

Checking Serial Independence of Residuals from a Nonlinear Model

Luca Bagnato and Antonio Punzo

Abstract In this paper the serial independence tests SIS (Serial Independence Simultaneous) and SICS (Serial Independence Chi-Square) are considered. These tests, proposed by the same authors, are here contextualized in the model validation phase for nonlinear models in which the underlying assumption of serial independence is tested on the estimated residuals. Simulations are used to explore the performance of the tests, in terms of size and power, once a linear/nonlinear model is fitted on the raw data. Results underline that both the tests are powerful against various types of alternatives.

Key words: Model validation, Nonlinear time series, Serial independence, χ^2 -test.

1 Introduction

In modeling observed time series the most popular approach consists in adopting the class of linear models. In this wide class, the well-known *white noise* process – characterized only by the properties of its first two moments – represents the building block and reflects information that is not directly observable. Nevertheless, the limitations of the linear models appear already in the classical paper [4]. Nowadays we know that there are nonstandard features, commonly referred as *nonlinear features* (see [3]) that, by definition, can not be captured in the linear frame. In attempt to overcome this problem, nonlinear models have been widely studied and developed

Luca Bagnato
Università di Milano-Bicocca, Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali;
e-mail: luca.bagnato@unimib.it

Antonio Punzo
Università di Catania, Dipartimento di Impresa, Culture e Società;
e-mail: antonio.punzo@unict.it

(for a discussion on the chief objectives which warrant the introduction of nonlinear models see, for example, [7]). For this kind of models, as indicated in [3], a white noise process is no longer a pertinent building block, as we have to look for measures beyond the second moments to characterize the nonlinear dependence structure. Thus, the white noise has to be replaced by a noise process composed by independent and identically distributed (*i.i.d.*) variables.

As well as for serial uncorrelatedness tests in the linear case, tests for validating serial independence are fundamental in the non linear approach. In detail, testing serial independence is often both a preliminary step carried out before modeling the data generating process (*model selection*) and a useful statistical tool, once a nonlinear model is fitted to the observed data, to check serial independence of the estimated residuals (*model validation*). This paper focuses on the second type of problem taking into account two tests proposed in [1] in the context of model selection. These tests are based on the well-known χ^2 -test and, consequently, are valid asymptotically. Natural fields of application, where a large number of observations is typically available, are: environment, finance, quality control, and so on.

The paper can be so schematically summarized. In Section 2 the problem of checking serial independence is contextualized in the model validation context, for a general nonlinear model, and the tests proposed in [1] are briefly summarized. The performance of these tests in this new context, in terms of size and power, is faced in Section 3 by simulations on some popular models. To find out more, in terms of graphical detecting of possible lag(s)-dependences, a simulation using the Serial Dependence Plot (SDP) presented in [1] is also illustrated. Finally, in Section 4, some conclusions are given.

2 Serial independence tests applied to residuals

Let $\{X_t\}$ be a stationary and ergodic time series. Let $(X_t, \dots, X_{t-n+1})$ be a single realization, of length n , of the process at the time t . In order to describe the process, suppose to use the nonlinear (parametric/nonparametric) model

$$X_t = f(\xi_t) + \sigma(\xi_t)\varepsilon_t, \quad (1)$$

where $\xi_t = (X_{t-1}, \dots, X_{t-d}, \varepsilon_{t-1}, \dots, \varepsilon_{t-v})$, $f(\cdot)$ and $\sigma(\cdot)$ are known/unknown functions of d past X_t 's and v past ε_t 's, with $d < n$ and $v < n$, and where $\{\varepsilon_t\}$ is a serial independent process. Model (1) can be viewed, in the nonparametric context, as a generalization of NARCH model adopted in [3] where $f(\cdot)$ and $\sigma(\cdot)$ depend only on the lagged variables $(X_{t-1}, \dots, X_{t-d})$. Naturally a simpler model can be used; here model (1) is considered to highlight the great flexibility in nonlinear model validation provided by the tests proposed in [1].

