TESTING FOR INDEPENDENCE AND REPRESENTING AUTO-DEPENDENCE

MATTEO PELAGATTI

ABSTRACT. VERY PRELIMINARY!

We introduce a nonparametric test for independence and some measures of dependence of two random variables. We propose graphical representations of these measures that describe the auto-dependence of a time series and help the identification of the possible data generating process. We derive the asymptotic distribution of the test under the null of independence and verify by Monte Carlo simulations that it represent a good approximation of the finite sample distribution.

We compare our test with the well-known Cramer-von Mises type empirical copula test and find it generally more powerful under a wide range of alternative hypotheses, that include some time series models wildly used in financial econometrics.

Finally, we exploit the measure of dependence implicitly defined in the test statistic to represent the autodependence of time series. We show that for a particular version of this autodependence function it is possible to derive its population values under popular time series models and use its sample version to infer the data generating process of a time series.

1. METHODOLOGY

Let X and Y be two random variables defined on the same probability space and denote with $\mathcal{G}$ the class of continuous square-integrable functions of X and with $\mathcal{H}$ the class of continuous square-integrable functions of Y:

$$\mathcal{G} := \big\{g(\cdot) : g \text{ is continuous and } \mathbb{E}g(X)^2 < \infty\big\}$$
$$\mathcal{H} := \big\{h(\cdot) : h \text{ is continuous and } \mathbb{E}h(X)^2 < \infty\big\}.$$

A well known characterisation of the stochastic independence of two random variables is based on the following result[1] (for instance Feller, 1971, Theorem at page 136):

$$X \perp Y \iff \mathbb{C}\text{or}\big[g(X), h(Y)\big] = 0, \quad \forall g \in \mathcal{G}, \forall h \in \mathcal{H}$$

where the symbol $\perp$ is used to denote independence.

[1]Actually, the classes of functions $\mathcal{G}$ and $\mathcal{H}$ could be further restricted to contain only *vanishing at infinity* functions.

Let F_x and F_y be the cumulative distribution functions (CDF) of X and Y, respectively, then the above result can also be stated as[2]

$$X \perp Y \iff \mathbb{C}\mathrm{or}\big[\bar{g}\big(F_x(X)\big), \bar{h}\big(F_y(Y)\big)\big] = 0, \quad \forall \bar{g} \in \mathcal{G}, \forall \bar{h} \in \mathcal{H},$$

where the domain of $\bar{h}$ and $\bar{g}$ is now the real interval $(0, 1)$.

The functions g and h and, even better, $\bar{g}$ and $\bar{h}$ can be approximated by linear combinations of a finite number of functions $g_0, g_1, \ldots, g_p$ and $h_0, h_1, \ldots, h_q$, such that for any positive ε_g and ε_h

$$\sup_x \left| \bar{g}(x) - \sum_{i=0}^{p} \alpha_i g_i(x) \right| < \varepsilon_g, \quad \sup_x \left| \bar{h}(x) - \sum_{i=0}^{q} \alpha_i h_i(x) \right| < \varepsilon_h,$$

where the functions g_i and h_i can be powers (by Weierstrass theorem), sinusoids (by Fourier theory), splines (Scherer, 1970), etc.

Consider the quantity

$$\varrho_{pq}(X, Y) := \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \mathbb{C}\mathrm{or}\left[\sum_{i=0}^{p} \alpha_i g_i(F_x(X)), \sum_{i=0}^{q} \alpha_i h_i(F_y(Y)) \right], \tag{1}$$

where the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ contain the coefficients of the linear combinations. Under the hypothesis $H_0 : X \perp Y$, we have $\varrho_{pq}(X, Y) = 0$.

Given $g_i(F_x(X))$ and $h_i(F_y(Y))$, the maximisation problem in equation (1) has a well known solution: $\varrho_{pq}(X, Y)$ is the first canonical correlation coefficient between the two random vectors $\boldsymbol{u} := \big[g_0(F_x(X)), \ldots, g_p(F_x(X))\big]^\top$ and $\boldsymbol{v} := \big[h_0(F_y(Y)), \ldots, h_q(F_y(Y))\big]^\top$, that is, the square of the largest eigenvalue of the matrix

$$\boldsymbol{\Sigma}_{uu}^{-1} \boldsymbol{\Sigma}_{uv} \boldsymbol{\Sigma}_{vv}^{-1} \boldsymbol{\Sigma}_{vu},$$

where the symbol $\boldsymbol{\Sigma}_{ab}$ denotes the covariance matrix of vector $\boldsymbol{a}$ with vector $\boldsymbol{b}$. The main idea is testing the null hypothesis by using the sample version of this canonical correlation analysis.

Remarks.

(1) In many cases, as for polynomial and trigonometric approximations, the order 0 functions can be dropped because they are just constants that do not effect the correlation value. They will be dropped in the rest of the paper.

(2) There is a measure of dependence called *maximal correlation coefficient* (Gebelein, 1941; Rényi, 1959) defined by

$$S(X, Y) = \sup_{g,h} \mathbb{C}\mathrm{or}[g(X), h(Y)],$$

where the sup is taken with respect to all the measurable functions for which $\mathbb{E}g(X)^2 < \infty$ and $\mathbb{E}h(Y)^2 < \infty$. According to Rényi (1959), $S(X, Y)$ is an excellent measure of dependence as it obeys all his seven axioms defining a measure of dependence.

[2]This is certainly true if F_x and F_y are continuous, in fact $g(x) = \bar{g}(F_x(x))$ and $\bar{g}(x) = g(F^{-1}(x))$, but some reasoning is needed if they are not continuous.

Analysing how our $\varrho_{pq}(X,Y)$ approximates $S(X,Y)$ as p and q diverge and how its sample version (i.e., the square root of the value λ_1 defined below) behaves as an estimator of $S(X,Y)$ goes beyond the scope of this paper, but will be investigated in a future work.

