Testing for serial independence: Beyond the Portmanteau approach

July 25, 2016

Abstract

Portmanteau tests are typically used to test serial independence even if, by construction, they are generally powerful only in presence of pairwise dependence between lagged variables. In this paper we present a simple statistic defining a new serial independence test which is able to detect more general forms of dependence. In particular, differently from the Portmanteau tests, the resulting test is powerful also under a dependent process characterized by pairwise independence. A diagram, based on p-values from the proposed test, is introduced to investigate serial dependence. Finally, the effectiveness of the proposal is evaluated in a simulation study and with an application on financial data. Both show that the new test, used in synergy with the existing ones, helps in the identification of the true data generating process.

Keywords: Chi-squared statistic, Serial dependence, Multi-way contingency table, Portmanteau approach, Nonlinear time series, Model diagnostic checking

1 Introduction

Investigating the temporal dependence structure, and testing for serial independence, are of fundamental importance in time series analysis. The autocorrelogram, which displays the strength of linear dependencies (autocorrelations) as a function of the time lags, and the testing procedures based on the autocorrelations (see King, 1987, for a survey), have been the primary tools for exploring and testing serial independence for many decades. Although these tools perform well when the serial dependence structure is linear and the innovations are Gaussian, they fail when studying a process with zero autocorrelation (see Hall and Wolff, 1995, for an example) and can behave rather poorly when applied to non-Gaussian and/or nonlinear time series (see the simulation results reported by Hallin and Mélard, 1988).

The considerations above have motivated the development of serial dependence diagrams (see Anderson and Vahid, 2005, Bagnato et al., 2012, Zhou, 2012), as well as serial independence tests (see Diks, 2009, pp. 6256–6257), which are powerful against general types of dependence (*omnibus* procedures); however, the majority of such tests and diagrams have a drawback: they are based on pairs of lagged variables and, consequently, they can fail in detecting kinds of dependence involving more than two lagged variables simultaneously.

Among the available proposals of the type described above, this paper focuses on the test and diagram proposed by Bagnato and Punzo (2010). The building block of these methods is the well-known Pearson χ^2 statistic computed on a pair of lagged variables in order to test independence for that lag (single-lag testing problem); the single-lag procedure is generalized to more than one lag via the classical "Portmanteau approach". However, the resulting "Portmanteau test" is blind when analyzing a dependent process characterized by pairwise independence.

We propose a simple test statistic, asymptotically distributed as a χ^2 under the null of serial independence; advantageously, the corresponding test is powerful with respect to general forms of dependence also involving more than one lag simultaneously. To investigate serial dependence, we provide a bar diagram with bars defined by the proposed test statistic.

The paper is organized as follows. In Section 2 we recall the classical pairwise approach to solve the problem of testing for serial independence. In Section 3 we outline and exemplify the drawbacks of the pairwise approach. In Section 4 we introduce the new serial independence test and we show how it can avoid some of the problems described in Section 3. In sections 5 and 6 we evaluate the effectiveness of our proposal by a simulation study and by an application to financial data, respectively. Conclusions are finally given in Section 7.

2 Pairwise approach for testing serial independence

Let $\{X_t\}_{t\in\mathbb{N}}$ be a real-valued and strictly stationary stochastic process with X_t having continuous density g. We want to solve the testing problem

$$\begin{aligned} &H_0 : \{X_t\}_{t\in\mathbb{N}} \text{ is an independent process} \\ &\qquad \textit{versus} \\ &H_1 : \{X_t\}_{t\in\mathbb{N}} \text{ is a dependent process.} \end{aligned} \tag{1}$$

Given its complexity, it is quite difficult to build a test statistic which is sensitive to general departures from H_0. Hence, the testing problem (1) is commonly handled by focusing on a "pairwise approach": the presence of serial independence is checked among pairs of lagged variables. Operationally, as well-explained in Robinson (1991) and Skaug and Tjøstheim (1993), the alternative hypothesis in (1) is substituted by the following simpler ones.

1. Single-lag testing problem: a particular lag $l \in \mathbb{N}_+$, with $\mathbb{N}_+ = \{1, 2, \ldots\}$, is chosen and the alternative is

$$H_1^{(l)} : f(x_{t-l}, x_t) \neq g(x_{t-l})\, g(x_t) \text{ over a subset of } \mathbb{R}^2 \text{ of non-null probability.} \quad (2)$$

2. Multiple-lag testing problem: a particular set of lags L (set of distinct positive natural numbers sorted in increasing order) is chosen and the alternative is

$$H_1^{(L)} : \quad f(x_{t-l}, x_t) \neq g(x_{t-l})\, g(x_t) \text{ over a subset of } \mathbb{R}^2 \text{ of non-null probability,}$$
$$\text{for some } l \in L. \quad (3)$$

With the sentence "over a subset of $\mathbb{R}^2$ of non-null probability" in (2) and (3) we means that the inequality postulated by the alternative must hold on at least a subset, say $\mathcal{S}$, of $\mathbb{R}^2$ such that the integral over $\mathcal{S}$, with respect to the measure induced by $g \cdot g$, is greater than zero. As said above, hypotheses (2) and (3) are simpler than hypothesis (1); there are cases in which (2) and (3) are false even if the underlying process is dependent.

In literature, a plethora of statistics to test H_0 in (1) against $H_1^{(l)}$ have been proposed (see Diks, 2009 for a review). The majority of these statistics are defined starting from a measure of discrepancy between $f(x_{t-l}, x_t)$ and $g(x_{t-l})\, g(x_t)$. The resulting tests are generally consistent against $H_1^{(l)}$ in the sense that their power tends to 1 if the sample size n tends to infinity when the data generating process exhibits some kind of dependence between X_{t-l} and X_t. These tests are usually extended to the multiple-lag testing problem mainly in two ways: the first one is to build a "Portmanteau" statistic defined as a convenient sum of the single-lag statistics; the second one consists in building a "simultaneous" test by means of p-value correction techniques based on Bonferroni's inequality and its extensions (see, e.g., Holm, 1979, Simes, 1986, Hochberg, 1988, and Hommel, 1988). As observed in Wright (1992), the p-value correction is straightforward, but the resulting test is generally very conservative (the effective level of the Type-I error probability is much less than the nominal one) and, consequently, suffers of a lack of power. For this reason, the Portmanteau approach is usually preferred and it will be the only one considered hereafter. Specifically, we will focus on Portmanteau procedures based on the autocorrelations and on the Pearson χ^2 statistic which are summarized below.

2.1 Single-lag testing problem

Let $(x_1, \ldots, x_n)$ be an observed time series of length n from $\{X_t\}_{t \in \mathbb{N}}$. To study the single-lag testing problem for the generic lag l, $l < n$, let's consider the $n_l = n - l$ pairs $\{(x_{i-l}, x_i)\}_{i=l+1}^{n} \equiv \{(x_j, x_{j+l})\}_{j=1}^{n_l}$. These pairs arise by the scheme in Table A.6 and they will be the starting point to define the test statistics to face the single-lag testing problem.

2.1.1 Sample autocorrelation function (ACF)

The testing problem for the single lag l is commonly addressed by measuring the lag-l autocorrelation ρ_l, i.e. the autocorrelation between X_t and X_{t-l}. Based on the observed time series and on the pairs highlighted in Table A.6, the sample lag-l autocorrelation is given by

$$\widehat{\rho}_l = \frac{\displaystyle\sum_{i=l+1}^{n} (x_{i-l} - \overline{x})(x_i - \overline{x})}{\displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2}, \tag{4}$$

where $\overline{x} = \sum_{i=1}^{n} x_i / n$ denotes the sample mean. Under H_0, i.e. for a serial independent process, $\sqrt{n}\widehat{\rho}_l$ asymptotically follows the standard normal distribution; thus, if the observed time series is sufficiently long, the null hypothesis of serial independence can be rejected at level α if $\widehat{\rho}_l > z_{1-\alpha/2}/\sqrt{n}$ or $\widehat{\rho}_l < -z_{1-\alpha/2}/\sqrt{n}$, $z_{1-\alpha/2}$ being the quantile of the standard normal distribution.

The test based on $\widehat{\rho}_l$ performs well against $H_1^{(l)} : \rho_l \neq 0$ (i.e. when the lag-l dependence is mainly linear or the process is Gaussian) but it can be inconsistent against alternatives in (2) characterized by a null lag-l autocorrelation (see Hall and Wolff, 1995, for an example) and it can behave rather poorly, both in terms of level and power, when applied to non-Gaussian and nonlinear time series (see the simulations reported by Hallin and Mélard, 1988).

In practice, the sample autocorrelation $\widehat{\rho}_l$ is computed for different, subsequent, values of l obtaining the sample autocorrelation function (ACF) which is usually depicted (along with the critical values $\pm z_{1-\alpha/2}/\sqrt{n}$) in the well known autocorrelogram.

2.1.2 Sample autodependence function (ADF)

Bagnato and Punzo (2010) proposed to classify the pairs highlighted in Table A.6 in a square contingency table (cf. Table 1) having marginal sets defined by $k \geq 2$ adjacent intervals $\left\{C_u^{(l)}\right\}_{u=1}^{k}$ (used to classify x_{i-l}) and $\left\{D_v^{(l)}\right\}_{v=1}^{k}$ (used to classify x_i). The absolute joint

Table 1: Contingency table for the single-lag testing problem for the generic lag l.

x_{i-l} \ x_i	$D_1^{(l)}$	$\cdots$	$D_v^{(l)}$	$\cdots$	$D_k^{(l)}$	Total
$C_1^{(l)}$	$n_{11}^{(l)}$	$\cdots$	$n_{1v}^{(l)}$	$\cdots$	$n_{1k}^{(l)}$	$n_{1+}^{(l)}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$C_u^{(l)}$	$n_{u1}^{(l)}$	$\cdots$	$n_{uv}^{(l)}$	$\cdots$	$n_{uk}^{(l)}$	$n_{u+}^{(l)}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$C_k^{(l)}$	$n_{k1}^{(l)}$	$\cdots$	$n_{kv}^{(l)}$	$\cdots$	$n_{kk}^{(l)}$	$n_{k+}^{(l)}$
Total	$n_{+1}^{(l)}$	$\cdots$	$n_{+v}^{(l)}$	$\cdots$	$n_{+k}^{(l)}$	n_l

frequencies in this table are defined by

$$\begin{aligned} n_{uv}^{(l)} &= \left|\left\{(x_{i-l}, x_i) : (x_{i-l}, x_i) \in C_u^{(l)} \times D_v^{(l)},\ i = l+1, \dots, n\right\}\right| \\ &= \left|\left\{(x_j, x_{j+l}) : (x_j, x_{j+l}) \in C_u^{(l)} \times D_v^{(l)},\ j = 1, \dots, n_l\right\}\right|, \quad u, v = 1, \dots, k, \end{aligned}$$

where $|A|$ denotes the cardinality of the set A, and the absolute marginal frequencies are given by

$$n_{u+}^{(l)} = \sum_{v=1}^{k} n_{uv}^{(l)}, \quad u = 1, \dots, k, \quad \text{and} \quad n_{+v}^{(l)} = \sum_{u=1}^{k} n_{uv}^{(l)}, \quad v = 1, \dots, k.$$

The marginal intervals $\left\{C_u^{(l)}\right\}_{u=1}^{k}$ and $\left\{D_v^{(l)}\right\}_{v=1}^{k}$ are data-dependent and are defined to yield uniform marginal distributions (*equi-frequent marginal intervals*). Operationally, the extremes of these equi-frequent marginal intervals correspond to the sample quantiles at levels $1/k, 2/k, \dots, (k-1)/k$ obtained from the observed sub-series $(x_1, \dots, x_{n-l})$ and $(x_{l+1}, \dots, x_n)$, respectively. The single-lag testing problem can be so handled using the

Pearson χ^2 statistic

$$\widehat{\delta}_l = \sum_{u=1}^{k}\sum_{v=1}^{k} \frac{\left(n_{uv}^{(l)} - \widehat{n}_{uv}^{(l)}\right)^2}{\widehat{n}_{uv}^{(l)}}, \tag{5}$$

where $\widehat{n}_{uv}^{(l)} = n_{u+}^{(l)} n_{+v}^{(l)}/n_l$ are the theoretical frequencies under the (null) hypothesis of independence of lag l, with $u, v = 1, \ldots, k$.

Remark 2.1. It is worth noting that $\widehat{\delta}_l$, defined with equi-frequent marginal intervals, can be equivalently computed on the observed ranks $(r_1, \ldots, r_n)$ of $(x_1, \ldots, x_n)$. This means that $\widehat{\delta}_l$ is a function of the observed serial empirical copula (see Genest and Rémillard, 2004, p. 338)

$$\widehat{c}_l(u_1, u_2) = \frac{1}{n_l} \sum_{i=l+1}^{n} \mathbb{I}\left\{\frac{r_{i-l}}{n_l+1} \leq u_1\right\} \mathbb{I}\left\{\frac{r_i}{n_l+1} \leq u_2\right\},$$

where $(u_1, u_2) \in [0,1]^2$, $\mathbb{I}\{\cdot\}$ is the indicator function, and the quantities $r_i/(n_l+1)$, $i = 1, \ldots, n$, represent the relative ranks.

It is well known that, if the n_l pairs $\{(X_{i-l}, X_i)\}_{i=l+1}^{n}$ were interpretable as a bivariate random sample from a bivariate distribution (as under the classical inferential paradigm), the asymptotic null distribution of $\widehat{\delta}_l$ should be the χ^2 with $(k-1)^2$ degrees of freedom (see Agresti, 2002, p. 79). For us this is not the case since, even under the hypothesis of serial independence, some of the n_l pairs have a common element and, consequently, they are dependent (e.g., if $l = 3$, then the pairs (X_1, X_4) and (X_4, X_7) are dependent). However, as outlined in Remark 2.1, $\widehat{\delta}_l$ is a function of the serial empirical copula and, consequently, the results in Genest and Rémillard (2004) assure that in our serial context, under the null hypothesis of serial independence, its asymptotic distribution does not change with respect to the classical inferential paradigm. Then, the limiting distribution of $\widehat{\delta}_l$ under the null hypothesis of serial independence is the χ^2 with $(k-1)^2$ degrees of freedom.

This fact allows to test the null hypothesis of serial independence using $\widehat{\delta}_l$ as test statistic: denoting with $\chi^2_{[\eta;q]}$ the q-quantile of the χ^2 distribution with η degrees of freedom, the null hypothesis is rejected at level α if $\widehat{\delta}_l > \chi^2_{[(k-1)^2;1-\alpha]}$.

To completely specify the test statistic $\widehat{\delta}_l$, the value of k must be selected. According to Bagnato et al. (2012), it is convenient to select k by matching the rules of Mann and Wald (1942) and Cochran (1954); in formula

$$k = \min\{k_s, k_p\}, \tag{6}$$

with

$$k_s = \left\lfloor \left(\frac{n_l}{5}\right)^{\frac{1}{2}} \right\rfloor \quad \text{and} \quad k_p = \left\lfloor 2^{\frac{11}{10}} \left(\frac{n_l - 1}{|z_{1-\alpha}|}\right)^{\frac{1}{5}} \right\rfloor,$$

with $\lfloor \cdot \rfloor$ denoting the floor function. k_s is the value for k related to the Cochran rule and requires that, under the assumption of serial independence, in each of the k^2 cells of the contingency table with equi-frequent marginal intervals, the expected number of frequencies is at least 5. This requirement aims to assure that the null distribution of $\widehat{\delta}_l$ is well approximated by the χ^2. k_p is the value of k prescribed by the Mann and Wald rule. This last value is the one that "maximizes" the power of the related independence test. The rule (6) has been introduced to assure both a good asymptotic approximation and a good power of the serial independence test.

Operationally, $\widehat{\delta}_l$ is computed for different, subsequent, values of l yielding the sample autodependence function (ADF) which is usually depicted (along with the critical value $\chi^2_{[(k-1)^2;1-\alpha]}$) in the autodependogram of Bagnato et al. (2012).

The following example gives an illustration of the procedure described above.

Example 2.1 (Contingency table)**.** Consider the observed time series, of length $n = 25$, reported in the last column of Tables A.7(a) and A.7(b). Suppose we are interested in solving the single-lag testing problems for lags $l = 2$ and $l = 3$.

