Improving the autodependogram using the Kulback-Leibler divergence

Luca Bagnato * Lucio De Capitani † Antonio Punzo ‡

Summary. The autodependogram is a graphical device recently proposed in the literature to analyze autodependencies. It is defined computing the classical Pearson χ^2-statistics of independence at various lags in order to point out the presence lag-depedencies. This paper proposes an improvement of this diagram obtained by substituting the χ^2-statistics with an estimator of the Kulback-Leibler divergence between the bivariate density of two delayed variables and the product of their marginal distributions. A simulation study, on well-established time series models, shows that this new autodependogram is more powerful than the previous one. An application to financial data is also shown.

Key words: Nonlinear time series, Kulback-Leibler divergence, Autocorrelogram, Autodependogram, Permutations, Gaussian kernel.

1 Introduction

Despite the fast advancement in nonlinear time-series models, there are few tools which can explore the complex dependence structures in nonlinear time-series like the autocorrelogram does for the linear ones. Starting from these considerations, Anderson and Vahid (2005), Bagnato *et al.* (2012), and Zhou (2012), have recently proposed graphical tools which are very similar, in aspect and intent, to the autocorrelogram but they are widely applicable to both linear and nonlinear time series. The diagram proposed in Bagnato *et al.* (2012) is called autodependogram and it is based on the lag-independence test proposed in Bagnato and Punzo (2010, see also Bagnato and Punzo, 2012) which is, roughly, a serial version of the well-known Pearson χ^2-test of independence. Analogously to the autocorrelogram, the autodependogram is obtained by representing the lags on the *x*-axis and the values of the χ^2-statistic, corresponding to each lag, on the *y*-axis. Moreover, it is furnished with a critical line which emphasizes the lags showing evidence against the null hypothesis of independence.

However, as concluded by Bagnato *et al.* (2013b) through a deep simulation study, in the serial context the χ^2-test is outperformed by other independence tests based on measures of

*Università Cattolica del Sacro Cuore, Dipartimento di Discipline Matematiche, Finanza Matematica e Econometria, Via Necchi 9, 20100, Milano, Italy, e.mail: luca.bagnato@unicatt.it

†Università di Milano-Bicocca, Dipartimento di Statistica e Metodi Quantitativi, Via Bicocca degli Arcimboldi 8, 20126, Milano, Italy e.mail: lucio.decapitani1@unimib.it

‡*Corresponding Author* - Università di Catania, Dipartimento di Economia e Impresa, Corso Italia 55, 95129, Catania, Italy, e.mail: antonio.punzo@unict.it

divergence between the estimated joint density and the product of the estimated marginal densities. Among them, the test based on the well-known Kullback-Leibler divergence appeared to be the best performer. This paper shows, by simulations, how the use of this test can substantially improve the performance of the autodependogram. The improvement is meant in terms of power of the autodependogram bars under the alternative hypothesis of lag-dependence. The size of the bars (under the null) is also taken into account in the simulations.

The paper is organized as follows. Section 2 recalls the autodependogram. In Section 3 the preliminary concepts for the definition of the new diagram, described in Section 4, are introduced. Performance and behavior of this diagram, in comparison with the autodependogram and the autocorrelogram, are studied in Section 5 and Section 6 via an application to a financial time series and a simulation study on well-known time series models, respectively. Conclusions are given in Section 7.

2 The autodependogram

Let $\{X_t\}_{t\in\mathbb{N}}$ represent a strictly stationary and ergodic stochastic process. Moreover, let $(X_1,\ldots,X_n)$ be an observed time series of length n from $\{X_t\}_{t\in\mathbb{N}}$. To study the generic dependence of lag r, let's consider the $n_r = n-r$ couples $\{(X_i,X_{i+r})\}_{i=1}^{n_r}$.

Bagnato and Punzo (2010) propose to group these couples $\{(X_i,X_{i+r})\}_{i=1}^{n_r}$ in a $k\times k$ contingency table and to test the independence of lag r through the well-known χ^2-test statistic:

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

where $n_{uv}^{(r)}$ denotes the number of couples in $\{(X_i,X_{i+r})\}_{i=1}^{n_r}$ belonging to the (u,v)-cell of the contingency table, $n_{u\cdot}^{(r)}$ ($n_{\cdot v}^{(r)}$) denotes the number of values in $\{X_i\}_{i=1}^{n_r}$ ($\{X_{i+r}\}_{i=1}^{n_r}$) belonging to the u-th row (v-th column) of the contingency table, and $\widehat{n}_{uv}^{(r)} = n_{\cdot v}^{(r)}n_{u\cdot}^{(r)}/n_r$ are the theoretical frequencies under the (null) hypothesis of independence of lag r. In Bagnato *et al.* (2012) it is shown by simulations that, also in the serial context, the limiting null distribution of $\widehat{\delta}_r$ is χ^2 with $(k-1)^2$ degrees of freedom. This fact allows to test the null hypothesis of independence for lag r using $\widehat{\delta}_r$ 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 $\widehat{\delta}_r > \chi^2_{[(k-1)^2;1-\alpha]}$.

In conformity with the *autocorrelogram* (ACF), the diagram obtained by plotting $\widehat{\delta}_r$ as a function of the time lags r, $r=1,\ldots,l$, is called *autodependogram* (ADF) by Bagnato *et al.* (2012). Naturally, the autodependogram can be applied to time series with missing data and can easily be re-adapted to evaluate cross-dependencies between two different time series.

To completely specify $\widehat{\delta}_r$, it is necessary to define the two marginal partitions in the contingency table. Bagnato *et al.* (2012) suggest to use the so-called *equifrequency intervals*, which assigns equal frequencies to each interval. Using the equifrequency intervals, only the value of k remains to be selected. In Bagnato *et al.* (2012) k is chosen such that

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

where $\lfloor \cdot \rfloor$ denotes the floor function while $z_{1-\alpha}$ stands for the $(1-\alpha)$-quantile of the standard normal distribution. The value of k_s is chosen to assure a sufficient adherence between the actual and nominal sizes of the test and it is a stronger version of the rule of Cochran (1954) requiring at least 5 expected frequencies in each cell of the contingency table. The value of k_p is an adaptation, to the contingency tables, of the well-known formula of Mann and Wald (1942) introduced in order to maximize the power of the χ^2-test of goodness-of-fit. Simulations in Bagnato *et al.* (2012, Section 5.1) have confirmed the validity of the rule (2).