Now, let $p = \max(d, v)$. Once estimated model (1), we can obtain the residual vector $(\hat{\varepsilon}_t, \dots, \hat{\varepsilon}_{t-n+p+1})$, of length $n - p$, in which the elements are obtained from the simple relations

$$\begin{aligned}\widehat{\varepsilon}_t &= \frac{X_t - \hat{f}(\boldsymbol{\xi}_t)}{\hat{\sigma}(\boldsymbol{\xi}_t)} \\ &\vdots \\ \widehat{\varepsilon}_{t-n+p+1} &= \frac{X_{t-n+p+1} - \hat{f}(\boldsymbol{\xi}_{t-n+p+1})}{\hat{\sigma}(\boldsymbol{\xi}_{t-n+p+1})}.\end{aligned}$$

Naturally, $(\widehat{\varepsilon}_t, \dots, \widehat{\varepsilon}_{t-n+p+1})$ can be seen as realization of $\{\varepsilon_t\}$.

Now, suppose we are interested in testing the underlying assumption of serial independence of $\{\varepsilon_t\}$ in model (1). In order to do this, we will follow the philosophy presented in [1] to check serial independence in model selection phase, conveniently re-adapted in this new context of model validation. Thus, choose an integer l , with $0 < l < n - p$, and consider Table 1 where, for each lag r , $r = 1, \dots, l$, the estimated residuals up to time $t - r$ are contained in the respective column. Moreover, consider

ε_t	Lagged residuals				
	ε_{t-1}	$\cdots$	ε_{t-r}	$\cdots$	ε_{t-l}
$\widehat{\varepsilon}_{t-n+p+1}$	$\widehat{\varepsilon}_{t-n+p+1}$				
$\widehat{\varepsilon}_{t-n+p+2}$					
$\vdots$					
$\widehat{\varepsilon}_{t-n+p+r+1}$	$\widehat{\varepsilon}_{t-n+p+r}$	$\cdots$	$\widehat{\varepsilon}_{t-n+p+1}$	$\cdots$	
$\vdots$	$\vdots$		$\vdots$		
$\widehat{\varepsilon}_{t-n+p+l+1}$	$\widehat{\varepsilon}_{t-n+p+l}$	$\cdots$	$\widehat{\varepsilon}_{t-n+p+l+1-r}$	$\cdots$	$\widehat{\varepsilon}_{t-n+p+1}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
$\widehat{\varepsilon}_{t-1}$	$\widehat{\varepsilon}_{t-2}$	$\cdots$	$\widehat{\varepsilon}_{t-r-1}$	$\cdots$	$\widehat{\varepsilon}_{t-l-1}$
$\widehat{\varepsilon}_t$	$\widehat{\varepsilon}_{t-1}$	$\cdots$	$\widehat{\varepsilon}_{t-r}$	$\cdots$	$\widehat{\varepsilon}_{t-l}$

Table 1: Lagged residuals ε_{t-r} , $r = 1, \dots, l$, $l < n - p$

$L = \{1, \dots, l\}$. Thus, the null hypothesis of interest can be expressed as

$$H_0 : \varepsilon_t \perp\!\!\!\perp \varepsilon_{t-r} \quad \text{for all } r \in L, \quad (2)$$

versus the alternative

$$H_1 : \varepsilon_t \not\perp\!\!\!\perp \varepsilon_{t-r} \quad \text{for some } r \in L. \quad (3)$$

Obviously, increasing l , the number of possible comparisons increases too. Before going on, for comprehension sake, consider $r = 1$; accepting $\varepsilon_t \perp\!\!\!\perp \varepsilon_{t-1}$, one implicitly accepts also that $\varepsilon_{t-1} \perp\!\!\!\perp \varepsilon_{t-2}$, $\varepsilon_{t-2} \perp\!\!\!\perp \varepsilon_{t-3}$, and so on. This issue can be easily comprehended considering adjacent columns in Table 1.

Thus, the “overall” hypotheses in (2) and (3) can be easily re-expressed as follows

$$H_0 : \bigcap_{r=1}^l H_0^{(r)} \quad \text{versus} \quad H_1 : \bigcup_{r=1}^l H_1^{(r)}, \quad (4)$$

where

$$H_0^{(r)} : \varepsilon_t \perp\!\!\!\perp \varepsilon_{t-r} \quad (5)$$

is the component null hypothesis concerning a sort of independence of lag r , while $H_1^{(r)}$ is the negation of (5), $r = 1, \dots, l$. The superscript “ (r) ” underlines that we are taking into account the comparison of lag r , that is, the residual variables ε_t and ε_{t-r} . In Table 1 the elements considered for this comparison are highlighted. In order to have the largest number of observations for each comparison, we have chosen to fix ε_t that, from a practical point of view, it is equivalent to fix the first column in Table 1.

