for the whole class of infinite autoregressive processes. In particular, in Bühlmann (1997) it is shown that if the AIC criterion for order selection is used then consistency is achieved for the arithmetic mean and a class of nonlinear statistics. Moreover, the method adapts automatically to the decay of the dependence structure of the process. Finally, the author remarks that the performance of the method is quite insensitive to choice of the criterion used for model selection as long as the order chosen is reasonable.

4. Asymptotic approximation

In this section we show that the asymptotic χ_1^2 approximation of the test based on $\hat{T}_k$ is not accurate even for long series. The test statistic is $n\hat{T}_k/\sigma_a^2$, where σ_a^2 is the asymptotic variance of $(\hat{T}_k)^{1/2}$. We generate 500 series with $n = 2000$ from the moving average process $X_t = 0.8\,\varepsilon_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim N(0,1)$ for all t. Recall that $\rho_1 = -0.488$ and $\sigma_a^2 = \sigma_p^2 + \sigma_u^2$. Now, σ_p^2 is the asymptotic variance of the parametric estimator $\hat{S}_1^{\mathrm{p}}$ and can be computed exactly for such process from the equation given in Proposition 2:

$$\sigma_p^2 = \left\{ \frac{\rho_1(2+\rho_1^2)}{(1-\rho_1^2)^{3/4}(4-\rho_1^2)^{3/2}} \right\}^2 (1+2\rho_1^2) = 0.04966.$$

As for σ_u^2, we derive an estimate for it through the Monte Carlo variance of $\hat{S}_1^{\mathrm{u}}$. The histogram of $n\hat{T}_1/\hat{\sigma}_a^2$ is shown in Figure 2(left) where we have superimposed a χ_1^2 density (solid red line).

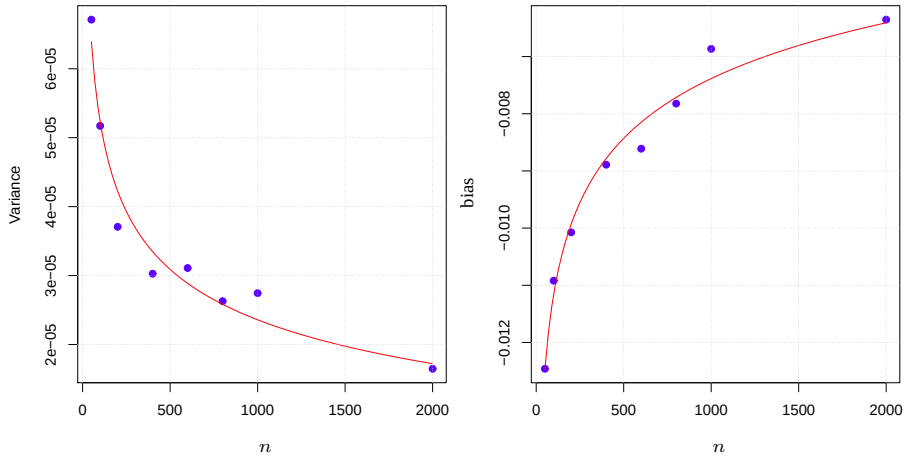

Fig. 3. Left: plot of the Monte Carlo variance $\hat{\sigma}_u^2$ as a function of the sample size n. The solid red line indicates a fit of the order $O(n^{-1/5})$. Right: the same but for the Monte Carlo bias for $\hat{S}_1^{\mathrm{u}}$. The solid red line indicates a fit of the order $O(n^{-1/10})$.

We are in the unrealistic scenario in which we know exactly the asymptotic variance of the parametric estimator and have a very good estimate for that of the non-parametric estimator; still, the approximation of the right tail of the distribution of the test statistic is not satisfactory. In fact, if we set the nominal level $\alpha = 0.05$ and use the χ^2_1 approximation we get an empirical size of 0.09. Moreover, the Kolmogorov-Smirnov test of equality of distributions rejects the null with a p-value $< 2.2e-16$. Similar considerations can be made for the asymptotic approximation of the unrestricted estimator $\hat{S}_k^{\mathrm{u}}$. In such case, by exploiting Eq. 2, the null of a linear Gaussian moving average process reduces to $H_0 : S_1 = 0.0366$ so that under H_0 we have:

$$\frac{\sqrt{n}(\hat{S}_1^{\mathrm{u}} - S_1)}{\sigma_u} \to N(0,1) \text{ in distribution}$$