If $l = 3$, the $n_3 = 22$ pairs to be considered are highlighted in Table A.7(a). To classify these pairs in the $k \times k$ contingency table reported in Table 1, the value of k first has to be determined. Based on (6) the value $k = 2$ is obtained, since, in this case, $k_s = 2$ and $k_p = 3$. The contingency table, with equi-frequent marginal intervals, is given in Table 2(a). As concerns the equi-frequent marginal intervals $C_1^{(3)} = (-\infty, 0.217)$ and

Table 2: Example 2.1: Contingency table for the testing problem of lag l.

(a) $l = 3$

x_{i-3} \ x_i	$(-\infty, 0.567)$	$[0.567, \infty)$	Total
$(-\infty, 0.217)$	4	7	11
$[0.217, \infty)$	7	4	11
Total	11	11	22

(b) $l = 2$

x_{i-2} \ x_i	$(-\infty, 0.593)$	$[0.593, \infty)$	Total
$(-\infty, 0.383)$	5	7	12
$[0.383, \infty)$	7	4	11
Total	12	11	23

$C_2^{(3)} = [0.217, \infty)$, the value 0.217 assures that both the marginal row frequencies are equal to $n_3/k = 11$ and it coincides with the median of the values in the first column of Table A.7(a). Similarly, the value 0.567 defining the equi-frequent marginal intervals $D_1^{(3)} = (-\infty, 0.567)$ and $D_2^{(3)} = [0.567, \infty)$ is the median of the shaded values in the last column of Table A.7(a). The four theoretical joint frequencies are $\widehat{n}_{uv}^{(l)} = 11 \cdot 11/22 = 5.5$, $u, v = 1, 2$, and the value of the test statistic in (5) is $\widehat{\delta}_l = 0.727$.

In the case $l = 2$, the pairs to be considered are highlighted in Table A.7(b) and there are $n_2 = 23$ of them. Also in this case, based on (6), $k_s = 2$ and $k_p = 3$; hence $k = 2$. The resulting 2×2 contingency table, with equi-frequent marginal intervals, is given in Table 2(b). In this case, the number of pairs ($n_2 = 23$) is not divisible by the number of marginal intervals ($k = 2$); thus, one of the two marginal intervals, the first, contains one pair more.

2.2 Multiple-lag testing problem via the Portmanteau approach

The statistical tests based on $\widehat{\rho}_l$ and $\widehat{\delta}_l$ are able to detect deviations from serial independence only when X_{t-l} and X_t are dependent. There are, indeed, dependent processes in which X_{t-l} and X_t are independent and, in such situations, the tests based on $\widehat{\delta}_l$ and $\widehat{\rho}_l$ can be inconsistent; this can happen, for example, when we use $\widehat{\delta}_1$ or $\widehat{\rho}_1$ but the underlying process is only characterized by dependence from the second lag onwards. This situation could arise when analyzing a seasonal time series with dependence only among the observations related to the same season. As previously mentioned, to partially overcome this problem, the most common technique is to build a "Portmanteau test" by summing simple transformations of the single-lag statistics related to all the lags l in L. The Portmanteau versions of the tests based on $\widehat{\rho}_l$ and $\widehat{\delta}_l$ will be summarized in Sections 2.2.1 and 2.2.2, respectively.

2.2.1 Portmanteau ACF

The most widespread Portmanteau test based on the ACF is the Ljung-Box test (Ljung and Box, 1978). It is based on the test statistic

$$\widehat{Q}_L = n\,(n+2) \sum_{l \in L} \frac{\widehat{\rho}_l^2}{n-l}, \tag{7}$$

which, under H_0, asymptotically follows the $\chi^2_{|L|}$ distribution. Thus, if the observed time series is sufficiently long, the null hypothesis of serial independence is rejected at level α if $\widehat{Q}_L > \chi^2_{[|L|;1-\alpha]}$. Note that, when the Ljung-Box test statistic in (7) is applied to test the serial independence of the residuals from a fitted autoregressive moving average (ARMA) model, the degrees of freedom need to be adjusted due to the estimation effect. In detail, for a fitted ARMA(p,q) model, where all the $p+q$ parameters are estimated, the degrees of freedom of the asymptotic χ^2 distribution should be set to $|L|-p-q$ (see Verbeek, 2000, Section 8.7.3). Note that the Ljung-Box test (hereafter referred to as Portmanteau ACF) is powerful only with respect to the alternative

$$H_1^{(L)} : \rho_l \neq 0 \text{ for some } l \in L,$$

which is a subset of the alternative hypothesis given in (3).

2.2.2 Portmanteau ADF

Starting from the single-lag statistic $\widehat{\delta}_l$, the Portmanteau statistic

$$\widehat{\Delta}_L = \sum_{l \in L} \widehat{\delta}_l \tag{8}$$

can be defined. As shown in Section 2.1, the asymptotic null distribution of $\widehat{\delta}_l$ is the χ^2 with $(k_l-1)^2$ degrees of freedom, where the subscript l in k_l highlights that the dimension of the (square) contingency table may depend on $l \in L$. Moreover, Bagnato and Punzo (2010) prove that, under H_0, the statistics $\widehat{\delta}_l$, $l \in L$, are independent and consequently $\widehat{\Delta}_L$ tends to the χ^2 with $\sum_{l \in L}(k_l-1)^2$ degrees of freedom. Therefore, the result of the test can be obtained by comparing either $\widehat{\Delta}_L$ with the critical value $\chi^2_{\left[\sum_{l \in L}(k_l-1)^2;1-\alpha\right]}$, or the p-value of $\widehat{\Delta}_L$ with α.

3 Problems of the pairwise approach

The single-lag and the Portmanteau statistics outlined in Sections 2.1 and 2.2, respectively, define tests which have low power and can be inconsistent with respect to alternatives belonging to H_1 in (1) but not included in (2) and (3), respectively. In particular, the statistical test based on $\widehat{\Delta}_L$ can be inconsistent for a serial dependent process such that

***case 1*:** X_{t-l} and X_t are dependent only for some $l \notin L$;

***case 2*:** X_{t-l} and X_t are independent for $l \in \mathbb{N}_+$ (serial dependence does not imply pairwise dependence).

The statistic $\widehat{Q}_L$ suffers from the same problems and, in addition, it is not powerful in capturing the presence of non-linear dependencies.

While ***case 1*** could be potentially faced by considering a convenient set L (set containing at least one lag l such that X_{t-l} and X_t are dependent), ***case 2*** can not be handled based on (2) and (3). From a practical point of view, ***case 1*** is not greatly relevant since temporal dependence is usually strong among variables which are temporally close and becomes negligible among variables temporally far. By contrast, if the adopted test is not consistent under ***case 2***, special kinds of dependence among temporally close variables might be undetectable.

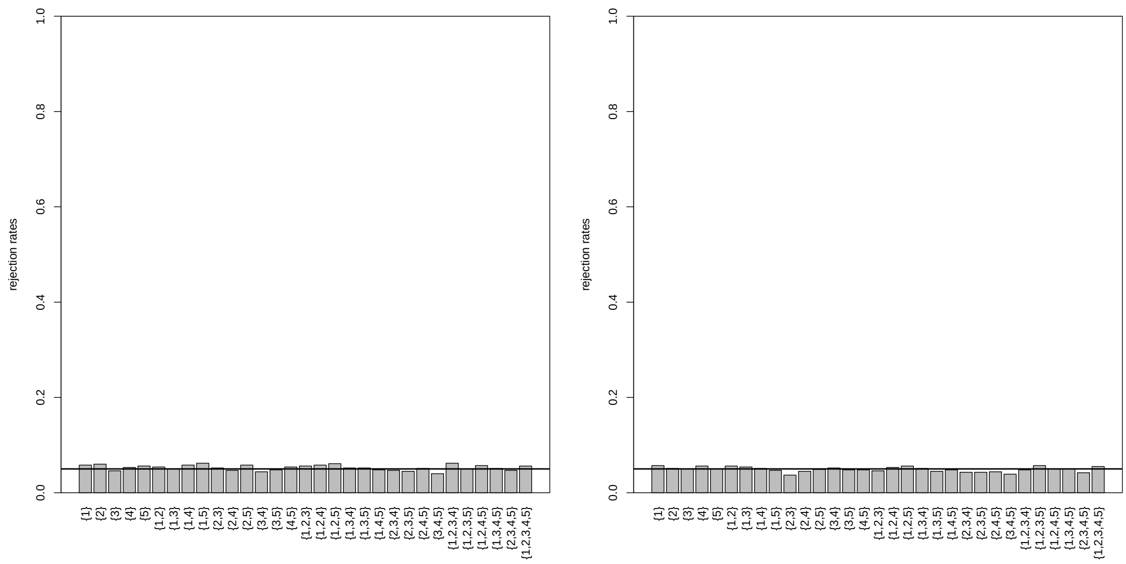
Example 3.1 (Motivation)**.** As an example of a situation of ***case 2***, consider the strictly stationary process

$$X_t = \operatorname{sign}\left(\varepsilon_{t-1}\varepsilon_{t-2}\right) + \varepsilon_t = \begin{cases} 1+\varepsilon_t & \text{if } \varepsilon_{t-1}\varepsilon_{t-2} \geq 0 \\ -1+\varepsilon_t & \text{if } \varepsilon_{t-1}\varepsilon_{t-2} < 0 \end{cases}, \tag{9}$$

where $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is a sequence of independent standard normal random variables. In Appendix B it is proved that $\{X_t\}_{t\in\mathbb{N}}$ is a dependent process with the following features: i) the random variables in (X_{t-l}, X_t) are independent for any l; ii) the random variables in (X_{t-2}, X_{t-1}, X_t) are not independent because X_t depends on the pair (X_{t-2}, X_{t-1}); iii) the random variables in (X_{t-3}, X_{t-1}, X_t) are not independent because X_t depends on the pair (X_{t-3}, X_{t-1}); iv) X_t is independent from all the pairs (X_{t-l_1}, X_{t-l_2}) with (l_1, l_2) different from $(2,1)$ and $(3,1)$.

To emphasize the blindness of the Portmanteau approach in this scenario, we compute $\widehat{Q}_L$ and $\widehat{\Delta}_L$ on one thousand time series of length $n = 1,000$ from (9). Rejection rates of the tests based on $\widehat{Q}_L$ and $\widehat{\Delta}_L$ are computed by fixing $\alpha = 0.05$. The value of k, for the definition of the contingency table related to $\widehat{\delta}_l$, is chosen based on the rule of thumb in (6). The obtained rejection rates, represented by vertical bars, are reported in the two diagrams of Figure 1; here, all the subsets of lags $L \in \mathcal{P}\left(\{1,2,3,4,5\}\right)$, $L \neq \emptyset$, are ordered lexicographically by size and displayed on the x-axis. With $\mathcal{P}\left(A\right)$ we denote the power set of A. To facilitate performance evaluation, a horizontal line is placed at $\alpha = 0.05$.

The rejection rates in Figure 1 are very close to the horizontal solid line, regardless of the test and the set L considered; while this is an expected result for the subsets composed of only one lag (i.e. $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, and $\{5\}$), this also shows that the Portmanteau



(a) Portmanteau ACF

(b) Portmanteau ADF

Figure 1: Example 3.1. Rejection rates, over 1,000 replications and with $n = 1,000$, from the tests based on $\widehat{Q}_L$ and $\widehat{\Delta}_L$, $L \in \mathcal{P}(\{1,2,3,4,5\})$, $L \neq \emptyset$.

multiple-lag testing procedures based on $\widehat{Q}_L$ and $\widehat{\Delta}_L$ are not able to detect the departure from the null hypothesis of serial independence. In the following section we propose a multiple-lag test which, as we will show in Section 4.3, is instead able to detect the underlying dependence induced by model (9).

4 Proposal for testing and investigating serial independence

Let $\boldsymbol{X}_{t-L}$ be the random vector containing the variables related to the set of lags L. As an example, if $L = \{1,3\}$, then $\boldsymbol{X}_{t-L} = (X_{t-3}, X_{t-1})$. The testing problem (1) can be alternatively formulated as follows

$$\begin{aligned} H_0 \quad &: \quad f(\boldsymbol{x}_{t-L}, x_t) = g(x_t) \prod_{l \in L} g(x_{t-l}) \text{ over a subset of } \mathbb{R}^{|L|+1} \text{ of probability one,} \\ &\quad\ \text{for any } L \\ &\qquad\qquad \textit{versus} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (10) \\ H_1 \quad &: \quad f(\boldsymbol{x}_{t-L}, x_t) \neq g(x_t) \prod_{l \in L} g(x_{t-l}) \text{ over a subset of } \mathbb{R}^{|L|+1} \text{ of non-null probability,} \\ &\quad\ \text{for some } L, \end{aligned}$$

where f is the joint density of $(\boldsymbol{X}_{t-L}, X_t)$.

To overcome the drawback of the Portmanteau approach described in Section 3 (i.e. the potential inconsistency under ***case 2***), in Section (4.1) we propose a χ^2 statistic defining a test which, for a fixed L, is consistent with respect to the alternative

$$H_1^{(L)} \quad : \quad f\left(\boldsymbol{x}_{t-L}, x_t\right) \neq g\left(x_t\right) h\left(\boldsymbol{x}_{t-L}\right) \text{ over a subset of } \mathbb{R}^{|L|+1} \text{ of non-null probability,}$$
$$\text{for some } L, \tag{11}$$

where h is the joint density of $\boldsymbol{X}_{t-L}$. Even if (11) is a subset of H_1 in (10), it is more general than (3) which is the alternative hypothesis of the Portmanteau approach: there are processes, such as the one described in Example 3.1, where (11) is true and (3) is false (i.e. ***case 2*** of Section 3).

In time series modeling, hypothesis (11) is particularly appealing and intuitively motivated since it describes a situation in which X_t depends on a certain set of lagged variables, i.e. $\boldsymbol{X}_{t-L}$. This is the most natural way of describing temporal dependence as demonstrated by the fact that common time series models have the form $X_t = m(\boldsymbol{X}_{t-L}, \epsilon_t, \epsilon_{t-1}, \ldots)$, where $m(\cdot)$ is a convenient parametric/nonparametric function, and $\{\epsilon_t\}_{t\in\mathbb{N}}$ is the innovation process.

4.1 Multiple-lag ADF

Let $\{C_u\}_{u=1}^k$ be a set of k equi-frequent marginal intervals for the observed time series $(x_1, \ldots, x_n)$. Moreover, let $l_{\max} = \max\{L\}$, $l_{\max} < n$, be the maximum lag belonging to L. To solve the multiple-lag testing problem having (11) as alternative hypothesis, once L is fixed, let's consider the set $\{(\boldsymbol{x}_{i-L}, x_i)\}_{i=l_{\max}+1}^n \equiv \{(\boldsymbol{x}_{j+l_{\max}-L}, x_{j+l_{\max}})\}_{j=1}^{n_L}$, of dimension $n_L = (n - l_{\max})$, composed by $(|L|+1)$-uples of variables. Here, $\boldsymbol{x}_{i-L}$ denotes the vector with elements x_{i-l}, $l \in L$. Consider the vector of indexes $\boldsymbol{v} = \left(v_1, \ldots, v_{|L|}\right) \in K^{|L|}$, with $K = \{1, \ldots, k\}$, which identifies the Cartesian product of equi-frequent intervals $\mathbf{C}_{\boldsymbol{v}} = C_{v_1} \times \cdots \times C_{v_{|L|}}$.