Since k is fixed and equal for each lag $r = 1, \ldots, l$, the *level*-α critical value $\chi^2_{\left[(k-1)^2;1-\alpha\right]}$ is used for each of the l tests of lag-independence. This allow to add an horizontal line, at height $\chi^2_{\left[(k-1)^2;1-\alpha\right]}$, to the autodependogram. In conformity with the autocorrelogram, the usefulness of such a line (hereafter referred to as level-α *critical line* or, simply, *critical line*) is clear since it provides a graphical counterpart of the acceptance/rejection regions of the χ^2-tests of lag-independence for each lag.

Although the autodependogram is very similar in aspect to the autocorrelogram, the former is able to capture various and general serial dependence structures while the latter points out only linear relationships. This feature motivates the adjective *omnibus* used in Bagnato *et al.* (2012). Obviously, this generality is paid for in terms of power, and hence in the resulting descriptive ability under certain dependence structures. For example, the autodependogram will be less informative with respect to the autocorrelogram when the lag-dependencies are linear, or with respect to a representation using rank correlation statistics when the lag-dependencies are monotonic.

3 Kullback-Leibler divergence in serial independence tests

In addition to the requirements introduced at the beginning of Section 2, suppose that X_1 is continuous with density g and support $\mathcal{S}$. Moreover, let f_r be the joint density function of (X_1, X_{1+r}). The presence of dependence for lag r can be checked by testing the statistical hypothesis

$$H_0^r: \; f_r(x,y) = g(x)\, g(y) \qquad \forall (x,y) \in \mathcal{S}^2. \tag{3}$$

To evaluate the discrepancy between $f_r(x,y)$ and $g(x)\, g(y)$, with the aim to obtain a test statistic for the testing problem (3), several divergence functionals can be considered (see, e.g., Diks, 2009). The results of a wide simulation study on several data generating processes conducted by Bagnato *et al.* (2013b) show that, among the most widespread divergences, the Kullback-Leibler (KL) functional

$$\Delta_r = \int_{\mathcal{S}^2} \log\left(\frac{f_r(x,y)}{g(x)g(y)}\right) f_r(x,y)\,dx\,dy, \qquad r = 1, \ldots, l, \tag{4}$$

seems to be the best performer; for examples of serial independence tests based on (4), see Robinson (1991) and Hong and White (2005). Naturally, the KL divergence functional satisfies the relations: (i) $\Delta_r \geq 0$; (ii) $\Delta_r = 0$ if and only if $f_r = g \cdot g$ almost surely. In particular, the greater the value of Δ_r, the stronger the dependence for lag r. So, intuition suggests that the testing problem (3) can be solved using an estimator $\widehat{\Delta}_r$ of Δ_r as test statistic and that the null hypothesis should be rejected for large values of $\widehat{\Delta}_r$.

Several methodologies are proposed in the literature to implement $\widehat{\Delta}_r$; the differences among the various approaches stem from the way: (i) the densities f_r and g are estimated (ii) the integral in (4) is computed (iii) the p-values are obtained. Among the alternatives considered by Bagnato *et al.* (2013b), the Gaussian kernel density estimator, the numerical integration (like in Granger *et al.*, 2004), and the permutation approach, appear to be the best solutions for (i), (ii) and (iii), respectively. Details on these settings are given below.

3.1 Gaussian kernel density estimator

The Gaussian Kernel (GK) estimator for the univariate density g is

$$\widehat{g}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x; X_i), \tag{5}$$

where $K_h(x; X_i) = (2\pi h^2)^{-1/2} \exp\left\{-\frac{1}{2}\left[h^{-1}(x - X_i)\right]^2\right\}$ and $h > 0$ is the bandwidth. Similarly, the GK estimator for the bivariate density f_r is

$$\widehat{f}_r(x,y) = \frac{1}{n-r}\sum_{i=1}^{n-r} K_h(x; X_i) K_h(y; X_{i+r}), \tag{6}$$

where K_h is defined as in (5). For simplicity, the bivariate kernel function in (6) is the product of two one-dimensional Gaussian kernels with equal bandwidths.

To apply the GK density estimator, a value for the bandwidth h needs to be chosen. With this aim, in literature, several data dependent procedures have been proposed. Examples are the Silverman's rule of thumb and the likelihood cross-validation method (see, e.g., Silverman, 1986, p. 52). The bandwidth obtained with the latter procedure, denoted with h_{LCV}, is particularly useful in this context because, as observed in Granger *et al.* (2004, p. 654), it produces optimal density estimators according to the KL criterion.

However, it is well-known that a bandwidth which is optimal for estimation is usually suboptimal for testing. In particular, although h_{LCV} suffices to establish consistency of the test statistics, this choice could not be optimal in terms of the power of the resulting tests. As observed by Anderson *et al.* (1994), in testing procedures a relative oversmoothing may be appropriate for some dependence functionals (test statistics). Nevertheless, the simulation results of Bagnato *et al.* (2013b) highlight that when the GK density estimator is adopted to define the estimator of Δ_r, the use of h_{LCV} is appropriate and, then, no oversmoothing is needed.

3.2 Estimation of the dependence functional

The simulation results of Bagnato *et al.* (2013b) point out that to estimate Δ_r a good solution consists plugging the GK density estimates of f_r and g into the definition of Δ_r obtaining the estimator

$$\widehat{\Delta}_r = \int_{S^2} \log\left(\frac{\widehat{f}_r(x,y)}{\widehat{g}(x)\,\widehat{g}(y)}\right) \widehat{f}_r(x,y)\,dxdy, \tag{7}$$

which can be computed by numerical integration (like in Granger *et al.*, 2004) or well approximated by a summation over a sufficiently fine grid of values.

In analogy to Bagnato *et al.* (2013b), the integral (7) is here approximated starting from a 100×100 grid of equally spaced values $\{(\tilde{x}_i, \tilde{y}_j) : i, j = 1, \ldots, 100\}$, obtaining the following estimator:

$$\widehat{\Delta}_r = 10^{-4} \sum_{i=1}^{100} \sum_{j=1}^{100} \log \left(\frac{\widehat{f}_r(x,y)}{\widehat{g}(x)\widehat{g}(y)} \right) \widehat{f}_r(x,y) \, \mathbb{1}\left((\tilde{x}_i, \tilde{y}_j) \in S_* \right), \tag{8}$$

where $\mathbb{1}(\cdot)$ is the indicator function, $\widehat{f}_r$ and $\widehat{g}$ are obtained through the GK estimator, and

$$S_* = \left\{ (\tilde{x}_i, \tilde{y}_j) : \widehat{f}(\tilde{x}_i, \tilde{y}_j) > 0, \, \widehat{g}(\tilde{x}_i) > 0, \, \widehat{g}(\tilde{y}_j) > 0 \right\}.$$

As concern the grid, the default settings of the `R` (R Development Core Team, 2012) package `sm` are followed:

$$\tilde{x}_i = \left(x_{(1)} - a \right) + (i-1) \frac{x_{(n)} - x_{(1)} + 2a}{99},$$

with $a = \left(x_{(n)} - x_{(1)} \right) / 4$ and $x_{(1)}$ ($x_{(n)}$) denoting the minimum (maximum) observed value. The grid for y is exactly the same.