As before, we use the Monte Carlo estimate for σ_u^2 and show in Figure 2(right) the histogram of $\sqrt{n}(\hat{S}_1^{\mathrm{u}} - S_1)/\hat{\sigma}_u$ for 500 realizations of the moving average process considered above. In such case, the bias of the nonparametric estimator is still present even for $n = 2000$. In Figure 3(left) we show the plot of the Monte Carlo variance $\hat{\sigma}_u^2$ as a function of the sample size n. The solid red line shows a superimposed fit of the order $O(n^{-1/5})$. Figure 3(right) shows the plot of the Monte Carlo bias for $\hat{S}_1^{\mathrm{u}}$ as a function of the sample size n. The solid red line shows a superimposed fit of the order $O(n^{-1/10})$. The two plots suggest that the rate of convergence for the bias is lower than that of the variance and this might be responsible of the poor approximation of the asymptotic test.

5. SIMULATION STUDY: FURTHER RESULTS

In this section we report further results of the simulation study. In Table 1 we show the empirical rejection percentages at nominal level $\alpha = 0.1$. The performance of the tests is consistent with the case $\alpha = 0.05$ and is an indication of the robustness of the results.

In Table 2 we show the empirical rejection percentages corrected for multiple testing through the Benjamini-Hochberg procedure at level $q = 0.1$. q is the control rate, namely, following the empirical Bayes interpretation, it is an estimate of the Bayes probability that a rejected null hy-

Table 1. *Rejection percentages at the nominal level $\alpha = 10\%$. The first three columns refer to $\hat{T}_k$ and the null H_0 of a linear Gaussian process; the last three columns refer to $\hat{S}_k^{\mathrm{u}}$ and the null H_0' of a generic linear process.*

		$\hat{T}_k$				$\hat{S}_k^{\mathrm{u}}$		
	n	50	100	200	n	50	100	200
	Model 1	75.0	23.0	15.8	Model 1	24.8	15.2	4.2
	Model 2	70.6	17.6	9.4	Model 2	9.4	4.0	3.6
	Model 3	60.3	27.4	13.6	Model 3	7.2	8.4	10.6
linear	Model 4	58.5	28.0	10.2	Model 4	8.4	7.6	7.2
Gaussian	Model 5	77.4	20.0	11.2	Model 5	13.8	6.6	6.4
	Model 6	88.4	23.6	13.2	Model 6	4.2	3.2	3.6
	Model 7	86.5	43.2	60.8	Model 7	37.0	23.2	15.2
	Model 8	85.4	44.3	66.8	Model 8	14.0	6.8	7.2
linear	Model 9	83.1	35.3	51.2	Model 9	21.4	23.8	28.2
non-Gaussian	Model 10	79.8	46.9	64.8	Model 10	14.2	15.8	21.8
	Model 11	87.2	37.7	52.6	Model 11	22.8	21.0	19.2
	Model 12	82.8	40.9	48.4	Model 12	10.2	9.8	10.8
	Model 13	75.1	69.4	89.6	Model 13	53.8	75.0	93.6
	Model 14	80.6	83.4	97.2	Model 14	52.6	77.0	95.2
	Model 15	82.7	89.6	98.4	Model 15	71.4	93.8	99.4
	Model 16	88.3	82.8	98.8	Model 16	69.0	92.6	99.2
	Model 17	83.6	47.0	64.4	Model 17	37.0	51.0	76.6
nonlinear	Model 18	85.2	78.0	93.0	Model 18	60.8	81.4	96.0
	Model 19	79.1	90.8	99.4	Model 19	49.8	74.2	97.0
	Model 20	88.7	76.2	88.0	Model 20	51.2	71.6	89.6
	Model 21	91.4	76.4	98.2	Model 21	52.6	81.6	98.4
	Model 22	99.3	100.0	100.0	Model 22	98.2	100.0	99.8
	Model 23	91.6	99.8	100.0	Model 23	91.4	100.0	100.0
	Model 24	99.0	100.0	100.0	Model 24	100.0	100.0	100.0

pothesis is actually correct. For more details see Benjamini & Hochberg (1995); Efron (2010). In contrast to the above performance of the metric entropy Granger et al. (2004), among many others, have shown the failure of correlation based methods to detect dependencies that are identified above for these same models and a number of others. In these other contexts, no other nonlinear method has shown more power than the metric entropy method, though some have matched its performance only in the case of some models. This suggests that the use of the metric entropy is a risk averse approach.