Following the notation of Section 2, define

$$\begin{aligned}
n_{u\boldsymbol{v}}^{(L)} &= \left|\left\{(\boldsymbol{x}_{i-L}, x_i) : (\boldsymbol{x}_{i-L}, x_i) \in \mathbf{C}_{\boldsymbol{v}} \times C_u,\ i = l_{\max}+1, \dots, n\right\}\right| \\
&= \left|\left\{(\boldsymbol{x}_{j+l_{\max}-L}, x_{j+l_{\max}}) : (\boldsymbol{x}_{j+l_{\max}-L}, x_{j+l_{\max}}) \in \mathbf{C}_{\boldsymbol{v}} \times C_u,\ j = 1, \dots, n_L\right\}\right|, \\
n_{u+}^{(L)} &= \sum_{\boldsymbol{v} \in K^{|L|}} n_{u\boldsymbol{v}}^{(L)}, \\
n_{+\boldsymbol{v}}^{(L)} &= \sum_{u=1}^{k} n_{u\boldsymbol{v}}^{(L)}.
\end{aligned}$$

In words, $n_{u\boldsymbol{v}}^{(L)}$ denotes the number of $(|L|+1)$-uples belonging to $\mathbf{C}_{\boldsymbol{v}} \times C_u$, $n_{u+}^{(L)}$ denotes the number of values in $\{x_i\}_{i=l_{\max}+1}^{n}$ belonging to C_u, and $n_{+\boldsymbol{v}}^{(L)}$ denotes the number of $|L|$-uples in $\{\boldsymbol{x}_{i-L}\}_{i=l_{\max}+1}^{n} \equiv \{\boldsymbol{x}_{j+l_{\max}-L}\}_{j=1}^{n_L}$ belonging to $\mathbf{C}_{\boldsymbol{v}}$.

The most intuitive way of handling the multiple-lag testing problem, having (11) as alternative hypothesis, is to consider a test statistic which measures the discrepancy between the empirical distribution of $(\boldsymbol{X}_{t-L}, X_t)$ and the theoretical distribution $g(x_t)\, h(\boldsymbol{x}_{t-L})$. Inspired by expression (11), it is natural to measure this discrepancy by using the "Pearson χ^2-like" statistic

$$\widehat{\delta}_L = \sum_{u=1}^{k} \sum_{\boldsymbol{v} \in K^{|L|}} \frac{\left(n_{u\boldsymbol{v}}^{(L)} - \widehat{n}_{u\boldsymbol{v}}^{(L)}\right)^2}{\widehat{n}_{u\boldsymbol{v}}^{(L)}}, \tag{12}$$

where $\widehat{n}_{u\boldsymbol{v}}^{(L)} = n_{u+}^{(L)} n_{+\boldsymbol{v}}^{(L)} / n_L$. Always based on Agresti (2002) and Genest and Rémillard (2004), the large sample null distribution of $\widehat{\delta}_L$ is well-approximated by the χ^2 with $\left(k^{|L|}-1\right)(k-1)$ degrees of freedom. Therefore, the result of the test can be obtained by comparing either $\widehat{\delta}_L$ with the critical value $\chi^2_{\left[\left(k^{|L|}-1\right)(k-1);1-\alpha\right]}$, or the p-value of $\widehat{\delta}_L$ with α.

Remark 4.1 (Justification of the test statistic (12))**.** An alternative statistic to (12) is the $(|L|+1)$-variate version of the commonly applied Pearson χ^2-statistic

$$\tilde{\delta}_L = \sum_{u=1}^{k} \sum_{\boldsymbol{v} \in K^{|L|}} \frac{\left(n_{u\boldsymbol{v}}^{(L)} - \tilde{n}_{u\boldsymbol{v}}^{(L)}\right)^2}{\tilde{n}_{u\boldsymbol{v}}^{(L)}}, \tag{13}$$

where $\tilde{n}_{u\boldsymbol{v}}^{(L)} = n_{u+}^{(L)} \prod_{i=1}^{|L|} n_{+v_i}^{(L)} / n_L$, with $(v_1, \dots, v_{|L|}) = \boldsymbol{v} \in K^{|L|}$. We choose $\widehat{\delta}_L$, instead of $\tilde{\delta}_L$, because it reflects and isolates the dependencies between X_t and $\boldsymbol{X}_{t-L}$, i.e., the dependence for the set of lags L. As an illustrative example of this advantage, consider the set of lags $L = \{1, 3\}$. In this case $\widehat{\delta}_{\{1,3\}}$ is sensitive to: *a)* the dependence between X_{t-1} and X_t;

b) the dependence between X_{t-3} and X_t; *c)* the dependence between the pair (X_{t-3}, X_{t-1}) and X_t. In other words, $\widehat{\delta}_{\{1,3\}}$ is sensitive to dependencies for lag-1, lag-3, and set of lags $\{1, 3\}$, respectively. To the contrary, $\tilde{\delta}_{\{1,3\}}$ is sensitive to any kind of dependence in the vector (X_{t-3}, X_{t-1}, X_t). Among these dependencies, there is also the lag-2 dependence implicitly reflected by the presence of the pair (X_{t-3}, X_{t-1}) in (X_{t-3}, X_{t-1}, X_t). Hence, if $\widehat{\delta}_{\{1,3\}}$ is used, a rejection of the null hypothesis of serial independence in correspondence to the set of lags $\{1, 3\}$ is certainly attributable to lag-dependencies involving the set $\{1, 3\}$ and its subsets. To the contrary, if $\tilde{\delta}_{\{1,3\}}$ is used, then rejection of the null hypothesis can be due to the lag-2 dependence, a kind of lag dependence that should not be considered when analyzing the set of lags $\{1, 3\}$.

4.2 Serial dependence diagram

Given a reference set $\mathcal{L}$ of (different) lags, a bar diagram may be defined by plotting the values of $\widehat{\delta}_L$ as a function of $L \in \mathcal{P}(\mathcal{L})$, $L \neq \emptyset$. This diagram may be used for a finer investigation of serial dependence.

The value of k for the definition of $\widehat{\delta}_L$ needs to be determined in advance. In particular, when $|L| = 1$, k is chosen according to the rule of thumb given in (6); in the other cases, k is selected through the following iterative procedure:

1. put $k = 3$;
2. compute all the expected cell counts under the null;
3. if the lowest cell count is lower than five, then $k = k - 1$ and the procedure ends; otherwise, put $k = k + 1$ and go to step 2.

The condition to be verified, in the third step, is a common stronger version of the rule proposed by Cochran (1954) aimed at preserving the size of the corresponding χ^2-test. Based on the procedure above, the value of k may change when L varies; in these terms, it should be denoted by k_L. Hence, the bars of the $\widehat{\delta}$-diagram may be not comparable between them because they are referred to χ^2-distributions with different degrees of freedom. A normalized bar diagram, representing the evidence of the presence of serial dependence, can be obtained by substituting $\widehat{\delta}_L$ with its p-value; hereafter, we will refer to this graphical representation as the bar diagram of the multiple-lag ADF. A horizontal dotted line is superimposed on the diagram corresponding to the desired significance level α, with $\alpha = 0.05$ the common choice.

4.3 Illustrative examples

In Example 4.1 we give an illustration for the definition of the multi-way contingency table introduced in Section 4.1, in Example 4.2 we provide an illustration of the proposed diagram, while in Example 4.3 we evaluate the ability of this diagram to detect complex dependence structures.

Example 4.1 (Three-way contingency table)**.** Consider the time series already analyzed in Example 2.1. Suppose we are interested in the multiple-lag testing problem for the set of lags $L = \{2, 3\}$. Hence, with respect to the notation of Section 4.1, we have $l_{\max} = 3$ and $|L| = 2$. The triplets to be considered are highlighted in Table A.8, and there are $n_{\{2,3\}} = 22$ of them. To define the three-way contingency table, we have to first determine the value of k. This is done by following the 3-step procedure described in Section 4.2. Starting with the value $k = 3$, the equi-frequent intervals defined on the observed time series are $(-\infty, 0.165)$, $[0.165, 0.891)$, and $[0.891, \infty)$. Due to the low value of $n_{\{2,3\}}$, we realize, without computations, that the lowest expected cell count under the null is lower than five; thus, $k = 2$ and the procedure ends. With $k = 2$, the equi-frequent marginal intervals defined on the whole observed time series are $(-\infty, 0.567)$ and $[0.567, \infty)$. Table 3 reports, in a flat version, the $2 \times 2 \times 2$ contingency table with both observed and expected (under the null) cell counts. As an example, the expected cell count 2.5, highlighted in the

Table 3: Example 4.1. Flat contingency table for the testing problem for the set of lags $L = \{2, 3\}$.

		observed cell count		expected cell count	
		x_i		x_i	
x_{i-3}	x_{i-2}	$(-\infty, 0.567)$	$[0.567, \infty)$	$(-\infty, 0.567)$	$[0.567, \infty)$
$(-\infty, 0.567)$	$(-\infty, 0.567)$	4	4	4	4
	$[0.567, \infty)$	1	4	2.5	2.5
$[0.567, \infty)$	$(-\infty, 0.567)$	2	2	2	2
	$[0.567, \infty)$	4	1	2.5	2.5
		11	11	11	11

last column of Table 3, is computed as $11\,(1+4)\,/22$, where the numbers in the numerator are highlighted in the flat table of the observed cell counts.

Example 4.2 (Bar diagram of the multiple-lag ADF)**.** To illustrate the behavior of the proposed bar diagram, a time series of length $n = 1,000$ is simulated from model (9). By considering $\mathcal{L} = \{1, 2, 3, 4, 5\}$, Figure 2 shows our diagram on the generated time series. To facilitate the interpretation of the results, a horizontal dotted critical line is placed at height $\alpha = 0.05$. Furthermore, black is used to color bars where the corresponding multiple-lag ADF yields rejection; white is used otherwise. Details on the selected values of k_L, and on the obtained p-values, are given in Table 4. Here, due to the course of dimensionality (Bellman, 1961), it easy to note how the value of k_L roughly decreases as the cardinality of L increases.

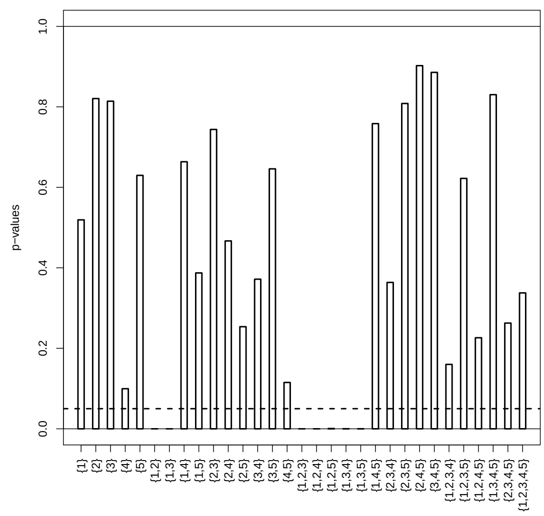


Figure 2: Example 4.2. Bar diagram of the multiple-lag ADF ($L \in \mathcal{P}(\mathcal{L})$, $L \neq \emptyset$ and $\mathcal{L} = \{1, 2, 3, 4, 5\}$) on a time series simulated from model (9). A horizontal dotted line is placed in correspondence of $\alpha = 0.05$.

By looking at the bar diagram in Figure 2, an underlying complex dependence structure appears. The first five bars show no evidence in favor of dependence on the single lags considered (see Appendix B.1 for theoretical support about this result): the p-values range from 0.82057 for lag $\{2\}$ to 0.09966 for lag $\{4\}$ (cf. Table 4). By considering the sets of cardinality 2, a high evidence in favor of dependence appears for the sets of lags $\{1, 2\}$ and $\{1, 3\}$ (practically null p-values), which are the active sets of lags based on the theoretical results given in Appendix B.2 and B.3, respectively. Such evidence is also shown for several

Table 4: Example 4.3. p-values and k_L, $L \in \mathcal{P}(\mathcal{L})$, $L \neq \emptyset$, and $\mathcal{L} = \{1, 2, 3, 4, 5\}$.

L	k_L	p-value	L	k_L	p-value
{1}	7	0.51920	{1,2,3}	3	0.00000
{2}	7	0.82057	{1,2,4}	3	0.00000
{3}	7	0.81397	{1,2,5}	3	0.00074
{4}	7	0.09966	{1,3,4}	3	0.00017
{5}	7	0.62975	{1,3,5}	3	0.00015
{1,2}	5	0.00000	{1,4,5}	3	0.75830
{1,3}	5	0.00000	{2,3,4}	3	0.36376
{1,4}	5	0.66355	{2,3,5}	3	0.80834
{1,5}	5	0.38729	{2,4,5}	3	0.90233
{2,3}	5	0.74383	{3,4,5}	3	0.88561
{2,4}	5	0.46686	{1,2,3,4}	2	0.16007
{2,5}	5	0.25374	{1,2,3,5}	2	0.62217
{3,4}	5	0.37176	{1,2,4,5}	2	0.22628
{3,5}	5	0.64592	{1,3,4,5}	2	0.83023
{4,5}	5	0.11516	{2,3,4,5}	2	0.26274
			{1,2,3,4,5}	2	0.33769

sets of lags of cardinality 3 containing either $\{1, 2\}$ or $\{1, 3\}$.

Example 4.3 (Power of the test)**.** By considering the data already presented in Example 3.1, we show how the multiple-lag ADF is able to detect the presence of serial dependence structures which are not captured by the Portmanteau ACF and by the Portmanteau ADF. Figure 3 displays the rejection rates related to the test based on $\widehat{\delta}_L$, $L \in \mathcal{P}(\mathcal{L})$, $L \neq \emptyset$, and $\mathcal{L} = \{1, 2, 3, 4, 5\}$. Similarly to Figure 1, which reports the results of the tests based on $\widehat{Q}_L$ and $\widehat{\Delta}_L$, the proposed test statistic $\widehat{\delta}_L$ has no power when the subsets $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, and $\{5\}$ are considered (coherently with the pairwise independence). Higher power is observed, apart from the set $\{1, 3, 4, 5\}$, for all the subsets of lags including either $\{1, 2\}$ or $\{1, 3\}$, while no dependencies are detected for other subsets of lags. These results are in line with the arguments presented in Appendix B.

Then, differently from the Portmanteau approaches, our test is able to capture the presence of dependence even in the context of pairwise independence. Further results about

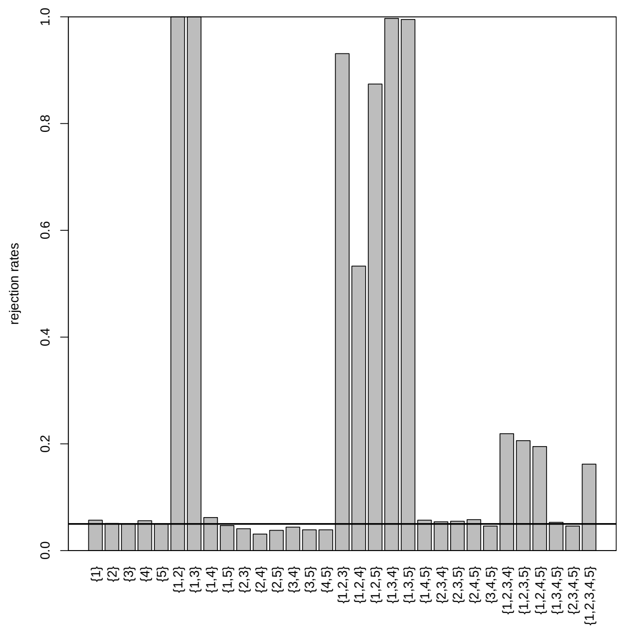


Figure 3: Example 4.3. Rejection rates, over 1,000 replications and with $n = 1,000$, from the test based on $\widehat{\delta}_L$, $L \in \mathcal{P}(\mathcal{L})$, $L \neq \emptyset$ and $\mathcal{L} = \{1, 2, 3, 4, 5\}$.

this aspect will be given in Section 5.3.

5 Simulation study

This section examines, via a Monte Carlo simulation study, the behavior of the test based on $\widehat{\delta}_L$ (hereafter simply denoted as "Multiple-lag ADF") in comparison with the tests based on $\widehat{Q}_L$ (Portmanteau ACF) and $\widehat{\Delta}_L$ (Portmanteau ADF), $L \in \mathcal{P}(\mathcal{L})$, $L \neq \emptyset$, $\mathcal{L} = \{1, 2, 3, 4, 5\}$, and $\alpha = 0.05$. The R-code (R Core Team, 2015) to obtain the bar diagrams of these three tests is available at `http://www.economia.unict.it/punzo`.

Table 5 shows the models, and the corresponding parameters specification, used in the simulation study. They include: three scenarios of serial independence (Section 5.1), four scenarios characterized by serial dependence of a purely linear type (Section 5.2), and three scenarios with nonlinear serial dependence (Section 5.3).