Suppose that $\{(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)\}$ is an i.i.d. sample for the pair (X,Y) and define the normalised ranks

$$U_{ni} = \frac{1}{n+1}\sum_{j=1}^{n} \mathbb{1}_{(-\infty,\,X_i]}(X_j), \qquad V_{ni} = \frac{1}{n+1}\sum_{j=1}^{n} \mathbb{1}_{(-\infty,\,Y_i]}(Y_j),$$

where $\mathbb{1}_A(\cdot)$ is an indicator for the set A. Define the $n\times p$ and $n\times q$ matrices

$$\mathbf{U} = \{g_j(U_{ni}) - \bar{g}_j\}, \qquad \mathbf{V} = \{h_k(V_{ni}) - \bar{h}_k\}$$

for $i=1,2,\ldots,n$, $j=1,2,\ldots,p$ and $k=1,2,\ldots,q$, with

$$\bar{g}_j = \frac{1}{n}\sum_{i=1}^{n} g_j(U_{ni}), \qquad \bar{h}_k = \frac{1}{n}\sum_{i=1}^{n} h_k(V_{ni}).$$

The columns of these two matrices contain the zero-mean versions of the basis functions applied to the normalised ranks of the sample observations. The sample canonical correlation analysis is carried out with respect to the variables in the columns of $\mathbf{U}$ and $\mathbf{V}$. Thus, let $\lambda_1,\lambda_2,\ldots,\lambda_r$, with $r=\min(p,q)$, be the non-increasingly ordered eigenvalues of the matrix

$$\mathbf{S}_{uu}^{-1}\mathbf{S}_{uv}\mathbf{S}_{vv}^{-1}\mathbf{S}_{vu}, \tag{2}$$

where

$$\mathbf{S}_{uu} := \frac{1}{n}\mathbf{U}^\top\mathbf{U}, \quad \mathbf{S}_{uv} := \frac{1}{n}\mathbf{U}^\top\mathbf{V}, \quad \mathbf{S}_{vv} := \frac{1}{n}\mathbf{U}^\top\mathbf{U}, \quad \mathbf{S}_{vu} := \mathbf{S}_{uv}^\top.$$

The test statistic we propose for the hypothesis $H_0 : X \perp Y$ is the classical Bartlett formula used in canonical correlation:

$$B_n = -(n-(p+q+3)/2)\sum_{j=1}^{r}\log(1-\lambda_j). \tag{3}$$

Proposition 1 (Asymptotic distribution under the null). *Let $g_i(\cdot)$, $i=1,2,\ldots,p$, and $h_j(\cdot)$, $j=1,2,\ldots,q$, be measurable functions that map $[0,1]$ to a bounded interval in $\mathbb{R}$ and such that*

$$\operatorname{rank}\mathbb{E}[\boldsymbol{g}\boldsymbol{g}^\top] = p, \qquad \operatorname{rank}\mathbb{E}[\boldsymbol{h}\boldsymbol{h}^\top] = q,$$

with

$$\boldsymbol{g} = [g_1(F_x(X)),\ldots,g_p(F_x(X))]^\top$$
$$\boldsymbol{h} = [h_1(F_y(Y)),\ldots,g_q(F_y(Y))]^\top;$$

then if $X \perp Y$

$$B_n \xrightarrow{d} \chi^2_{pq},$$

with χ^2_{pq} *chi-square distribution with* $p \cdot q$ *degrees of freedom.*

Proof. The result obtains as straightforward application of Theorem 1 of Bai and He (2004). □

Notice that the test statistic, under the null, is *distribution free* because it depends on the observations of the two independent random variables only though their respective ranks. Thus, under the null, the joint distribution of the ranks of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ is given by the product of two (discrete) uniform distributions, whatever the distribution of X and Y. So, the asymptotic distribution of the test statistic is really not necessary as critical values can be obtained by Monte Carlo methods for a set of sample sizes and significance levels. However, if the asymptotic distribution is accurate also for typical sample sizes, the application of the test by practitioners becomes simpler. As the next section illustrates, the asymptotic distribution seems to be very accurate also for finite samples.

2. Monte Carlo results for finite sample properties

In all the simulations we set $p = q$ (i.e., same order in truncating the expansions of $\bar{g}(\cdot)$ and $\bar{h}(\cdot)$) and the results (i.e., kernel density plots, Monte Carlo sizes and powers) are always based on 50,000 replications.

We provide simulations for tests based on three type of function approximations. The same functions are applied to the variables U_{ni} and V_{ni}. For $j = 1, 2, \ldots, p$ and $x \in (0, 1)$:

Polynomial: $g_j(x) = h_j(x) = x^j$;

Fourier:

$$g_j(x) = h_j(x) = \begin{cases} \cos\left((j+1)\pi x\right) & \text{if } j \text{ is odd,} \\ \sin\left(j\pi x\right) & \text{if } j \text{ is even;} \end{cases}$$

Spline: $g_j(x) = h_j(x) =$ the j-th cubic spline basis function computed over the grid $\{1/(n+1), 2/(n+2), \ldots, n/(n+1)\}$ where n is the sample size (cf. Hastie et al., 2009, Chapter 5). Notice that this basis is not available if $p < 3$.

2.1. Null distribution and size.

For a nominal size of 5%, Table 1 reports the actual size of the proposed test for various basis functions, different orders of the expansions, p, and sample sizes, n.

The actual size seems to be rather accurate for all basis functions and sample sizes. For $n = 50$ and $p = 5$ the tests are just slightly oversized.

By observing Figures 1–3 that compare the empirical (kernel) distributions of the test statistics under different orders and sample sizes with the asymptotic distribution provided by Proposition , it appears clear how accurate the asymptotic approximation is.

TABLE 1. Actual size for the asymptotic 5% critical value

Basis	$n = 50$	$n = 100$	$n = 200$	$n = 400$
Polynomial				
$p = 2$	0.050	0.049	0.052	0.051
$p = 3$	0.051	0.049	0.050	0.052
$p = 5$	0.061	0.055	0.052	0.050
Fourier				
$p = 2$	0.054	0.051	0.051	0.051
$p = 3$	0.053	0.051	0.051	0.049
$p = 5$	0.055	0.053	0.050	0.048
Spline				
$p = 3$	0.051	0.049	0.050	0.052
$p = 5$	0.058	0.055	0.052	0.049

2.2. **Power.** The power of the proposed test statistic against several alternatives is compared to the power of the Cramér-von Mises type statistic based on the empirical copula (Hoeffding, 1948; Blum et al., 1961; Deheuvels, 1980, 1981; Genest and Rémillard, 2004; Genesta et al., 2006)

$$T_n := \int_{(0,1)^2} \mathbb{C}_n(u,v)^2 du dv = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} D_n(U_{ni}, U_{nj}) D_n(V_{ni}, V_{nj}),$$

where $\mathbb{C}_n(u,v)$ is the *centred empirical copula process*, U_{ni}, V_{nj} are the normalised ranks defined above and

$$D_n(s,t) := \frac{2n+1}{6n} + \frac{(n+1)s^2 - s}{2n} + \frac{(n+1)t^2 - t}{2n} - \max(s,t).$$

We used the implementation in the package *copula* (Yan, 2007; Kojadinovic and Yan, 2010; Hofert et al., 2015) of R (R Core Team, 2015).