3.3 Computing p-values

Once the test statistic $\widehat{\Delta}_r$ is defined, the p-value of the corresponding test has to be computed. Among the various proposals, the simulation study of Bagnato *et al.* (2013b) suggests that the Permutation approach represents a good compromise between simplicity and performance. It exploits the fact that, conditionally on the observed data $x_1, \ldots, x_n$, each of the possible $n!$ permutations is equally likely under the assumption of serial independence (see Diks, 2009). In detail, let $\widehat{\Delta}_r^{(0)}$ denote the value assumed by $\widehat{\Delta}_r$ for the observed data. Analogously, let $\widehat{\Delta}_r^{(b)}$ be the dependence functional estimate obtained from a random permutation of the original data, with $b = 1, \ldots, B$. Under the assumption of serial independence, $\widehat{\Delta}_r^{(1)}, \ldots, \widehat{\Delta}_r^{(B)}$ are equally likely and the p-value can be defined as in Diks and Panchenko (2007):

$$\widehat{q}_r = \frac{\# \left\{ \widehat{\Delta}_r^{(s)} : \widehat{\Delta}_r^{(s)} > \widehat{\Delta}_r^{(0)}; s = 0, 1, \ldots, B \right\} + L}{B+1}, \qquad r = 1, \ldots, l, \tag{9}$$

where L is defined as follows. Let $Z = \# \left\{ \widehat{\Delta}_r^{(s)} : \widehat{\Delta}_r^{(s)} = \widehat{\Delta}_r^{(0)}; s = 0, 1, \ldots, B \right\}$ denote the number of ties plus one. If $Z = 1$ then $L = 1$, while if $Z > 1$ then L is drawn form the discrete uniform distribution on $\{1, \ldots, Z\}$. The above procedure for computing the p-value leads to a randomized test having an exact level-α if the null is rejected whenever $\widehat{q}_r \leq \alpha$ and $0 < \alpha = c/(B+1) < 1$ for some integer c. However note that, the probability of the event $L > 1$ in (9) is practically null for GK because, in this case, ties can be observed only if the same permutation is drawn (or if a particular regularity is observed in the original series).

4 The KL-autodependogram

The autodependogram bars allow to rank the corresponding lags in terms of evidence of the presence of dependence. In these terms, an "equivalent" diagram could be obtained substituting $\widehat{\delta}_r$ with

$$\widetilde{p}_r = 1 - \widehat{p}_r, \qquad r = 1, \ldots, l, \tag{10}$$

where $\widehat{p}_r$ denotes the p-value associated to $\widehat{\delta}_r$:

$$\widehat{p}_r = 1 - F_{(k-1)^2}(\widehat{\delta}_r) \ ,$$

being F_η the cumulative χ^2 distribution with η degrees of freedom. Also in this case a level-α critical line at height $1-\alpha$ can be added to the resulting diagram. Note that, $\widetilde{p}_r$ unbalances the two decision regions by mapping the rejection one only on $(1-\alpha, 1]$. To face this issue, as suggested by Bagnato *et al.* (2013a), it is possible to introduce a particular monotonic decreasing transformation $p_r^* = g(\widehat{p}_r)$, on $[0,1]$, such that $g(\alpha) = 1/2$. In this way acceptance and rejection regions are mapped on $(1/2, 1]$ and $[0, 1/2]$, respectively. Here, like in Bagnato *et al.* (2013a), the following function is used

$$p_r^* = \begin{cases} \frac{2\alpha - \widehat{p}_r}{2\alpha} & \text{if } \widehat{p}_r < \alpha \\ \frac{1-\widehat{p}_r}{2(1-\alpha)} & \text{if } \widehat{p}_r \geq \alpha \end{cases} \ , \qquad r = 1, \ldots, l. \tag{11}$$

This transformations only modifies the aspect of the autodependogram leaving unchanged its power in detecting autodependencies. We propose a alternative graphical devices called KL-autodependogram (KL-ADF) which is built by substituting $\widehat{p}_r$ in (11) with the p-value $\widehat{q}_r$ defined in (9). This leads to the estimator

$$q_r^* = \begin{cases} \frac{2\alpha - \widehat{q}_r}{2\alpha} & \text{if } \widehat{q}_r < \alpha \\ \frac{1-\widehat{q}_r}{2(1-\alpha)} & \text{if } \widehat{q}_r \geq \alpha \end{cases} \ , \qquad r = 1, \ldots, l. \tag{12}$$

The diagram obtained by plotting the values of q_r^* as a function of r, $r = 1, \ldots, l$, will be denoted as KL-autodependogram hereafter.

5 An application to financial data

In this section an application to a financial time series is considered. The R-code to obtain the two diagrams of p_r^* and q_r^*, $r = 1, \ldots, l$, is available at `http://www.economia.unict.it/punzo`. In particular, we consider daily returns of the stock market index for the Borsa Italiana (FTSE MIB), spanning the period from 5 May 2011 to 27 November 2012 ($n = 400$ observations downloadable from `http://finance.yahoo.com/`; see Figure 1). The analysis is carried out by considering the sample of $n = 400$ observations but also the three most recent subsamples of size $n = 300$, $n = 200$, and $n = 100$.

For each considered value of n, Figure 2 shows the autocorrelogram on the squared raw data, and both the modified autodependogram (p_r^*) and the KL-autodependogram (q_r^*), $r = 1, \ldots 15$, computed on the raw data. The critical lines in the diagrams individuate the critical region of probability $\alpha = 0.05$.