In order to test $H_0^{(r)}$, the first step is to create the bivariate joint distribution as shown in Table 2, where k is the number of classes for each variable and $a_0^{(r)}, a_1^{(r)}, \dots, a_k^{(r)}$, $a = c, d$, are cutoff points defining *equipfrequency classes*.

ε_t	ε_{t-r}	$d_0^{(r)} \dashv d_1^{(r)} \dots d_{j-1}^{(r)} \dashv d_j^{(r)} \dots d_{k-1}^{(r)} \dashv d_k^{(r)}$
$c_0^{(r)} \dashv c_1^{(r)}$	$n_{11}^{(r)} \dots n_{1j}^{(r)} \dots n_{1k}^{(r)}$	
$\vdots$	$\vdots$	
$c_{i-1}^{(r)} \dashv c_i^{(r)}$	$n_{i1}^{(r)} \dots n_{ij}^{(r)} \dots n_{ik}^{(r)}$	
$\vdots$	$\vdots$	
$c_{k-1}^{(r)} \dashv c_k^{(r)}$	$n_{k1}^{(r)} \dots n_{kj}^{(r)} \dots n_{kk}^{(r)}$	
		$n^{(r)}$

Table 2: Bivariate frequencies table related to $H_0^{(r)}$.

Extreme cutoff points are defined as

$$\begin{aligned} c_0^{(r)} &= \min \{ \widehat{\varepsilon}_t, \dots, \widehat{\varepsilon}_{t-n+p+r+1} \} & c_k^{(r)} &= \max \{ \widehat{\varepsilon}_t, \dots, \widehat{\varepsilon}_{t-n+p+r+1} \} \\ d_0^{(r)} &= \min \{ \widehat{\varepsilon}_{t-r}, \dots, \widehat{\varepsilon}_{t-n+p+1} \} & d_k^{(r)} &= \max \{ \widehat{\varepsilon}_{t-r}, \dots, \widehat{\varepsilon}_{t-n+p+1} \}, \end{aligned}$$

which are easily interpretable as highlighted in Table 1.

Now, a simple and well-known χ^2 -test of independence can be adopted to test the null hypothesis (5) with reference to Table 2. In particular we have denoted the test statistic for $H_0^{(r)}$ by

$$T_r = \sum_{i=1}^k \sum_{j=1}^k \frac{[n_{ij}^{(r)} - \hat{n}_{ij}^{(r)}]^2}{\hat{n}_{ij}^{(r)}}, \quad (6)$$

where $n_{ij}^{(r)}$ are the observed frequencies and $\hat{n}_{ij}^{(r)}$ are the theoretical frequencies under the null hypothesis, $i, j = 1, \dots, k$. It is well known that under $H_0^{(r)}$ the test statistic (6) is asymptotically distributed as a χ^2 with $(k-1)^2$ degrees of freedom.

How it is easy to note from Table 1, when we compare ε_t with ε_{t-r} , we have $n^{(r)} = n - p - r$ observations in Table 2. Since we are building equifrequency classes, k is the only parameter to be chosen in order to construct Table 2. It may be selected so that the dependence structure between the two variables in Table 2 is maintained. Obviously, in choosing k , the number of observations has to be also taken into account. Indeed, in order to make valid the above asymptotic result, the well-known rule of thumb $\hat{n}_{ij}^{(r)} \geq 5$ for each $i, j = 1, \dots, k$ and $r = 1, \dots, l$, has to hold. This means that the number of observations, in comparison of lag r , has to satisfy the inequality $n^{(r)} \geq 5k^2$. Consequently, the number of classes has to be chosen such that $k \leq \sqrt[2]{n^{(r)}/5}$. As highlighted by simulations in [1], $k = 5$ reveals to be a good choice when $n \geq 125 + p + l$.

Recalling that, as shown in [1], test statistics T_u and T_v are independent for any $u \neq v$, it is possible to summarize the proposed tests in the following way:

SIS test: the first, named in [1] as Serial Independence Simultaneous (SIS) test, takes advantage of the union-intersection (UI) principle introduced in [5]. In detail, the overall null hypothesis H_0 is accepted against the alternative if, and only if, all $H_0^{(r)}$, $r = 1, \dots, l$, are simultaneously accepted. In other words, taking p -values into account, the null hypothesis is accepted if, and only if, the component test with the smallest p -value is accepted.