The power is evaluated with respect to the following types of dependences between X and Y.

Correlation (COR): X and Y are standard normal with correlation 0.2.

Product when large (PRO): X and Z are independent standard normal and $Y := Z \cdot \max(1, X)$.

ARCH(1) (ARC): X and Z are independent standard normal and $Y := Z \cdot (0.5X^2)$.

Same sign if small, opposite sign if large (SIG): X is a uniform random variable on $[-1, 1]$, Z is an exponential random variable and

$$Y := \begin{cases} Z \cdot \text{sign}(X) & \text{if } X > 1.5 \\ Z \cdot \text{sign}(-X) & \text{otherwise.} \end{cases}$$

Stochastic Volatility (SVT): Define Z_t to be the following process

$$H_t = 0.8H_{t-1} + \eta_t$$
$$Z_t = \varepsilon_t \cdot \exp\left(H_t/2\right)$$

with $\{\eta_t\}$ i.i.d. sequence with standard normal marginal distribution and $\{\varepsilon_t\}$ i.i.d. sequence with Student's t_5 distribution, $\eta_t \perp \varepsilon_s$ for all t and s. Set $Y := Z_t$ and $X := Z_{t-1}$.

GARCH(1,1) (GAR): Define Z_t to be the following process

$$\sigma_t^2 = 10^{-6} + 0.1Z_{t-1}^2 + 0.8\sigma_{t-1}^2$$
$$Z_t = \varepsilon_t \sigma_t$$

with $\{\eta_t\}$ i.i.d. sequence with Student's t_5 distribution. Set $Y := Z_t$ and $X := Z_{t-1}$.

Table 2 reports the rejection rates of the tests under the above alternative hypotheses and different sample sizes. If we exclude the case of jointly normal distributed random variables with 0.2 correlation, tests such as Poly3 (polynomial expansion with $p = 3$) or Spline3 (cubic spline basis with $p = 3$) are uniformly more powerful than the classic Cramér-von Mises test based on the empirical copula. The last column of Table 2 reports the mean rejection rates computed over our choice of dependence structures. These numbers seem to suggest the use of our test based on the polynomial expansion or on the spline basis in both cases with $p = 3$. Of course this conclusion is true for the particular choice of dependence structures used in our simulation: suggestions on different types of dependence are welcome.

3. Extensions for time series: auto-dependence functions

Suppose that we observe a finite realisation of the stationary process $\{X_t\}$, then a simple way to represent the dependence between X_t and X_{t-k} is by computing the square root of the largest eigenvalue of the matrix in equation (2), where the covariance matrices are computed between the p functions of the ranks of $\{X_{k+1}, X_{k+2}, \ldots, X_n\}$ and of $\{X_1, X_2, \ldots, X_{n-k}\}$.

For example, Figure 4 depicts the sample maximum correlation coefficient of X_t with X_{t-k} plotted against the lag k. The graph on the top is based on a polynomial approximation of order 2, while the bottom plot is computed using a polynomial approximation of order 5. The horizontal dashed line represents the 95% quantile of the bootstrapped distribution of the sample maximum correlation under independence.

Of course the test statistic (3) can be computed for all pairs X_t and X_{t-k} and, under the null of serial independence of X_t, these statistics, say $B_n(k)$ are all asymptotically independent $\chi^2_{p^2}$ random variables.

TABLE 2. Powers of the tests under various alternative hypothesis (highest rejection rates in boldface)

n	Test	COR	PRO	ARC	SIG	SVT	GAR	Mean
50	Copula	0.23	0.06	0.29	0.81	0.07	0.05	0.25
50	Poly2	0.15	0.16	1.00	0.23	0.29	0.07	0.32
50	Poly3	0.12	0.17	1.00	1.00	0.24	0.07	**0.43**
50	Poly5	0.09	0.21	1.00	1.00	0.20	0.06	**0.43**
50	Fourier2	0.07	0.07	1.00	0.96	0.22	0.06	0.40
50	Fourier3	0.06	0.07	1.00	0.93	0.18	0.06	0.38
50	Fourier5	0.06	0.07	1.00	0.97	0.13	0.06	0.38
50	Spline3	0.12	0.17	1.00	1.00	0.24	0.07	**0.43**
50	Spline5	0.10	0.20	1.00	1.00	0.21	0.06	**0.43**
100	Copula	0.43	0.07	0.90	1.00	0.09	0.05	0.42
100	Poly2	0.29	0.26	1.00	0.32	0.59	0.11	0.43
100	Poly3	0.22	0.27	1.00	1.00	0.50	0.10	**0.52**
100	Poly5	0.15	0.33	1.00	1.00	0.41	0.10	0.50
100	Fourier2	0.09	0.08	1.00	1.00	0.49	0.07	0.46
100	Fourier3	0.08	0.10	1.00	1.00	0.42	0.08	0.45
100	Fourier5	0.08	0.10	1.00	1.00	0.30	0.07	0.42
100	Spline3	0.22	0.27	1.00	1.00	0.50	0.10	**0.52**
100	Spline5	0.17	0.32	1.00	1.00	0.44	0.10	0.51
200	Copula	0.72	0.08	1.00	1.00	0.13	0.06	0.50
200	Poly2	0.56	0.45	1.00	0.47	0.90	0.17	0.59
200	Poly3	0.45	0.47	1.00	1.00	0.83	0.15	**0.65**
200	Poly5	0.29	0.56	1.00	1.00	0.74	0.16	0.63
200	Fourier2	0.15	0.12	1.00	1.00	0.83	0.10	0.53
200	Fourier3	0.11	0.17	1.00	1.00	0.78	0.11	0.53
200	Fourier5	0.12	0.18	1.00	1.00	0.64	0.10	0.50
200	Spline3	0.45	0.47	1.00	1.00	0.83	0.15	**0.65**
200	Spline5	0.35	0.57	1.00	1.00	0.78	0.16	0.64
400	Copula	0.95	0.13	1.00	1.00	0.25	0.06	0.57
400	Poly2	0.89	0.75	1.00	0.74	1.00	0.30	0.78
400	Poly3	0.81	0.78	1.00	1.00	0.99	0.24	**0.80**
400	Poly5	0.61	0.86	1.00	1.00	0.97	0.26	0.78
400	Fourier2	0.27	0.21	1.00	1.00	0.99	0.17	0.61
400	Fourier3	0.19	0.34	1.00	1.00	0.98	0.19	0.62
400	Fourier5	0.21	0.36	1.00	1.00	0.95	0.16	0.61
400	Spline3	0.81	0.78	1.00	1.00	0.99	0.24	**0.80**
400	Spline5	0.70	0.87	1.00	1.00	0.98	0.27	**0.80**

Thus, a Box-Pierce type test can be obtained by summing the first K statistics:

$$\tilde{Q}_n(K) := \sum_{k=1}^{K} B_n(k) \xrightarrow{d} \chi^2_{Kp^2}.$$

For example, the $\tilde{Q}_n(30)$ statistics computed using the values in the plots of Figure 4 are 1528 and 2349 with 5% critical values 146 and 815, respectively: in both cases the null of serial independence is strongly rejected.