Concerning the independence cases, data are randomly generated from the standard Gaussian (denoted with ε_t), from the Student-t with 3 degrees of freedom (denoted with u_t) and from the Cauchy (denoted with v_t). The Gaussian noise ε_t is always used for

Table 5: Models, and corresponding parameters specification, adopted in the simulation study. The first column reports the section in which the simulation is discussed.

	Model	Parameter specification
Section 5.1	Gaussian	$X_t = \varepsilon_t$
	Student-t (3 d.f.)	$X_t = u_t$
	Cauchy	$X_t = \upsilon_t$
Section 5.2	MA(1)	$X_t = 0.2\varepsilon_{t-1} + \varepsilon_t$
	MA(3)	$X_t = 0.2\varepsilon_{t-3} + \varepsilon_t$
	AR(1)	$X_t = 0.3X_{t-1} + \varepsilon_t$
	AR(3)	$X_t = 0.3X_{t-3} + \varepsilon_t$
Section 5.3	GARCH(1,1)	$X_t = \sigma_t\varepsilon_t$, with $\sigma_t^2 = 0.01 + 0.2X_{t-1}^2 + 0.5\sigma_{t-1}^2$
	Bilinear AR(2)	$X_t = 0.5X_{t-2}\varepsilon_{t-1} + \varepsilon_t$
	Multiplicative MA(1)	$X_t = 0.3\varepsilon_{t-1}\varepsilon_{t-2} + \varepsilon_t$

the remaining models. As regards the scenarios related to the linear dependence, the well known AR and MA models are considered. The nonlinear models taken into account are: the GARCH(1,1) that is characterized by a quadratic form of dependence, by zero correlation, and by a decaying memory structure; the Bilinear AR(2) that has a complex nonlinear and non-monotonic form of dependence but no autocorrelation structure beyond lag zero; and the Multiplicative MA(1) which is characterized by independence from the third lag (included) onward. Finally, to evaluate the estimation effect, the competing tests are applied to residuals from different fitted models.

For each of the 10 models in Table 5, 1,000 samples, each of size $n = 800$, are randomly generated. In particular, for the 7 models characterized by serial dependence, a time series of length 900 is initially generated, but only the final 800 observations are used in order to mitigate the impact of initial values.

5.1 Results under serial independence

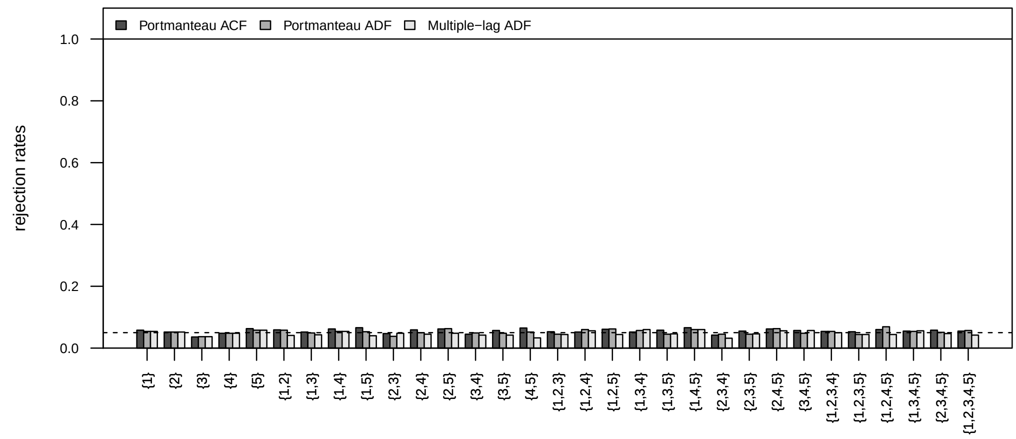
Figure 4 shows the results, in terms of rejection rates, under serial independence (H_0). When the noise is Gaussian (cf. Figure 4(a)), all the competing approaches, as expected, maintain the size. Similar results are obtained under the Student-t noise (cf. Figure 4(b)). When the noise is Cauchy distributed (cf. Figure 4(c)), the test based on $\widehat{Q}_L$ is conservative; this is due to the fact that assumptions on which the asymptotic χ^2 distribution of $\widehat{Q}_L$ is based are not satisfied because the Cauchy distribution has no moments of any order (see Romano and Thombs, 1996, p. 590, for details). To the contrary the ADF-based approaches, that are the tests based on $\widehat{\Delta}_L$ and $\widehat{\delta}_L$, maintain the size; the robustness towards distributions with non-existing moments is a clear advantage of the ADF-based techniques.

5.2 Results under serial linear dependence

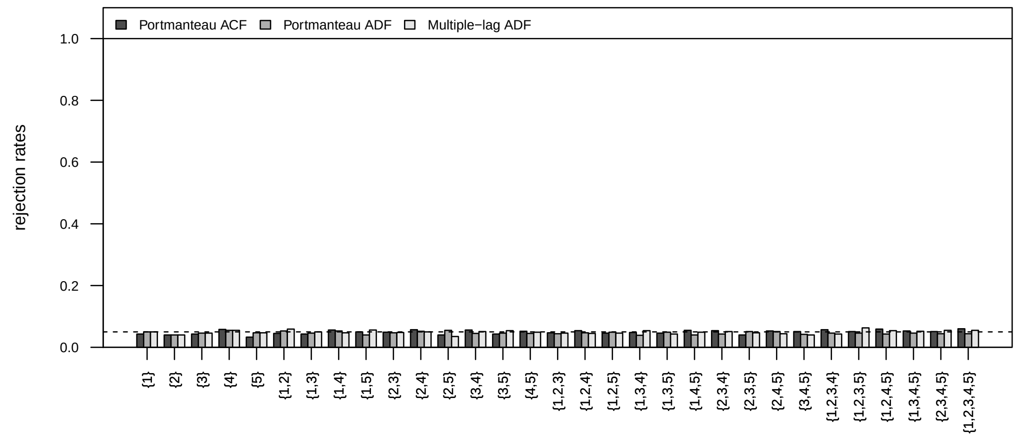
Figures 5–8 show the diagrams related to the obtained results under linear serial dependence for each of the four linear models in Table 5. The first diagram of each Figure reports the rejection rates of tests computed on the raw data, while the second diagram reports the rejection rates of the tests computed on residuals from the fitted true model. Parameter estimates, for linear models, are obtained via the maximum likelihood approach by using the `arma()` function included in the R-package **tseries** (Trapletti et al., 2015).

The first diagram in the Figures 5–8 clearly shows a higher performance of the Portmanteau ACF in detecting the linear serial dependence implicit in the MA and AR models. This result is rather expected because the Portmanteau ACF is an instrument specifically conceived to detect the presence of linear serial dependence. However, also the Portmanteau ADF and the Multiple-lag ADF are powerful enough in detecting dependencies. Their performance is quite similar even if the Multiple-lag ADF exhibits a higher persistence in power when the cardinality of the set of lags L increases.

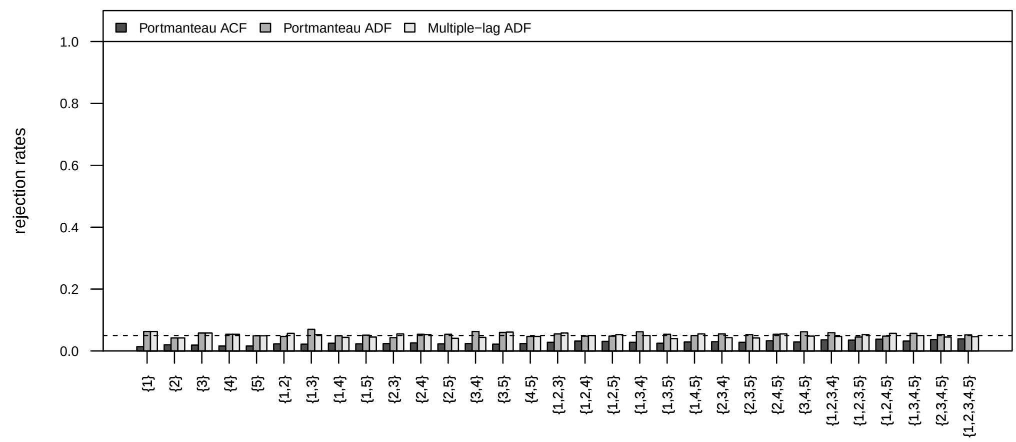
Concerning the behavior of the tests applied on residuals from the fit of the corresponding true model, we expect that the estimation removes all the dependence in the time series, i.e. the tests on residuals should have a power equal to α for all the considered sets of lags. The obtained results (see the second diagram of Figures 5–8) confirm this fact for the Portmanteau ADF and the Multiple-lag ADF. By contrast, the Portmanteau ACF has some problems. Specifically, considering for example the residuals from the fitted AR(1)



(a) Gaussian

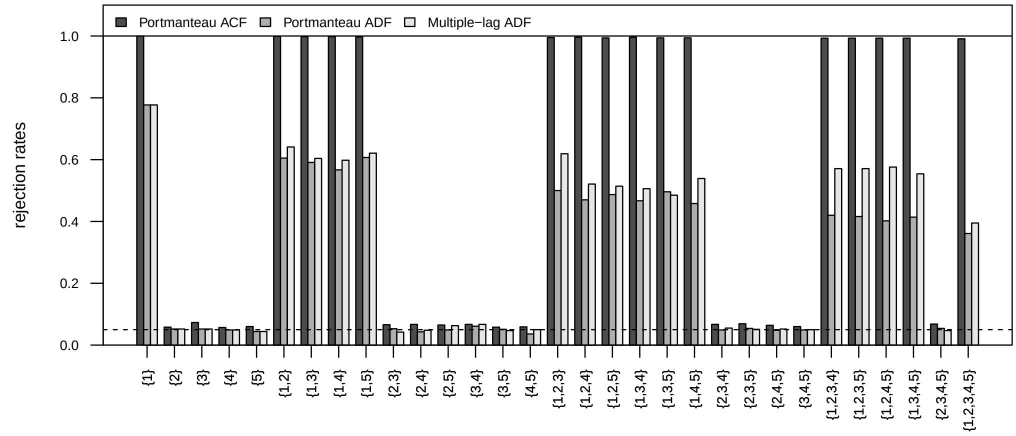


(b) Student-t (3 d.f.)

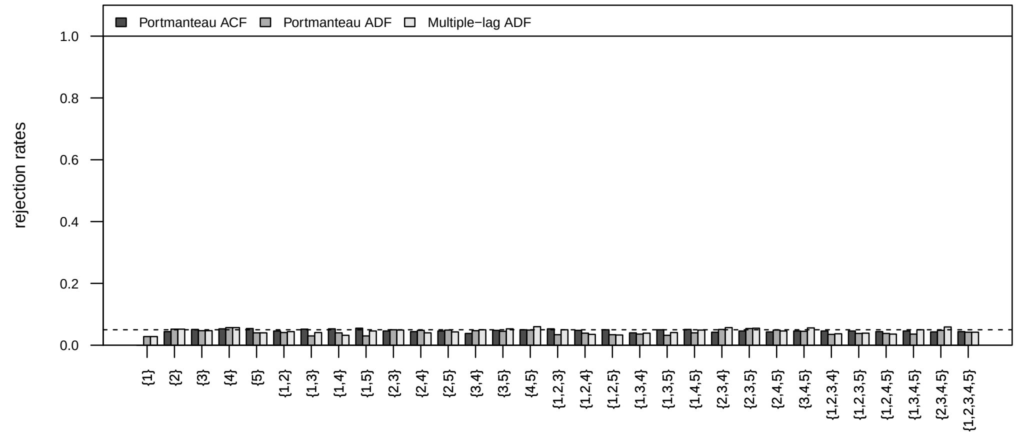


(c) Cauchy

Figure 4: H_0. Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).



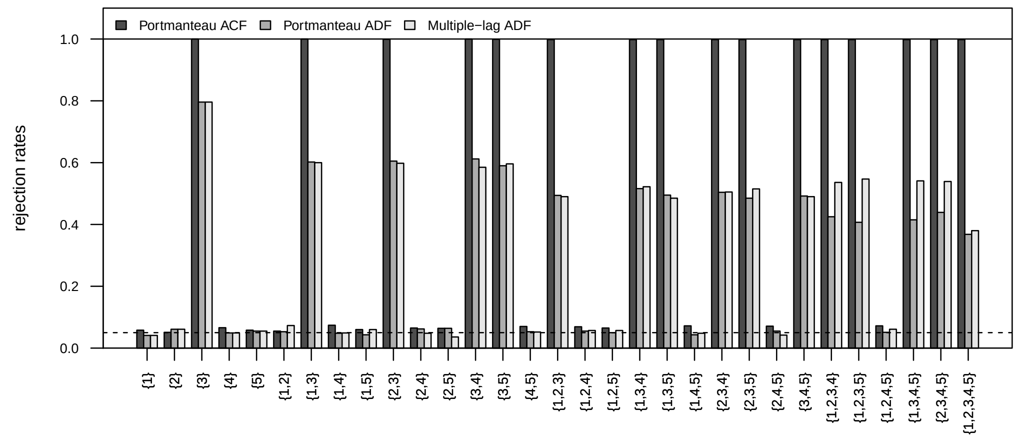
(a) raw data



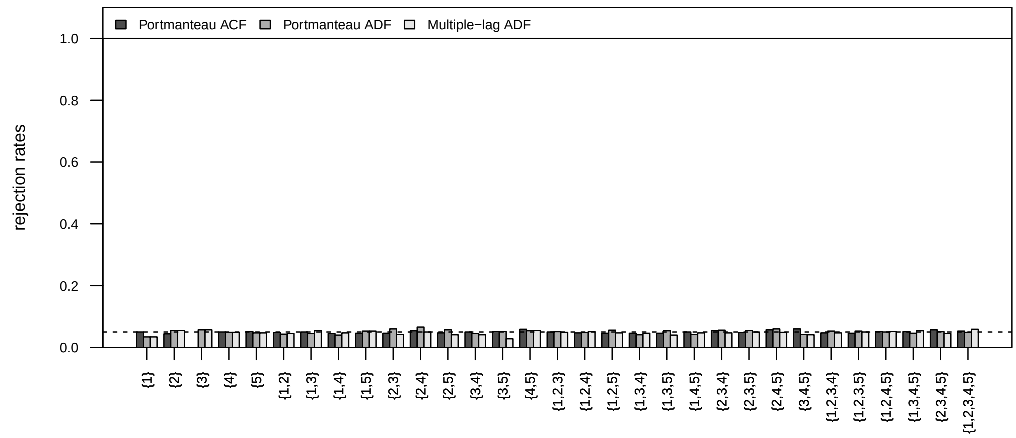
(b) MA(1) residuals

Figure 5: MA(1). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).

model (see Figure 7(b)), the lag-1 autocorrelation between the residuals is structurally very close to 0 due to the well known estimation effect and, consequently, the rejection rate corresponding to lag-1 is roughly 0 (see the bar related to lag-1 in Figure 7(b)). Moreover, the correction for the estimation effect recalled in Section 2.2.1, cannot be applied in this context since $|L| - p = 1 - 1 = 0$ and, therefore, the autocorrelation test based on $\widehat{\rho}_1$ (which is equivalent to the Portmanteau ACF test with $L = \{1\}$) has no sense in this case (see, again, Verbeek, 2000, Section 8.7.3). Similar considerations can be made for lag-1 in Figure 5(b) and lag-3 in Figures 6(b) and 8(b). Concluding, the ADF-based approaches do not require any correction for the estimation effect regardless of the particular set of lags L, and this is a further clear advantage for them.