Exploiting the independence between test statistics, the significance levels at which each individual test of independence of lag r has to be performed – in order to preserve the significance level α (chosen in advance) for the overall null hypothesis in (4) – are fixed at $1 - (1 - \alpha)^{1/l}$ for any $r = 1, \dots, l$ (see [6]);

SICS test: the second, named in [1] as Serial Independence Chi-Square (SICS) test, use the (positive) test statistic

$$T = \sum_{r=1}^l T_r = \sum_{r=1}^l \sum_{i=1}^k \sum_{j=1}^k \frac{[n_{ij}^{(r)} - \hat{n}_{ij}^{(r)}]^2}{\hat{n}_{ij}^{(r)}}, \quad (7)$$

assuming small values under H_0 . Taking advantage of the independence between test statistics, the asymptotic distribution of T results to be a χ^2 with $l(k-1)^2$ degrees of freedom.

3 Simulation study

In order to evaluate the performance of both SIS and SICS tests in the model validation context, their empirical size and power have been analyzed by a simulation study performed using the R environment.

The models used to generate the raw series are of type: nonlinear moving average (4)-(6), linear autoregressive (7), nonlinear autoregressive (8), sign autoregressive (9), bilinear (10), ARCH(1) (11), GARCH(1,1) (12), threshold autoregressive (13),

logistic map (14), Hénon map (15) and Hénon map with dynamic noise (16). Details on the adopted models, used also in [2], are contained in the second column of Table 3.

Model	Specification	Test	raw data				AR(1)				ARMA(1,1)-GARCH(1,1)			
			Lag l				Lag l				Lag l			
			2	3	4	5	2	3	4	5	2	3	4	5
(4)	$X_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$	SIS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		SICS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(5)	$X_t = \varepsilon_t + 0.6\varepsilon_{t-1}^2 + 0.6\varepsilon_{t-2}^2$	SIS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		SICS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(6)	$X_t = \varepsilon_t + 0.8\varepsilon_{t-1}\varepsilon_{t-2}$	SIS	0.566	0.513	0.471	0.445	0.563	0.492	0.464	0.421	0.204	0.178	0.167	0.151
		SICS	0.676	0.575	0.502	0.456	0.650	0.563	0.498	0.456	0.214	0.201	0.160	0.151
(7)	$X_t = 0.3X_{t-1} + \varepsilon_t$	SIS	1.000	1.000	1.000	1.000	0.061	0.047	0.046	0.054	0.032	0.037	0.038	0.039
		SICS	1.000	1.000	1.000	0.998	0.045	0.050	0.044	0.046	0.046	0.047	0.050	0.049
(8)	$X_t = 0.8 X_{t-1} ^{0.5} + \varepsilon_t$	SIS	1.000	1.000	1.000	1.000	0.702	0.650	0.615	0.586	0.665	0.621	0.584	0.561
		SICS	1.000	0.998	0.994	0.993	0.648	0.532	0.452	0.394	0.593	0.491	0.423	0.378
(9)	$X_t = \text{sign}(X_{t-1}) + \varepsilon_t$	SIS	1.000	1.000	1.000	1.000	0.196	0.169	0.149	0.139	0.091	0.073	0.065	0.062
		SICS	1.000	1.000	1.000	1.000	0.237	0.218	0.214	0.195	0.089	0.082	0.076	0.081
(10)	$X_t = 0.6\varepsilon_{t-1}X_{t-2} + \varepsilon_t$	SIS	0.784	0.751	0.726	0.702	0.768	0.736	0.712	0.686	0.267	0.259	0.228	0.213
		SICS	0.824	0.774	0.725	0.669	0.807	0.755	0.690	0.623	0.248	0.253	0.222	0.211
(11)	$X_t = \sqrt{h_t}\varepsilon_t, \quad h_t = 1 + 0.4X_{t-1}^2$	SIS	0.885	0.817	0.799	0.781	0.842	0.804	0.784	0.759	0.048	0.045	0.044	0.037
		SICS	0.814	0.731	0.659	0.617	0.797	0.709	0.642	0.585	0.034	0.037	0.035	0.042
(12)	$X_t = \sqrt{h_t}\varepsilon_t, \quad h_t = 0.01 + 0.8h_{t-1} + 0.15X_{t-1}^2$	SIS	0.583	0.608	0.612	0.616	0.574	0.594	0.598	0.613	0.027	0.041	0.042	0.048
		SICS	0.670	0.734	0.766	0.789	0.652	0.717	0.751	0.775	0.037	0.045	0.043	0.040
(13)	$X_t = [-0.5 + 0.9I_{[0, \infty)}(X_{t-1})]X_{t-1} + \varepsilon_t$	SIS	0.994	0.993	0.991	0.987	0.982	0.971	0.964	0.961	0.978	0.972	0.965	0.964
		SICS	0.986	0.957	0.939	0.919	0.957	0.912	0.881	0.839	0.958	0.900	0.853	0.807
(14)	$X_t = 4X_{t-1}(1 - X_{t-1}), \quad 0 < X_t < 1$	SIS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		SICS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(15)	$X_t = 1 + 0.3X_{t-2} - 1.4X_{t-1}^2$	SIS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		SICS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(16)	$X_t = Z_t + \sigma\varepsilon_t, \quad Z_t = 1 + 0.3Z_{t-2} - 1.4Z_{t-1}^2$	SIS	0.989	0.993	0.992	0.993	0.974	0.985	0.984	0.985	0.174	0.244	0.226	0.201
		SICS	0.991	0.997	0.995	0.995	0.975	0.992	0.986	0.985	0.180	0.299	0.300	0.293