Figure 5 depicts the auto-dependence function using polynomial expansions of order 2 and 5 for the stock returns of Microsoft. There is a clear dependence that lasts for more than 100 days, as the virtually zero p-values of both $\tilde{Q}_n(100)$ statistics confirm.

4. Inferring the model from the auto-dependence type

Since many time series models are built for the conditional moments of the observations, we may build an auto-dependence function using just the first p powers of the time series, without taking the ranks. Of course, in this case we have to assume that at least $2p$ moments of the time series exist. For example, if we set $p = 2$ it is not hard to compute the population canonical correlations between (X_t, X_t^2) and (X_{t-k}, X_{t-k}^2) for any integer k.

In order to eloquently synthesise the information contents of the various canonical correlation analyses we can use a plot like that in Figure 6, where the two sequences of auto-canonical correlations are depicted, respectively, in the two plots. The positive and negative heights of the bars represent the magnitude of the canonical correlation, while their colours represent the relative magnitude of the squared canonical coefficients: in particular the colours in the positive bars represent the $\boldsymbol{x}_t$-side coefficients, while those in the negative bars depict the $\boldsymbol{x}_{t-k}$-side coefficients. In formulas, if we name $\rho_{j,k}$, $j = 1, \ldots, p$ the canonical correlations at lag k, $\alpha_{i,j,k}$ the j-th canonical coefficient relative to the i-th power of X_t and $\beta_{i,j,k}$ the j-th canonical coefficient of the i-th power of X_{t-k}, then the multi-colour bars represent (as change of colour) the cumulative sums of

$$\rho_{j,k} \frac{\alpha^2_{i,j,k}}{\sum_i \alpha^2_{i,j,k}} \text{ (for positive bars)}, \qquad -\rho_{j,k} \frac{\beta^2_{i,j,k}}{\sum_i \beta^2_{i,j,k}} \text{ (for negative bars)}.$$

Notice that $\rho_{j,k}$ is always nonnegative, but $\alpha_{i,j,k}$ and $\beta_{i,j,k}$ can take positive and negative values.

AR(1)

For a fourth-order stationary AR(1) process with coefficient $\phi \in$

$(-1,1)$, if we set $\boldsymbol{x}_{t-k} := [X_{t-k}, X_{t-k}^2]^\top$, we have

$$\boldsymbol{R}_k := \mathbb{C}\mathrm{or}\left[\boldsymbol{x}_t, \boldsymbol{x}_{t-k}\right] = \begin{bmatrix} \phi^k & \phi^k\lambda \\ \phi^{2k}\lambda & \phi^{2k} \end{bmatrix},$$

where $\lambda := \mathrm{skew}(X_t)/\sqrt{\mathrm{kurt}(X_t)-1}$. If we consider a symmetric AR(1) process (i.e., $\lambda = 0$) and define the coefficient matrices $\boldsymbol{A}_k := [\alpha_{i,j,k}]_{i,j\in\{1,\ldots,p\}}$ and $\boldsymbol{B}_k := [\beta_{i,j,k}]_{i,j\in\{1,\ldots,p\}}$, the canonical correlation analysis yields

$$\rho_{1,k} = \phi^k, \quad \rho_{2,k} = \phi^{2k}, \quad \boldsymbol{A}_k = \boldsymbol{B}_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The corresponding autodependogram is depicted in Figure 6

The formulas for a skewed AR(1) process can be found in the Appendix, but the relative autodependogram can be seen is Figure 7

MA(1)

For a fourth-order stationary MA(1), $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, with ε_t i.i.d. such that $\mathbb{E}\varepsilon_t = 0$, $\mathbb{E}\varepsilon_t^2 = 1$, $\mathbb{E}\varepsilon^3 = m_3$ and $\mathbb{E}\varepsilon^4 = m_4$, we have the following cross-correlation matrices

$$\boldsymbol{R}_0 = \begin{bmatrix} 1 & \frac{(1+\theta^3)m_3}{\sqrt{(1+\theta^2)\gamma_\theta}} \\ \frac{(1+\theta^3)m_3}{\sqrt{(1+\theta^2)\gamma_\theta}} & 1 \end{bmatrix}, \quad \boldsymbol{R}_1 = \begin{bmatrix} \frac{\theta}{1+\theta^2} & \frac{\theta m_3}{\sqrt{(1+\theta^2)\gamma_\theta}} \\ \frac{\theta^2 m_3}{\sqrt{(1+\theta^2)\gamma_\theta}} & \frac{\theta^2(m_4-1)}{\gamma_\theta} \end{bmatrix},$$

$\boldsymbol{R}_{k\in 2,3,\ldots} = \boldsymbol{0}$, with $\gamma_\theta := (m_4-1)(1+\theta^4)+4\theta^2$. For a symmetric MA(1) (i.e., $m_3 = 0$) the canonical canonical correlation analysis yields

$$\rho_{1,1} = \frac{\theta}{(1+\theta)}, \quad \rho_{2,1} = \left(\frac{1+\theta^4}{\theta^2} + \frac{4}{m_4-1}\right)^{-1} \boldsymbol{A}_1 = \boldsymbol{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

while for $k = 2, 3, \ldots$, the canonical correlations are zero and the coefficients arbitrary. It is interesting to notice that, differently from the AR(1) case, for the MA(1) the kurtosis of the driving white noise process influences the second canonical correlation: higher kurtosis generates larger correlation. The formulas for the skewed MA(1) are too involved to be interpreted and so we illustrate one case just through its autodependogram.