As regards the case $n = 100$, it is possible to see from Figure 2(a) as a slight linear dependence of the squared data is detected by the autocorrelogram for the first and the sixth lag. The diagram of p_r^* in Figure 2(b) does not show these lag-dependences, while the KL-autodependogram in Figure 2(c) detects the dependence of lag 1. Generally, regardless from n, the KL-autodependogram is always roughly in agreement with the autocorrelogram on the squared raw data (see, for example, Figure 2(j) and Figure 2(l)). Furthermore, it tends to detect more lag-dependecies than the diagram of p_r^* (compare the second and the third column of

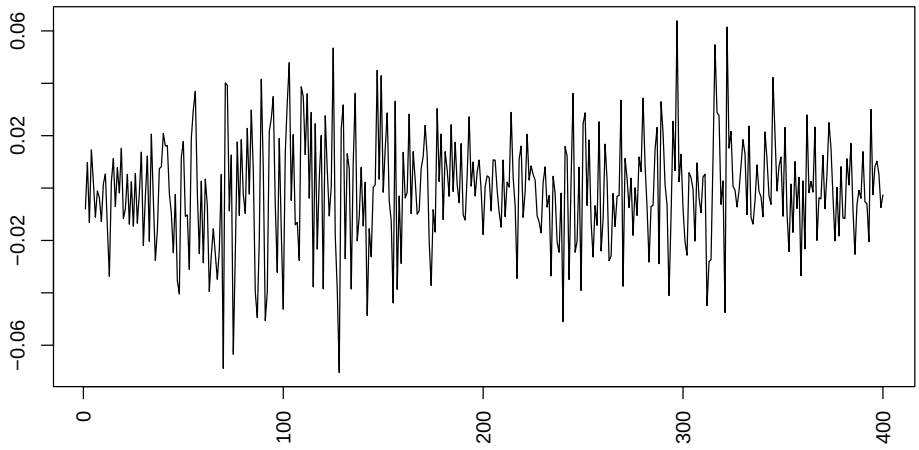


Figure 1: Daily returns of the stock market index for the Borsa Italiana (FTSE MIB), spanning from 5 May 2011 to 27 November 2012.

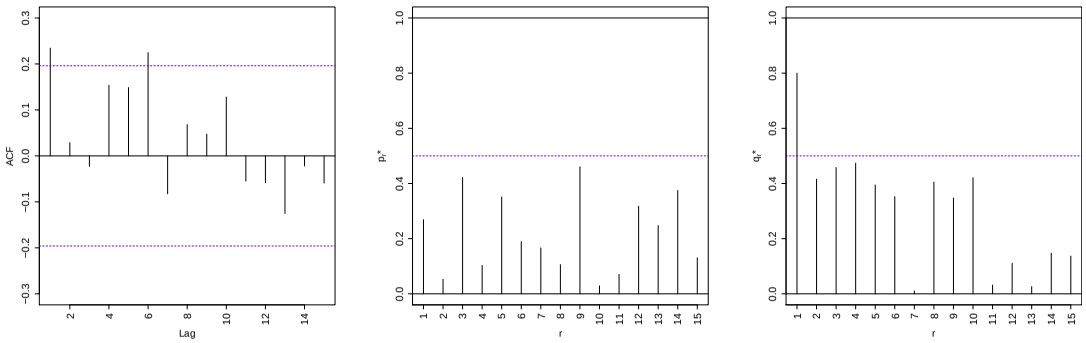
diagrams in Figure 2). Both these aspects will be better investigated in the simulation study described in the next section and, for this reason, we prefer to postpone the related comments there (where the data generating process is known).

6 A simulation study

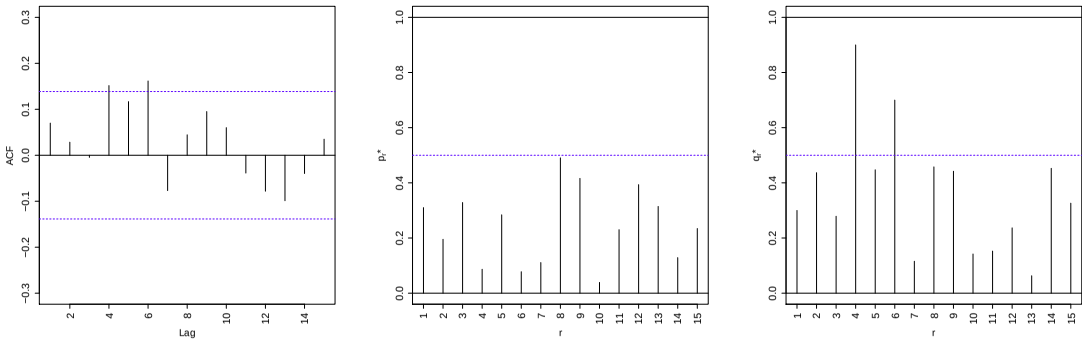
In this section a simulation study, covering several well-known time series models, is presented to show the gain in performance of the KL-autodependogram (simply abbreviated as KL-ADF) with respect to the autocorrelogram and the autodependogram.

6.1 Design and computational details

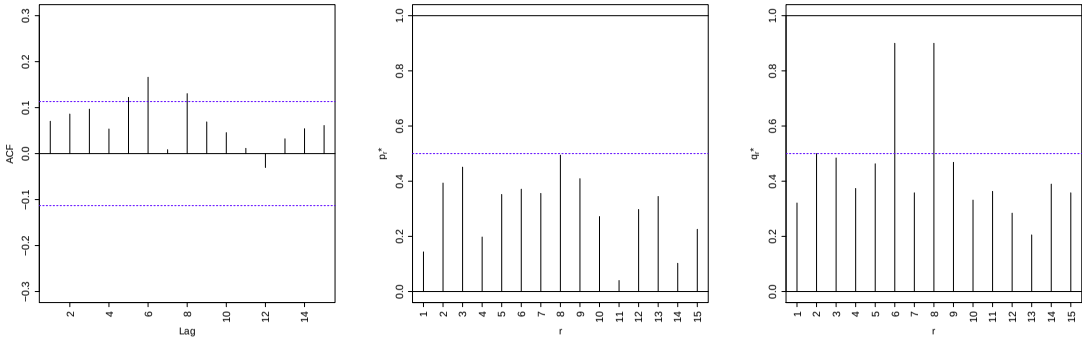
Table 1 shows the data generating models, and the corresponding parameters specification, used in the simulation study. They include: one scenario of serial independence (M1), two scenarios characterized by serial dependence of a purely linear type (M2-M3), and nine scenarios with nonlinear serial dependence (M4-M12). Concerning the independence case, data are randomly generated from the standard Gaussian, denoted with ε_t. The Gaussian noise ε_t is always used for the remaining models. As regards the scenarios related to the linear dependence, the well-known AR(1) and MA(1) models are considered. The nonlinear models taken into account are: the Quadratic MA(1) that have zero-correlation and a quadratic form of dependence only in correspondence to lag 1; the ARCH and the GARCH that, like the Quadratic MA, are characterized by a quadratic form of dependence and by zero correlation but, differently from the Quadratic MA, they have 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; some well-known extensions of the GARCH model. All the simulations are performed considering $l = 10$ lags and $\alpha = 0.05$. Differently from Bagnato *et al.* (2013b), that only consider the sample size $n = 100$, here also the value $n = 400$ is taken into account. For each of the 12 models in Table 1, and for each size n, one thousand samples are randomly generated. In particular, for the 11 models characterized by serial dependence, a time series of length $n+100$