As concerns the bandwidth selection criterion, the results suggest that the reference criterion should be paired with the bootstrap scheme, while the cross validation criterion has to be preferred when using the surrogate based test. In particular, the reference bandwidth leads to a severe oversize when used with surrogate data whereas the bootstrap approach has low power when paired with the likelihood cross validation. This latter result may be due to the residual based nature of the sieve bootstrap. We do not generally recommend the centered residual based bootstrap technique in conjunction with cross validation. Recall that in this method, one is centering the residuals around their means, and the likelihood cross validation is seemingly doing a better job of detecting the removal of dependence structure in the centered residuals. In the nonparametric estimation approach, cross validation is well capable of detecting irrelevant regressors; see Li & Racine (2007). The re-centering bootstrap is also a good method for removing estimation uncer-

Table 2. *Rejection percentages of the BH(q) procedure at $q = 10\%$. The first three columns refer to $\hat{T}_k$ and the null H_0 of a linear Gaussian process; the last three columns refer to $\hat{S}_k^{\mathrm{u}}$ and the null H_0' of a generic linear process.*

		$\hat{T}_k$				$\hat{S}_k^{\mathrm{u}}$		
	n	50	100	200	n	50	100	200
	Model 1	72.5	10.6	7.4	Model 1	5.2	1.2	0.0
	Model 2	67.5	8.2	4.1	Model 2	0.8	0.2	0.2
	Model 3	50.2	14.6	6.5	Model 3	0.0	0.8	1.0
linear	Model 4	54.4	12.0	6.9	Model 4	0.0	0.4	0.2
Gaussian	Model 5	73.2	10.4	3.8	Model 5	1.8	0.8	0.8
	Model 6	78.2	9.6	3.6	Model 6	0.6	0.0	0.2
	Model 7	83.6	33.0	50.1	Model 7	13.4	6.0	2.6
	Model 8	80.7	39.9	55.2	Model 8	3.0	0.6	1.8
linear	Model 9	80.0	30.2	41.5	Model 9	4.8	4.2	7.0
non-Gaussian	Model 10	77.2	39.8	57.1	Model 10	2.2	3.8	3.8
	Model 11	83.1	31.3	46.2	Model 11	5.2	5.8	3.2
	Model 12	78.0	32.1	44.6	Model 12	2.4	1.2	2.2
	Model 13	65.1	52.2	80.6	Model 13	25.0	47.0	79.8
	Model 14	78.1	67.2	93.8	Model 14	21.2	42.0	78.0
	Model 15	77.2	79.6	95.2	Model 15	33.0	77.0	97.8
	Model 16	83.6	68.2	95.8	Model 16	31.6	71.4	96.4
	Model 17	80.2	29.8	47.2	Model 17	8.0	21.8	49.8
nonlinear	Model 18	82.3	59.2	84.0	Model 18	32.2	56.8	85.0
	Model 19	76.4	78.6	97.0	Model 19	11.4	29.0	77.6
	Model 20	83.1	64.0	76.6	Model 20	16.4	30.2	51.8
	Model 21	87.2	62.4	93.8	Model 21	16.2	54.6	89.8
	Model 22	97.7	100.0	100.0	Model 22	48.0	97.4	99.6
	Model 23	88.8	98.6	100.0	Model 23	58.6	88.2	100.0
	Model 24	93.7	100.0	100.0	Model 24	100.0	100.0	100.0

tainty, perhaps due to removal of omitted variables. In our context, omitted variables are indeed the nonlinear terms! Hence the lack of power to detect them with cross validation.

As with the traditional correlogram, the results of the tests can be depicted in an appealing graphical fashion. In Figs. 4–6 we show the plots for some of the processes studied. For instance, in Figure 4 we show with the black solid line the value of $\hat{T}_k$, $k = 1, \ldots, 5$ computed over a realization of Model 1, Gaussian autoregressive process, left, and Model 15, nonlinear moving average process, right. The green/light gray and the blue/dark gray dashed lines represent the quantiles of the distribution of the surrogates at the 95% and 99% levels, respectively. In some instances the tests manage to identify correctly the lags at which nonlinear dependence is known to be present for different models, see e.g. the nonlinear dependence at lag 2 of Model 15, see Figure 4, right. In general, the identification of the lags at which a significant nonlinear dependence is expected is not straightforward. In our experience, such exercise is informative for processes with a moving-average-like structure.