(a) raw data



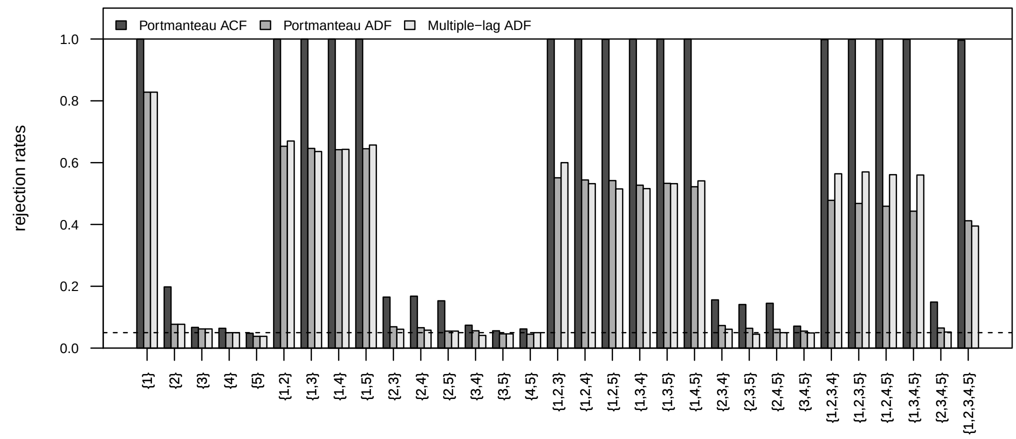
(b) MA(3) residuals

Figure 6: MA(3). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).

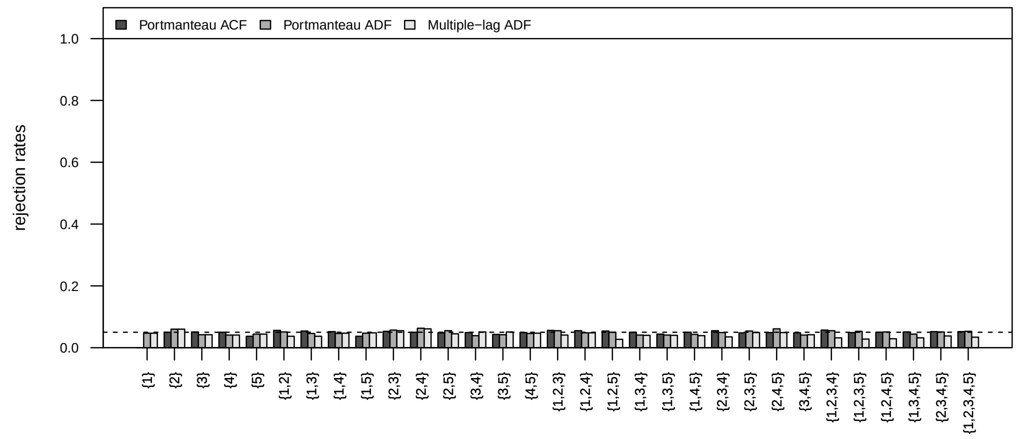
5.3 Results under serial nonlinear dependence

The results obtained under the nonlinear models are depicted in Figure 9 for the GARCH(1,1), in Figures 10–11 for the Bilinear AR(2), and in Figures 12–13 for the Multiplicative MA(1).

Concerning the results related to the GARCH(1,1) model, the performance of the considered testing procedures is evaluated on: raw data (Figure 9(a)), residuals from a GARCH(1,0) model (Figure 9(b)), and residuals from a GARCH(1,1) model (Figure 9(c)). Parameters estimates for GARCH models are obtained by using the conditional maximum likelihood (CML) method and the function `garchFit()` included in the **fGarch** package (Wuertz and Chalabi, 2013). In practice, GARCH models are usually applied on financial



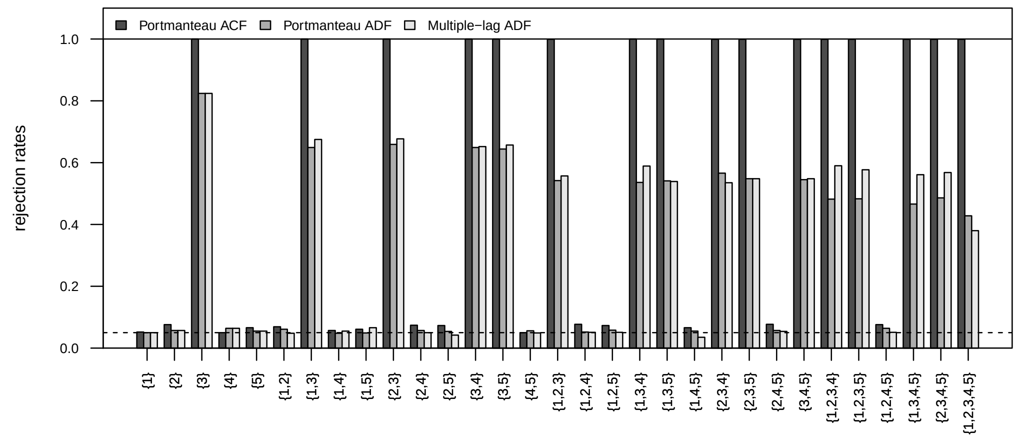
(a) raw data



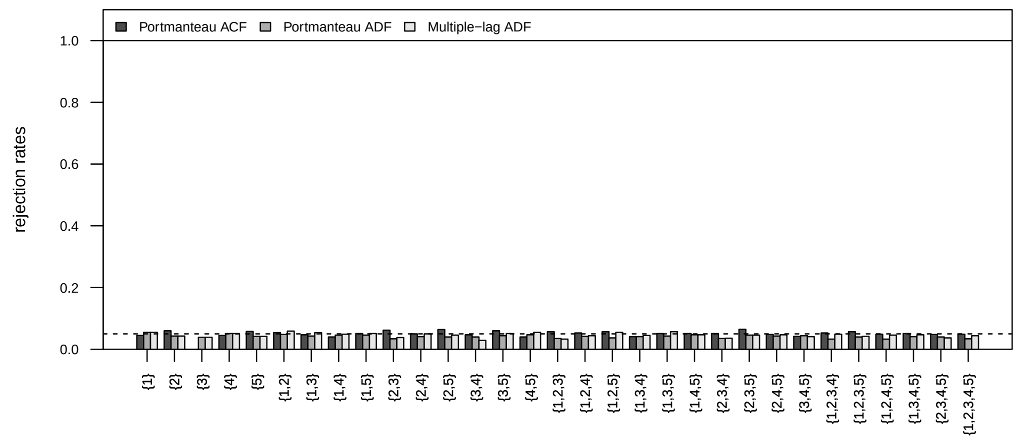
(b) AR(1) residuals

Figure 7: AR(1). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).

returns which commonly present linear dependence among the squared values. Accordingly, under this model, we follow the usual practice of applying the Portmanteau ACF on the squared raw series/residuals. Under this scenario we expect that: the Portmanteau ACF on squared raw series should be the best performer since the GARCH model can be viewed as an ARMA model on the squared series; the power observed on raw data should significantly decrease when the residuals from the GARCH(1,0) are analyzed; the power of all the tests should be equal to α when applied on residuals from the GARCH(1,1) model. The simulation results almost confirm these expectations. They reveal that the Portmanteau ADF and the Multiple-lag ADF show no power in detecting the residual dependence after the estimation of the GARCH(1,0) model. This suggests that, under the GARCH model,



(a) raw data

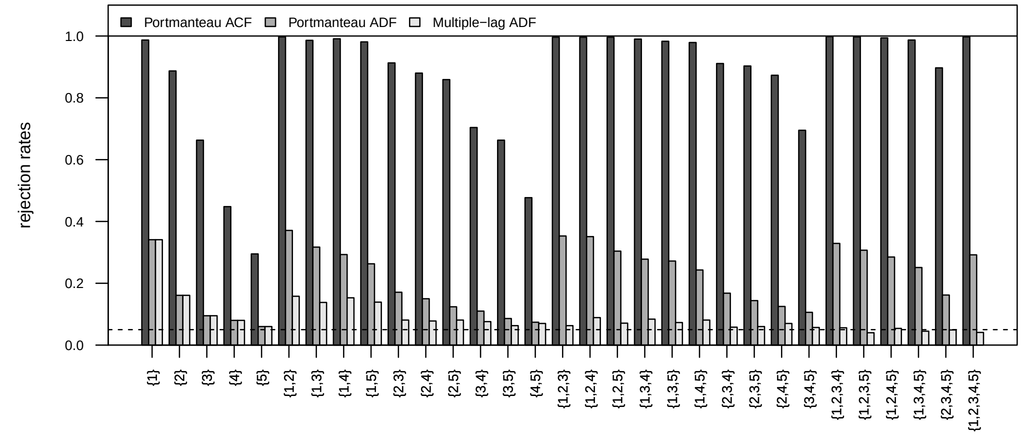


(b) AR(3) residuals

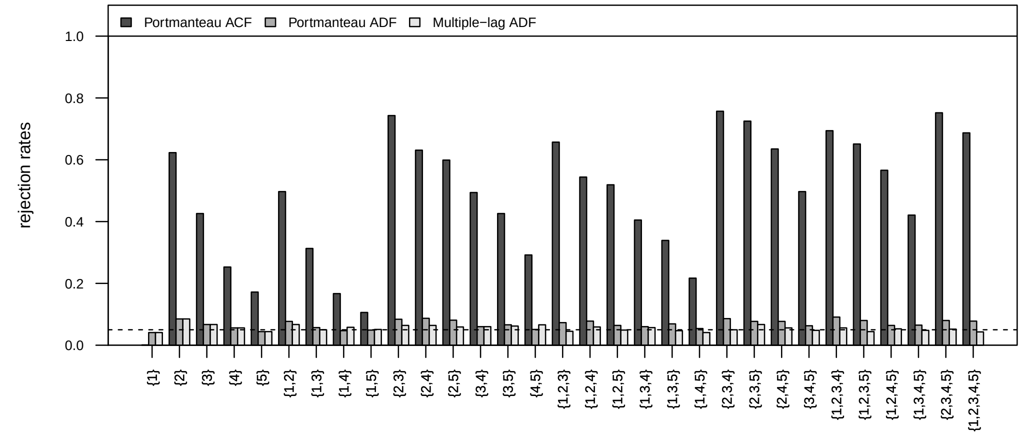
Figure 8: AR(3). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).

the ADF based procedures are mainly sensitive to the dependence due to the ARCH components and substantially blind to the purely GARCH part. Concerning the Portmanteau ACF, as for the AR and MA models discussed in the previous section, the distortion due to the estimation effect is clearly visible on the first lag of Figure 9(b) and for all the sets of lags in Figure 9(c).

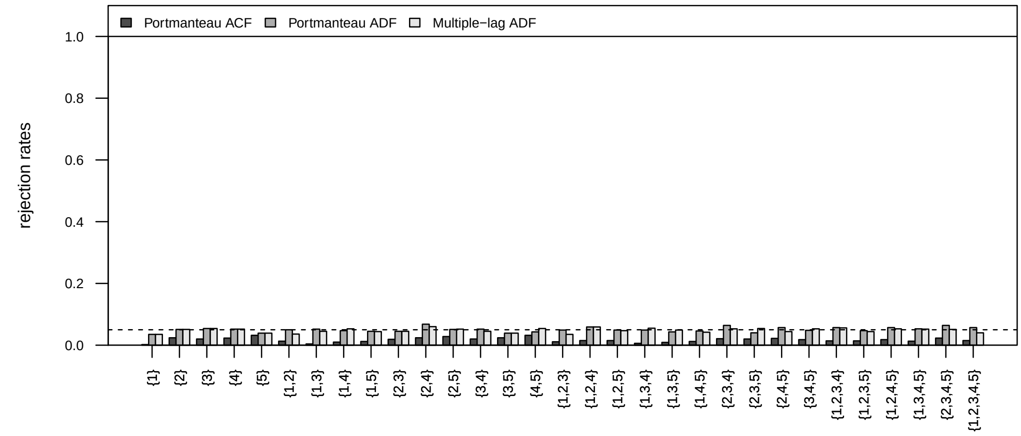
For the Bilinear AR(2) model, the performance of the considered tests is evaluated on: raw data (Figure 10(a)), residuals from an ARMA(1,1) model (Figure 10(b)), residuals from a GARCH(1,1) model (Figure 11(a)), and residuals from the correctly specified model (Figure 11(b)). Based on Rao (1981), we implemented a specific R code to obtain the CML estimates of the parameters for the Bilinear AR(2). We expect that: due to the



(a) Raw data

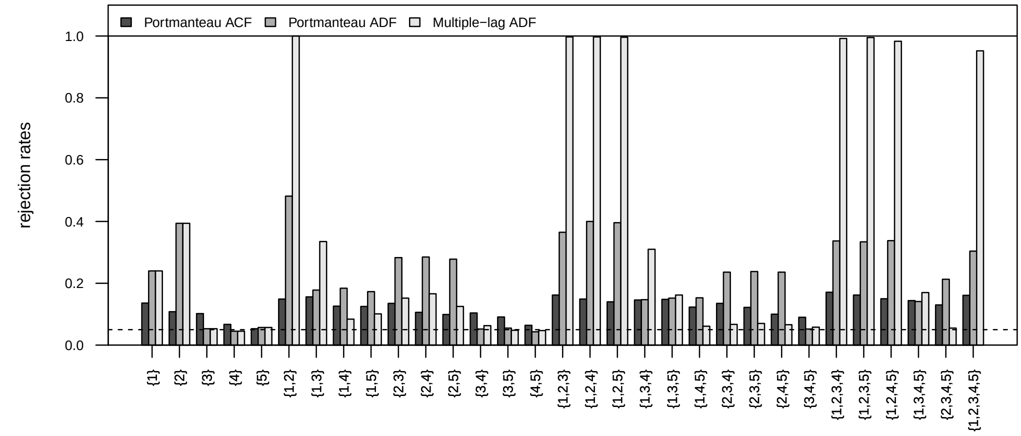


(b) GARCH(1,0) residuals

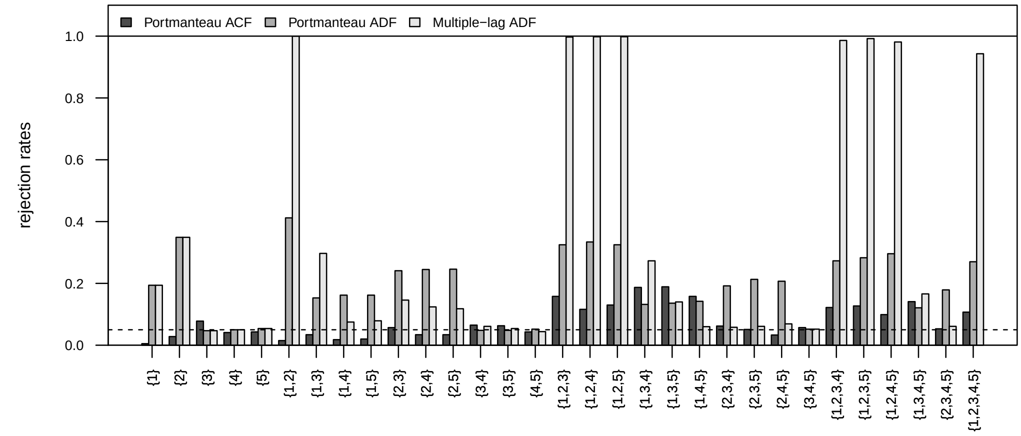


(c) GARCH(1,1) residuals

Figure 9: GARCH(1,1). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).