GARCH(1,1)

Le us consider the GARCH(1,1) process

$$\begin{aligned} Y_t &= \sigma_t Z_t, \qquad Z_t \sim NID(0,1) \\ \sigma_t^2 &= \omega + \alpha Y_{t-1}^2 + \beta\sigma_{t-1}^2, \end{aligned}$$

where $\alpha > 0$, $\beta > 0$ and the conditions of the existence of the fourth moment hold. Let us define

$$W_{1,t} = \frac{Y_t}{\sqrt{E[Y_t^2]}}$$
$$W_{2,t} = \frac{Y_t^2 - E[Y_t^2]}{\sqrt{E(Y_t^2 - E[Y_t^2])^2}},$$

and set $W_t = [W_{1,t}\ W_{2,t}]^\top$ then

$$\Sigma_{11} := E[W_t W_t^\top] =: \Sigma_{22}$$

and

$$\Sigma_{12}(k) = E[W_t W_{t-k}^\top] = \begin{bmatrix} 0 & 0 \\ 0 & \frac{\alpha(1-\alpha\beta-\beta^2)(\alpha+\beta)^{k-1}}{(1-2\alpha\beta-\beta^2)} \end{bmatrix}.$$

The canonical correlations between W_t and W_{t-k} are

$$\rho_1(k) = \frac{\alpha(1-\alpha\beta-\beta^2)(\alpha+\beta)^{k-1}}{(1-2\alpha\beta-\beta^2)}, \qquad \rho_2(k) = 0,$$

and the x and y-side eigenvectors are $[0\ 1]^\top$ and $[1\ 0]^\top$. Thus, for the GARCH(1,1) model only the first canonical correlation is nonzero for all k and only the second moment coefficients receive a nonzero weight. The first canonical correlations are proportional to $(\alpha+\beta)^{k-1}$, where the fourth-moment existence condition guarantees that $0 < \alpha + \beta < 1$.

References

Bai, Z. and X. He (2004). A chi-square test for dimensionality with non-Gaussian data. *Journal of Multivariate Analysis 88*, 109–117.

Blum, J. R., J. Kiefer, and M. Rosenblatt (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist. 32*(2), 485–498.

Deheuvels, P. (1980). Non parametric tests of independence. In *Nonparametric Asymptotic Statistics, Proceedings of the Conference held in Rouen in 1979*, Volume 821 of *Lecture Notes in Mathematics*, pp. 95–107. Springer.

Deheuvels, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *Journal of Multivariate Analysis 11*, 102–113.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications* (2nd ed.). John Wiley & Sons.

Gebelein, H. (1941). Das statistische problem der korrelation als variations- und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung. *Zeitschrift fr Angewandte Mathematik und Mechanik 21*(6), 364–379.

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

Genesta, C., J.-F. Quessy, and B. Rémillard (2006). Local efficiency of a Cramérvon Mises test of independence. *Journal of Multivariate Analysis 97*, 274–294.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Element of Statistical Learning*. Springer.

Hoeffding, W. (1948, 12). A non-parametric test of independence. *The Annals of Mathematical Statistics 19*(4), 546–557.

Hofert, M., I. Kojadinovic, M. Maechler, and J. Yan (2015). *copula: Multivariate Dependence with Copulas*. R package version 0.999-14.

Kojadinovic, I. and J. Yan (2010). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software 34*(9), 1–20.

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

Rényi, A. (1959). On measures of dependence. *Acta Mathematica 10*(3), 441–451.

Scherer, K. (1970). On the best approximation of continuous functions by splines. *SIAM Journal of Numerical Analysis 7*(3), 418–423.

Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software 21*(4), 1–21.

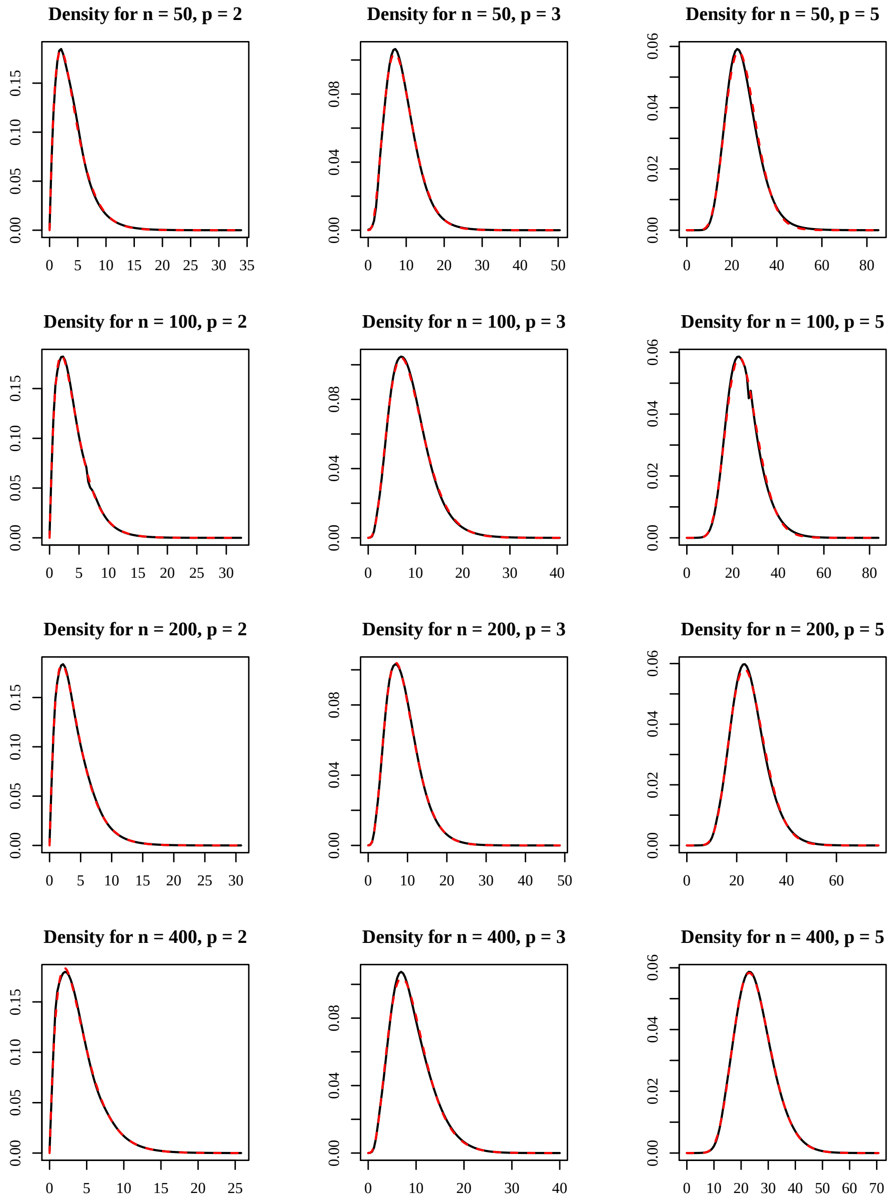

FIGURE 1. Comparison of kernel densities of the simulated B_n statistic based on *polynomial* basis functions with the asymptotic distribution for different sample sizes n and orders $p = q$.