Table 3: Simulated rejection rates, of SIS and SICS tests, for raw data (see [1]) and for estimated residuals of an AR(1) and an ARMA(1,1)-GARCH(1,1); nominal size $\alpha = 0.05$ and $l = 2, \dots, 5$. In the models specification $\{\varepsilon_t\}$ represents a sequence of independent standard normals. Sample size $n = 1000$ and number of simulations equal to 1000 for each independently realized time series from each process.

Note that we have chosen to adopt the same models, and the same model numbering, as in [1] in order to make comparisons easier. Details on simulation setting can be easily obtained from the caption of Table 3. The study has also the objective to evaluate the performance of the tests at the varying of l between 2 and 5.

Table 3 shows rejection rates, of both tests, applied to raw data and estimated residuals of an AR(1) and of what we denote as ARMA(1,1)-GARCH(1,1). With the term ARMA(1,1)-GARCH(1,1) we mean that a GARCH(1,1) is fitted on the residuals arising from an ARMA(1,1) estimated on the raw data. As regards the residuals from the AR(1), the observed size of both tests results to be close to the nominal size for model (7), that is an AR(1). In the other cases, the power of both tests drops compared to the same tests based on raw data, which indicates that some of the dependence structure is captured by the AR(1) model. As regards the residuals arising from the ARMA(1,1)-GARCH(1,1), the observed size of both tests results to be close to the nominal size above all for model (7), (11), and (12). In fact, these can be considered as a particular case of an ARMA(1,1)-GARCH(1,1). In the other cases, the power of both tests drops compared to the same tests based on estimated residuals from the AR(1), which indicates that further dependence

structure is captured by the ARMA(1,1)-GARCH(1,1) model with respect to the AR(1).