6. COMPUTATIONAL COMPLEXITY

The computational complexity of the algorithms depend upon several factors such as the bandwidth selection method, the bootstrap scheme, the number of bootstrap replications (B), the number of lags (k_{max}), the sample size (n). It is possible to show that the computational complexity of

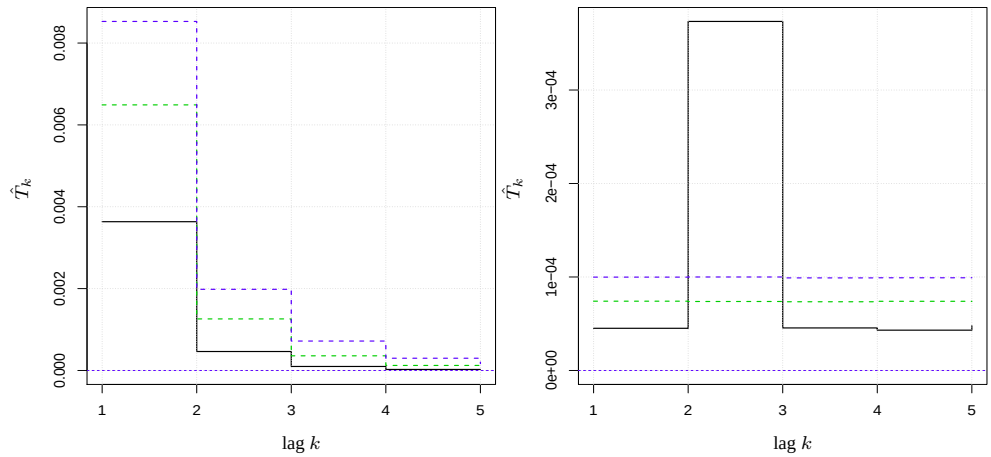

Fig. 4. left: mean $\hat{T}_k$, $k = 1, \ldots, 5$ for Model 1 (Gaussian autoregressive process, black solid line). The rejection bands at 95% and 99% level (green/light gray and blue/dark gray dashed line, respectively) are obtained from the quantiles of the surrogates distribution. right the same but for Model 15 (nonlinear moving average).

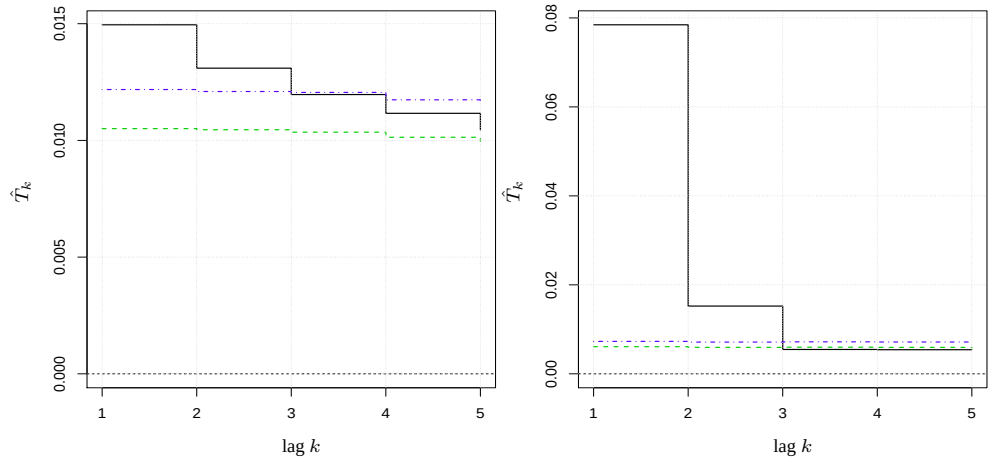

Fig. 5. The same as Figure. 4 but for a GARCH(1,1) process (left, Model 17) and for the chaotic logistic map (right, Model 24).

the tests based on surrogate data coupled with the MLCVcriterion is of the order of $O(k_{max}Bn^2)$ whereas that of bootstrap tests with the normal reference method has order $O(k_{max}Bn)$. This is shown in Figure 7 where we plot the time (in minutes) spent for the computation of the tests as a function of n for the two methods; the impact of k_{max} and B is linear so that we have fixed them.

REFERENCES

BENJAMINI, Y. & HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, pp. 289–300.

BERKOWITZ, J. & KILIAN, L. (2000). Recent developments in bootstrapping time series. *Economet. Rev.* **19**, 1–48.

BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3**, 123–148.