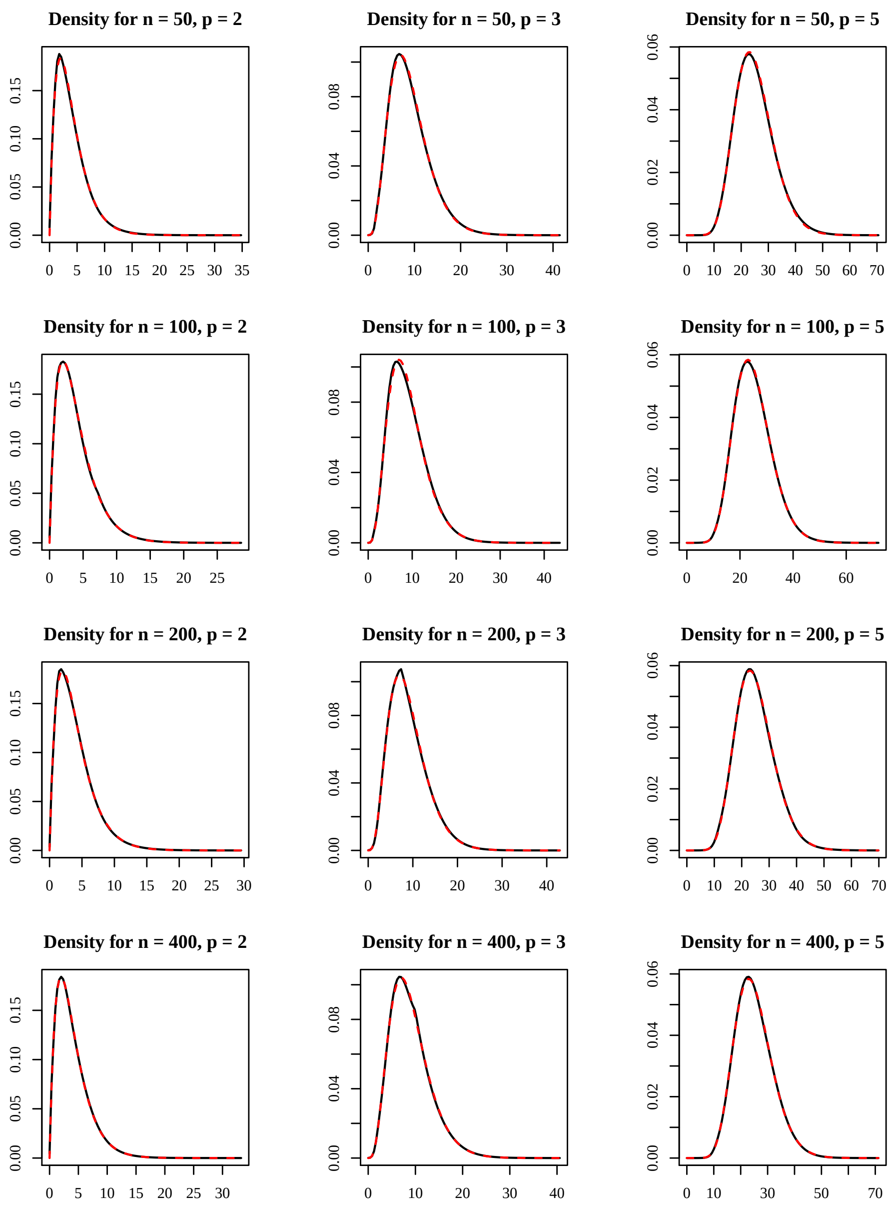

FIGURE 2. Comparison of kernel densities of the simulated B_n statistic based on *Fourier* basis functions with the asymptotic distribution for different sample sizes n and orders $p = q$.