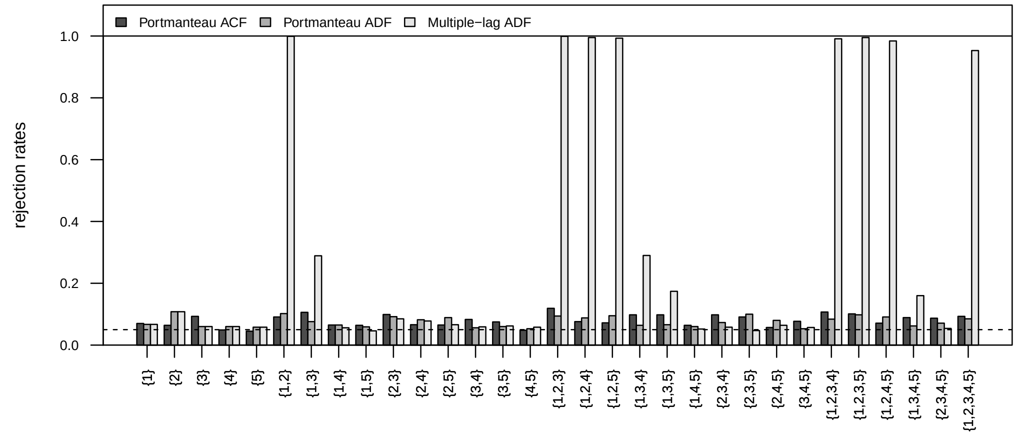
(a) Raw data



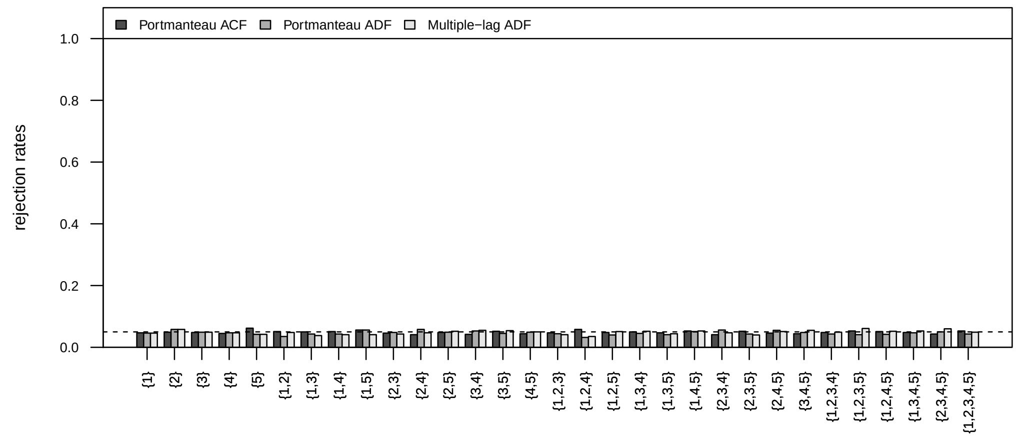
(b) ARMA(1,1) residuals

Figure 10: Bilinear AR(2). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).

nonlinear nature of the dependence of the generating model (which is characterized by serial uncorrelation; refer to Section 5), the power observed on the raw data should not significantly change when residuals from the fitted ARMA(1,1) model are analyzed; the power of all the tests should be equal to α when applied on residuals from the correctly specified model. The latter expectation is corroborated by the simulation results. The former expectation is only confirmed when analyzing the results of the Portmanteau ADF and the Multiple-lag ADF, while the behavior of the Portmanteau ACF on the residuals from the ARMA(1,1) model is affected by the estimation effect which is particularly evident when the correction of the degrees of freedom can not be applied (i.e. on the first 15 bars in Figure 10(b)). We also note that, even if the Bilinear AR(2) model is an uncorrelated



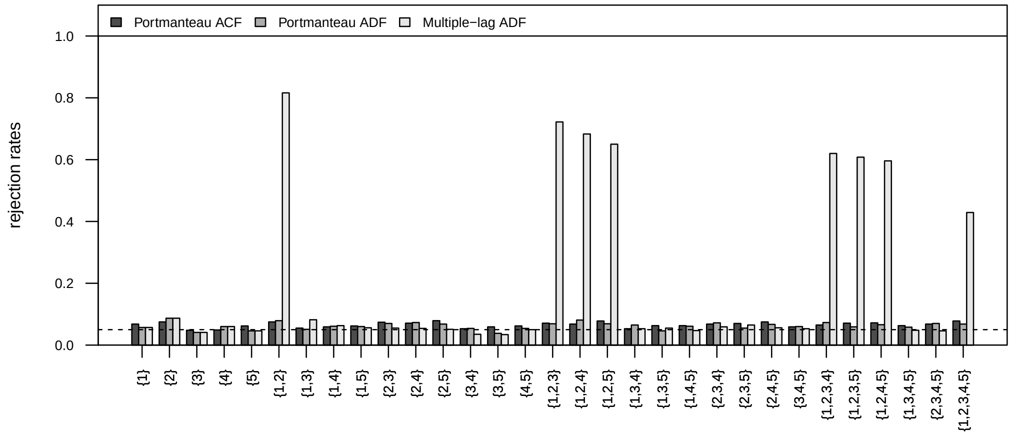
(a) GARCH(1,1) residuals



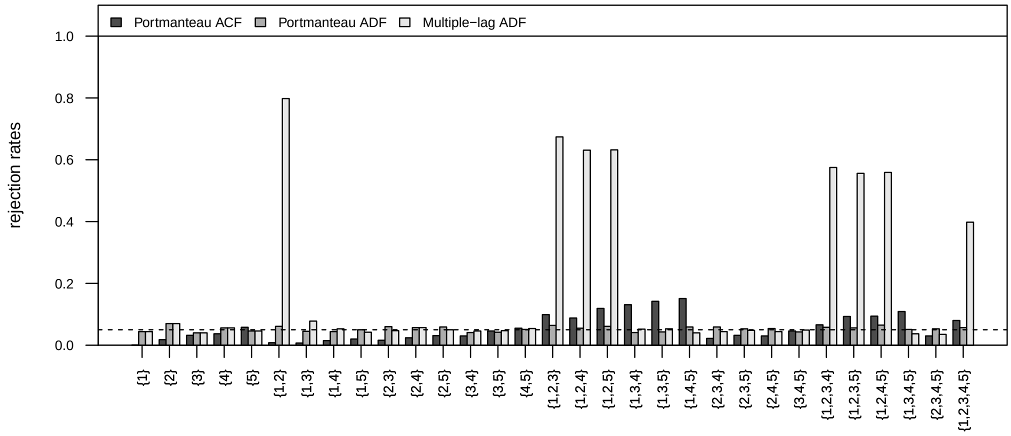
(b) Bilinear AR(2) residuals

Figure 11: Bilinear AR(2). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).

process, the Portmanteau ACF test has a power that is slightly greater than α. This fact is not surprising and is a simple consequence of the fact that the test statistic $\widehat{Q}_L$ is asymptotically χ^2 only under the null hypothesis of serial independence and not simply under the assumption of serial uncorrelation (see Genest and Rémillard, 2004, for details). Another interesting observation regards the analysis of the residuals from the GARCH(1,1) model: Figure 11(a) shows that the power of the Portmanteau ACF and of the Portmanteau ADF are substantially equal to α regardless of L. Operationally, if we use these tests on observed time series from a Bilinear AR(2), we will often erroneously tend to use the GARCH(1,1) as the model to represent these temporal dynamics. To the contrary, the power of the multiple-lag ADF still remains very high in correspondence to some sets of



(a) Raw data

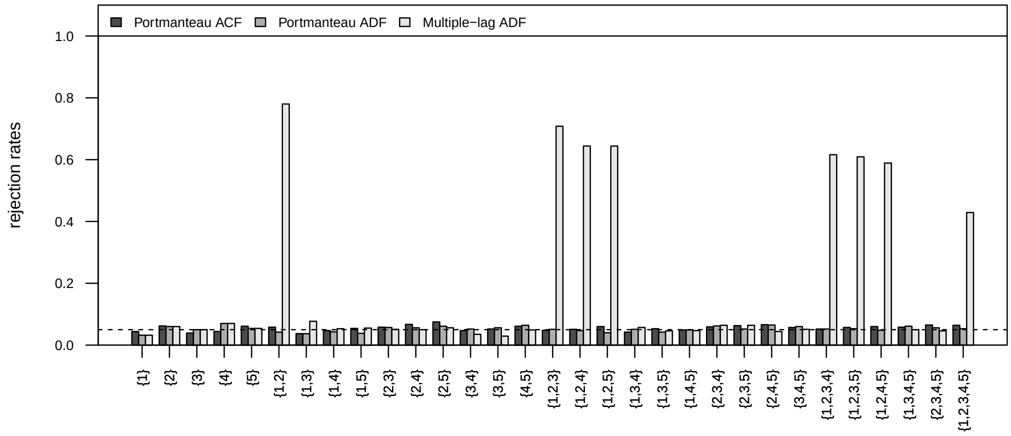


(b) ARMA(1,1) residuals

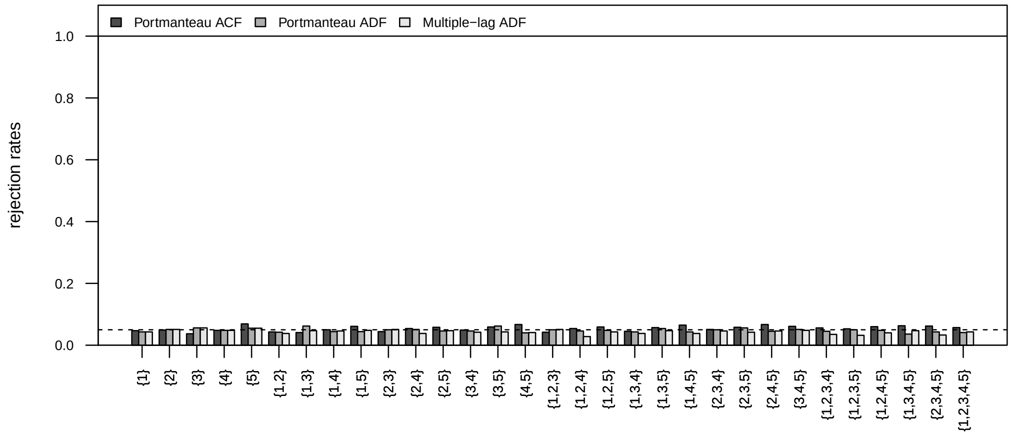
Figure 12: Multiplicative MA(1). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).

lags. This result emphasizes the practical relevance of the multiple-lag ADF which is the only graphical device, among the considered ones, able to capture the residual underlying serial dependence.

Similar conclusions are obtained when analyzing the Multiplicative MA(1) model (Figures 12–13). For the estimation of the correctly specified model, the CML is used (refer to Figure 13(b)). In this case, the result is even more extreme since the Portmanteau ACF and Portmanteau ADF have a power equal to α even on the raw data.



(a) GARCH(1,1) residuals



(b) Multiplicative MA(1) residuals

Figure 13: Multiplicative MA(1). Simulated rejection rates, over 1,000 replications and with $n = 800$, for the tests based on: $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF).

6 Real data application

In this section an application to a financial time series is considered. In particular, we consider the `SMI` dataset included in the R-package **SDD** (Bagnato et al., 2015). The series consists of $n = 660$ daily returns of the Swiss Market Index spanning the period from August 12$^{\text{th}}$, 2009, to March 6$^{\text{th}}$, 2012; see Figure 14.

Figure 15 displays a 4×3 matrix of diagrams; each of them reports, by column, the p-values of the multiple-lag tests based on $\widehat{Q}_L$ (Portmanteau ACF), $\widehat{\Delta}_L$ (Portmanteau ADF), and $\widehat{\delta}_L$ (Multiple-lag ADF), $L \in \mathcal{P}(\mathcal{L})$, $L \neq \emptyset$, $\mathcal{L} = \{1, 2, 3, 4, 5\}$. The three tests are applied to the raw series (first row) and to residuals from three nonlinear models (second–fourth

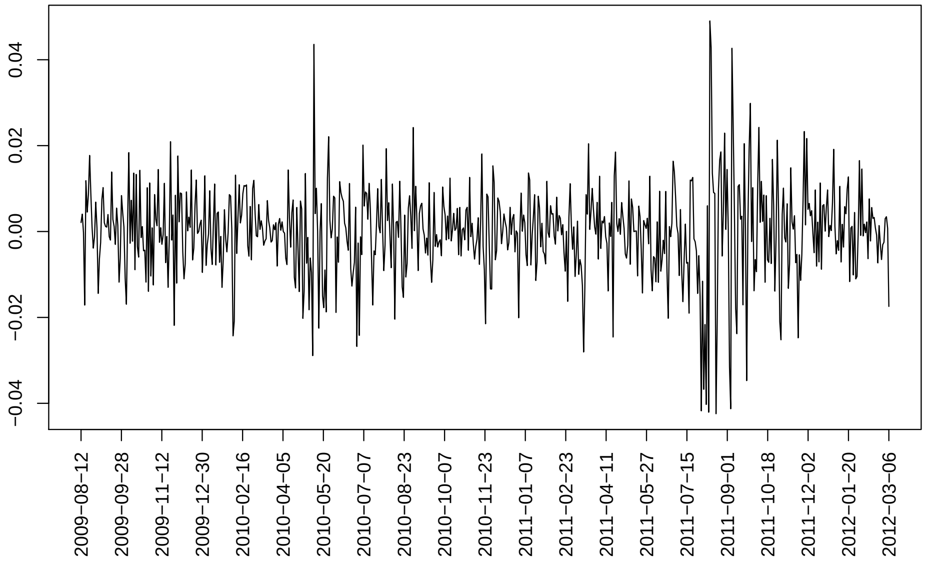


Figure 14: Daily returns of the Swiss Market Index (SMI), spanning from August 12^{th}, 2009, to March 6^{th}, 2012.

rows) which will be detailed below. Note that, due to the nature of the dependence usually observed for financial returns, $\widehat{Q}_L$ is computed on the squared series.

6.1 Results on the raw series

By looking at the first row of diagrams in Figure 15, an underlying dependence structure is detected by all the considered approaches. In particular a clear autocorrelation, for all the considered subsets of lags, is highlighted by the Portmanteau ACF computed over the squared series.

6.2 Results on residuals on the GARCH model

To capture the autocorrelation on the squared series, a GARCH(1,1) model has been adopted; it is widely used to model financial time series (see, e.g., Bollerslev et al., 1992). The GARCH(1,1) model is estimated, with the CML approach, using the `garch()` function of the R-package `tseries`. As we can see from the second row of diagrams in Figure 15, the Portmanteau ACF does not display any significant linear dependence among the squared residuals. To the contrary, the remaining diagrams in the second row suggest the presence of dependence, for example on the single lags $\{1\}$, $\{2\}$, and $\{3\}$. This suggests the use of a more sophisticated nonlinear model able to capture the underlying dependence structure.

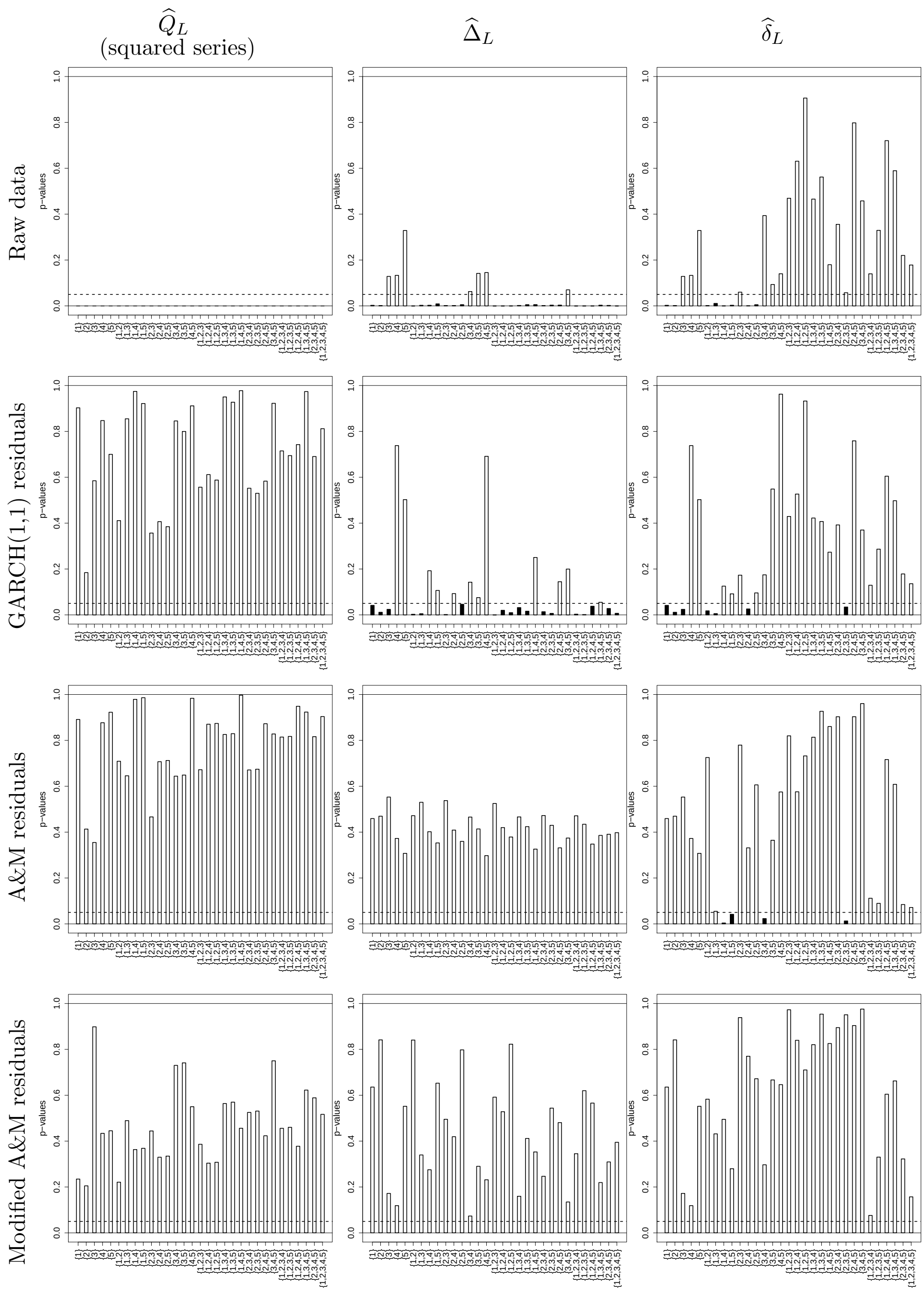


Figure 15: `SMI` dataset: matrix of diagrams of p-values for the test statistics $\widehat{Q}_L$ (first column), $\widehat{\Delta}_L$ (second column), and $\widehat{\delta}_L$ (third column), $L \in \mathcal{P}(\mathcal{L})$, $L \neq \emptyset$ and $\mathcal{L} = \{1, 2, 3, 4, 5\}$; $\alpha = 0.05$. The three tests are applied to: raw data (first row), residuals from the GARCH(1,1) (second row), residuals from the A&M model in (14) (third row), and residuals from the modified A&M model in (15) (fourth row). $\widehat{Q}_L$ is applied on the squared series.