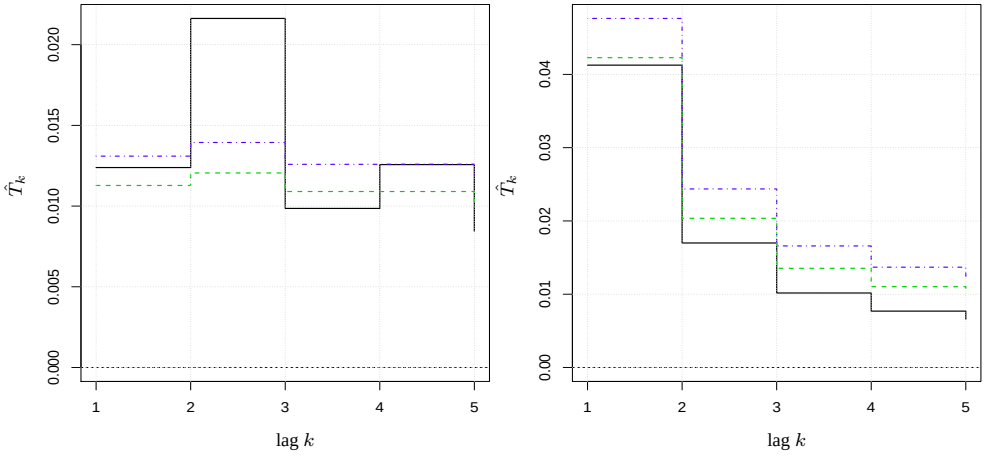

Fig. 6. The same as Figure 4 but for a Bilinear process (left, Model 13) and for the SETAR model (right, Model 19).

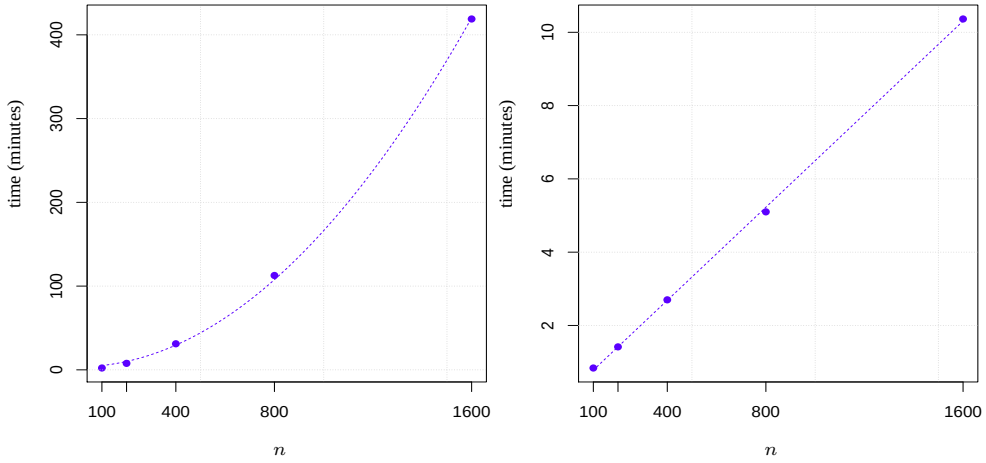

Fig. 7. Computational complexity: time (in minutes) versus length of the series n for the surrogate scheme + MLCVcriterion (left) with quadratic fit (dashed line) and bootstrap scheme + reference criterion (right) with linear fit (dashed line).

EFRON, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge: Cambridge University Press.

GRANGER, C. W. J., MAASOUMI, E. & RACINE, J. (2004). A dependence metric for possibly nonlinear processes. *J. Time Ser. Anal.* **25**, 649–669.

LI, Q. & RACINE, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.

MAASOUMI, E. (1993). A compendium to information theory in economics and econometrics. *Economet. Rev.* **12**, 137–181.

POLITIS, D. (2003). The impact of bootstrap methods on time series analysis. *Statist. Sci.* **18**, 219–230. Silver anniversary of the bootstrap.

RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10**, 441–451.

SCHREIBER, T. & SCHMITZ, A. (1996). Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.* **77**, 635–638.

SCHREIBER, T. & SCHMITZ, A. (2000). Surrogate time series. *Physica D* **142**, 346–382.

VIDAL, R. V. V. (1993). Applied simulated annealing. vol. 396 of *Lect. Notes in Econom. and Math. Syst.* Berlin: Springer Verlag.