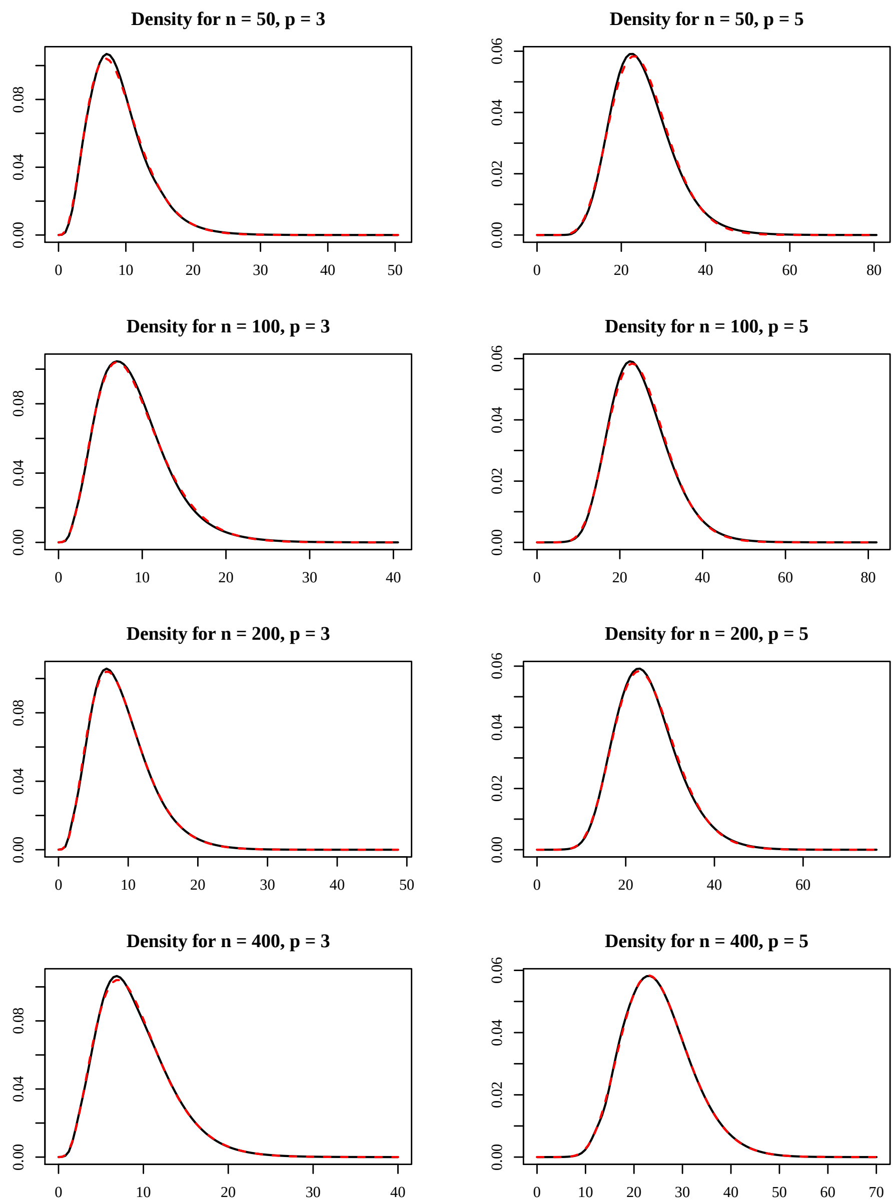

FIGURE 3. Comparison of kernel densities of the simulated B_n statistic based on cubic *spline* basis functions with the asymptotic distribution for different sample sizes n and orders $p = q$.

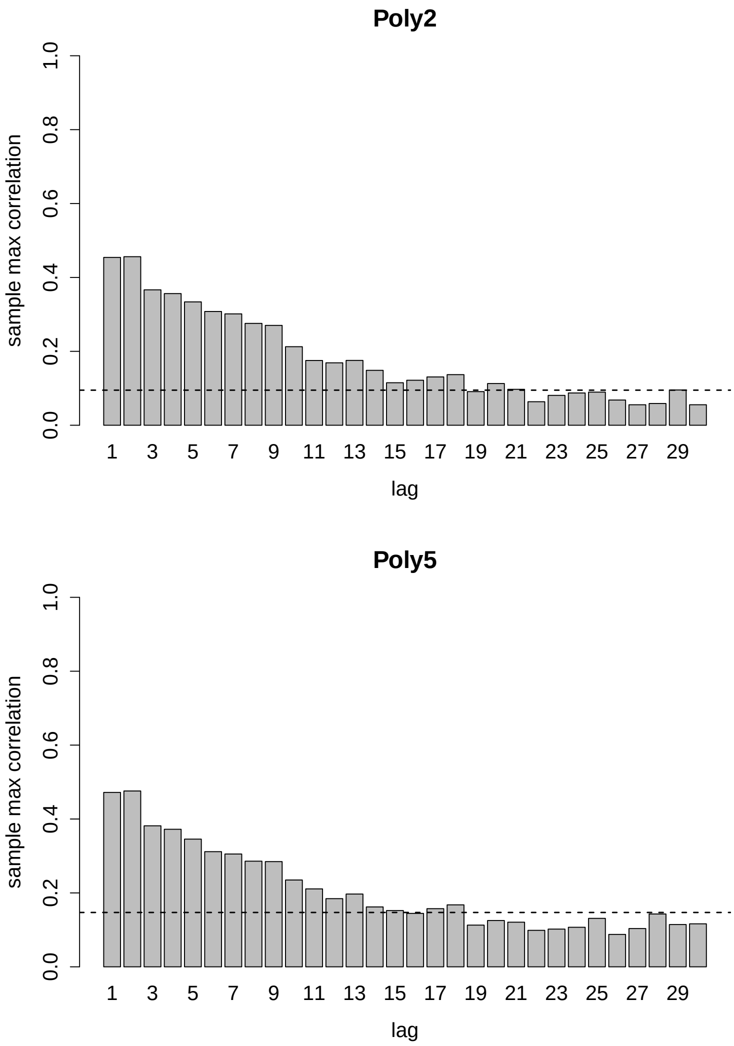

FIGURE 4. Auto-dependence of a stochastic volatility model represented though the sample maximum correlation coefficient plotted against the lag.

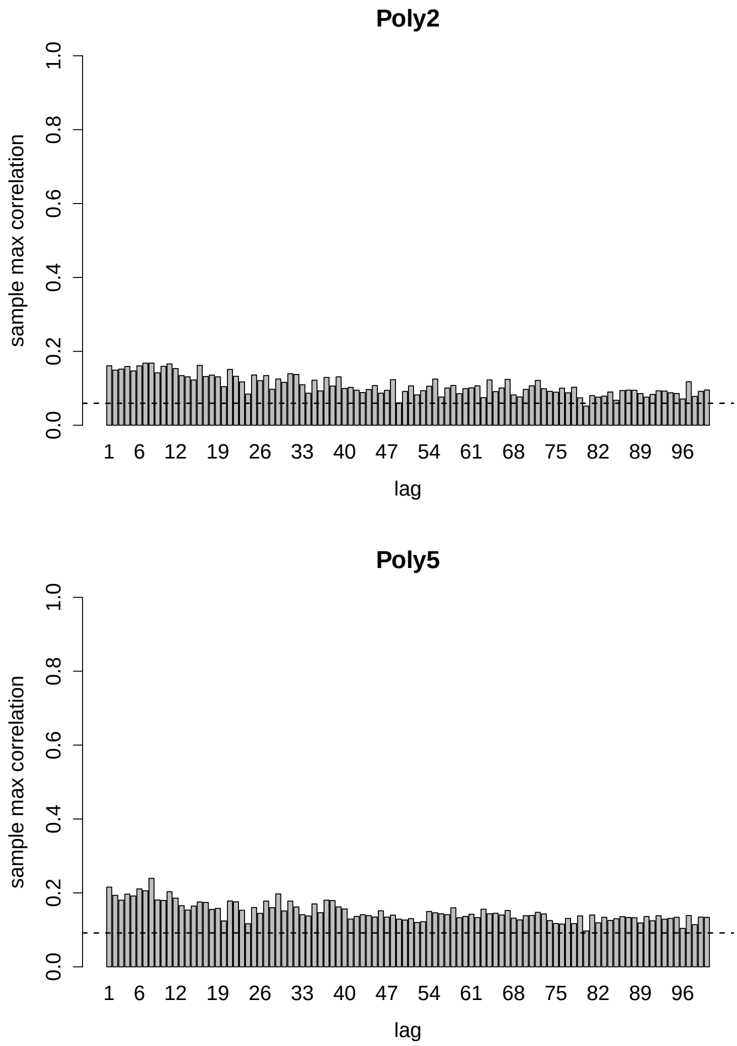

FIGURE 5. Auto-dependence of the daily stock returns of Microsoft in the period 2005-04-22 to 2015-04-24.