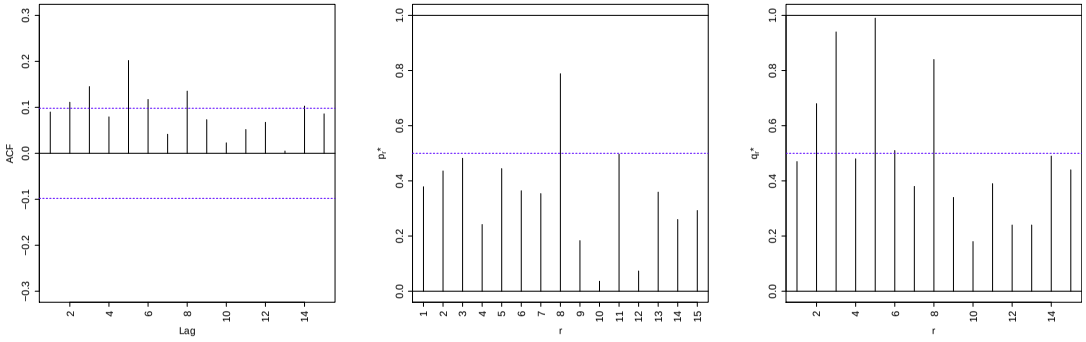
(a) ACF - $n = 100$ (b) p_r^* - $n = 100$ (c) q_r^* - $n = 100$



(d) ACF - $n = 200$ (e) p_r^* - $n = 200$ (f) q_r^* - $n = 200$



(g) ACF - $n = 300$ (h) p_r^* - $n = 300$ (i) q_r^* - $n = 300$



(j) ACF - $n = 400$ (k) p_r^* - $n = 400$ (l) q_r^* - $n = 400$

Figure 2: Autocorrelogram on the squared raw data (on the left), diagram of p_r^* on the raw data (on the middle), and KL-autodependogram on the raw data (on the right). Data are daily returns of the stock market index for the Borsa Italiana (FTSE MIB) from 5 May 2011 to 27 November 2012 ($n = 400$). The most recent subsamples of size $n = 300$, $n = 200$, and $n = 100$, are also considered. For all the diagrams, $l = 15$ and $\alpha = 0.05$.

is initially generated, but only the final n observations are used in order to mitigate the impact of initial values. As concern the autodependogram, the rule (2) provides $k = 4$ for $n = 100$ and $k = 6$ for $n = 400$. As concern the KL-autodependogram, the p-values $\widehat{q}_r$ in (9) have been obtained by considering $B = 99$ permutations.

Table 1: Models, and corresponding parameters specification, adopted in the simulation study.

Model	Name	Equation
M1	*i.i.d.*	$X_t = \varepsilon_t$
M2	AR(1)	$X_t = .25X_{t-1} + \varepsilon_t$
M3	MA(1)	$X_t = .25\varepsilon_{t-1} + \varepsilon_t$
M4	Multiplicative MA(1)	$X_t = .8\varepsilon_{t-1}\varepsilon_{t-2} + \varepsilon_t$
M5	Threshold AR(1)	$X_t = X_{t-1}\left[-.5 + .9\ \mathbb{1}\left(X_{t-1} \geq 0\right)\right]$
M6	Bilinear AR(2)	$X_t = .8X_{t-2}\varepsilon_{t-1} + \varepsilon_t$
M7	Quadratic MA(1)	$X_t = .25\varepsilon_{t-1}^2 + \varepsilon_t$
M8	ARCH(1)	$X_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = .01 + .5X_{t-1}^2$
M9	GARCH$(1,1)$	$X_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = .01 + .125X_{t-1}^2 + .8\sigma_{t-1}^2$
M10	E-GARCH$(1,1)$	$X_t = \sigma_t\varepsilon_t, \quad \log\left(\sigma_t^2\right) = .01 + .7\log\left(\sigma_{t-1}^2\right) - .3\varepsilon_{t-1} + .7\left(\left\|\varepsilon_{t-1}\right| - E\left\|\varepsilon_{t-1}\right|\right)$
M11	Threshold GARCH$(1,1)$	$X_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = .25 + .6\sigma_{t-1}^2 + .5X_{t-1}^2\mathbb{1}\left(\varepsilon_{t-1} < 0\right) + .2X_{t-1}^2\mathbb{1}\left(\varepsilon_{t-1} \geq 0\right)$
M12	GARCH$(1,1)$-M	$X_t = .003 + 2\sigma_t^2 + a_t, \quad a_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = .0002 + .13a_{t-1}^2 + .81\sigma_{t-1}^2$

6.2 Results

Figure 3 shows, in each plot, the rejection rates for the KL-ADF, the ADF, and the ACF obtained under M1. Results indicates that the size is maintained, for each considered lag and for both the values of n, by all the considered statistics. Analogously, Figure 4 displays the rejection rates obtained under M2-M3 which are characterized by linear serial dependence. Although the behavior of the three diagrams is common, with higher power when n increases, the autocorrelogram has naturally the best performance. This is due to the fact that, differently from the KL-autodependogram and the autodependogram which are *omnibus* tools, the autocorrelogram is a "directional" diagram of serial dependence which is intended to detect linear serial dependencies. Thus, it is naturally more powerful under this condition. However, between the other two competitors, the KL-ADF clearly outperforms the ADF.

Finally, Figure 5-7 show the rejection rates for M4-M12. Also in this case, the statistics have the same overall behavior; moreover, the power improves in line with n. Nevertheless, it is evident the supremacy of the KL-ADF in all the considered scenarios. A very interesting fact is that, apart from M5, M10, and M11, in the case $n = 100$ the ADF works worst than the ACF although this ranking overturns in the case $n = 400$. This suggest that the ADF needs of a sufficiently large sample size to outperforms the ACF for nonlinear data generating processes. This aspect is not highlighted in Bagnato *et al.* (2012).