6.3 Results on residuals on the A&M model

We consider the nonparametric A&M model proposed by Yang et al. (1999). It consists of a nonparametric autoregression with multiplicative structure for the conditional variance and additive structure for the conditional mean. As usual in the financial context, only the part of the model concerning the volatility is retained, that is

$$X_t = \sigma\left(X_{t-1}, \ldots, X_{t-d}\right)\varepsilon_t, \quad \text{with} \quad \sigma^2\left(X_{t-1}, \ldots, X_{t-d}\right) = c\prod_{j=1}^{d} \exp\left\{f_j\left(X_{t-j}\right)\right\}, \qquad (14)$$

where c is a positive constant and $\left\{f_j\left(X_{t-j}\right)\right\}_{j=1}^{d}$ are unknown functions describing the different impact of the lagged variables on the conditional variance. The process $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is assumed to be i.i.d. such that $E\left(\varepsilon_t\right) = E\left(\varepsilon_t^3\right) = 0$, $E\left(\varepsilon_t^2\right) = 1$, $E\left(\varepsilon_t^4\right) < \infty$, and ε_t is independent of X_{t-j} for $j > 0$. The backfitting algorithm on the additive model resulting from an opportune log-transformation of the series, is used to estimate the multiplicative model (14). This algorithm is implemented using the `gam()` function of the R-package **gam** (Hastie, 2013) and each additive component is fitted with smoothing splines.

Model (14) is fitted to the raw data with $d \in \{1, 2, 3, 4, 5\}$. Among these fitted models, the best one, in terms of reducing the number of black bars of the considered diagrams, is the model with $d = 4$ (see the third row in Figure 15). However, while the Portmanteau ACF and the Portmanteau ADF do not underline any kind of dependence, the p-values related to $\widehat{\delta}_L$ highlight that a residual dependence exists on the sets of lags $\{1, 4\}$, $\{1, 5\}$, $\{3, 4\}$, and $\{2, 3, 5\}$.

6.4 Results on residuals on the modified A&M model

In the fashion of multiplicative MA and bilinear AR models, we modify model (14) by adding suitable multiplicative interactions to the conditional variance based on the active sets of lags $\{1, 4\}$, $\{1, 5\}$, and $\{3, 4\}$. After some trials, we identify the following model

$$X_t = \sigma\left(X_{t-1}, \ldots, X_{t-5}\right)\varepsilon_t, \qquad (15)$$

where $\sigma\left(X_{t-1}, \ldots, X_{t-5}\right)$ is given by

$$c\prod_{j=1}^{4} \exp\left\{f_j\left(X_{t-j}\right)\right\} \exp\left\{f_5\left(X_{t-1}X_{t-4}\right)\right\} \exp\left\{f_6\left(X_{t-1}X_{t-5}\right)\right\} \exp\left\{f_7\left(X_{t-3}^2 X_{t-4}^2\right)\right\}.$$

The model above is able to capture all the remaining dependence, as we can note by the fourth row of diagrams in Figure 15.

7 Conclusions

Serial independence is typically tested by statistics which only account for pairwise dependencies (i.e. those involving pairs of lagged variables); a classical example, considered as a benchmark herein, is the Portmanteau approach that, as shown in Section 3.1 and 5.3, can be blind with respect to non-pairwise dependencies. A test statistic, which overcomes this problem, is proposed in this paper. It is simple to compute, being based on a multi-way contingency table, and has an asymptotic χ^2 distribution with known degrees of freedom under the null hypothesis of serial independence. Simulation results showed that the corresponding serial independence test maintains the size even when the data generating process does not have moments and that it is powerful for a wide variety of linear and nonlinear data generating processes. Meaningful are the results related to the Bilinear AR and Multiplicative MA models (see Section 5.3) where the new test statistic is able to reveal dependencies that are not perceived by the classical Portmanteau tests. The simulation study also highlights that, when applied on residuals from fitted models, the new test does not require any correction for the estimation effect (differently from the commonly used Ljung-Box test which is based on the Portmanteau approach). The application on the financial data of Section 6 demonstrates how the proposed test can be used, in synergy with the classical Portmanteau tests, in a finer identification of the true data generating process.

Appendix

A Some tables from Sections 2.1 and 4.3

In the following, we collect some Tables pertaining to Sections 2.1 and 4.3.

Table A.6: Preliminary scheme to determine the pairs to be considered for the single-lag testing problem for the generic lag l.

i	x_{i-l}	x_{i-l+1}	$\cdots$	x_{i-1}	x_i
1					x_1
2				x_1	x_2
$\vdots$			$\iddots$	$\vdots$	$\vdots$
l		x_1	$\cdots$	x_{l-1}	x_l
$l+1$	x_1	x_2	$\cdots$	x_l	x_{l+1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	x_{i-l}	x_{i-l+1}	$\cdots$	x_{i-1}	x_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n-1$	x_{n-1-l}	x_{n-l}	$\cdots$	x_{n-2}	x_{n-1}
n	x_{n-l}	x_{n-l+1}	$\cdots$	x_{n-1}	x_n

Table A.7: Example 2.1: Preliminary scheme to determine the pairs to be considered for the testing problem of lag l.

(a) $l = 3$

i	x_{i-3}	x_{i-2}	x_{i-1}	x_i
1				0.217
2			0.217	-0.542
3		0.217	-0.542	0.891
4	0.217	-0.542	0.891	0.596
5	-0.542	0.891	0.596	1.636
6	0.891	0.596	1.636	0.689
7	0.596	1.636	0.689	-1.281
8	1.636	0.689	-1.281	-0.213
9	0.689	-1.281	-0.213	1.897
10	-1.281	-0.213	1.897	1.777
11	-0.213	1.897	1.777	0.567
12	1.897	1.777	0.567	0.016
13	1.777	0.567	0.016	0.383
14	0.567	0.016	0.383	-0.045
15	0.016	0.383	-0.045	0.034
16	0.383	-0.045	0.034	0.169
17	-0.045	0.034	0.169	1.165
18	0.034	0.169	1.165	-0.044
19	0.169	1.165	-0.044	-0.100
20	1.165	-0.044	-0.100	-0.283
21	-0.044	-0.100	-0.283	1.541
22	-0.100	-0.283	1.541	0.165
23	-0.283	1.541	0.165	1.308
24	1.541	0.165	1.308	1.288
25	0.165	1.308	1.288	0.593

(b) $l = 2$

i	x_{i-2}	x_{i-1}	x_i
1			0.217
2		0.217	-0.542
3	0.217	-0.542	0.891
4	-0.542	0.891	0.596
5	0.891	0.596	1.636
6	0.596	1.636	0.689
7	1.636	0.689	-1.281
8	0.689	-1.281	-0.213
9	-1.281	-0.213	1.897
10	-0.213	1.897	1.777
11	1.897	1.777	0.567
12	1.777	0.567	0.016
13	0.567	0.016	0.383
14	0.016	0.383	-0.045
15	0.383	-0.045	0.034
16	-0.045	0.034	0.169
17	0.034	0.169	1.165
18	0.169	1.165	-0.044
19	1.165	-0.044	-0.100
20	-0.044	-0.100	-0.283
21	-0.100	-0.283	1.541
22	-0.283	1.541	0.165
23	1.541	0.165	1.308
24	0.165	1.308	1.288
25	1.308	1.288	0.593

Table A.8: Example 4.1. Preliminary scheme to determine the triples to be considered for the multiple-lag testing problem for the set of lags $L = \{2, 3\}$.

i	x_{i-3}	x_{i-2}	x_{i-1}	x_i
1				0.217
2			0.217	-0.542
3		0.217	-0.542	0.891
4	0.217	-0.542	0.891	0.596
5	-0.542	0.891	0.596	1.636
6	0.891	0.596	1.636	0.689
7	0.596	1.636	0.689	-1.281
8	1.636	0.689	-1.281	-0.213
9	0.689	-1.281	-0.213	1.897
10	-1.281	-0.213	1.897	1.777
11	-0.213	1.897	1.777	0.567
12	1.897	1.777	0.567	0.016
13	1.777	0.567	0.016	0.383
14	0.567	0.016	0.383	-0.045
15	0.016	0.383	-0.045	0.034
16	0.383	-0.045	0.034	0.169
17	-0.045	0.034	0.169	1.165
18	0.034	0.169	1.165	-0.044
19	0.169	1.165	-0.044	-0.100
20	1.165	-0.044	-0.100	-0.283
21	-0.044	-0.100	-0.283	1.541
22	-0.100	-0.283	1.541	0.165
23	-0.283	1.541	0.165	1.308
24	1.541	0.165	1.308	1.288
25	0.165	1.308	1.288	0.593

B Model (9): theoretical results about serial dependence

To ease the reader, we recall the definition of model (9)

$$X_t = \text{sign}\left(\varepsilon_{t-1}\varepsilon_{t-2}\right) + \varepsilon_t, \tag{B.16}$$

where $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is a sequence of independent standard normals. The process defined by (B.16) is strictly stationary (it is a measurable transformation of the strictly stationary process $\{\varepsilon_t\}$) and the marginal distribution of X_t is given by

$$\begin{aligned} P(X_t \leq x) &= \frac{1}{2}P(X_t \leq x|\varepsilon_{t-1}\varepsilon_{t-2} < 0) + \frac{1}{2}P(X_t \leq x|\varepsilon_{t-1}\varepsilon_{t-2} > 0) \\ &= \frac{1}{2}P(\varepsilon_t \leq x+1) + \frac{1}{2}P(\varepsilon_t \leq x-1) \\ &= \frac{1}{2}\left[\Phi(x+1) + \Phi(x-1)\right], \end{aligned} \tag{B.17}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal. Expression (B.17) clearly exhibits that X_t is a mixture of two normal distributions with unit variance and means -1 and $+1$.

B.1 Pairwise independence

The pairwise independence in model (B.16) is directly proved by showing that

$$P(X_{t-l} \leq x \cap X_t \leq y) = P(X_{t-l} \leq x)P(X_t \leq y), \qquad \forall (x,y) \in \mathbb{R}^2 \text{ and } l = 1, 2, \ldots.$$

The condition above is trivially verified if $l \geq 3$, but the cases $l = 1$ and $l = 2$ remain to be analyzed. Only the case $l = 1$ is here considered because $l = 2$ can be handled in a similar way.

Let E_i, $i = 1, \ldots, 16$, be the events corresponding to all the possible signs configurations of the random variables $(\varepsilon_{t-3}, \varepsilon_{t-2}, \varepsilon_{t-1}, \varepsilon_t)$ defining X_{t-1} and X_t. Thanks to the law of total probability we have that

$$P\left(X_{t-1} \leq x \cap X_t \leq y\right) = \sum_{i=1}^{16} P\left(X_{t-1} \leq x \cap X_t \leq y \cap E_i\right).$$

Now, assuming that

$$E_1 = \left\{\varepsilon_{t-3} > 0 \cap \varepsilon_{t-2} > 0 \cap \varepsilon_{t-1} > 0 \cap \varepsilon_t > 0\right\},$$

it is easy to show that

$$\{X_{t-1} \leq x \cap X_t \leq y \cap E_1\} \equiv \begin{cases} \tilde{E}_1 & \text{if } \min(x-1, y-1) > 0 \\ \emptyset & \text{otherwise} \end{cases},$$

where

$$\tilde{E}_1 = \{\varepsilon_{t-3} > 0 \cap \varepsilon_{t-2} > 0 \cap 0 < \varepsilon_{t-1} < (x-1) \cap 0 < \varepsilon_t < (y-1)\}.$$

Hence, $P(X_{t-1} \leq x \cap X_t \leq y \cap E_1)$ is equal to $P_1 = P(\tilde{E}_1) = \frac{1}{4}\left[\Phi(x-1) - \frac{1}{2}\right]\left[\Phi(y-1) - \frac{1}{2}\right]$ if *condition* $\min(x-1, y-1) > 0$ is satisfied and 0 otherwise. In Table B.9 all the events E_i, $i = 1, \ldots, 16$, are reported along with the corresponding non-trivial representation $\tilde{E}_i$ of the set $\{X_{t-1} \leq x \cap X_t \leq y \cap E_i\}$ obtained under an appropriate *condition*.