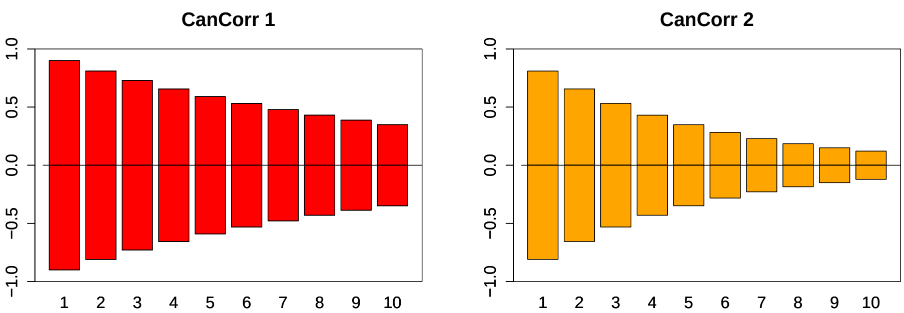

FIGURE 6. Population autodependogram of a symmetric AR(1) with coefficient 0.9

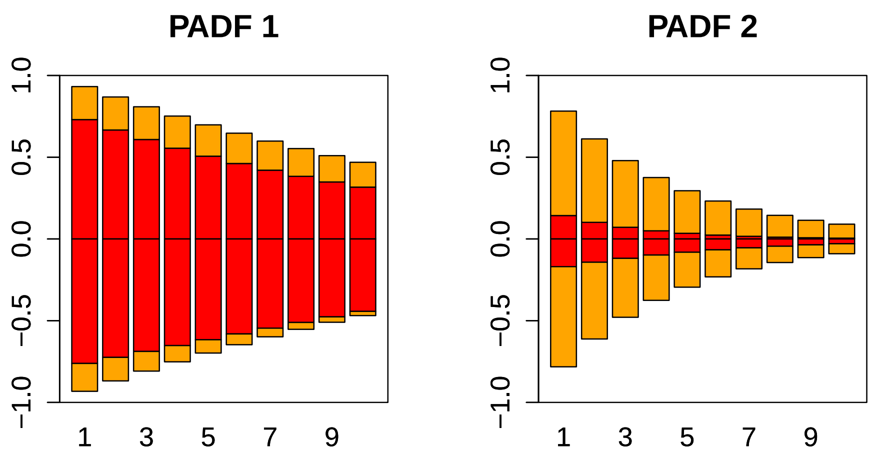

FIGURE 7. Population autodependogram of a skewed AR(1) with coefficients $\phi = 0.9$ and $\lambda = 0.8$.

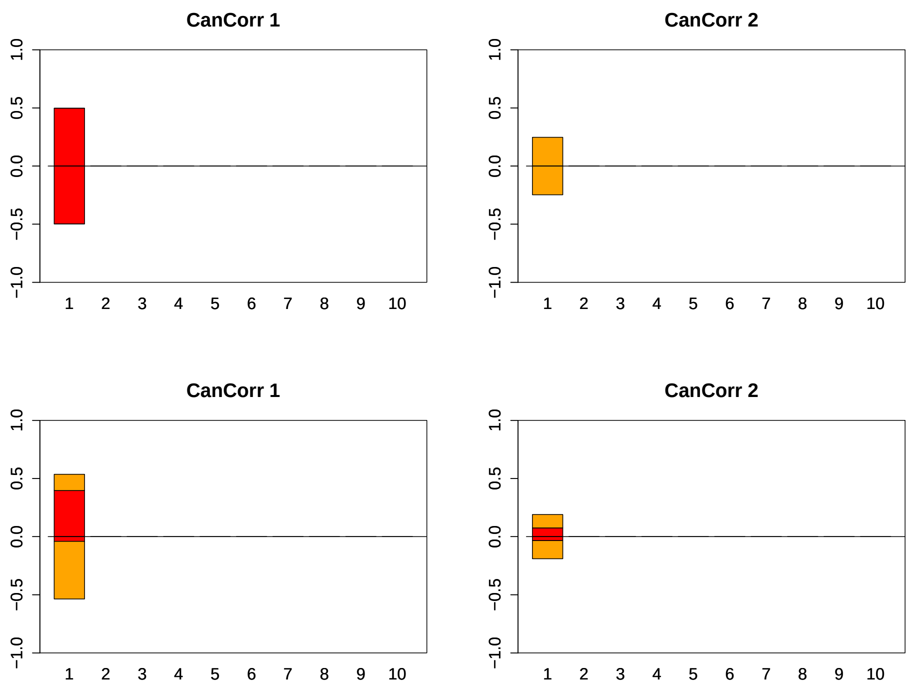

FIGURE 8. Autodependogram for (top) a Gaussian MA(1) with $\theta = 0.9$, and (bottom) a skewed MA(1) with $m_3 = 1$ and $m_4 = 5$.

Appendix A. Cumbersome formulas

Canonical correlation analysis of a skewed forth-order stationary AR(1) process with coefficient ϕ, $\lambda := \text{skew}/\sqrt{\text{kurt} - 1}$

(4)

$$\rho^2_{1,k} = \frac{\phi^{2k}\left(1 - 2\lambda^2\phi^k + \phi^{2k} + \left(1 - \phi^k\right)\sqrt{1 + \phi^k\left(2 - 4\lambda^2 + \phi^k\right)}\right)}{2\left(1 - \lambda^2\right)},$$

$$\rho^2_{2,k} = \frac{\phi^{2k}\left(1 - 2\lambda^2\phi^k + \phi^{2k} - \left(1 - \phi^k\right)\sqrt{1 + \phi^k\left(2 - 4\lambda^2 + \phi^k\right)}\right)}{2\left(1 - \lambda^2\right)},$$

$$\boldsymbol{A}_k := \begin{bmatrix} \frac{-1-\phi^k+\sqrt{1+\phi^k\left(2-4\lambda^2+\phi^k\right)}}{2\lambda} & \frac{-1-\phi^k-\sqrt{1+\phi^k\left(2-4\lambda^2+\phi^k\right)}}{2\lambda}, \\ 1 & 1 \end{bmatrix}$$

$$\boldsymbol{B}_k := \begin{bmatrix} -\frac{\phi^{-k}\left(-1-\phi^k+\sqrt{1+\phi^k\left(2-4\lambda^2+\phi^k\right)}\right)}{2\lambda} & \frac{\phi^{-k}\left(1+\phi^k+\sqrt{1+\phi^k\left(2-4\lambda^2+\phi^k\right)}\right)}{2\lambda} \\ 1 & 1 \end{bmatrix}.$$