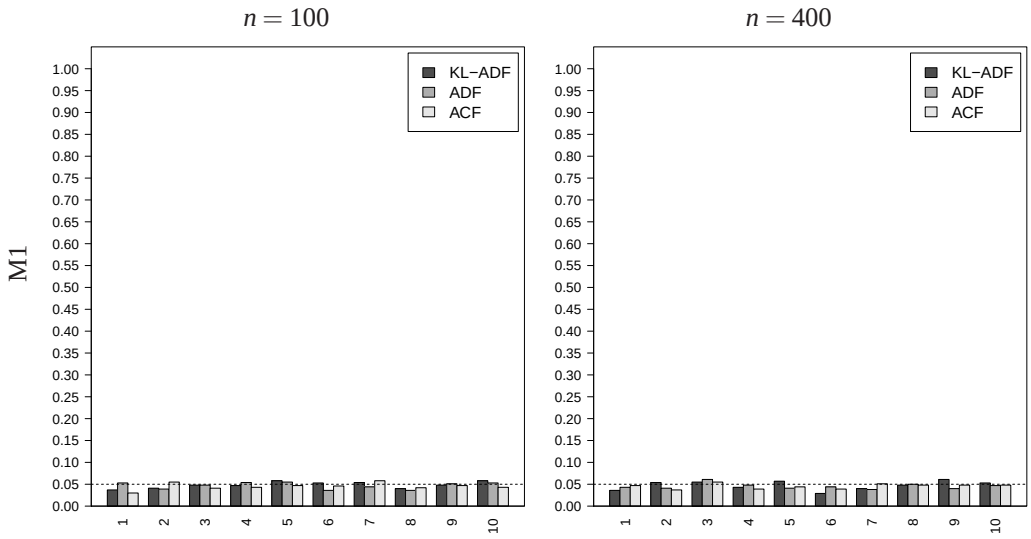


Figure 3: Simulated rejection rates, over 1000 replications, for model M1 in Table 1. KL-ADF, ADF, and ACF are compared.

7 Conclusions

A graphical device, named KL-autodependogram, has been proposed to detect autodependencies. For each lag, the dependence is assessed through a test based on the Kullback-Leibler divergence between the estimated joint density and the product of the estimated marginal densities. The Gaussian kernel density technique is used to estimate these densities and a permutation approach is adopted to compute *p*-values. This new *omnibus* tool has been shown to provide a substantial improvement with respect to the recent proposed autodependogram. This also implies that, on linear time series, the KL-autodependogram provides results more similar, in terms of power, to the autocorrelogram. These results have been shown by simulations but also corroborated via a real application to a financial time series.

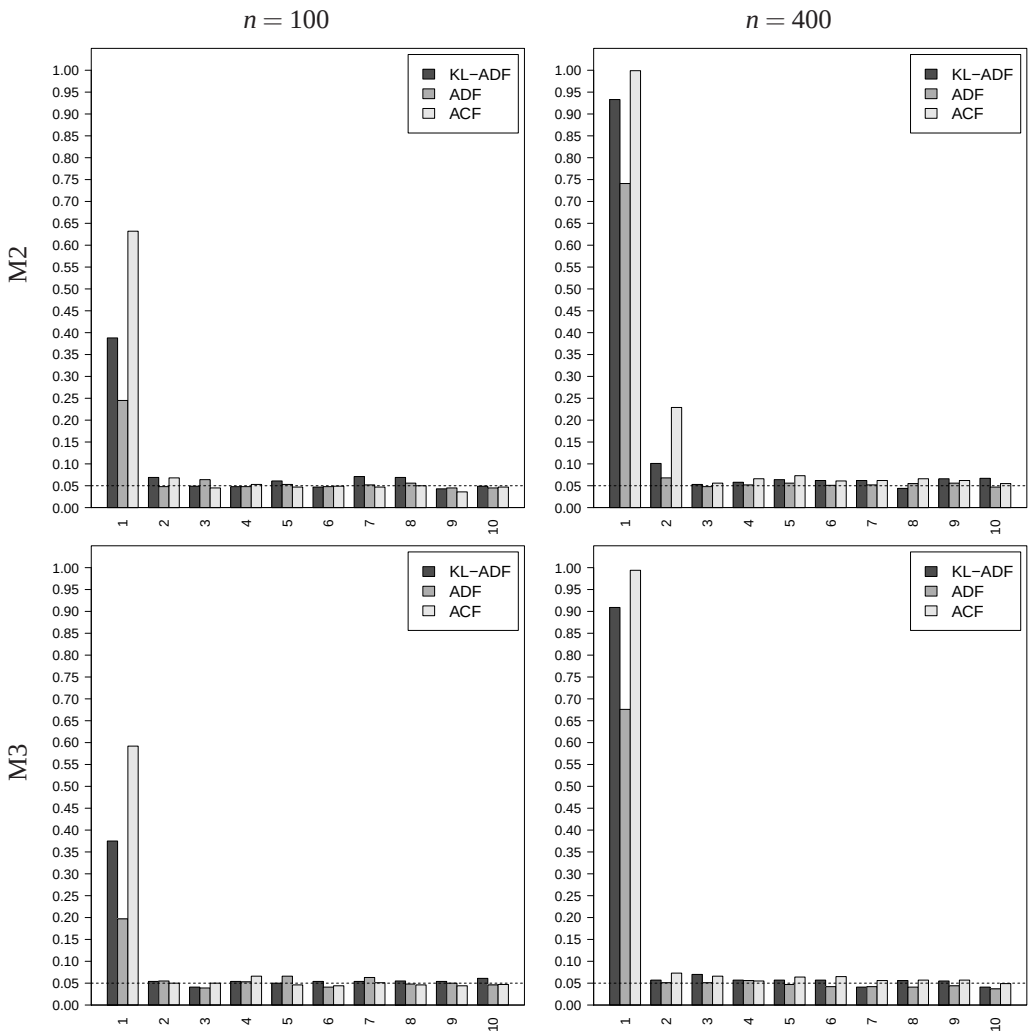


Figure 4: Simulated rejection rates, over 1000 replications, for models M2 and M3 in Table 1. KL-ADF, ADF, and ACF are compared.