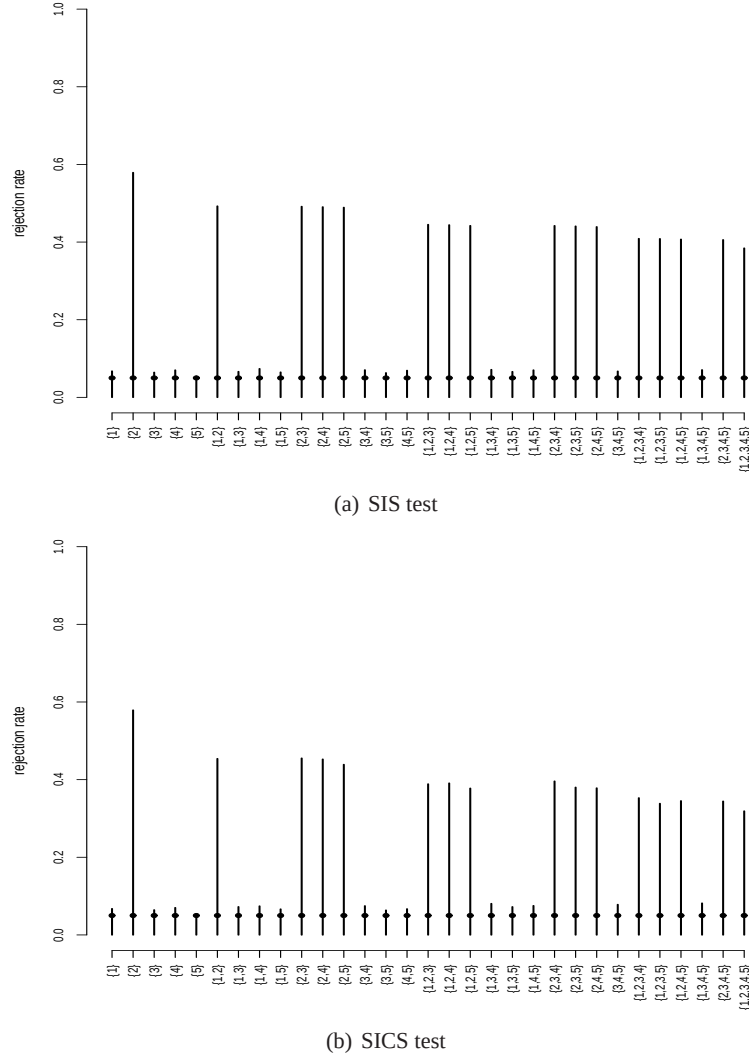


Fig. 1: Rejection rates, at the nominal level $\alpha = 0.05$, constructed on 10000 residual series obtained fitting an ARCH(1) model on series, of length $n = 1000$, generated from an ARCH(2).

Whenever the null hypothesis of serial independence is rejected on the estimated residuals, it could be useful to discover the causes in order to improve the adopted model. With this aim, it could be useful to take into account the Serial Dependence

Plot (SDP) presented in [1]. To ascertain this suggestion, we have chosen to generate 10000 series, of length $n = 1000$, from an ARCH(2) of equation $X_t = \sqrt{h_t}\varepsilon_t$, $h_t = 1 + 0.4X_{t-1}^2 + 0.3X_{t-2}^2$. On the generated series we have fitted an ARCH(1). The simulated rejection rates, at the nominal level $\alpha = 0.05$, of both tests are graphically represented in Fig. 1 by the SDP. Here, the rejection rates are plotted as function of the all possible significant subsets of lags $S \subseteq L$, once fixed $l = 5$. Thus, for example, considering the subset $S = \{2, 4\}$, the simultaneous serial dependence between ε_t and ε_{t-2} , and between ε_t and ε_{t-4} , is investigated. Also, considering the subset $S = \{1, 2, 5\}$, the simultaneous serial dependence between ε_t and ε_{t-1} , between ε_t and ε_{t-2} , and between ε_t and ε_{t-5} is taken into account. Note that a black point on each bar highlights the nominal level $\alpha = 0.05$. It easy to note as in Fig. 1(a) and Fig. 1(b) the higher bar is in correspondence to the set $S = \{2\}$ representing the type of lag-dependence that is not captured from the ARCH(1).

4 Concluding remarks

In this paper the two tests for serial independence proposed in [1] are applied in the context of nonlinear model validation. A wide simulation study has been also performed to investigate their performance. Both tests, which exploit the well-known χ^2 -test of independence, show rejection rates, under null hypothesis, which are very close to the nominal size chosen in advance. They appear also to be powerful against various types of alternatives.

Furthermore, through the Serial Dependence Plot, these tests reveal how the model could be improved indicating possible lag(s) dependence to insert in the model. Further developments will be devoted to check the performance of the tests in the validation of nonparametric structure like, for example, the additive one.

References

1. L. Bagnato and A. Punzo. On the use of χ^2 -test to check serial independence. Technical report, Università di Milano-Bicocca, 2010.
2. C. Diks. Nonparametric tests for independence. Technical report, University of Amsterdam, 2008.
3. F. Jianqing and Y. Qiwei. *Nonlinear time series: nonparametric and parametric methods*. Springer, 2003.
4. P.A.P. Moran. The statistical analysis of the Canadian lynx cycle, i: Structure and prediction. *Australian Journal of Zoology*, 1:163–173, 1953.
5. S.N. Roy. On a heuristic method of test construction and its use in multivariate analysis. *Annals of Mathematical Statistics*, 24:220–238, 1953.
6. Juliet Popper Shaffer. Multiple hypothesis testing. *Annual Reviews in Psychology*, 46(1):561–584, 1995.
7. D. Tjøstheim. Non-linear time series: A selective review. *Scandinavian Journal of Statistics*, 21:97–130, 1994.