i	E_i				$\tilde{E}_i$				*condition*	P_i
	ε_{t-3}	ε_{t-2}	ε_{t-1}	ε_t	ε_{t-3}	ε_{t-2}	ε_{t-1}	ε_t		
1	$+$	$+$	$+$	$+$	$+$	$+$	$\in (0, x-1)$	$\in (0, y-1)$	$\min(x-1; y-1) > 0$	$\frac{1}{4}\left[\Phi(x-1) - \frac{1}{2}\right]\left[\Phi(y-1) - \frac{1}{2}\right]$
2	$+$	$+$	$+$	$-$	$+$	$+$	$\in (0, x-1)$	$< \min(0; y-1)$	$(x-1) > 0$	$\frac{1}{4}\left[\Phi(x-1) - \frac{1}{2}\right]\min\left[\Phi(y-1); \frac{1}{2}\right]$
3	$+$	$+$	$-$	$+$	$+$	$+$	$< \min(0; x-1)$	$\in (0, y+1)$	$(y+1) > 0$	$\frac{1}{4}\min\left[\Phi(x-1); \frac{1}{2}\right]\left[\Phi(y+1) - \frac{1}{2}\right]$
4	$+$	$+$	$-$	$-$	$+$	$+$	$< \min(0; x-1)$	$< \min(0; y+1)$	*none*	$\frac{1}{4}\min\left(\Phi(x-1); \frac{1}{2}\right]\min\left(\Phi(y+1); \frac{1}{2}\right]$
5	$+$	$-$	$+$	$+$	$+$	$-$	$\in (0, x+1)$	$< \min(0; y+1)$	$\min(x+1; y+1) > 0$	$\frac{1}{4}\left[\Phi(x+1) - \frac{1}{2}\right]\left[\Phi(y+1) - \frac{1}{2}\right]$
6	$+$	$-$	$+$	$-$	$+$	$-$	$\in (0, x+1)$	$< \min(0; y+1)$	$(x+1) > 0$	$\frac{1}{4}\left[\Phi(x+1) - \frac{1}{2}\right]\min\left[\Phi(y+1); \frac{1}{2}\right]$
7	$+$	$-$	$-$	$+$	$+$	$-$	$< \min(0; x+1)$	$< \min(0; y-1)$	$(y-1) > 0$	$\frac{1}{4}\min\left[\Phi(x+1); \frac{1}{2}\right]\left[\Phi(y-1) - \frac{1}{2}\right]$
8	$+$	$-$	$-$	$-$	$+$	$-$	$< \min(0; x+1)$	$< \min(0; y-1)$	*none*	$\frac{1}{4}\min\left[\Phi(x+1); \frac{1}{2}\right]\min\left[\Phi(y-1); \frac{1}{2}\right]$
9	$-$	$+$	$+$	$+$	$-$	$+$	$\in (0, x+1)$	$\in (0, y-1)$	$\min(x+1; y-1) > 0$	$\frac{1}{4}\left[\Phi(x+1) - \frac{1}{2}\right]\left[\Phi(y-1) - \frac{1}{2}\right]$
10	$-$	$+$	$+$	$-$	$-$	$+$	$\in (0, x+1)$	$< \min(0; y-1)$	$(x+1) > 0$	$\frac{1}{4}\left[\Phi(x+1) - \frac{1}{2}\right]\min\left[\Phi(y-1); \frac{1}{2}\right]$
11	$-$	$+$	$-$	$+$	$-$	$+$	$< \min(0; x+1)$	$\in (0, y+1)$	$(y+1) > 0$	$\frac{1}{4}\min\left[\Phi(x+1); \frac{1}{2}\right]\left[\Phi(y+1) - \frac{1}{2}\right]$
12	$-$	$+$	$-$	$-$	$-$	$+$	$< \min(0; x+1)$	$< \min(0; y+1)$	*none*	$\frac{1}{4}\min\left[\Phi(x+1); \frac{1}{2}\right]\min\left[\Phi(y+1); \frac{1}{2}\right]$
13	$-$	$-$	$+$	$+$	$-$	$-$	$\in (0, x-1)$	$\in (0, y+1)$	$\min(x-1; y+1) > 0$	$\frac{1}{4}\left[\Phi(x-1) - \frac{1}{2}\right]\left[\Phi(y+1) - \frac{1}{2}\right]$
14	$-$	$-$	$+$	$-$	$-$	$-$	$\in (0, x-1)$	$< \min(0; y+1)$	$(x-1) > 0$	$\frac{1}{4}\left[\Phi(x-1) - \frac{1}{2}\right]\min\left[\Phi(y+1); \frac{1}{2}\right]$
15	$-$	$-$	$-$	$+$	$-$	$-$	$< \min(0; x-1)$	$\in (0, y-1)$	$(y-1) > 0$	$\frac{1}{4}\min\left[\Phi(x-1); \frac{1}{2}\right]\left[\Phi(y-1) - \frac{1}{2}\right]$
16	$-$	$-$	$-$	$-$	$-$	$-$	$< \min(0; x-1)$	$< \min(0; y-1)$	*none*	$\frac{1}{4}\min\left[\Phi(x-1); \frac{1}{2}\right]\min\left[\Phi(y-1); \frac{1}{2}\right]$

Table B.9: Definition of the set $\tilde{E}_i$ along with the related *condition* and probability P_i, $i = 1, \ldots, 16$.

Now, by using the information in Table B.9, it is possible to verify that $P(X_{t-1} \leq x \cap X_t \leq y) = P(X_{t-1} \leq x) P(X_t \leq y)$ for all $(x, y) \in \mathbb{R}^2$. For example, assume that $x + 1 < 0$ and $y + 1 < 0$ and, consequently, $x - 1 < 0$ and $y - 1 < 0$. From the conditions in Table B.9 it follows that

$$\begin{aligned}
P(X_{t-1} \le x \cap X_t \le y) &= \sum_{i=1}^{16} P(X_{t-1} \le x \cap X_t \le y \cap E_i) \\
&= \sum_{i \in \{4,8,12,16\}} P_i \\
&= \left\{ \frac{1}{2} \left[\Phi(x-1) + \Phi(x+1) \right] \right\} \left\{ \frac{1}{2} \left[\Phi(y-1) + \Phi(y+1) \right] \right\} \\
&= P\left(X_{t-1} \le x\right) P\left(X_t \le y\right).
\end{aligned}$$

B.2 Dependence on the set of lags $\{1, 2\}$

To prove the dependence of X_t from (X_{t-1}, X_{t-2}) it is sufficient to prove that X_t and $Y_t = X_{t-1}X_{t-2}$ are correlated, that is

$$\begin{aligned}
\mathrm{Cov}(X_t, Y_t) &= \mathrm{E}(X_t Y_t) - \mathrm{E}(X_t)\mathrm{E}(Y_t) \\
&= \mathrm{E}(X_t Y_t) \\
&= \mathrm{E}(X_t X_{t-1} X_{t-2}) \\
&= \mathrm{E}\left\{ \left[\mathrm{sign}(\varepsilon_{t-1}\varepsilon_{t-2}) + \varepsilon_t \right] \left[\mathrm{sign}(\varepsilon_{t-2}\varepsilon_{t-3}) + \varepsilon_{t-1} \right] \left[\mathrm{sign}(\varepsilon_{t-3}\varepsilon_{t-4}) + \varepsilon_{t-2} \right] \right\} \\
&= \mathrm{E}[\varepsilon_{t-1}\varepsilon_{t-2} \mathrm{sign}(\varepsilon_{t-1}\varepsilon_{t-2})] \\
&= \mathrm{E}\left(|\varepsilon_{t-1}\varepsilon_{t-2}| \right) > 0.
\end{aligned}$$

B.3 Dependence on the set of lags $\{1, 3\}$

To prove the dependence of X_t from (X_{t-1}, X_{t-3}) it is sufficient to prove that X_t and $Z_t = X_{t-1}^2 X_{t-3}$ are correlated, that is

$$\begin{aligned}
\mathrm{Cov}(X_t, Z_t) &= \mathrm{E}(X_t Z_t) - \mathrm{E}(X_t)\mathrm{E}(Z_t) \\
&= \mathrm{E}(X_t Z_t) \\
&= \mathrm{E}(X_t X_{t-2}^2 X_{t-3}) \\
&= \mathrm{E}\left\{ \left[\mathrm{sign}(\varepsilon_{t-1}\varepsilon_{t-2}) + \varepsilon_t \right] \left[\mathrm{sign}(\varepsilon_{t-2}\varepsilon_{t-3}) + \varepsilon_{t-1} \right]^2 \left[\mathrm{sign}(\varepsilon_{t-4}\varepsilon_{t-5}) + \varepsilon_{t-3} \right] \right\} \\
&= 2\mathrm{E}[\varepsilon_{t-1}\varepsilon_{t-3} \mathrm{sign}(\varepsilon_{t-1}\varepsilon_{t-2}) \mathrm{sign}(\varepsilon_{t-2}\varepsilon_{t-3})] \\
&= 2\mathrm{E}[\varepsilon_{t-1}\varepsilon_{t-3} \mathrm{sign}(\varepsilon_{t-1}\varepsilon_{t-3})] \\
&= \mathrm{E}\left(|\varepsilon_{t-1}\varepsilon_{t-3}| \right) > 0.
\end{aligned}$$

B.4 Independence on the set of lags $\{l_1, l_2\}$ different from $\{1,2\}$ and $\{1,3\}$

Here, it is shown that X_t is independent from (X_{t-2}, X_{t-3}). A similar procedure can be followed to show that X_t is independent from (X_{t-l_1}, X_{t-l_2}) for all the pairs (l_1, l_2) different from $(1,2)$ and $(1,3)$. Following Kallenberg (2009), the independence between X_t and (X_{t-2}, X_{t-3}) can be proved showing that $E(X_t^p X_{t-2}^q X_{t-3}^s) = E(X_t^p)E(X_{t-2}^q)E(X_{t-3}^s)$ for all $p, q, s \in \mathbb{N}_+$. Using the notation $S_t = \text{sign}(\varepsilon_t)$, it is possible to note that $\text{sign}(\varepsilon_t \varepsilon_{t-l}) = \text{sign}(\varepsilon_t)\text{sign}(\varepsilon_{t-l})$ and $E(S_t^j) = 0$ if j is odd. From the binomial theorem it results that

$$E\left(X_t^p X_{t-2}^q X_{t-3}^s\right) = E\left[\sum_{j=0}^{p} \binom{p}{j} S_{t-1}^{p-j} S_{t-2}^{p-j} \varepsilon_t^j \sum_{h=0}^{q} \binom{q}{h} S_{t-3}^{q-h} S_{t-4}^{q-h} \varepsilon_{t-2}^h \sum_{k=0}^{m} \binom{m}{k} S_{t-4}^{m-k} S_{t-5}^{m-k} \varepsilon_{t-3}^k\right]$$
$$= \sum_{j=0}^{p}\sum_{h=0}^{q}\sum_{k=0}^{m} \binom{p}{j}\binom{q}{h}\binom{m}{k} E(S_{t-5}^{m-k})E(S_{t-4}^{q-h+m-k})E(S_{t-3}^{q-h}\varepsilon_{t-3}^k)E(S_{t-2}^{p-j}\varepsilon_{t-2}^h)E(S_{t-1}^{p-j})E(\varepsilon_t^j).$$

The expression above reveals that the addends in the triple summation with at least one of the indexes j, $p-j$, $m-k$ odd, are null. Moreover, the addends with $q-h$ odd and $m-k$ even are null too since, in this case, the exponent $q-h+m-k$ of S_{t-4} is odd. Then: all the addends in the triple summation with at least one of the indexes j, $p-j$, $m-k$, and $q-h$ odd, are null. The remaining addends are characterized by even values of j, $p-j$, $m-k$, and $q-h$. By noting that $S_t^a = 1$ with probability one if a is even, these remaining addends have the following form

$$E(\varepsilon_{t-3}^k)E(\varepsilon_{t-2}^h)E(\varepsilon_t^j).$$

The addends of the form described above are null if one of the indexes k, h, and j is odd. Now, note that if one of the exponents, p, q, or m is odd, then also all the addends with k, h, and j even are null since, in this case, at least one indexes among $p-j$, $m-k$, and $q-h$ is odd. Consequently

$$E(X_t^p X_{t-2}^q X_{t-3}^s) = 0 = E(X_t^p)E(X_{t-2}^q)E(X_{t-3}^s) \text{ if one of the values } p,\, q, \text{ or } m \text{ is odd.} \tag{B.18}$$

To the contrary, if all the exponents p, q, and m are even, then $p-j$, $m-k$, and $q-h$ are all even if and only if j, k, h are all even. Consequently:

$$E(X_t^p X_{t-2}^q X_{t-3}^s) = E\left[\sum_{j=0}^{p/2}\binom{p}{2j}\sum_{h=0}^{q/2}\binom{q}{2h}\sum_{k=0}^{m/2}\binom{m}{2k}E(\varepsilon_{t-3}^{2k})E(\varepsilon_{t-2}^{2h})E(\varepsilon_t^{2j})\right]$$
$$= E\left[\sum_{j=0}^{p}\binom{p}{j}\sum_{h=0}^{q}\binom{q}{h}\sum_{k=0}^{m}\binom{m}{k}E(\varepsilon_{t-3}^{k})E(\varepsilon_{t-2}^{h})E(\varepsilon_t^{j})\right]$$
$$= E\left\{\left[\sum_{j=0}^{p}\binom{p}{j}\varepsilon_t^j\right]\left[\sum_{h=0}^{q}\binom{q}{h}\varepsilon_{t-2}^h\right]\left[\sum_{k=0}^{m}\binom{m}{k}\varepsilon_{t-3}^k\right]\right\}$$
$$= E\left[(1+\varepsilon_t)^p(1+\varepsilon_{t-2})^q(1+\varepsilon_{t-3})^m\right] = E\left[(1+\varepsilon_t)^p\right]E\left[(1+\varepsilon_{t-2})^q\right]E\left[(1+\varepsilon_{t-3})^m\right] \ .$$

Using, once again, the binomial theorem, it is possible to prove that, if j is even, thus $E(X_t^j) = E[(1+\varepsilon_t)^j]$. If p, q, and k are all even, then $E\left(X_t^p X_{t-2}^q X_{t-3}^s\right) = E\left(X_t^p\right)E\left(X_{t-2}^q\right)E\left(X_{t-3}^s\right)$. This last case, in addition to (B.18), prove that $E\left(X_t^p X_{t-2}^q X_{t-3}^s\right) = E\left(X_t^p\right)E\left(X_{t-2}^q\right)E\left(X_{t-3}^s\right)$ for all $p, q, s \in \mathbb{N}_+$.

References

Agresti, A. (2002). Categorical data analysis. *A John Wiley and Sons, Inc. Publication, Hoboken, New Jersey, USA.*

Anderson, H. M. and F. Vahid (2005). Nonlinear correlograms and partial autocorrelograms. *Oxford Bulletin of Economics and Statistics 67*, 957–982.

Bagnato, L., L. De Capitani, A. Mazza, and A. Punzo (2015). **SDD**: An R package for serial dependence diagrams. *Journal of Statistical Software 64*(Code Snippet 2), 1–19.

Bagnato, L. and A. Punzo (2010). On the use of χ^2-test to check serial independence. *Statistica & Applicazioni VIII*(1), 57–74.

Bagnato, L., A. Punzo, and O. Nicolis (2012). The autodependogram: a graphical device to investigate serial dependences. *Journal of Time Series Analysis 33*(2), 233–254.

Bellman, R. (1961). *Adaptive Control Process.* Princeton University Press, Princeton.

Bollerslev, T., R. Chou, and K. Kroner (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics 52*(1), 5–59.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics 10*(4), 417–451.

Diks, C. (2009). Nonparametric tests for independence. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science*, pp. 6252–6271. New York: Springer.

Genest, C. and B. Rémillard (2004). Test of independence and randomness based on the empirical copula process. *Test 13*(2), 335–369.

Hall, P. and R. Wolff (1995). On the strength of dependence of a time series generated by a chaotic map. *Journal of Time Series Analysis 16*(6), 571–583.

Hallin, M. and G. Mélard (1988). Rank-based tests for randomness against first-order serial dependence. *Journal of the American Statistical Association 83*(404), 1117–1128.

Hastie, T. (2013). ***gam**: Generalized Additive Models.* R package, Version 1.12, available at `http://CRAN.R-project.org/package=gam`.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika 75*(4), 800–802.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika 75*(2), 383–386.

Kallenberg, W. C. (2009). Estimating copula densities, using model selection techniques. *Insurance: mathematics and economics 45*(2), 209–223.

King, M. (1987). Testing for autocorrelation in linear regression models: a survey. In M. L. King and D. E. A. Giles (Eds.), *Specification analysis in the linear model*, London, pp. 19–73. Routledge Kegan & Paul.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

Mann, H. B. and A. Wald (1942). On the choice of the number of class intervals in the application of the chi square test. *The Annals of Mathematical Statistics 13*(3), 306–317.

Rao, T. S. (1981). On the theory of bilinear time series models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 244–255.

Robinson, P. M. (1991). Consistent nonparametric entropy-based testing. *The Review of Economic Studies 58*(3), 437–453.

Romano, J. P. and L. A. Thombs (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association 91*(434), 590–600.

R Core Team (2015). R*: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika 73*(3), 751–754.

Skaug, H. J. and D. Tjøstheim (1993). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika 80*(3), 591–602.

Trapletti, A., K. Hornik, and B. LeBaron (2015). ***tseries****: Time Series Analysis and Computational Finance.* R package, Version 0.10-34, available at `http://CRAN.R-project.org/package=tseries`.

Verbeek, M. (2000). *A Guide to Modern Econometrics.* New York: Wiley.

Wright, S. P. (1992). Adjusted p-values for simultaneous inference. *Biometrics 48*(4), 1005–1013.

Wuertz, D. and Y. Chalabi (2013). ***fGarch****: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling.* R package, Version 3010.82, available at `http://CRAN.R-project.org/package=fgarch`.

Yang, L., W. Härdle, and J. P. Nielsen (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis 20*(5), 579–604.

Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis 33*(3), 438–457.