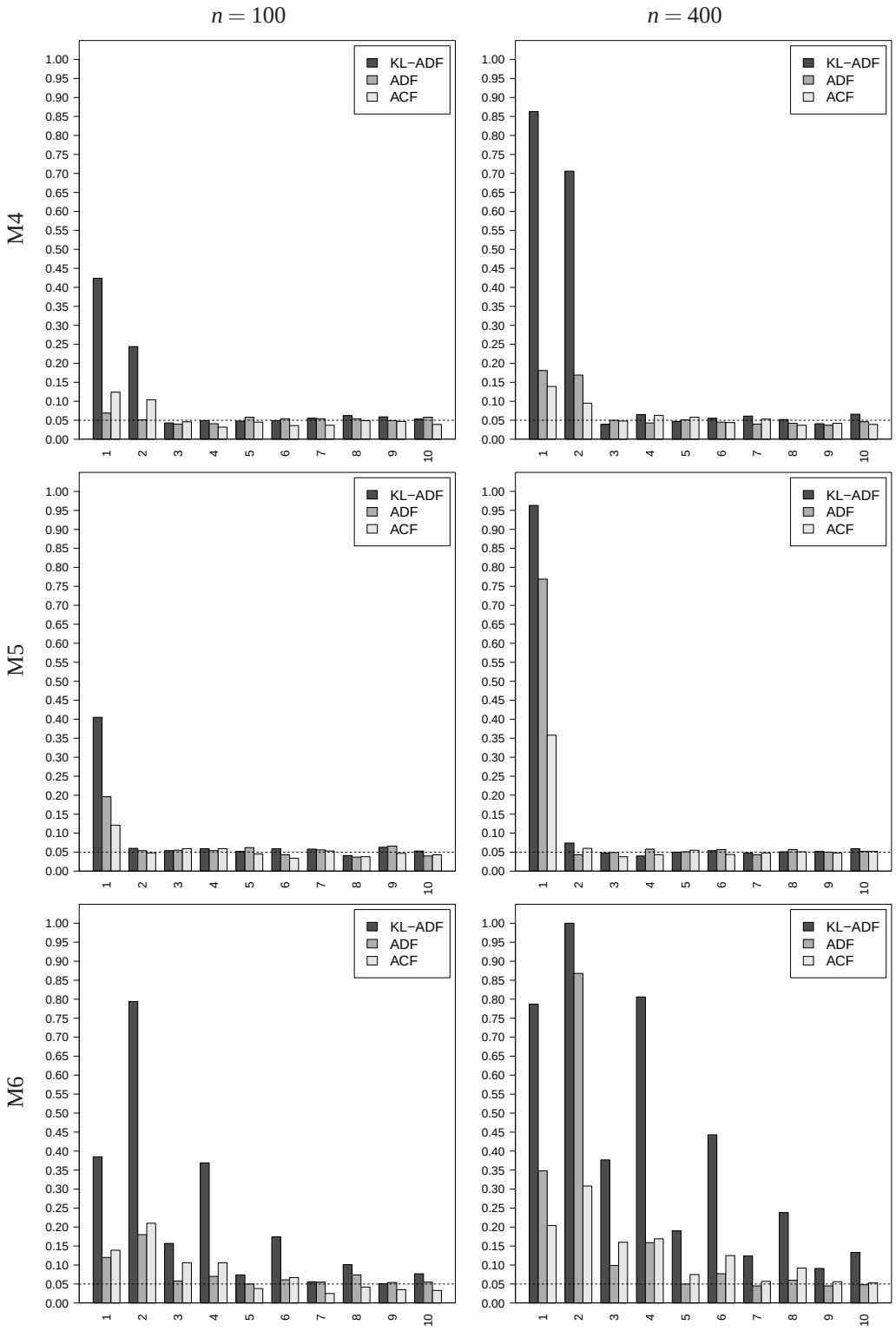


Figure 5: Simulated rejection rates, over 1000 replications, for models M4-M6 in Table 1. KL-ADF, ADF, and ACF are compared.

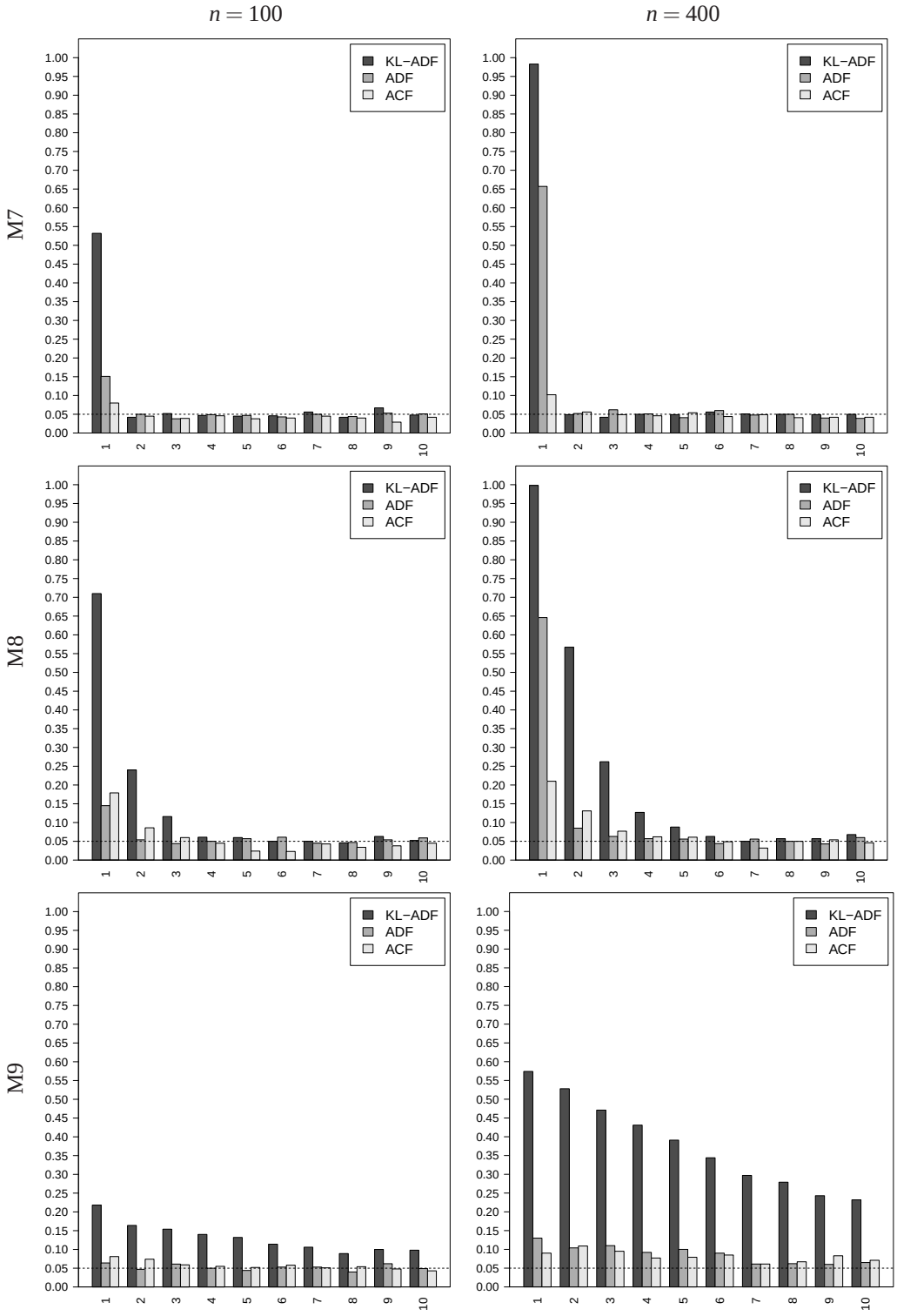


Figure 6: Simulated rejection rates, over 1000 replications, for models M7-M9 in Table 1. KL-ADF, ADF, and ACF are compared.

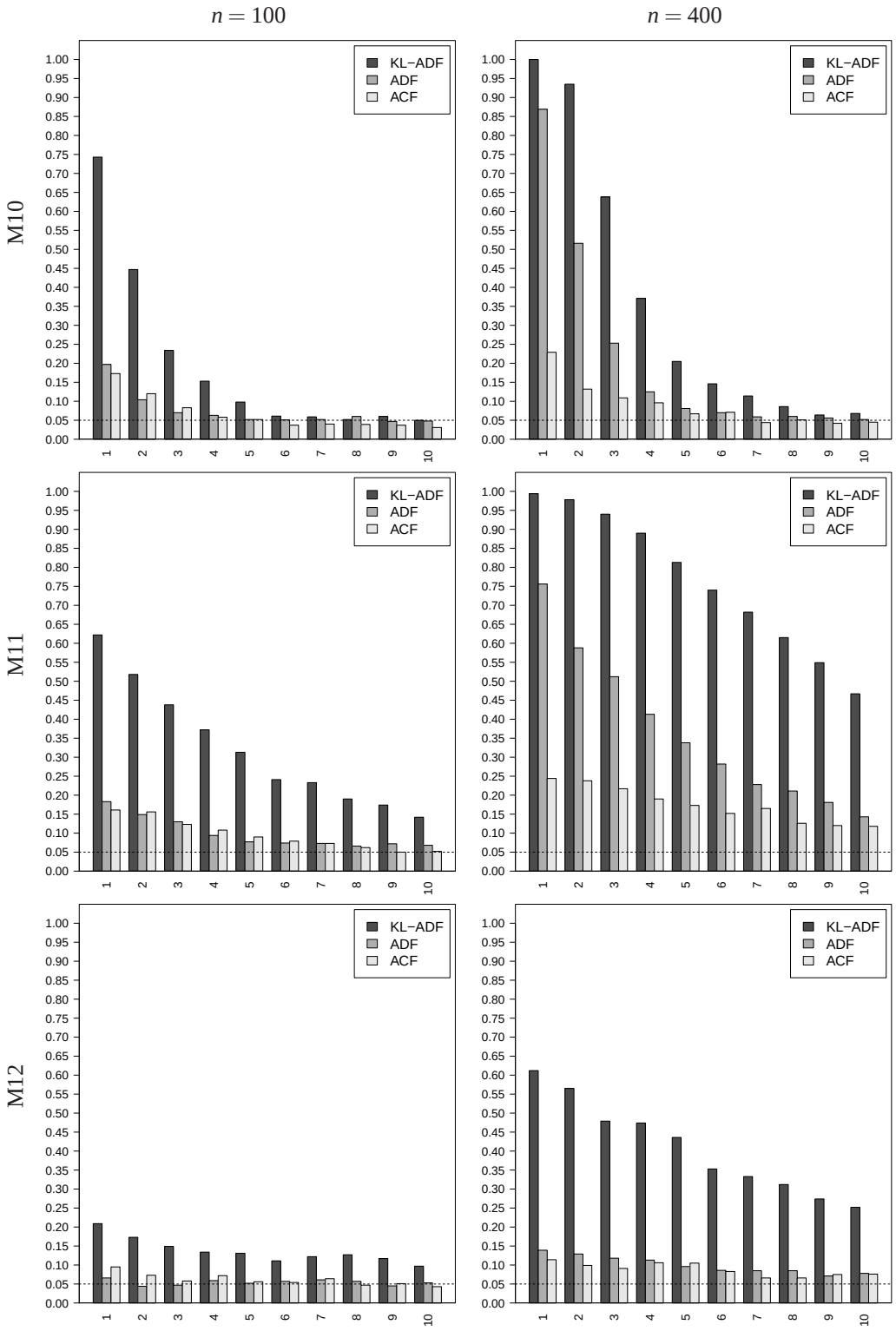


Figure 7: Simulated rejection rates, over 1000 replications, for models M10-M12 in Table 1. KL-ADF, ADF, and ACF are compared.

References

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

Anderson, N., Hall, P., and Titterington, D. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariate Analysis*, **50**(1), 41–54.

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

Bagnato, L. and Punzo, A. (2012). Checking serial independence of residuals from a nonlinear model. In W. A. Gaul, A. Geyer-Schulz, L. Schmidt-Thieme, and J. Kunze, editors, *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, volume XIV of *Studies in Classification, Data Analysis and Knowledge Organization*, pages 203–211, Berlin-Heidelberg. Springer-Verlag.

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

Bagnato, L., De Capitani, L., and Punzo, A. (2013a). Detecting serial dependencies with the reproducibility probability autodependogram. *Advances in Statistical Analysis*. DOI: 10.1007/s10182-013-0208-y.

Bagnato, L., De Capitani, L., and Punzo, A. (2013b). Testing serial independence via density-based measures of divergence. *Methodology and Computing in Applied Probability*. DOI: 10.1007/s11009-013-9320-4.

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, editor, *Encyclopedia of Complexity and Systems Science*, pages 6252–6271. Springer, New York.

Diks, C. and Panchenko, V. (2007). Nonparametric tests for serial independence based on quadratic forms. *Statistica Sinica*, **17**(1), 81–89.

Granger, C., Maasoumi, E., and Racine, J. (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis*, **25**(5), 649–669.

Hong, Y. and White, H. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica*, **73**(3), 837–901.

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

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

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

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall Ltd.

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